Noncommutation Requirements and Field Dependence of Sources of Tensor Fields

K. RAMAN*
Physics Department, Brown University, Providence, Rhode Island 02912†

(Received 25 February 1970)

Starting with the noncommutation requirements among different components of a conserved source tensor for a massive spin-2 field, we discuss the resulting noncommutation requirements between the source and the field variables, and the field dependence of the source that this necessitates.

1. INTRODUCTION

In theories involving vector fields and currents, an important consideration is the noncommutation among certain components of the currents and its relation to the field dependence of the currents.[1-6] For instance, the vacuum expectation value (VEV) of the equal-time commutator (ETC) $[j_0(\mathbf{x}, t), j_k(\mathbf{y}, t)]$ cannot vanish in a Lorentz-invariant theory with a positive-definite metric. This in turn implies that $\delta j_k(\mathbf{x})/\delta A_l(\mathbf{y})$ must be nonvanishing, where $j_\mu(x)$ is the source of the vector field $A_\mu(x)$.

In this paper we briefly discuss the corresponding requirements for the sources of massive spin-2 fields and their field dependence. These requirements are important in theories involving spin-2 fields, e.g., in attempts to formulate tensor-meson dominance of matrix elements of the energy–momentum tensor by spin-2 mesons.[7] The related question of the metric dependence of the stress tensor for a system involving matter fields coupled to the gravitational field has been discussed in detail by Boulware and Deser.[8]

In Sec. 2, we write down the noncommutation requirements among different components of the source tensor $J_{\mu\nu}$. In Sec. 3 we note the noncommutation requirements that this implies between the source $J_{\mu\nu}$ and the field variables, and the field dependence of the source that this necessitates. We indicate how the explicit field dependence can be determined in specific models.

2. NONCOMMUTATION REQUIREMENTS AMONG SOURCE COMPONENTS

We consider a massive spin-2 meson field $U_{\mu\nu}$ coupled to a divergenceless source tensor $J_{\mu\nu}$:

$$\partial_\mu J_{\mu\nu} = 0. \tag{2.1}$$

We first note the noncommutation requirements following from (2.1), Lorentz invariance, and a positive-definite metric.

For a divergenceless, symmetric tensor $J_{\mu\nu}$, we may write the decomposition

$$J_{\mu\nu}(x) = \tilde{J}_{\mu\nu}(x) + \tfrac{1}{3} d_{\mu\nu}(\partial) j(x), \tag{2.2}$$

where

$$j(x) \equiv J^\sigma{}_\sigma(x) \tag{2.3}$$

is the trace of $J_{\mu\nu}(x)$ and where $\tilde{J}_{\mu\nu}(x)$ is traceless as well as divergenceless. In (2.2), we have used the notation

$$d_{\mu\nu}(\partial) = \eta_{\mu\nu} - \partial_\mu \partial_\nu (\Box^2)^{-1}, \tag{2.4}$$

where $\eta_{\mu\nu}$ is the Minkowski metric $(1, -1, -1, -1)$. As usual, we use μ and ν to denote Lorentz indices 0, 1, 2, 3, and i, j, k, and l the spatial indices.

We can write, in general,

$$\langle 0| [\tilde{J}_{\mu\nu}(x), J_{\lambda\sigma}(y)] |0\rangle = \mathfrak{D}_{\mu\nu,\lambda\sigma}(\partial) g(x - y), \tag{2.5}$$

where $g(x - y)$ is a Lorentz-invariant function of the 4-vector $(x - y)$ and

$$\mathfrak{D}_{\mu\nu,\lambda\sigma}(\partial) = \tfrac{1}{2}(d_{\mu\lambda} d_{\nu\sigma} + d_{\mu\sigma} d_{\nu\lambda}) - \tfrac{1}{3} d_{\mu\nu} d_{\lambda\sigma}. \tag{2.6}$$

It follows that

$$\langle 0| [\tilde{J}_{\mu\nu}(x), j(y)] |0\rangle = 0 \tag{2.7}$$

and

$$\begin{aligned}\langle 0| [J_{\mu\nu}(x), J_{\lambda\sigma}(y)] |0\rangle &= \langle 0| [\tilde{J}_{\mu\nu}(x), \tilde{J}_{\lambda\sigma}(y)] |0\rangle \\ &\quad + \tfrac{1}{9} d_{\mu\nu}(\partial_x) d_{\lambda\sigma}(\partial_y) \langle 0| [j(x), j(y)] |0\rangle.\end{aligned} \tag{2.8}$$

The following results are then obtained.

Theorem 1: If

$$\langle 0| [J^\dagger_{00}(\mathbf{x}), J_{k0}(\mathbf{y})] |0\rangle = 0, \tag{2.9}$$

then $J_{\mu\nu}(x) |0\rangle = 0$. (Here and in the following, commutators of operators with arguments $\mathbf{x}$, $\mathbf{y}$, etc., denote ETC's.)

Theorem 2: If

$$\langle 0| [J^\dagger_{0k}(\mathbf{x}), J_{mn}(\mathbf{y})] |0\rangle = 0, \tag{2.10}$$

then $J_{\mu\nu}(x) |0\rangle = 0$.

For a local source operator $J_{\mu\nu}(x)$, the result $J_{\mu\nu}(x)\,|0\rangle = 0$ would imply that $J_{\mu\nu}(x) = 0$,[9] which would mean that the relations (2.9) and (2.10) are not compatible with a theory with a nontrivial interaction.

Proof: The proof of Theorem 1 is direct. For instance, suppose the relation (2.10) holds. Operating with

$$\frac{\partial}{\partial x_k}\frac{\partial}{\partial y_m}\frac{\partial}{\partial y_n}$$

and using (2.1) and (2.8), one obtains

$$\langle 0|\,[\partial_0 \tilde{J}^{\dagger}_{00}(\mathbf{x}), \partial_0^2 \tilde{J}_{00}(\mathbf{y})]\,|0\rangle + \tfrac{1}{9}\langle 0|\,[\partial_0\nabla^2\zeta^{\dagger}(x), \partial_0^2\nabla^2\zeta(\mathbf{y})]\,|0\rangle = 0, \quad (2.11)$$

where $\zeta(x) = (\Box^2)^{-1}(x)$.[10] From (2.11), it follows that

$$\langle 0|\,\tilde{J}^{\dagger}_{00}(\mathbf{x})P_0^3\tilde{J}_{00}(\mathbf{y})\,|0\rangle + \tfrac{1}{9}\langle 0|\,\zeta^{\dagger}(\mathbf{x})\mathbf{P}^2P_0^3\mathbf{P}^2\zeta(\mathbf{y})\,|0\rangle = 0. \quad (2.12)$$

Positive definiteness now implies that

$$\tilde{J}_{00}(x)\,|0\rangle = 0, \quad \zeta(x)\,|0\rangle = 0 \quad (2.13)$$

and therefore that

$$J_{00}(x)\,|0\rangle = 0, \quad (2.14)$$

which leads to

$$J_{\mu\nu}(x)\,|0\rangle = 0 \quad (2.15)$$

in a Lorentz-invariant theory. Theorem 2 may be proved similarly.

Theorems 1 and 2 are contained in the spectral representations written by Boulware and Deser[8] for the VEV's of the ETC's in (2.9) and (2.10) for a divergenceless, symmetric tensor. We have here shown how they may be obtained directly[11]; we shall use them to derive the results of the next section.

Another result that may be obtained similarly is the following:

Theorem 3: The vacuum expectation value of the ETC

$$[J^{\dagger}_{k0}(\mathbf{x}), j(\mathbf{y})] \quad (2.16)$$

cannot vanish unless $j(y)\,|0\rangle = 0$.

Proof: If this VEV did vanish, one would obtain

$$(\Box^2_x)^{-1}\langle 0|\,[\nabla^2_x\partial_0 j^{\dagger}(\mathbf{x}), j(\mathbf{y})]\,|0\rangle = 0, \quad (2.17)$$

which gives

$$\langle 0|\,\{j^{\dagger}(\mathbf{x})\mathbf{P}^2P_0 j(\mathbf{y}) + j(\mathbf{y})\mathbf{P}^2P_0 j(\mathbf{x})\}\,|0\rangle = 0. \quad (2.18)$$

Positive definiteness then implies that $j(x)\,|0\rangle = 0$.

A difference between the results obtained here for $J_{\mu\nu}$ and those holding for a vector current j_μ is that, whereas the latter follow in the same way for a nonconserved vector current, there is an important difference between a divergenceless source tensor $J_{\mu\nu}$ and a source tensor with nonvanishing divergence.

For a source tensor $J_{\mu\nu}$ for which $\partial_\mu\partial_\nu J_{\mu\nu}(x) = \chi(x) \neq 0$ and $J_\sigma{}^\sigma(x) = j(x) \neq 0$, there are two scalar operators $\chi(x)$ and $j(x)$ contained in the decomposition of $J_{\mu\nu}(x)$. Since the VEV $\langle 0|\,\chi^{\dagger}(x)j(y)\,|0\rangle$ would, in general, not vanish and since its sign is not definite, results such as those noted here will not follow directly for a source tensor with a nonvanishing divergence $\partial_\mu\partial_\nu J_{\mu\nu}(x)$.

3. FIELD DEPENDENCE OF THE SOURCE

We now examine the implications of the noncommutation properties of $J_{\mu\nu}(x)$ discussed in the last section for the field dependence of the sources.

We consider a massive spin-2 field described by the field operators $U_{\mu\nu}(x)$ and $\Pi^{\lambda}{}_{\mu\nu}(x)$, interacting with a divergenceless source $J_{\mu\nu}(x)$, and described by the following equations[12-16]:

$$\partial_\rho\Pi^{\rho}{}_{\mu\nu} - \tfrac{1}{2}(\partial_\mu\Pi_\nu + \partial_\nu\Pi_\mu) - \tfrac{1}{2}\eta_{\mu\nu}\partial_\lambda(\tilde{\Pi}_\lambda - \Pi_\lambda) = \tfrac{1}{2}m^2(U_{\mu\nu} - \eta_{\mu\nu}u) - \tfrac{1}{2}gJ_{\mu\nu}, \quad (3.1)$$

$$(\Pi^{\nu}{}_{\lambda}{}^{\mu} - \tfrac{1}{2}\delta^{\mu}_{\lambda}\tilde{\Pi}^{\nu}) + (\Pi^{\mu}{}_{\lambda}{}^{\nu} - \tfrac{1}{2}\delta^{\nu}_{\lambda}\tilde{\Pi}^{\mu}) - \eta^{\mu\nu}\Pi_\lambda = \partial_\lambda W^{\mu\nu} - \tfrac{1}{2}\delta_\lambda{}^{\mu}(\partial_\sigma W^{\nu\sigma}) - \tfrac{1}{2}\delta_\lambda{}^{\nu}(\partial_\sigma W^{\mu\sigma}), \quad (3.2)$$

where

$$W^{\mu\nu} \equiv U^{\mu\nu} - \tfrac{1}{2}\eta^{\mu\nu}u, \quad \Pi_\mu \equiv \Pi^{\alpha}_{\alpha\mu}, \quad \tilde{\Pi}^{\mu} \equiv \Pi^{\mu}_{\alpha\alpha}, \quad (3.3)$$

and

$$\partial_\mu J^{\mu\nu} = 0. \quad (3.4)$$

Equations (3.1)–(3.4) lead to the following equations:

$$\partial^\mu(U_{\mu\nu} - \eta_{\mu\nu}u) = 0, \quad (3.5)$$

$$\Pi^{\lambda}{}_{\mu\nu} = \tfrac{1}{2}[\partial_\mu U_{\lambda\nu} + \partial_\nu U_{\lambda\mu} - \partial_\lambda U_{\mu\nu}], \quad (3.6)$$

$$u = \frac{-g}{3m^2}j, \quad j \equiv J^{\sigma}{}_{\sigma}, \quad (3.7)$$

$$(\Box^2 + m^2)U_{\mu\nu} - [\partial_\mu\partial_\nu + m^2\eta_{\mu\nu}]u = gJ_{\mu\nu}. \quad (3.8)$$

From Eqs. (3.1)–(3.8), we may separate out the following constraint equations[17]:

$$gJ_{00} = \partial_m\partial_j U^{\mathrm{T}}_{mj} - (\tfrac{2}{3}\nabla^2 - m^2)U_{mm}, \quad (3.9)$$

$$gJ_{k0} = m^2U_{k0} - 2\partial_m\Pi^0_{mk} + 2\partial_k\Pi^0_{mm}. \quad (3.10)$$

In Eq. (3.9), U^{T}_{mj} denotes the transverse part of U_{mj}:

$$U^{\mathrm{T}}_{mj} = U_{mj} - \tfrac{1}{3}\delta_{mj}U_{ll}. \quad (3.11)$$

From these equations, we may derive the following noncommutation requirements of the source with the field variables:

Theorem 4: If

$$\langle 0|\,[J^{\dagger}_{k0}(\mathbf{x}),\, U^{\mathrm{T}}_{ml}(\mathbf{y})]\,|0\rangle = 0 \tag{3.12a}$$

and

$$\langle 0|\,[J^{\dagger}_{k0}(\mathbf{x}),\, U_{mm}(\mathbf{y})]\,|0\rangle = 0, \tag{3.12b}$$

then it follows that $J_{\mu\nu}(x)\,|0\rangle = 0$.

This is obtained by using Theorem 1 and the constraint equation (3.9) to express J_{00} in Theorem 1 in terms of U_{ml}.

Theorem 5: If

$$\langle 0|\,[J^{\dagger}_{k0}(\mathbf{x}),\, U_{mm}(\mathbf{y})]\,|0\rangle = 0 \tag{3.13a}$$

and

$$\langle 0|\,[J^{\dagger}_{00}(\mathbf{x}),\, \Pi^{0}_{ml}(\mathbf{y})]\,|0\rangle = 0, \tag{3.13b}$$

then $J_{\mu\nu}(x)\,|0\rangle = 0$.

To obtain this, we first use Theorem 1, together with (3.10), to obtain the following result: If

$$\langle 0|\,[J^{\dagger}_{00}(\mathbf{x}),\, U_{k0}(\mathbf{y})]\,|0\rangle = 0$$

and

$$\langle 0|\,[J^{\dagger}_{00}(\mathbf{x}),\, \Pi^{0}_{ml}(\mathbf{y})]\,|0\rangle = 0,$$

then $J_{\mu\nu}(x)\,|0\rangle = 0$.

Using Lorentz covariance, we may write

$$\langle 0|\,[J^{\dagger}_{\mu\nu}(x),\, (U_{\lambda\sigma}(y) - \eta_{\lambda\sigma}u(y))]\,|0\rangle = \mathfrak{D}_{\mu\nu,\lambda\sigma}(\partial)F_1(x-y) + d_{\mu\nu}(\partial)d_{\lambda\sigma}(\partial)F_2(x-y), \tag{3.14}$$

where F_1 and F_2 are Lorentz-invariant functions of the 4-vector $(x - y)$. From (3.14), it is seen that the relation

$$\langle 0|\,[J^{\dagger}_{00}(x),\, U_{k0}(y)]\,|0\rangle = 0$$

implies (3.13a). Theorem 5 follows.

Theorem 6: If

$$\langle 0|\,[J^{\dagger}_{kl}(\mathbf{x}),\, \Pi^{0}_{mj}(\mathbf{y})]\,|0\rangle = 0 \tag{3.15}$$

and

$$\langle 0|\,[J^{\dagger}_{m0}(\mathbf{x}),\, (U_{kl}(\mathbf{y}) - \eta_{kl}u(\mathbf{y}))]\,|0\rangle = 0, \tag{3.16}$$

then $J_{\mu\nu}(x)\,|0\rangle = 0$.

To prove this, we use Theorem 2 and the constraint equation (3.10) to show that, if (3.15) holds and if $\langle 0|\,[J^{\dagger}_{kl}(\mathbf{x}),\, U_{m0}(\mathbf{y})]\,|0\rangle = 0$, then $J_{\mu\nu}(x)\,|0\rangle = 0$. Using (3.14), we then obtain Theorem 6.

We finally note the following result:

Theorem 7: If

$$\langle 0|\,[j^{\dagger}(\mathbf{x}),\, U_{k0}(\mathbf{y})]\,|0\rangle = 0, \tag{3.17}$$

then $j(x)\,|0\rangle = 0$.

Proof: Using (3.14) and (3.7), we can show that (3.17) implies that

$$\langle 0|\,[J^{\dagger}_{k0}(\mathbf{x}), j(\mathbf{y})]\,|0\rangle = 0. \tag{3.18}$$

Theorem 3 then leads to Theorem 7.

The noncommutation requirements between the source and field may be used for obtaining information about the form of the field dependence of the source. For instance, we may write the spectral representation

$$\langle 0|\,[J^{\dagger}_{\mu\nu}(x), J_{\lambda\sigma}(y)]\,|0\rangle = \int_0^{\infty} ds[\sigma_2(s)\mathfrak{D}_{\mu\nu,\lambda\sigma}(\partial) + \sigma_0(s)\,d_{\mu\nu}(\partial)\,d_{\lambda\sigma}(\partial)]\Delta(x-y,s) \tag{3.19}$$

and evaluate the spectral functions in terms of a truncated sum over intermediate states.

We then obtain

$$\langle 0|\,[j^{\dagger}(\mathbf{x}),\, U_{k0}(\mathbf{y})]\,|0\rangle = \frac{g}{m^2}\langle 0|\,[j^{\dagger}(\mathbf{x}),\, J_{k0}(\mathbf{y})]\,|0\rangle = -\frac{g}{m^2}\int \frac{ds}{s}\,3\sigma_0(s)\partial_k\delta(x-y). \tag{3.20}$$

Expressing U_{0k} in terms of the dynamical variable π_{jl},

$$U_{0k} = -\frac{1}{m^2}\partial_l\pi_{kl} + \frac{2}{3m^4}\partial_k\partial_m\partial_l\pi_{ml} + \frac{g}{m^2}\left(J_{0k} - \frac{2}{3m^2}\partial_k\partial_m J_{0m}\right), \tag{3.21}$$

we may obtain from (3.20) the relation

$$\langle 0|\left[j^{\dagger}(\mathbf{x}),\, \partial_l\left(\pi_{kl}(\mathbf{y}) - \frac{2}{3m^4}\eta_{kl}\partial_j\partial_m\pi_{jm}(\mathbf{y})\right)\right]|0\rangle = \frac{-2g}{3m^2}\langle 0|\,[j^{\dagger}(\mathbf{x}),\, \partial_k\partial_m J_{0m}(\mathbf{y})]\,|0\rangle = \frac{-g}{3m^2}\int\frac{ds}{s}\,\sigma_0(s)\nabla^2\partial_k\delta(x-y), \tag{3.22}$$

which constrains the dependence of $j(x)$ on $U^{\mathrm{T}}_{kl}(x)$.

We remark that Eq. (3.22) may be written as a differential equation for the function

$$\langle 0|\,\frac{\delta j(\mathbf{x})}{\delta U^{\mathrm{T}}_{jm}(\mathbf{y})}\,|0\rangle,$$

which could, in principle, be solved in particular models to find the VEV of this functional derivative.

We finally note that, in a Lagrangian theory, care must be taken in relating the source tensor $J^{\mu\nu}$ to the interaction terms in the Lagrangian. For instance, the interaction term in the Lagrangian cannot be merely of the form $U^{\mu\nu}J^{\mu\nu}$, with $J^{\mu\nu}$ taken as a divergenceless tensor constructed from a set of matter fields ϕ such as, for instance, the stress tensor for the field ϕ. Such an interaction term cannot lead to a source $J^{\mu\nu}$ in the field equations that would satisfy the field-dependence requirements; further, this interaction term would make $J^{\mu\nu}$ nonconserved.

In this paper we have examined the constraints on the field dependence of the source following because the source is divergenceless. Further constraints are imposed by the hypothesis of a field-source identity[7]; these will be discussed elsewhere.

We hope to examine the explicit nature of the field dependence in particular models in a separate work.

* Work supported in part by the U.S. Atomic Energy Commission (Report No. NYO-2262TA-214).

† Address during 1970–71: Physics Department, Wesleyan University, Middletown, Conn. 06457.

[1] J. Schwinger, Phys. Rev. Letters **3**, 296 (1959); Phys. Rev. **130**, 406 (1963).
[2] K. Johnson, Nucl. Phys. **25**, 431 (1961).
[3] S. Okubo, Nuovo Cimento **44**, A1015 (1966).
[4] L. S. Brown, Phys. Rev. **150**, 1338 (1966).
[5] D. G. Boulware and S. Deser, Phys. Rev. **151**, 1278 (1966).
[6] Y. Frishman, Phys. Rev. **159**, 1428 (1967).
[7] K. Raman, Phys. Rev., in press.
[8] D. G. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967).
[9] P. Federbush and K. Johnson, Phys. Rev. **120**, 1926 (1960).
[10] The inverse d'Alembertian operator can be defined in terms of an appropriate Green's function, with suitable boundary conditions.
[11] By not using the Källén–Lehmann representation, we avoid the question of the asymptotic behavior of the spectral functions. The methods used here are similar to those of Ref. 6.
[12] More details of these are given in Ref. 7. For a discussion of the theory of free massive spin-2 fields, see Chang (Ref. 13) and Deser, Trubatch, and Trubatch (Ref. 14).
[13] S. J. Chang, Phys. Rev. **148**, 1259 (1966).
[14] S. Deser, J. Trubatch, and S. Trubatch, Can. J. Phys. **44**, 1715 (1968).
[15] We here derive the relations in terms of $U_{\mu\nu}$ and $\Pi^{\lambda}{}_{\mu\nu}$, these being symmetric in μ and ν. Analogous relations may be obtained in a theory expressed in terms of a symmetric $U_{\mu\nu}$ and a field variable $G^{\mu[\nu\lambda]}$ antisymmetric in ν and λ, e.g., as used in Ref. 13 or Ref. 16 below.
[16] K. Raman, Nuovo Cimento **55A**, 650 (1968), and Erratum, to be published.
[17] The conjugate dynamical variables may be chosen as U^{T}_{ij} and
$$\pi_{ij} = -2\Pi^{0}_{ij} + 2(\nabla^2 - \tfrac{3}{2}m^2)^{-1}(\partial_i\partial_j - \tfrac{1}{2}m^2\eta_{ij})\Pi^{0}_{mm};$$
this may be shown as in Ref. 13. For an interacting field, there is the additional independent quantity $u = U^{\sigma}{}_{\sigma}$, which is given directly in terms of the source by Eq. (3.7). Note that the relation between the constraint variables and the above dynamical variables will now involve the source $J_{\mu\nu}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Reduction of a Class of $O(4, 2)$ Representations with Respect to $SO(4, 1)$ and $SO(3, 2)$

A. O. BARUT*
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado 80302
AND
A. BÖHM†
Physics Department, University of Texas, Austin, Texas

(Received 12 March 1970)

A complete classification of representations of $SO(4, 2)$ with infinitesimal generators S_{AB}, characterized by the *representation relation* $\{S_{AB}, S^{A}{}_{C}\} = -2ag_{BC}$, and their extension by parity have been determined. The possible values of a are $a = 1 - S^2$, $S = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots$. These representations have then been reduced according to the two chains $SO(4, 2) \supset SO(4, 1) \supset SO(4) \supset SO(3)$ and $SO(4, 2) \supset SO(3, 2) \supset SO(3) \otimes SO(2)$. The equivalence of these representations with the oscillator representations is established.

1. INTRODUCTION

A class of unitary irreducible representations of the group $O(4, 2)$ [$SO(4, 2)$ extended with parity] have found important applications in the dynamical problems in atomic and particle physics (Sec. 7). The problem of reduction of these representations with respect to the two chains [given in the abstract or in the Eqs. (2.5) and (2.6) below] arise in these applications, and is also of mathematical interest. The purpose of this paper is to give an algebraic characterization of the class of representations ("representation relation"), and to present the rather remarkable features that occur when these representations are reduced according to the above chains. In

which could, in principle, be solved in particular models to find the VEV of this functional derivative.

We finally note that, in a Lagrangian theory, care must be taken in relating the source tensor $J^{\mu\nu}$ to the interaction terms in the Lagrangian. For instance, the interaction term in the Lagrangian cannot be merely of the form $U^{\mu\nu}J^{\mu\nu}$, with $J^{\mu\nu}$ taken as a divergenceless tensor constructed from a set of matter fields ϕ such as, for instance, the stress tensor for the field ϕ. Such an interaction term cannot lead to a source $J^{\mu\nu}$ in the field equations that would satisfy the field-dependence requirements; further, this interaction term would make $J^{\mu\nu}$ nonconserved.

In this paper we have examined the constraints on the field dependence of the source following because the source is divergenceless. Further constraints are imposed by the hypothesis of a field-source identity[7]; these will be discussed elsewhere.

We hope to examine the explicit nature of the field dependence in particular models in a separate work.

* Work supported in part by the U.S. Atomic Energy Commission (Report No. NYO-2262TA-214).

† Address during 1970–71: Physics Department, Wesleyan University, Middletown, Conn. 06457.

[1] J. Schwinger, Phys. Rev. Letters **3**, 296 (1959); Phys. Rev. **130**, 406 (1963).
[2] K. Johnson, Nucl. Phys. **25**, 431 (1961).
[3] S. Okubo, Nuovo Cimento **44**, A1015 (1966).
[4] L. S. Brown, Phys. Rev. **150**, 1338 (1966).
[5] D. G. Boulware and S. Deser, Phys. Rev. **151**, 1278 (1966).
[6] Y. Frishman, Phys. Rev. **159**, 1428 (1967).
[7] K. Raman, Phys. Rev., in press.
[8] D. G. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967).
[9] P. Federbush and K. Johnson, Phys. Rev. **120**, 1926 (1960).
[10] The inverse d'Alembertian operator can be defined in terms of an appropriate Green's function, with suitable boundary conditions.
[11] By not using the Källén–Lehmann representation, we avoid the question of the asymptotic behavior of the spectral functions. The methods used here are similar to those of Ref. 6.
[12] More details of these are given in Ref. 7. For a discussion of the theory of free massive spin-2 fields, see Chang (Ref. 13) and Deser, Trubatch, and Trubatch (Ref. 14).
[13] S. J. Chang, Phys. Rev. **148**, 1259 (1966).
[14] S. Deser, J. Trubatch, and S. Trubatch, Can. J. Phys. **44**, 1715 (1968).
[15] We here derive the relations in terms of $U_{\mu\nu}$ and $\Pi^{\lambda}{}_{\mu\nu}$, these being symmetric in μ and ν. Analogous relations may be obtained in a theory expressed in terms of a symmetric $U_{\mu\nu}$ and a field variable $G^{\mu[\nu\lambda]}$ antisymmetric in ν and λ, e.g., as used in Ref. 13 or Ref. 16 below.
[16] K. Raman, Nuovo Cimento **55A**, 650 (1968), and Erratum, to be published.
[17] The conjugate dynamical variables may be chosen as U^{T}_{ij} and
$$\pi_{ij} = -2\Pi^{0}_{ij} + 2(\nabla^2 - \tfrac{3}{2}m^2)^{-1}(\partial_i\partial_j - \tfrac{1}{2}m^2\eta_{ij})\Pi^{0}_{mm};$$
this may be shown as in Ref. 13. For an interacting field, there is the additional independent quantity $u = U^{\sigma}{}_{\sigma}$, which is given directly in terms of the source by Eq. (3.7). Note that the relation between the constraint variables and the above dynamical variables will now involve the source $J_{\mu\nu}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Reduction of a Class of $O(4, 2)$ Representations with Respect to $SO(4, 1)$ and $SO(3, 2)$

A. O. BARUT*
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado 80302
AND
A. BÖHM†
Physics Department, University of Texas, Austin, Texas

(Received 12 March 1970)

A complete classification of representations of $SO(4, 2)$ with infinitesimal generators S_{AB}, characterized by the *representation relation* $\{S_{AB}, S^{A}{}_{C}\} = -2ag_{BC}$, and their extension by parity have been determined. The possible values of a are $a = 1 - S^2$, $S = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots$. These representations have then been reduced according to the two chains $SO(4, 2) \supset SO(4, 1) \supset SO(4) \supset SO(3)$ and $SO(4, 2) \supset SO(3, 2) \supset SO(3) \otimes SO(2)$. The equivalence of these representations with the oscillator representations is established.

1. INTRODUCTION

A class of unitary irreducible representations of the group $O(4, 2)$ [$SO(4, 2)$ extended with parity] have found important applications in the dynamical problems in atomic and particle physics (Sec. 7). The problem of reduction of these representations with respect to the two chains [given in the abstract or in the Eqs. (2.5) and (2.6) below] arise in these applications, and is also of mathematical interest. The purpose of this paper is to give an algebraic characterization of the class of representations ("representation relation"), and to present the rather remarkable features that occur when these representations are reduced according to the above chains. In

Sec. 6 we establish the equivalence of these representations with those obtained by means of boson creation operators (oscillator representations).

2. CHARACTERIZATION OF $SO(4, 2)$: REPRESENTATIONS

A. Notations

The following notations are used for $SO(4, 2)$: The generators of $SO(4, 2)$ are

$$S_{AB}, \quad A, B, C = 5, 0, 1, 2, 3, 4. \tag{2.1}$$

They obey the commutation relations

$$[S_{AB}, S_{CD}] = -i(g_{AC}S_{BD} + g_{BD}S_{AC} - g_{BC}S_{AD} - g_{AD}S_{BC}), \tag{2.2}$$

with $g_{55} = +1$, $g_{00} = +1$, $g_{11} = -1$, $g_{22} = -1$, $g_{33} = -1$, and $g_{44} = -1$. We define

$$\Gamma_a = S_{5a}, \quad a = 0, 1, 2, 3, 4, \tag{2.3}$$

and use the letters a, b, c, d for the indices 0, 1, 2, 3, 4. There are two $SO(4, 1)$ subgroups

$$SO(4, 1)_{S_{ab}} \quad \text{and} \quad SO(4, 1)_{\Gamma_a, S_{\alpha\beta}}, \quad \alpha, \beta = 1, 2, 3, 4.$$

We indicate the generators of a subgroup as subscripts. We use the letters $\alpha, \beta, \gamma, \cdots$ for the indices 1, 2, 3, 4. $SO(4)$ is generated by $S_{\alpha\beta}$. For the indices 1, 2, 3 we use the letters $i, j, k, \cdots$. $SO(3)$ generated by S_{ij} is the rotation group. For the indices 0, 1, 2, 3 we use the letters $\mu, \nu, \delta, \cdots$. $SO(3, 1)_{S_{\mu\nu}}$ generated by $S_{\mu\nu}$ is the Lorentz group. There is another $SO(3, 1)$ subgroup: $SO(3, 1)_{\Gamma_i, S_{ij}}$. For the indices 5, 0, 1, 2, 3 we use the letters $\xi, \eta, \zeta, \cdots$. There is one $SO(3, 2)$ subgroup $SO(3, 2)_{S_{\xi\eta}}$.

B. Representation Relation

The first problem solved in this paper is to find all unitary irreducible representations of $SO(4, 2)$ which fulfill the additional condition (representation relation)[1]

$$\{S_{AB}, S^A{}_C\} = -2ag_{BC}, \tag{2.4}$$

where a is a number. It will turn out that only for special values of a are there nontrivial representations. Our task is therefore to determine the possible values of a and to give a complete classification of the corresponding irreducible representations.

The second problem solved is to reduce the representations characterized by (2.4) according to the reduction chains

$$SO(4, 2)_{501234} \supset SO(3, 2)_{50123} \supset SO(3)_{123} \otimes SO(2)_{50}, \tag{2.5}$$

$$SO(4, 2)_{501234} \supset SO(4, 1)_{01234} \supset SO(4)_{1234} \supset SO(3)_{123}. \tag{2.6}$$

Although there are two distinct $SO(4, 1)$ subgroups, it is sufficient to consider in the reduction one of them; the other reduction is algebraically identical to this one.

C. Algebraic Relations which Follow from the Representation Relation (2.4)

From Eq. (2.4) it follows immediately that

$$\tfrac{1}{2}S_{AB}S^{AB} = -3a, \tag{2.7}$$

so that a can only be real and is essentially the quadratic Casimir operator of $SO(4, 2)$. From (2.4) and the definition (2.3) it follows further that

$$\{\Gamma_b, \Gamma_c\} + \{S_{ab}, S^a{}_c\} = -2ag_{bc}, \quad a, b = 0, 1, 2, 3, 4. \tag{2.8}$$

Equation (2.7) can be rewritten as

$$\Gamma_b\Gamma^b + \tfrac{1}{2}S_{ab}S^{ab} = -3a, \tag{2.9}$$

and from (2.8) we obtain

$$\Gamma_b\Gamma^b + S_{ab}S^{ab} = -5a, \tag{2.10}$$

so that (2.9) and (2.10) give

$$\tfrac{1}{2}S_{ab}S^{ab} = -2a \tag{2.11}$$

and

$$\Gamma_b\Gamma^b = -a. \tag{2.12}$$

Thus, the second-order Casimir operator of the subgroup $SO(4, 1)_{01234}$ is a constant. We must also evaluate the fourth-order Casimir operator

$$W = -\tfrac{1}{64}\varepsilon^{abcde}\varepsilon_a{}^{fghi}S_{bc}S_{de}S_{fg}S_{hi} \tag{2.13}$$

of $SO(4, 1)_{01234}$ for the representations characterized by (2.4). A lengthy but straightforward calculation using (2.2) and (2.8) gives

$$W = a(1 - a), \tag{2.14}$$

i.e., also a constant.

In a similar way we evaluate the second- and fourth-order Casimir operators of the subgroup $SO(3, 2)_{50123}$ and obtain

$$\tfrac{1}{2}S_{\xi\eta}S^{\xi\eta} = -2a, \quad \xi, \eta = 5, 0, 1, 2, 3, \tag{2.15}$$

and

$$-P_1 = -\tfrac{1}{64}\varepsilon^{\kappa\xi\eta\zeta\epsilon}\varepsilon_\kappa{}^{\xi'\eta'\zeta'\epsilon'}S_{\xi\eta}S_{\zeta\epsilon}S_{\xi'\eta'}S_{\zeta'\epsilon'} = a(1 - a). \tag{2.16}$$

Finally, we derive two relations involving the $SO(4)_{\alpha\beta} \otimes SO(2)_{\Gamma_0}$ and $SO(3, 1)_{\mu\gamma} \otimes SO(1, 1)_{\Gamma_4}$ subgroups. From (2.8) we obtain

$$\Gamma_0{}^2 + S_{a0}S^{a0} = -a, \tag{2.17}$$

and with (2.11)

$$S_{a0}S^{a0} + \tfrac{1}{2}S_{\alpha\beta}S^{\alpha\beta} = -2a, \quad \alpha, \beta = 1, 2, 3, 4. \tag{2.18}$$

We find

$$\Gamma_0^2 = \tfrac{1}{2}S_{\alpha\beta}S^{\alpha\beta} + a. \tag{2.19}$$

Similarly, from (2.4) we have

$$\{S_{4\xi}, S^4{}_\eta\} + \{S_{\zeta\xi}, S^\zeta{}_\eta\} = -2ag_{\xi\eta} \tag{2.20}$$

so that

$$\Gamma_4^2 - S_{\zeta 5}S^{\zeta 5} = a.$$

From this and (2.15) we obtain

$$\Gamma_4^2 = -a - \tfrac{1}{2}S_{\mu\nu}S^{\mu\nu}. \tag{2.21}$$

3. REDUCTION WITH RESPECT TO $SO(4, 1)_{01234}$

According to (2.11) and (2.13), both the second- and fourth-order Casimir operators of the subgroup $SO(4, 1)_{01234}$ are constants for representations of $SO(4, 2)$ characterized by (2.4). We suspect therefore that these irreducible representations of $SO(4, 2)$ will remain irreducible also under the subgroup $SO(4, 1)$. However, so far we only know that the irreducible representations into which it reduces must have the same values of the second- and fourth-order Casimir operators. If the irreducible representations remain irreducible under $SO(4, 1)$, then the $SO(4) \supset SO(3) \supset SO(2)$ basis of $SO(4, 1)$ is already a complete basis of the $SO(4, 2)$ irreducible representation. We shall therefore investigate this point, making use of the complete classification of the irreducible representations of $SO(4, 1)$ given by Newton[2] and by Dixmier.[3]

The irreducible representations of $SO(4)_{1234}$ are characterized by two numbers (k_0, n), where $|k_0|$ is integer or half-integer and n is a natural number.[4] These two numbers are related to the values of the Casimir operators by

$$\begin{aligned} \tfrac{1}{2}S_{\alpha\beta}S^{\alpha\beta} &= k_0{}^2 + (|k_0| + n)^2 - 1, \\ \tfrac{1}{8}\varepsilon^{\alpha\beta\gamma\delta}S_{\alpha\beta}S_{\gamma\delta} &= k_0(|k_0| + n). \end{aligned} \tag{3.1}$$

From (2.19) and (3.1) we see that

$$\text{spectrum } \Gamma_0^2 = a - 1 + k_0{}^2 + (|k_0| + n)^2, \tag{3.2}$$

so that the spectrum of Γ_0 (up to a sign) is determined by the spectrum of $SO(4)_{1234}$ in the irreducible representation of $SO(4, 2)$. We shall now compare our irreducible representations characterized by (2.11) and (2.14) with the complete list of irreducible representations of $SO(4, 1)$ of Newton[2] and as corrected by Dixmier.[3] For this purpose we divide our representations into subclasses:

(1) $a = 1$. Then $W = 0$ [Eq. (2.14)], and $Q \equiv -\frac{1}{2}S_{ab}S^{ab} = 2$. This is a class I representation of Newton,[2] and its reduction with respect to $SO(4)$ is given by

$$I^{(a=1)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus\, (k_0 = 0, n). \tag{3.3}$$

There are no other representations in other classes with this value of a.

(2) $a = 0$. Then $W = 0$ and $Q = 0$. There is one class II representation with these values of W and Q and with the $SO(4)$ reduction

$$\text{II}^{(a=0)} \xrightarrow[SO(4)]{} \sum_{n=2,3,4\cdots}^{\infty} \oplus\, (k_0 = 0, n), \tag{3.4}$$

and there are representations in the class IVa and IVb (which were not in the original listing of Newton and have been later added by Dixmier) with the reductions

$$\text{IV}_\text{a}^{(a=0)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus\, (k_0 = -1, n), \tag{3.5}$$

$$\text{IV}_\text{b}^{(a=0)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus\, (k_0 = +1, n). \tag{3.6}$$

Finally, there is a further representation of $SO(4, 1)$ with $Q = 0$, $W = 0$, i.e., $a = 0$. This is the class III representation,[2] with $S = 1$, with the reduction

$$\text{III}^{(a=0)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus \Big((k_0 = -1, n) \oplus (k_0 = +1, n) \times \sum_{n=2,3\cdots}^{\infty} \oplus\, (k_0 = 0, n) \Big). \tag{3.7}$$

(3) $0 < a < 1$. Then $W > 0$, $0 < Q < 2$. There is a class III representation of this kind whenever a is such that

$$a(1 - a) = S(S + 1)2a + (S - 1)S(S + 1)(S + 2), \quad S = \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots \tag{3.8a}$$

and

$$\begin{aligned} 2a &> \tfrac{9}{4} - (S + \tfrac{1}{2})^2, \quad \text{for } S = \text{integer}, \\ 2a &> \tfrac{5}{2} - (S + \tfrac{1}{2})^2, \quad \text{for } S = \text{half-integer}. \end{aligned} \tag{3.8b}$$

Equation (3.8a) can be satisfied with a real a only for two values of S:

$$\begin{aligned} &S = \tfrac{1}{2}, \quad a = \tfrac{3}{4} \text{ or } a = -\tfrac{5}{4}, \\ &S = 1, \quad a = -3 \text{ or } a = 0. \end{aligned} \tag{3.9}$$

All these values are excluded—the first one ($S = \frac{1}{2}$, $a = \frac{3}{4}$) by (3.8b). Consequently, there is no class III representation in the range $0 < a < 1$.

There are, however, other representations of $SO(4, 1)$ with the eigenvalues of $Q = \frac{3}{2}$ and $W = \frac{3}{16}$, i.e., with $a = \frac{3}{4}$ in this range. These are representations in the class IVa and IVb (which were also not in the original listing of Newton, but were later added by

Dixmier). Their reduction is

$$\mathrm{IV}_{\mathrm{a}}^{(a=\frac{3}{4})} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus (k_0 = -\tfrac{1}{2}, n),$$
$$\mathrm{IV}_{\mathrm{b}}^{(a=\frac{3}{4})} \xrightarrow[SO(4)]{} \sum_{n=1,2\cdots} \oplus (k_0 = +\tfrac{1}{2}, n). \quad (3.10)$$

(4) $a > 1$. Then $Q > 2$ and $W < 0$, and there are no unitary irreducible representations of $SO(4, 1)$ with these values of Q and W.

(5) $a < 0$. Then $Q < 0$ and $W < 0$. These representations can only be of class IV, and there is a class IV representation of this kind whenever a is such that

$$2a = -t(t-1) - (S-1)(S+2),$$
$$a(1-a) = -t(t-1)S(S+1), \quad (3.11)$$

where

$$S = \tfrac{3}{2}, 2, \tfrac{5}{2}, 3, \cdots$$

and

$$0 < t \leq S.$$

From (3.11) we obtain $S = t$; hence the only values of a for which there are unitary representations of $SO(4, 1)$ of class IV are

$$a = 1 - S^2, \quad S = \tfrac{3}{2}, 2, \tfrac{5}{2}, \cdots,$$

or

$$a = -\tfrac{5}{4}, -3, -\tfrac{21}{4}, \cdots. \quad (3.12)$$

The reduction of these representations reads

$$\mathrm{IV}_{\mathrm{a}}^{(a=1-S^2)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus (k_0 = -S, n) \quad (3.13)$$

and

$$\mathrm{IV}_{\mathrm{b}}^{(a=1-S^2)} \xrightarrow[SO(4)]{} \sum_{n=1,2,3\cdots}^{\infty} \oplus (k_0 = S, n),$$
$$S = \tfrac{3}{2}, 2, \tfrac{5}{2}, \cdots. \quad (3.14)$$

Herewith we have obtained a complete classification of irreducible representations of $SO(4, 2)$ characterized by the additional representation relation (4). Collecting all the cases, we see that the spectrum of a is

$$a = 1 - S^2, \quad S = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots. \quad (3.15)$$

For $S = 0$ ($a = 1$) there is only one representation of $SO(4, 1)$; hence the $SO(4, 2)$ representation remains irreducible when restricted to the $SO(4, 1)$ subgroup. This fact is very well known. For $S = \frac{3}{2}, 2, \frac{5}{2}, \cdots$, $a = -\frac{5}{4}, -3, -\frac{21}{4}, \cdots$, there are two $SO(4, 1)$ representations for the same value of S [Eqs. (3.13) and (3.14)], and the question arises whether they belong to the same irreducible representation of $SO(4, 2)$ or whether they belong to inequivalent irreducible representations of $SO(4, 2)$; in the latter case $SO(4, 2)$ would again remain irreducible under $SO(4, 1)$. These two representations of $SO(4, 1)_{01234}$ differ in the sign of k_0. Now there is no operator in $SO(4, 2)$ which changes the value (including the sign) of k_0. First of all, $S_{0\alpha}$ does not change k_0, because $S_{0\alpha}$ is a generator of $SO(4, 1)_{01234}$; then $S_{5\alpha}$ is an $SO(4)$ vector operator equivalent to[5] $S_{0\alpha}$ and consequently does not change k_0; finally, S_{50} commutes with all of $SO(4)$. Thus,

$$k_0 \quad \text{and} \quad \operatorname{sgn}(k_0) \quad (3.16)$$

is an $SO(4, 2)$ invariant. Hence for each value of $S = \frac{3}{2}, 2, \frac{5}{2}, 3, \cdots$ there are *two* inequivalent representations of $SO(4, 2)$,

$(S, \operatorname{sgn} k_0 = -1)$ with the reduction (3.13),

$(S, \operatorname{sgn} k_0 = +1)$ with the reduction (3.14),

and both of these representations remain irreducible under $SO(4, 1)$.

For $S = \frac{1}{2}$, there are two $SO(4, 1)$ representations: $(S = \frac{1}{2}, k_0 = -\frac{1}{2})$ and $(S = \frac{1}{2}, k_0 = +\frac{1}{2})$ [Eqs. (3.10)]. These two representations of $SO(4, 1)$ must extend—if they extend at all—to inequivalent representations of $SO(4, 2)$. Now in these representations the relation $a = 1 - k_0^2$ is fulfilled, so that by (3.2) we are led to a correct spectrum of Γ_0. And because Γ_0 is the only generator that lies outside $SO(4, 1)$, we have no further restrictions, from which we conclude that these two representations extend to $SO(4, 2)$ and remain irreducible under $SO(4, 1)$.

For $S = 1$, there are four $SO(4, 1)_{01234}$ representations:

$$(S = 1, k_0 = 0), \quad (S = 1, k_0 = -1),$$
$$(S = 1, k_0 = +1), \quad \begin{pmatrix} & & 1 \\ S = 1, k_0 = & 0 & \\ & & -1 \end{pmatrix}.$$

By the same argument as used for the $S = \frac{1}{2}$ case, we conclude that these representations of $SO(4, 1)$ must extend, if they extend at all, to inequivalent representations of $SO(4, 2)$. However, now the representations

$$(S = 1, k_0 = 0) \quad \text{and} \quad \begin{pmatrix} & & 1 \\ S = 1, k_0 = & 0 & \\ & & -1 \end{pmatrix}$$

do not extend to a representation of $SO(4, 2)$. This follows from (3.2) and the fact that the difference of two eigenvalues μ of Γ_0 must be an integer. For the case $a = 0$, $k_0 = 0$, we would obtain from (3.2)

$$\mu = \pm(n^2 - 1)^{\frac{1}{2}}, \quad n = 2, 3, 4, \cdots,$$

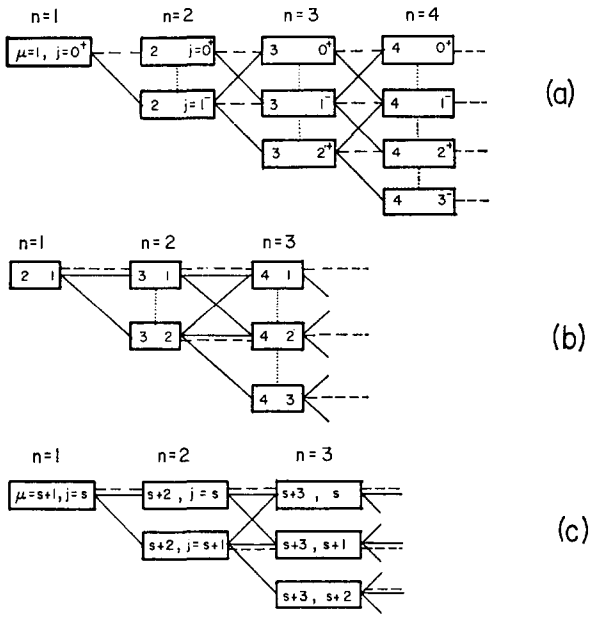

FIG. 1. Multiplicity patterns of $SO(4, 2)$ representations. Every $SO(4)$ irreducible representation (IR) that occurs in one representation of $SO(4, 2)$ is represented by a column of boxes, each box representing the $SO(3)_{S_i}$ representation which is contained in this $SO(4)$ IR. Solid lines between the boxes indicate the nonzero matrix elements of S_{5i} and S_{0i} that transform between the $SO(3)$ representations; a broken line indicates the nonzero matrix elements of S_{54} and S_{04}, and a dotted line indicates the nonzero matrix elements of S_{4i}. The first number in each box gives the eigenvalue μ of Γ_0 [Eq. (3.19)]: (a) the representation $(S = 0, k_0 = 0)^+$; (b) $(S = 1, k_0 = -1)^+$ and $(S = 1, k_0 = +1)^+$; (c) $(S, k_0 = +S)^+$ and $(S, k_0 = -S)^+$, $S = \frac{3}{2}, 2, \frac{5}{2}, \cdots$. For the IR's $(S, \operatorname{sgn} k_0)^-$ the reduction is the same as for IR's $(S, \operatorname{sgn} k_0)^+$, only μ has to be replaced by $-\mu$ in each box.

so that μ cannot change in integer steps. For the same reason also the representation

$$\begin{pmatrix} & 1 \\ S = 1, k_0 = & 0 \\ & -1 \end{pmatrix}$$

is excluded. The two remaining ones,

$$(S = 1, k_0 = -1) \quad \text{and} \quad (S = 1, k_0 = +1),$$

extend to inequivalent representations of $SO(4, 2)$ and remain irreducible under $SO(4, 1)$. There is a fourth irreducible representation of $SO(4, 1)$ for which $a = 0$; this is the 1-dimensional trivial representation $SO(4, 1) \to 1$ which has the "$SO(4)$ reduction"

$$(S = 1, k_0 = 0)^{\text{trivial}} \xrightarrow[SO(4)]{} (k_0 = 0, n = 1)$$
$$\xrightarrow[SO(3)]{} (j = 0) \xrightarrow[SO(2)]{} (j_3 = 0).$$

To summarize, in all the representations $(S, k_0 = \pm S)$, $S = 0, \frac{1}{2}, 1, \frac{3}{2}, \frac{5}{2}, \cdots$, we have

$$|k_0| = S = j_{\min} \tag{3.17}$$

and

$$1 - a = S^2 = k_0^2. \tag{3.18}$$

Each of these $SO(4, 1)$ representations not only extend to $SO(4, 2)$, but they also exactly extend to *two* inequivalent irreducible representations of $SO(4, 2)$. This additional doubling is due to the sign of Γ_0: If we insert (3.18) into (3.2), we find

$$\text{spectrum } \Gamma_0 = \mu = \pm(S + n), \quad n = 1, 2, 3, \cdots. \tag{3.19}$$

There is no operator in $SO(4, 2)$ which changes the sign of $(S + n)$ and consequently the sign of μ. Thus

$$\operatorname{sgn} \mu \tag{3.20}$$

is another invariant of our $SO(4, 2)$ representations. We denote by $(S, \operatorname{sgn} k_0)^+$ and $(S, \operatorname{sgn} k_0)^-$, the irreducible representations of $SO(4, 2)$ which contain the $(S, \operatorname{sgn} k_0)$ representation of $SO(4, 1)$ and for which $\operatorname{sgn} \mu = +1$ and $\operatorname{sgn} \mu = -1$, respectively.

For the following considerations a graphical representation of our results will be very useful (Fig. 1).

4. REDUCTION WITH RESPECT TO $SO(3, 2)_{50123}$

In this section we consider the reduction according to the chain (2.5). According to Eqs. (2.15) and (2.16), only those IR's of $SO(3, 2)$ which have the same value of the fourth- and second-order Casimir operator will occur in the reduction. However, contrary to the $SO(4, 1)$ reduction discussed in the previous section, the $SO(3, 2)$ IR's are characterized by *three* numbers, and we cannot expect that the IR's of $SO(4, 2)$ remain always irreducible also under $SO(3, 2)$.

The reduction into IR's of $SO(3, 2)$ can be easily read off from the graphical representation of Fig. 1. In $SO(3, 2)$ we do not have the operators S_{4i}, S_{45}, and S_{40}. Thus, the connection between the boxes given by broken and dotted lines does not exist for the $SO(3, 2)$ operators. Furthermore, in contrast to the $SO(4, 1)$ reductions, the representations $(S, \operatorname{sgn} k_0)^+$ and $(S, \operatorname{sgn} k_0)^-$ of $SO(4, 2)$ contain inequivalent $SO(3, 2)$ representations, differing by the sign of μ; we shall call them $(S, \operatorname{sgn} k_0)^+_{SO(3,2)}$ and $(S, \operatorname{sgn} k_0)^-_{SO(3,2)}$, respectively.

The resultant reduction is shown in Fig. 2 for the $(S, \operatorname{sgn} k_0)^+$ class. The reduction of the IR's $(S, \operatorname{sgn} k_0)^-$ is completely analogous; one has only to replace everywhere μ by $-\mu$.

To summarize: We have seen that the IR's $(S, k_0 = \pm S)^\pm$, $S = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$, of $SO(4, 2)$ characterized by (2.4) (where a can only be $a = 1 - S^2$) reduce with respect to $SO(3, 2)$ into a subclass of singleton representations of Ehrman.[7] Except for $S = 0$, they remain irreducible under $SO(3, 2)$. For $S = 0$ the

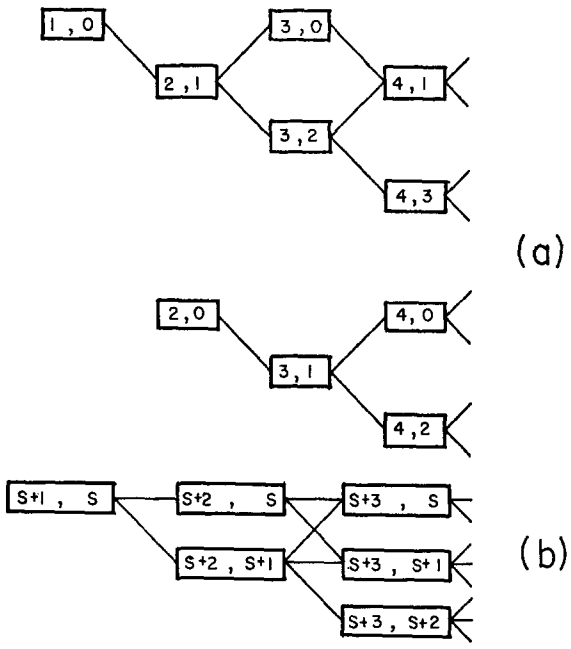

FIG. 2. Multiplicity pattern[6] of $SO(3, 2)$ representations into which the IR's of $SO(4, 2)$ reduce. (a) The representation $(S = 0, k_0 = 0)^+$ given in Fig. 1(a) reduces into *two* inequivalent IR's of $SO(3, 2)$: $(S = 0, k_0 = 0)^+ \rightarrow (S = 0, k_0 = 0, \mu_{\min} = 1) \oplus (S = 0, k_0 = 0, \mu_{\min} = 2)$. (b) The IR's $(S, k_0 = \pm S)^+$ of $SO(4, 2)$ $(S = \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots)$ remain irreducible under $SO(3, 2)$ and contain the IR of $SO(3, 2)$ shown.

representation reduces into the direct sum of two inequivalent IR's of $SO(3, 2)$ with the same value of the two Casimir operators. These singleton representations are those which reduce continually with respect to $SO(3, 1)$.[6]

5. EXTENSION BY PARITY

We now want to extend $SO(4, 2)$ by the parity P. Because $SO(3)_{S_{ij}}$ is the rotation group and $SO(3,1)_{S_{\mu\nu}}$ is the Lorentz group, we must have

$$[P, S_{ij}] = 0, \tag{5.1}$$

$$\{P, S_{0j}\} = 0, \tag{5.2}$$

and consequently [because of equivalence of the $SO(3, 1)$]

$$\{P, \Gamma_i\} = 0. \tag{5.2'}$$

If we further assume that also the S_{4i} are vectors (rather than pseudovectors)

$$\{P, S_{4i}\} = 0, \tag{5.3}$$

then we remain with the $SO(2, 1)_{S_{05}S_{04}S_{54}}$ group of scalar operators

$$[P, SO(2, 1)_{S_{03}S_{04}S_{54}}] = 0. \tag{5.4}$$

So Γ_0, $\mathbf{S}^2$, and S_3 can be simultaneously diagonalized with P. However, because of (5.3) and

$$\sum_i S_{4i}S_i = k_0(|k_0| + n), \tag{5.5}$$

it follows that P changes the sign of k_0. Consequently, only the IR's $(S = 0, k_0 = 0)^{\pm}$ can extend to representations of $SO(4, 2)$ and P.

For the IR's $(S = 0, k_0 = 0)$, the $SO(4)$ states are already P eigenstates because $k_0 = 0$. So to each box in Fig. 1(a) corresponds a definite parity, and, if we choose the P eigenvalue of the lowest state to be $+1$, we obtain the parity assignment as given in the upper right corner of each box in Fig. 1(a).

In all the other IR's $(S, k_0 = \pm S)^{\pm}$, adjoining of P leads to parity doubling: An IR of $(S, |k_0|)^{\pm}$ of $\{SO(4, 2)$ and $P\}$ reduces with respect to the proper $SO(4, 2)$ into the direct sum of two $(S, k_0)^{\pm}$:

$$(S, |k_0|)^{\pm}_{\{SO(4,2),P\}} \xrightarrow[SO(4,2)]{} (S, k_0 = S)^{\pm} \otimes (S, k_0 = -S)^{\pm}. \tag{5.6}$$

To obtain parity eigenstates we have to take the linear combinations

$$|\mu j j_3 \pm; S\rangle = |\mu j j_3; S\, k_0 = +S\rangle \pm |\mu j j_3; S, k_0 + -S\rangle. \tag{5.7}$$

6. RELATION TO BOSON FORMALISM

In this section we give an explicit form of the representations of $SO(4, 2)$ satisfying (2.4) in terms of two pairs of boson creation and annihilation operators (a_1, a_2) and (b_1, b_2). Many of the results obtained can also be explicitly verified in terms of this realization, and we establish the equivalence of the representation relation with this explicit form:

$$\begin{aligned}
S_{ij} &= \tfrac{1}{2}(a^+\sigma_k a + b^+\sigma_k b), \quad (ijk\text{-cyclic}),\\
S_{i4} &= -\tfrac{1}{2}(a^+\sigma_i a - b^+\sigma_i b),\\
S_{i0} &= -\tfrac{1}{2}(a^+\sigma_i C b^+ - aC\sigma_i b),\\
S_{i5} &= \tfrac{1}{2}i(a^+\sigma_i C b^+ + aC\sigma_i b),\\
S_{45} &= \tfrac{1}{2}(a^+ C b^+ + aCb),\\
S_{40} &= \tfrac{1}{2}i(a^+ C b^+ - aCb),\\
S_{05} &= \tfrac{1}{2}(a^+a + b^+b + 2),
\end{aligned} \tag{6.1}$$

where σ_i are the Pauli matrices and C the antisymmetric matrix

$$C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

These operators act on the states in the $SO(3)^{(1)} \otimes SO(3)^{(2)}$ basis of $SO(4, 1)$:

$$|j_1 m_1 j_2 m_2\rangle = N a_1^{+j_1+m_1} a_2^{+j_1-m_1} b_1^{+j_2+m_2} b_2^{+j_2-m_2} |0\rangle,$$
$$N^{-2} = (j_1 + m_1)!\,(j_1 - m_1)!\,(j_2 + m_2)!\,(j_2 - m_2)!. \tag{6.2}$$

The operator, which has the eigenvalues k_0 and which is an $SO(4, 2)$ invariant [see Eqs. (3.1) and

(3.16)], is represented by

$$K_0 = \tfrac{1}{2}(a^+a - b^+b), \tag{6.3}$$

has eigenvalues $k_0 = j_1 = j_2$, and commutes with all S_{AB}.

Parity operator (see Sec. 5) is represented by

$$P: a^+ \to b^+, \quad b^+ \to -a^+ \tag{6.4}$$

and changes the sign of k_0.

The representation with the opposite sign of the eigenvalues of Γ_0 can be obtained by replacing

$$j_1 \to -1 - j_1, \quad j_2 \to -1 - j_2 \tag{6.5}$$

and applying P.

From (6.1), we obtain after some calculation, e.g.,

$$\begin{aligned} S_{A5}S^A{}_5 &= \tfrac{1}{4}[(a^+a)^2 + (b^+b)^2 - 2a^+ab^+b - 4] \\ &= -S_{A4}S^A{}_4, \end{aligned} \tag{6.6}$$

which on states (6.2) gives immediately the value $(k_0^2 - 1)$. By symmetry, the same is true for the sum of squares of all other rows (or columns) of the matrix (S_{AB}).

Similarly, we find, after some calculations,

$$\{\Gamma_\mu^{(1)}, \Gamma^{\mu(2)}\} = 0, \tag{6.7}$$

where $\Gamma_\mu^{(1)} = S_{5\mu}$ and $\Gamma_\mu^{(2)} = S_{4\mu}$, and by symmetry the remaining equations of (2.4) are satisfied. These "oscillator-like" representations have been discussed in different forms in Refs. 8–11.

7. GUIDE TO APPLICATIONS

The representation ($S = 0$, $k_0 = 0$) is realized in the conformal interpretation of $O(4, 2)$ to describe massless spin-zero particles[12] and, in the dynamical group interpretation of $O(4, 2)$, to describe the rest frame states of H-atom[9] and mesons.[13] The existence of Γ_μ is crucial in the calculation of transition probabilities. The representation ($S = \frac{1}{2}, k_0 = \pm\frac{1}{2}$) has been used in the $O(4, 2)$ hadron model and accounts for the dipole form factor of the proton.[10] The representations ($S, k_0 = \pm S$) occur in the dyonium model, an atom formed out of two-spinless particles having both electric and magnetic charges.[11] Matrix elements for some finite transformations for these representations have been given in Refs. 9–11.

ACKNOWLEDGMENTS

One of us (A. O. B.) is grateful to Center for Particle Theory at the University of Texas for its hospitality, where part of this work was done. We should like to thank F. Schwarz for valuable suggestions and for the writing of the Appendix.

APPENDIX

In this appendix we give explicit expressions for the infinitesimal generators. Clearly, it is sufficient to give the generators $S_{i+1,i}$, $i = 1, \cdots, 5$, because the other generators can be determined from these by the relation (2.2). We take the generators S_{21}, S_{32}, and S_{43} from[4]

$$S_{21}\,|a; k_0, n, j, j_3\rangle = j_3\,|a; k_0, n, j, j_3\rangle, \tag{A1}$$

$$\begin{aligned} S_{32}\,|a; k_0, n, j, j_3\rangle &= \tfrac{1}{2}[(j + j_3 + 1)(j - j_3)]^{\frac{1}{2}} \\ &\quad \times |a; k_0, n, j, j_3 + 1\rangle \\ &\quad + \tfrac{1}{2}[(j + j_3)(j - j_3 + 1)]^{\frac{1}{2}} \\ &\quad \times |a; k_0, n, j, j_3 - 1\rangle, \end{aligned} \tag{A2}$$

$$\begin{aligned} S_{43}\,|a; k_0, n, j, j_3\rangle &= [(j + j_3 + 1)(j - j_3 + 1)]^{\frac{1}{2}} \\ &\quad \times C_{j+1}\,|a; k_0, n, j + 1, j_3\rangle \\ &\quad + j_3 A_j\,|a; k_0, n, j, j_3\rangle \\ &\quad - [(j - j_3)(j + j_3)]^{\frac{1}{2}} \\ &\quad \times C_j\,|a; k_0, n, j - 1, j_3\rangle. \end{aligned} \tag{A3}$$

Here

$$A_j = \frac{k_0(|k_0| + n)}{j(j + 1)},$$

$$C_j = \frac{1}{j}\left(\frac{(j^2 - k_0^2)[j^2 - (|k_0| + n)^2]}{4j^2 - 1}\right)^{\frac{1}{2}}.$$

The expression S_{54} we take from[14]

$$\begin{aligned} S_{54}\,|a; k_0, n, j, j_3\rangle &= \tfrac{1}{2}a(k_0, n)[(k_0 + j + 1)(k_0 - j)]^{\frac{1}{2}} \\ &\quad \times |a; k_0 + 1, n, j, j_3\rangle \\ &\quad + \tfrac{1}{2}b(k_0, n)[(j - n)(j + n + 1)]^{\frac{1}{2}} \\ &\quad \times |a; k_0, n + 1, j, j_3\rangle \\ &\quad - \tfrac{1}{2}c(k_0, n)[(j + n)(j - n + 1)]^{\frac{1}{2}} \\ &\quad \times |a; k_0, n - 1, j, j_3\rangle \\ &\quad - \tfrac{1}{2}d(k_0, n)[(k_0 + j)(k_0 - j - 1)]^{\frac{1}{2}} \\ &\quad \times |a; k_0 - 1, n, j, j_3\rangle. \end{aligned} \tag{A4}$$

Here

$$a(k_0, n) = \left(\frac{(k_0 - r)(k_0 + r + 1)[k_0(k_0 + 1) + \sigma]}{(k_0 - n)(k_0 - n + 1)(k_0 + n)(k_0 + n + 1)}\right)^{\frac{1}{2}},$$

$$b(k_0, n) = \left(\frac{(r - n)(r + n + 1)[n(n + 1) + \sigma]}{(k_0 - n - 1)(k_0 - n)(k_0 + n)(k_0 + n + 1)}\right)^{\frac{1}{2}},$$

$$c(k_0, n) = -b(k_0, n - 1), \quad d(k_0, n) = a(k_0 - 1, n).$$

The matrix elements can also be derived easily using Eqs. (6.1) and (6.2); the basis $|j_1 m_1 j_2 m_2\rangle$ can be transformed to the $|jj_3\rangle$ basis by $3j$-symbols. r and σ are

connected with the Casimir operators through

$$+r(r+1)+2+\sigma = 2a,$$
$$+r(r+1)\sigma = a(a-1).$$

For all the representations considered in this paper we have $r = S = |k_0| = +(1-a)^{\frac{1}{2}}$. $S_{50} = \Gamma_0$ is given by (3.2):

$$S_{50}\,|a_v; k_0 n_v; jj_3\rangle = \pm[(a-1+k_0^2)+(|k_0|+n)^2]^{\frac{1}{2}}\,|a_0; k_0 n_0; jj_3\rangle. \tag{A5}$$

* Supported in part by the Air Force Office of Scientific Research under Grant AF-AFOSR-30-67.

† Supported in part by the National Science Foundation.

[1] The characterization of a representation by a relation of the form (2.4) is very convenient. For Dirac representation of $SO(4, 2)$, for example, we have $S_{34} = -2S_{05}S_{12}$ or $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}$. Equation (2.4) is also a generalization of the relation $\{J_{ij}, J^i{}_k\} = ag_{jk}$ for $SU(2)$, which by virtue of the spectrum of J_3 gives only the 2-dimensional representation for $a = \frac{1}{2}$.

[2] T. D. Newton, Ann. Math. **51**, 730 (1950).

[3] T. Dixmier, Bull. Soc. Math. France **89**, 9 (1961).

[4] For example, A. Böhm, Nuovo Cimento **43**, 665 (1966), Appendix A. We adopt the convention that k_0 can be positive and negative and $c = k_0 + n$ positive only, so that $k_0 = j_1 - j_2$ and $|k_0| + n = j_1 + j_2 + 1$, where j_1 and j_2 are the characters of the $SO(3)^{(i)}$ in the Lie algebra decomposition $SO(4) = SO(3)^{(1)} \times SO(3)^{(2)}$.

[5] We can repeat the same considerations as above for the $SO(4, 1)_{51234}$ subgroup, replacing everywhere $S_{0\alpha}$ by S_{56}, and are led to the same conclusions for $S_{5\alpha}$ that we have for $S_{0\alpha}$.

[6] The irreducible representations of $SO(3, 1)_{0123}$ are characterized by two numbers (k_0', c) [M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964)] so that $\frac{1}{2}S_{\mu\nu}S^{\mu\nu} = c^2 + k_0'^2 - 1$ and $\frac{1}{8}\epsilon_{\mu\nu\lambda\delta}S^{\mu\nu}S^{\lambda\rho} = k_0' c$. Hence, from (2.21), spectrum $\Gamma_4{}^2 = -a + 1 - c^2 - k_0'^2$. Because k_0', is the lowest spin, $k_0' = k_0$. Hence, $k_0'^2 = 1 - a^2$ and therefore spectrum $\Gamma_4{}^2 = -c^2$. Because Γ_4 is a noncompact generator of a $SO(2, 1)$ subgroup, we know from general theorems that $-\infty <$ spectrum $\Gamma_4 < \infty$. Then spectrum $\Gamma_4{}^2 \geq 0$, and consequently $c^2 < 0$. Thus c is pure imaginary, and only the principal series representations (k_0', c) of $SO(3, 1)_{\mu\nu}$ appear in the reduction of our $SO(4, 2)$ representations with respect to $SO(3, 1)$. The spectrum of (ic) and therefore the reduction of these $SO(4, 2)$ representations with respect to $SO(3, 1)$ are continuous.

[7] J. B. Ehrman, Thesis, Princeton, 1954; in particular, Figs. 7-5, 7-7, and 7-15.

[8] J. Olszewski, Acta Phys. Polonica **30**, 105 (1966). See also I. Todorov, *Proceedings of the Coral Gables Conference on Fundamental Interactions* (Gordon & Breach, New York, 1969), Vol. I.

[9] A. Malkin and V. I. Man'ko, ZhETF Pis. Red. [JETP Lett.] **2**, 146 (1966); A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967); **157**, 1180 (1967); **160**, 1149 (1967); C. Fronsdal, *ibid.* **156**, 1665 (1967); Y. Nambu, Progr. Theoret. Phys. (Kyoto) Suppl. **37**, 368 (1966).

[10] A. O. Barut and H. Kleinert, Phys. Rev. **161**, 1464 (1967); H. Kleinert, *ibid.* **163**, 1807 (1967); A. O. Barut, D. Corrigan, and H. Kleinert, Phys. Rev. Letters **20**, 167 (1968); Phys. Rev. **167**, 1527 (1968); A. O. Barut, Phys. Letters **26B**, 308 (1968).

[11] A. O. Barut, in *Proceedings of the Coral Gables Conference on Fundamental Interactions* (Gordon & Breach, New York, 1970); Phys. Rev. (to be published).

[12] See, e.g., G. Mack and I. Todorov, ICTP, IC/68/86, Trieste, Italy; L. Gross, J. Math. Phys. **5**, 687 (1964).

[13] A. O. Barut, Nucl. Phys. **B4**, 455 (1968); A. O. Barut and K. C. Tripathy, Nucl. Phys. **B7**, 125 (1968).

[14] S. Ström, Arkiv Fysik **30**, 455 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Approach to Equilibrium in a Simple Model

CHARLES RADIN
Department of Physics and Astronomy, The University of Rochester
Rochester, New York 14627

(Received 13 March 1970)

The time evolution of a class of generalized quantum Ising models (with various long-range interactions, including Dyson's $1/r^\alpha$) has been studied from the C^*-algebraic point of view. We establish that: (1) All $\langle A\rangle_t$ are weakly almost periodic in time; (2) there exists a unique averaging procedure over time; (3) the time evolution in the thermodynamical limit can be locally implemented by effective Hamiltonians in the algebra of quasilocal observables; (4) there exists a specific connection between the spectral properties of the time evolution of the initial state and the approach to equilibrium; (5) there are examples in which the time evolution is not G-Abelian.

1. INTRODUCTION

A general class of Ising-type 2-body interactions on an infinite lattice of spins is considered, with the time behavior of the model being the object of study. It is shown that there is a canonical time average on the states of the system which gives a manageable prescription for determining the equilibria for time-developing states. Three subclasses of models are then singled out for a detailed study of the approach to equilibrium which they produce—finite-range interactions and infinite-range interactions decreasing polynomially and exponentially with respect to distance. With the results of a free-induction relaxation experiment in mind, a class of physically significant states is studied, yielding the corresponding equilibria and the detailed rate of approach to

connected with the Casimir operators through

$$+r(r+1)+2+\sigma = 2a,$$
$$+r(r+1)\sigma = a(a-1).$$

For all the representations considered in this paper we have $r = S = |k_0| = +(1-a)^{\frac{1}{2}}$. $S_{50} = \Gamma_0$ is given by (3.2):

$$S_{50}\,|a_v; k_0 n_v; jj_3\rangle = \pm[(a-1+k_0^2)+(|k_0|+n)^2]^{\frac{1}{2}}\,|a_0; k_0 n_0; jj_3\rangle. \tag{A5}$$

* Supported in part by the Air Force Office of Scientific Research under Grant AF-AFOSR-30-67.

† Supported in part by the National Science Foundation.

[1] The characterization of a representation by a relation of the form (2.4) is very convenient. For Dirac representation of $SO(4, 2)$, for example, we have $S_{34} = -2S_{05}S_{12}$ or $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}$. Equation (2.4) is also a generalization of the relation $\{J_{ij}, J^i{}_k\} = ag_{jk}$ for $SU(2)$, which by virtue of the spectrum of J_3 gives only the 2-dimensional representation for $a = \frac{1}{2}$.

[2] T. D. Newton, Ann. Math. **51**, 730 (1950).

[3] T. Dixmier, Bull. Soc. Math. France **89**, 9 (1961).

[4] For example, A. Böhm, Nuovo Cimento **43**, 665 (1966), Appendix A. We adopt the convention that k_0 can be positive and negative and $c = k_0 + n$ positive only, so that $k_0 = j_1 - j_2$ and $|k_0| + n = j_1 + j_2 + 1$, where j_1 and j_2 are the characters of the $SO(3)^{(i)}$ in the Lie algebra decomposition $SO(4) = SO(3)^{(1)} \times SO(3)^{(2)}$.

[5] We can repeat the same considerations as above for the $SO(4, 1)_{51234}$ subgroup, replacing everywhere $S_{0\alpha}$ by S_{56}, and are led to the same conclusions for $S_{5\alpha}$ that we have for $S_{0\alpha}$.

[6] The irreducible representations of $SO(3, 1)_{0123}$ are characterized by two numbers (k_0', c) [M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964)] so that $\frac{1}{2}S_{\mu\nu}S^{\mu\nu} = c^2 + k_0'^2 - 1$ and $\frac{1}{8}\epsilon_{\mu\nu\lambda\delta}S^{\mu\nu}S^{\lambda\rho} = k_0' c$. Hence, from (2.21), spectrum $\Gamma_4{}^2 = -a + 1 - c^2 - k_0'^2$. Because k_0', is the lowest spin, $k_0' = k_0$. Hence, $k_0'^2 = 1 - a^2$ and therefore spectrum $\Gamma_4{}^2 = -c^2$. Because Γ_4 is a noncompact generator of a $SO(2, 1)$ subgroup, we know from general theorems that $-\infty <$ spectrum $\Gamma_4 < \infty$. Then spectrum $\Gamma_4{}^2 \geq 0$, and consequently $c^2 < 0$. Thus c is pure imaginary, and only the principal series representations (k_0', c) of $SO(3, 1)_{\mu\nu}$ appear in the reduction of our $SO(4, 2)$ representations with respect to $SO(3, 1)$. The spectrum of (ic) and therefore the reduction of these $SO(4, 2)$ representations with respect to $SO(3, 1)$ are continuous.

[7] J. B. Ehrman, Thesis, Princeton, 1954; in particular, Figs. 7-5, 7-7, and 7-15.

[8] J. Olszewski, Acta Phys. Polonica **30**, 105 (1966). See also I. Todorov, *Proceedings of the Coral Gables Conference on Fundamental Interactions* (Gordon & Breach, New York, 1969), Vol. I.

[9] A. Malkin and V. I. Man'ko, ZhETF Pis. Red. [JETP Lett.] **2**, 146 (1966); A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967); **157**, 1180 (1967); **160**, 1149 (1967); C. Fronsdal, *ibid.* **156**, 1665 (1967); Y. Nambu, Progr. Theoret. Phys. (Kyoto) Suppl. **37**, 368 (1966).

[10] A. O. Barut and H. Kleinert, Phys. Rev. **161**, 1464 (1967); H. Kleinert, *ibid.* **163**, 1807 (1967); A. O. Barut, D. Corrigan, and H. Kleinert, Phys. Rev. Letters **20**, 167 (1968); Phys. Rev. **167**, 1527 (1968); A. O. Barut, Phys. Letters **26B**, 308 (1968).

[11] A. O. Barut, in *Proceedings of the Coral Gables Conference on Fundamental Interactions* (Gordon & Breach, New York, 1970); Phys. Rev. (to be published).

[12] See, e.g., G. Mack and I. Todorov, ICTP, IC/68/86, Trieste, Italy; L. Gross, J. Math. Phys. **5**, 687 (1964).

[13] A. O. Barut, Nucl. Phys. **B4**, 455 (1968); A. O. Barut and K. C. Tripathy, Nucl. Phys. **B7**, 125 (1968).

[14] S. Ström, Arkiv Fysik **30**, 455 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Approach to Equilibrium in a Simple Model

CHARLES RADIN
Department of Physics and Astronomy, The University of Rochester
Rochester, New York 14627

(Received 13 March 1970)

The time evolution of a class of generalized quantum Ising models (with various long-range interactions, including Dyson's $1/r^\alpha$) has been studied from the C^*-algebraic point of view. We establish that: (1) All $\langle A\rangle_t$ are weakly almost periodic in time; (2) there exists a unique averaging procedure over time; (3) the time evolution in the thermodynamical limit can be locally implemented by effective Hamiltonians in the algebra of quasilocal observables; (4) there exists a specific connection between the spectral properties of the time evolution of the initial state and the approach to equilibrium; (5) there are examples in which the time evolution is not G-Abelian.

1. INTRODUCTION

A general class of Ising-type 2-body interactions on an infinite lattice of spins is considered, with the time behavior of the model being the object of study. It is shown that there is a canonical time average on the states of the system which gives a manageable prescription for determining the equilibria for time-developing states. Three subclasses of models are then singled out for a detailed study of the approach to equilibrium which they produce—finite-range interactions and infinite-range interactions decreasing polynomially and exponentially with respect to distance. With the results of a free-induction relaxation experiment in mind, a class of physically significant states is studied, yielding the corresponding equilibria and the detailed rate of approach to

equilibrium for each type of interaction. Finally, the time behavior of each model is linked to specific spectral properties of the corresponding effective Hamiltonians.

The framework for the models is that of the C^*-algebra approach. The theories of invariant means on groups and asymptotic probability distributions are also employed.

2. THE GENERALIZED ISING MODEL

At each site i in a ν-dimensional lattice $\mathbf{Z}^\nu$, associate a 2-dimensional complex Euclidean space $\mathbf{C}_i^2$ (spin space). Let F be the set of all finite subsets of $\mathbf{Z}^\nu$. Then, for each finite volume $V \in F$, consider the direct product space $\otimes_{i\in V} \mathbf{C}_i^2$ of Murray and von Neumann,[1] and let $\mathfrak{A}(V)$ be defined as $B(\otimes_{i\in V} \mathbf{C}_i^2)$, the set of all bounded operators on $\otimes_{i\in V} \mathbf{C}_i^2$. $\mathfrak{A}(V)$ is the set of observables pertinent to the volume V. It is a concrete C^*-algebra with respect to the usual operator norm, denoted $\| \ \|$, and adjoint, denoted *. For any two volumes V, $V' \in F$, satisfying $V \subset V'$, there is a natural mathematical way to imbed $\mathfrak{A}(V)$ in $\mathfrak{A}(V')$—symbolically, for $A \in \mathfrak{A}(V)$ let $\bar{A} = A \otimes_{i\in V'/V} I_i \in \mathfrak{A}(V')$, where I_i is the identity operator in $\mathfrak{A}(i)$ and $V'/V = \{j \in \mathbf{Z}^\nu \mid j \in V', j \notin V\}$. It is easy to check that this imbedding is norm preserving and, in fact, a $*$-isomorphism of the C^*-algebra $\mathfrak{A}(V)$ onto a subalgebra of $\mathfrak{A}(V')$. To obtain all local observables, we follow the prescription of Takeda,[2] which essentially involves the construction of a $*$-algebra $\mathfrak{A}^0$ defined as the union $\bigcup_{V\in F} \mathfrak{A}(V)$ with "equivalent" elements identified. This normed algebra is not complete, but upon completion it is a C^*-algebra, denoted $\mathfrak{A}$, consisting of the so-called quasilocal observables.[3] We denote by $\mathfrak{S}$ the set of all states on $\mathfrak{A}$.

To simplify notation throughout this paper, we identify $\mathfrak{A}(V)$ with its image in $\mathfrak{A}^0$ or $\mathfrak{A}$ and also with the matrix algebra $GL(2^{N(V)}, \mathbf{C})$, where $N(V)$ is the number of sites in V. For example, $\sigma_x^i \in \mathfrak{A}(i) \subset \mathfrak{A}^0 \subset \mathfrak{A}$ for the Pauli matrix σ_x.

We now turn our attention to the dynamics of the system.[4] With each $V \in F$, we associate an energy observable $H_V \in \mathfrak{A}(V)$ defined by

$$H_V = \tfrac{1}{2} \sum_{(j,k)\in V\times V} \epsilon_{jk}\sigma_z^j\sigma_z^k,$$

where $V \times V$ is the Cartesian product of V with itself. To make H_V self-adjoint, we require that ϵ_{jk} be real; for homogeneity and isotropy we require that ϵ_{jk} be a function only of the Euclidean distance $|j-k|$ between j and k, i.e., $\epsilon_{jk} = \epsilon(|j-k|)$. To avoid self-interaction, we assume that $\epsilon_{jj} = 0$ and, for stability, we require that the total energy at any site due to interaction with the entire lattice be finite, i.e., $\sum_{j\in\mathbf{Z}^\nu} |\epsilon(|j|)| < \infty$. We call this the generalized Ising model (GIM).

For $A \in \mathfrak{A}^0$, we can give the dynamics as follows. Consider $\alpha_t^V(A) = \exp(iH_Vt)A \exp(-iH_Vt)$. It is easy to see that the α_t^V are $*$-automorphisms of $\mathfrak{A}^0$ and form a group with the multiplication

$$(\alpha_{t_1}^V \cdot \alpha_{t_2}^V)(A) = \alpha_{t_1}^V(\alpha_{t_2}^V(A)) = \alpha_{t_1+t_2}^V(A),$$

i.e.,

$$\alpha_{t_1}^V\alpha_{t_2}^V = \alpha_{t_1+t_2}^V.$$

Clearly, we need to take the infinite volume limit to get the full dynamics. Therefore, for a local observable $A \in \mathfrak{A}(V_1)$, consider $\alpha_t^{V_2}(A)$ with $V_2 \supset V_1$. By Magnus' formula,[5] we have

$$\begin{aligned}\alpha_t^{V_2}(A) &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} [H_{V_2}, [\cdots, [H_{V_2}, A]\cdots]] \\ &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \sum_{(j_1,k_1)\in V_2\times V_2} \cdots \sum_{(j_n,k_n)\in V_2\times V_2} \\ &\quad \times [\Phi_{j_nk_n}, [\cdots, [\Phi_{j_1k_1}, A]\cdots]],\end{aligned}$$

where $\Phi_{jk} = \frac{1}{2}\epsilon_{jk}\sigma_z^j\sigma_z^k$ and $[\ ,\ ]$ denotes the commutator. It is clear that, if j_1 and $k_1 \in V_2/V_1$, then $[\Phi_{j_1k_1}, A] = 0$. Therefore, we may restrict the relevant summation index to $(j_1, k_1) \in V_2 \times V_2/[(V_2/V_1) \times (V_2/V_1)]$. For any two subsets W_1 and W_2 of $\mathbf{Z}^\nu$, let $W_2 \div W_1$ denote $W_2 \times W_2/[(W_2/W_1) \times (W_2/W_1)]$, a subset of $\mathbf{Z}^\nu \times \mathbf{Z}^\nu$. Now consider $[\Phi_{j_1k_1}, A]$ in more detail. This operator can have at most σ_z's at the sites outside V_1. Therefore, if $(j_2, k_2) \in (V_2/V_1) \times (V_2/V_1)$, then $[\Phi_{j_2k_2}, [\Phi_{j_1k_1}, A]] = 0$. By induction, we see that we can restrict all n summation indices to $(j_i, k_i) \in V_2 \div V_1$. Now we bring these summation symbols back inside the brackets to get

$$\alpha_t^{V_2}(A) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \Bigg[\sum_{(j_n,k_n)\in V_2\div V_1} \Phi_{j_nk_n}, \Bigg[\cdots, \Bigg[\sum_{(j_1,k_1)\in V_2\div V_1} \Phi_{j_1k_1}, A\Bigg]\cdots\Bigg]\Bigg].$$

At this point, we take the infinite volume limit since, by the stability condition, the net $\sum_{(j,k)\in V_2\div V_1} \Phi_{jk}$ has a norm limit in $\mathfrak{A}$ as $V_2 \to \infty$, namely $\tilde{H}_{V_1} \equiv \sum_{(j,k)\in\mathbf{Z}^\nu\div V_1} \Phi_{jk}$. By Magnus' formula, we get

$$\operatorname{norm-}\lim_{V_2\to\infty} \alpha_t^{V_2}(A) \equiv \alpha_t(A) = \exp(i\tilde{H}_{V_1}t)A \times \exp(-i\tilde{H}_{V_1}t).$$

To see that $\{\alpha_t^V(A) \mid V \in F\}$ is Cauchy for all $A \in \mathfrak{A}$, we use the inequality

$$\begin{aligned}\|\alpha_t^{V'}(A) - \alpha_t^V(A)\| &\le \|\alpha_t^{V'}(A_0) - \alpha_t^V(A_0)\| \\ &\quad + \|\alpha_t^{V'}(A - A_0)\| + \|\alpha_t^V(A_0 - A)\|.\end{aligned}$$

Taking $A_0 \in \mathfrak{A}^0$ and using $\|\alpha_t^V\| = 1$ gives the result. At this point, it is easy to show that the α_t form a group of $*$-automorphisms of $\mathfrak{A}$. The fact that $\|\alpha_t(A) - \alpha_{t_0}(A)\| \underset{t \to t_0}{\to} 0$ for all $A \in \mathfrak{A}$ is easily checked (on $\mathfrak{A}^0$ first). We collect our results up to this point as

Proposition 1: If

$$H_V = \tfrac{1}{2} \sum_{(j,k) \in V \times V} \epsilon(|j-k|)\sigma_z^j \sigma_z^k, \qquad \epsilon(|j-k|)$$

is real, $\epsilon(|0|) = 0$, and $\sum_{j \in \mathbf{Z}^\nu} |\epsilon(|j|)| < \infty$, then, for all $A \in \mathfrak{A}$, the net $\alpha_t^V(A) = e^{iH_V t} A e^{-iH_V t}$ has a norm limit in $\mathfrak{A}$ as $V \to \infty$, denoted $\alpha_t(A)$. The set $\{\alpha_t \mid t \in \mathbf{R}\}$ forms a strongly continuous group of $*$-automorphisms of $\mathfrak{A}$ satisfying $\alpha_{t_1}\alpha_{t_2} = \alpha_{t_1+t_2}$. Furthermore, for any

$$A \in \mathfrak{A}(V), \quad \alpha_t(A) = e^{i\tilde{H}_V t} A e^{-i\tilde{H}_V t},$$

where

$$\tilde{H}_V \equiv \tfrac{1}{2} \sum_{(j,k) \in \mathbf{Z}^\nu \div V} \epsilon(|j-k|)\sigma_z^j \sigma_z^k.$$

As an application of this proposition, we obtain

$$\alpha_t(\sigma_x^j) = \sigma_x^j \cos(2P_j t) - \sigma_y \sin(2P_j t), \tag{1}$$

$$\alpha_t(\sigma_y^j) = \sigma_y^j \cos(2P_j t) + \sigma_x^j \sin(2P_j t), \tag{2}$$

where

$$P_j = \sum_{k \in \mathbf{Z}^\nu} \epsilon(|k-j|)\sigma_z^k.$$

3. EQUILIBRIUM

We now have a time development and wish to investigate the approach to equilibrium. A first step in this direction is to answer the following question: Given a nonequilibrium state ρ on $\mathfrak{A}$, what should be the corresponding equilibrium state $\tilde{\rho}$? A useful tool for investigating this problem is contained in a paper of Emch, Knops, and Verboven.[6] We first give some necessary background.

Let G be a topological group and define the normed linear space $CB(G)$ as the set of all bounded, continuous, complex-valued functions on G, with pointwise addition and scalar multiplication and sup norm. A mean on $CB(G)$ is by definition a linear form M on $CB(G)$ which satisfies

(i) $M(\bar{f}) = \overline{M(f)}$, where the overbar denotes complex conjugation for all f in $CB(G)$,

(ii) $\inf_{x \in G} |f(x)| \leq M(f) \leq \sup_{x \in G} |f(x)|$ for all real-valued f in $CB(G)$.

This is clearly a mathematical translation of the heuristic concept that M averages over the group.[7] M is called a left invariant mean if $M(L_y f) = M(f)$ for all $y \in G$ and all $f \in CB(G)$, where the translation $L_y f \in CB(G)$ is defined by $L_y f(x) = f(yx)$. If $\mathfrak{A}$ is a C^*-algebra with unit and if $\{\alpha_g \mid g \in G\}$ is a strongly continuous[8] group of $*$-automorphisms of $\mathfrak{A}$, then,[9] for any left invariant mean M on $CB(G)$ and state ρ on $\mathfrak{A}$, the form $M\rho$ defined on $\mathfrak{A}$ by $M\rho(A) = M[\rho(\alpha_x A)]$ is a state on $\mathfrak{A}$, invariant in the sense that $M\rho(\alpha_x A) = M\rho(A)$, for all $A \in \mathfrak{A}$, $x \in G$. Using $G = \mathbf{R}$ interpreted as time development, we see that $M\rho$ is a time average of the states $\alpha_t^* \rho$ defined by $\alpha_t^* \rho(A) = \rho(\alpha_t A)$. Therefore, each invariant mean[10] M could be used to project a given state onto possibly different equilibria. Since there are[11] many invariant means on $CB(\mathbf{R})$, the question of the uniqueness of this prescription arises. We now wish to investigate this question in the case of the GIM. To do so, we need some definitions.

The class $AP(\mathbf{R})$ of almost-periodic functions on the real line can be defined as the subset of $CB(\mathbf{R})$ of all f such that the set of translates $\{L_t f \mid t \in \mathbf{R}\}$ is precompact in the norm topology. The set of weakly almost-periodic functions $W(\mathbf{R})$ consists of the subset of $CB(\mathbf{R})$ of all f such that $\{L_t f \mid t \in \mathbf{R}\}$ is precompact in the weak topology. Since the weak topology is weaker than the norm topology, $AP(\mathbf{R}) \subset W(\mathbf{R})$. $W(\mathbf{R})$ plays an important role in the theory of invariant means since[12] all invariant means on $CB(\mathbf{R})$ coincide on the subspace $W(\mathbf{R})$; furthermore, they can be taken in the form

$$Mf = \lim_{T \to \infty} T^{-1} \int_0^T f(t)\, dt.$$

In this connection we now prove:

Proposition 2: In the GIM, $\rho[\alpha_t(A)]$ is a weakly almost-periodic function of $t \in \mathbf{R}$, for all $A \in \mathfrak{A}$ and $\rho \in \mathfrak{S}$.

Proof: First, let $A' = \sigma_{l_1}^{j_1} \cdots \sigma_{l_n}^{j_n}$, where $l_k = x$, y, or z, and all the j_k are distinct sites of the lattice. Then $\alpha_t(A') = \alpha_t(\sigma_{l_1}^{j_1}) \cdots \alpha_t(\sigma_{l_n}^{j_n})$. Using

$$\cos(B) = \tfrac{1}{2}[\exp(iB) + \exp(-iB)]$$

and

$$\sin(B) = [\exp(iB) - \exp(-iB)]/2i$$

in (1) and (2), where $B \in \mathfrak{A}$, we put $\alpha_t(A')$ in the form of a finite linear sum of terms such as

$$\exp(iQ_1 t)\sigma_{l_1}^{j_1} \cdots \exp(iQ_n t)\sigma_{l_n}^{j_n},$$

where, for $l_k = z$, $Q_k = 0$ and, otherwise, $Q_k = \pm 2P_{l_k}$. We move the exponentials to the right by noticing that

$$\exp(iQ_m t)\sigma_{l_k}^{j_k} = \sigma_{l_k}^{j_k} \exp(iR_m t),$$

where $R_m = \sigma_{l_k}^{j_k} Q_m \sigma_{l_k}^{j_k}$. Note that, independently of l_k, R_m only has σ_z's in it, i.e., $R_m = \sum_j a_j \sigma_z^j$ (a finite sum with $a_j \in \mathbf{R}$). Therefore, we also see that all the

R_m are self-adjoint and commute. By moving all the exponentials to the right in this way, we obtain $\alpha_t(A')$ as a finite linear sum of terms of the form $\sigma_{l_1}^{j_1} \cdots \sigma_{l_n}^{j_n} \exp(iSt)$ with S self-adjoint. If $\rho \in \mathfrak{S}$, by considering the GNS representation Π_ρ associated with ρ, with cyclic vector Φ_ρ in $\mathcal{H}_\rho$, we see that $\rho[\alpha_t(A')]$ is a finite linear sum of functions of t of the form

$$(\Phi_\rho, \Pi_\rho[\sigma_{l_1}^{j_1}] \cdots \Pi_\rho[\sigma_{l_n}^{j_n}] \exp(i\Pi_\rho[S]t)\Phi_\rho)$$

which, upon taking adjoints, becomes

$$(\Psi_\rho, \exp(i\Pi_\rho[S]t)\Phi_\rho).$$

But

$$(\Psi_\rho, \exp(i\Pi_\rho[S]t)\Phi_\rho) \in W(\mathbf{R}),$$

since[13] $\{\exp(i\Pi_\rho[S]t) \mid t \in \mathbf{R}\}$ is precompact in the weak operator topology. Therefore, $\rho[\alpha_t(A')] \in W(\mathbf{R})$, since $W(\mathbf{R})$ is a linear space. For an A'' equal to a finite linear sum of A's of the above form, the same result follows by linearity again. For arbitrary $A \in \mathfrak{A}$, take $A_n \to A$, as $n \to \infty$ (norm topology), with A_n of the latter form. Then we have

$$\|\alpha_t(A_n) - \alpha_t(A)\| = \|\alpha_t(A_n - A)\| = \|A_n - A\|.$$

Hence, $\rho[\alpha_t(A_n - A)] \to 0$, as $n \to \infty$, uniformly in $t \in \mathbf{R}$. Therefore, $\rho[\alpha_t(A)]$ is the limit of a sequence $\rho[\alpha_t(A_n)]$ of functions in $W(\mathbf{R})$, converging in the sup norm. Hence, $\rho[\alpha_t(A)] \in W(\mathbf{R})$, since[14] $W(\mathbf{R})$ is a closed subspace of $CB(\mathbf{R})$. QED

Note that a shorter proof of Proposition 2, which does not rely on the local implementation of Proposition 1, can be obtained by observing that

$$\begin{aligned} \exp(iH_V t)A' \exp(-iH_V t) &= A' \exp(iA'H_V A't) \\ &\quad \times \exp(-iH_V t) \\ &= A' \exp(iA'H_V A't \\ &\quad - iH_V t). \end{aligned}$$

However, parts of the given proof are needed below.

We have proven that a canonical time average[15] $M\rho$ exists for every initial state ρ. We use "canonical" to emphasize the uniqueness of M. Given this association between states and equilibrium states, we consider the following question: Can one ensure that $M\rho'$ will be "close" to $M\rho$ by taking ρ' sufficiently "close" to ρ? We answer this for the three simplest topologies.

Lemma 1: In the GIM, with $\mathfrak{S}$ in its norm topology, the mapping $M: \mathfrak{S} \to \mathfrak{S}$ is continuous.

Proof: Let ϕ, $\phi_n \in \mathfrak{S}$, with $\phi_n \to \phi$, as $n \to \infty$, in norm. For $x \in \mathfrak{A}_1$, the closed unit ball of $\mathfrak{A}$,

$$\phi_n[\alpha_t(x)] \to \phi[\alpha_t(x)],$$

as $n \to \infty$, uniformly in $t \in \mathbf{R}$ and uniformly in $x \in \mathfrak{A}_1$. Therefore, given $\epsilon > 0$, there exists $N > 0$ such that, for all $n \geq N$,

$$|\phi_n[\alpha_t(x)] - \phi[\alpha_t(x)]| < \epsilon, \quad \text{for all } t \in \mathbf{R}, \ x \in \mathfrak{A}_1.$$

Therefore,

$$\left| \frac{1}{T} \int_0^T \phi_n[\alpha_t(x)]\, dt - \frac{1}{T} \int_0^T \phi[\alpha_t(x)]\, dt \right| \leq \frac{1}{T} \int_0^T \epsilon\, dt = \epsilon,$$

independently of $T > 0$. Hence, $M\phi_n \to M\phi$, as $n \to \infty$, in norm, i.e.,

$$\lim_{n \to \infty} \sup_{x \in \mathfrak{A}_1} \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T (\phi_n[\alpha_t(x)] - \phi[\alpha_t(x)])\, dt \right| = 0.$$

QED

Corollary: In the GIM, with $\mathfrak{S}$ in its weak topology, the mapping $M: \mathfrak{S} \to \mathfrak{S}$ is continuous.

The proof is immediate from Dunford and Schwartz.[16]

Proposition 3: In the GIM with dimension $\nu = 1$, let $\epsilon(|j|) = 1/\xi^{|j|}$ for $j \neq 0$, $\xi > 2$. Then, with $\mathfrak{S}$ in its w^*-topology, the mapping $M: \mathfrak{S} \to \mathfrak{S}$ is not continuous.

Proof: A proof by contradiction is immediate from the following two facts: On the one hand, we exhibit a state ϕ such that $M\phi(\sigma_x^0) \neq 0$; on the other hand, we exhibit a subset U of $\mathfrak{S}$ which is w^*-dense and for which $M\rho(\sigma_x^0) = 0$ for all ρ in U. To this end, let ρ be the product state $\otimes_{j \in \mathbf{Z}} \hat{f}_j$, where $\hat{f}_j$ is the state on $\mathfrak{A}(j)$ defined by any normalized vector f_j which satisfies

$$\begin{aligned} \sigma_z^j f_j &= f_j, \quad \text{if } j > 0, \\ \sigma_z^j f_j &= -f_j, \quad \text{if } j < 0, \\ \sigma_x^0 f_0 &= f_0. \end{aligned}$$

Now let $\phi = M\rho$. Then, ϕ is time invariant and

$$\begin{aligned} \phi(\sigma_x^0) &= M[\rho(\alpha_t[\sigma_x^0])] \\ &= M[\rho(\sigma_x^0 \cos[2P_0 t] - \sigma_y^0 \sin[2P_0 t])]. \end{aligned}$$

If Π_ρ is the GNS representation of $\mathfrak{A}$ corresponding to ρ, with cyclic vector Φ_ρ, it is easy to see that $\Pi_\rho(P_0)\Phi_\rho = 0$ from cancellations. Hence, $\rho(P_0^n) = 0$ for all $n \in \mathbf{N}$, $n \neq 0$. Therefore, $\rho[\alpha_t(\sigma_x^0)] = \rho(\sigma_x^0) = 1$ for all $t \in \mathbf{R}$. Hence, $M\phi(\sigma_x^0) = \phi(\sigma_x^0) = 1$. This concludes the first part of the proof. Now, from Dixmier,[17] we know that the set U of vector states of

any nonnull representation of the simple, antiliminal[18] C^*-algebra $\mathfrak{A}$ is w^*-dense in $\mathfrak{S}$. Consider the GNS representation generated by $\rho = \otimes_{j\in\mathbf{Z}} \hat{f}_j$, where $\sigma_z^j f_j = f_j$ for all $j \in \mathbf{Z}$. Choose the orthonormal basis for $\mathcal{H}_\rho$ consisting of $\{\Psi_I \mid I \in F\}$, where

$$\Psi_I \equiv \sum_{i\in I} \Pi_\rho(\sigma_x^i)\Phi_\rho \quad \text{for} \quad I \in F, \quad I \neq \varnothing,$$

and

$$\Psi_\varnothing \equiv \Phi_\rho.$$

Now $\alpha_t(\sigma_x^0) = \sigma_x^0 \cos(2P_0 t) - \sigma_y^0 \sin(2P_0 t)$. Note that $\Pi_\rho(P_0)\Psi_I = p_I\Psi_I$, where $p_I \equiv \sum_{j\in\mathbf{Z}} \epsilon(|j|)g_j$ and

$$g_j = +1, \quad \text{if} \quad j \in \mathbf{Z}/I,$$
$$= -1, \quad \text{if} \quad j \in I.$$

Therefore,

$$\Pi_\rho(P_0^m)\Psi_I = (p_I)^m\Psi_I$$

and

$$(\Psi_I, \Pi_\rho(P_0^m)\Psi_j) = \delta_{I,J}p_I^m.$$

If Ω is a unit vector in $\mathcal{H}_\rho$, let $\Omega = \sum_{I\in F} \omega_I\Psi_I$. Hence,

$$(\Omega, \Pi_\rho(P_0^m)\Omega) = \sum_{I,J\in F} \bar{\omega}_I\omega_J(\Psi_I, \Pi_\rho(P_0^m)\Psi_J)$$
$$= \sum_{I\in F} \bar{\omega}_I\omega_I p_I^m.$$

Similarly,

$$(\Omega, \Pi_\rho[\sigma_x^0 P_0^m]) = \sum_{I\in F} \bar{\omega}_{I_0}\omega_I p_I^m$$

where $I_0 = I/\{0\}$ if $0 \in I$, $I_0 = I \cup \{0\}$ if $0 \notin I$. Therefore,

$$(\Omega, \Pi_\rho(\sigma_x^0)\cos[2\Pi_\rho(P_0)t]\Omega)$$
$$= \left(\Omega, \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} \Pi_\rho(\sigma_x^0 P_0^{2n})\Omega\right)$$
$$= \sum_{I\in F} \bar{\omega}_{I_0}\omega_I \cos(2p_I t).$$

By taking finite sums in $I \in F$, we can approximate $h(t) \equiv (\Omega, \Pi_\rho(\sigma_x^0)\cos[2\Pi_\rho(P_0)t]\Omega)$ uniformly in $t \in \mathbf{R}$. On the finite sums $h_N(t)$, since no p_I can vanish for the given interactions, we have $Mh_N = 0$. Therefore, $Mh = 0$, since M is continuous on $CB(\mathbf{R})$. Similarly, $M[(\Omega, \Pi_\rho(\sigma_y^0)\sin(2\Pi_\rho[P_0]t)\Omega)] = 0$, so that

$$[M\hat{\Omega}](\sigma_x^0) = 0$$

for all vector states $\hat{\Omega}$ corresponding to the representation Π_ρ. QED

Note that the above proof can also be used to show that the GIM is not always G-Abelian in time, i.e., need not satisfy the condition

$$M[\phi([\alpha_t(A), B])] = 0 \quad \text{for all } A, B \in \mathfrak{A}$$
$$\text{and all } \phi \in \mathfrak{S}_I,$$

where $\mathfrak{S}_I$ is the set of time invariant states. To see this, just use the ϕ of the above proof and $A = \sigma_z^0$, $B = \sigma_y^0$. This example reinforces the doubts one might have of the validity of the assumption that general systems are G-Abelian in time and, hence, justifies the attempt to avoid the assumption. Compare in this respect Knops[19] and Emch, Knops, and Verboven.[20] See also Araki.[21]

Proposition 3 is rather disconcerting in that one has good reason for taking the w^*-topology on $\mathfrak{S}$ as the most physical one. The proposition might, however, be an indication of the fact that $\mathfrak{S}$ itself is bigger than actually needed for physical purposes.[22] It is, furthermore, conceivable that M is w^*-continuous on a w^*-dense subset $\mathfrak{S}_0$ of $\mathfrak{S}$, where $\mathfrak{S}_0$ itself contains all physically accessible states.

4. TEMPERATURE STATES

It is desirable, for the consistency of the approach used in this paper, to establish the existence of infinite volume limits of the usual canonical equilibrium ensembles since, in the present theory, these limits should play the role of states describable by a temperature. Specifically, the question is whether one can take a limit of the states[23] ρ_β^V defined on $\mathfrak{A}(V)$ by

$$\rho_\beta^V(A) = \mathrm{Tr}_V(A \exp[-\beta H_V])/\mathrm{Tr}_V(\exp[-\beta H_V]),$$

where Tr_V is the usual normalized trace state on $\mathfrak{A}(V)$. Araki has shown that, for a 1-dimensional lattice and any finite-range interaction, such a limit does exist.[24] By restricting ourselves to ferromagnetic Ising-type interactions, we obtain the same conclusion for infinite-range interactions in ν dimensions.[25]

Proposition 4: In the GIM, assume that $\epsilon(|j|) \leq 0$ for all $j \in \mathbf{Z}^\nu$. Then, extending the canonical ensemble ρ_β^V to the state $\tilde{\rho}_\beta^V \equiv \rho_\beta^V \otimes_{j\in\mathbf{Z}^\nu/V} \mathrm{Tr}_j$ on $\mathfrak{A}$, we see that there exists a state ρ_β^∞ on $\mathfrak{A}$ defined as $w^*\text{-}\lim_{V\to\infty} \tilde{\rho}_\beta^V$.

Proof: The proof consists of reducing the problem to the finite-volume subalgebras where generalized Griffiths inequalities can be used. Consider, then, any three nonempty elements V_1, V_2, and V_3 of F such that $V_1 \subset V_2 \subset V_3$. Introduce the following two interactions on $\mathfrak{A}(V_3)$:

$$H_3 = \tfrac{1}{2} \sum_{(j,k)\in V_3\times V_3} \epsilon(|j-k|)\sigma_z^j\sigma_z^k$$

and

$$H_2 = \tfrac{1}{2} \sum_{(j,k)\in V_2\times V_2} \epsilon(|j-k|)\sigma_z^j\sigma_z^k.$$

Note that H_3 can be obtained from H_2 by adding ferromagnetic bonds. If, for each $\rho \in \mathfrak{S}$, we define the

state $\rho \mid \mathfrak{A}(V)$ on $\mathfrak{A}(V)$ by restriction, then we have

$$\tilde{\rho}_\beta^{V_3}(\cdot) \mid \mathfrak{A}(V_1) = \mathrm{Tr}_{V_3}(\cdot \exp[-\beta H_3])/\mathrm{Tr}_{V_3}(\exp[-\beta H_3]) \mid \mathfrak{A}(V_1)$$

and

$$\tilde{\rho}_\beta^{V_2}(\cdot) \mid \mathfrak{A}(V_1) = \mathrm{Tr}_{V_2}(\cdot \exp[-\beta H_2])/\mathrm{Tr}_{V_2}(\exp[-\beta H_2]) \mid \mathfrak{A}(V_1).$$

We now show that, in fact,

$$\tilde{\rho}_\beta^{V_2}(\cdot) \mid \mathfrak{A}(V_1) = \mathrm{Tr}_{V_3}(\cdot \exp[-\beta H_2])/\mathrm{Tr}_{V_3}(\exp[-\beta H_2]) \mid \mathfrak{A}(V_1). \tag{3}$$

To see this, introduce the following orthonormal basis in $\otimes_{i\in V_3} \mathbf{C}_i^2$:

$$\begin{aligned} e_1 &= f_1^1 \otimes f_2^1 \otimes \cdots \otimes f_{N(V_3)}^1, \\ e_2 &= f_1^2 \otimes f_2^1 \otimes \cdots \otimes f_{N(V_3)}^1, \\ e_3 &= f_1^1 \otimes f_2^2 \otimes \cdots \otimes f_{N(V_3)}^1, \\ e_4 &= f_1^2 \otimes f_2^2 \otimes \cdots \otimes f_{N(V_3)}^1, \\ &\vdots \\ e_{2^{N(V_3)}} &= f_1^2 \otimes f_2^2 \otimes \cdots \otimes f_{N(V_3)}^2, \end{aligned}$$

where f_i^j is a fixed normalized vector in $\mathbf{C}_i^2$ satisfying $\sigma_z^i f_i^1 = f_i^1$ and $\sigma_z^i f_i^2 = -f_i^2$. If $A \in \mathfrak{A}(V_1)$, then $\exp(-\beta H_2)A \in \mathfrak{A}(V_2)$. Calculating in the above basis gives, for any $B \in \mathfrak{A}(V_2)$, we obtain

$$\mathrm{Tr}_{V_3}(B) = 2^{N(V_3)-N(V_2)} \mathrm{Tr}_{V_2}(B).$$

Therefore, (3) follows. Now we need some further notation before we can continue with the proof. For each triple $A = (A_1, A_2, A_3)$ where the $A_i \in F$ are pairwise disjoint, define σ^A as $(\prod_{i\in A_1} \sigma_x^i)(\prod_{j\in A_2} \sigma_y^j) \times (\prod_{k\in A_3} \sigma_z^k)$, where $\prod_{\ell\in\varnothing} B_\ell$ is defined to be the identity I. Note that $\mathfrak{A}^0$ is the linear manifold generated by the set of all σ^A. Now, if $A_1 \cup A_2 \neq \varnothing$, then

$$\tilde{\rho}_\beta^V(\sigma^A) = 0 \tag{4}$$

for all $V \in F$ such that $V \supset A_1 \cup A_2 \cup A_3$ since in the above basis the diagonal elements of $\exp(-\beta H_V)\sigma^A$ are all zero. Kelly and Sherman[26] have shown that, by increasing the number of bonds, we have

$$\tilde{\rho}_\beta^{V_3}(\sigma^A) \geq \tilde{\rho}_\beta^{V_2}(\sigma^A)$$

if

$$A = (\varnothing, \varnothing, A_3) \quad \text{and} \quad A_3 \subset V_2 \subset V_3. \tag{5}$$

Combining (4) and (5) with another theorem of Kelly and Sherman, which says that $\tilde{\rho}_\beta^{V'}(\sigma^A) \geq 0$ if $\sigma^A \in \mathfrak{A}(V) \subset \mathfrak{A}(V')$ and $A = (\varnothing, \varnothing, A_3)$, we have, for $V' \supset V$ and $\sigma^A \in \mathfrak{A}(V)$,

$$0 \leq \tilde{\rho}_\beta^{V'}(\sigma^A) \leq 1 \quad \text{and} \quad \tilde{\rho}_\beta^{V'}(\sigma^A) \geq \tilde{\rho}_\beta^V(\sigma^A),$$

so that $\tilde{\rho}_\beta^V(\sigma^A)$ is an increasing function of V, bounded above by 1. Therefore, $\lim \tilde{\rho}_\beta^V(\sigma^A)$, as $V \to \infty$, exists if V increases by inclusion. Define $\rho_\beta^\infty(\sigma^A)$ as the limit. Considering ρ_β^∞ as a functional on the $*$-subalgebra $\mathfrak{A}^0$ of $\mathfrak{A}$, ρ_β^∞ is clearly linear and bounded, with norm 1. When ρ_β^∞ is extended to $\mathfrak{A}$, it still has norm 1 and satisfies $\rho_\beta^\infty(I) = 1$. Therefore, it is a state on $\mathfrak{A}$.

QED

5. APPROACH TO EQUILIBRIUM

Now that we have shown the existence of a canonical association between arbitrary initial states and equilibrium states, and also that at least some of these equilibrium states are reasonable, we would like to investigate the association in more detail. One reason for this is to examine the question of recurrences. We are motivated in this approach by a paper of Emch,[27] where the following experiment is considered.

A CaF_2 crystal is placed in a magnetic field (thus determining the z direction), and allowed to reach thermal equilibrium. Then, an rf pulse is applied which turns the net nuclear magnetization to the x direction. The magnetization in the x direction is then measured as a function of time, and the result is an oscillatory function which damps to the equilibrium value of zero.[28]

Emch assumes an interaction of the form

$$H_V = \left(\tfrac{1}{2} \sum_{(j,k)\in V\times V} \epsilon(|j-k|)\sigma_z^j \sigma_z^k\right) - B \sum_{j\in V} \sigma_z^j$$

on a finite 1-dimensional volume V. As the state representing the system after the application of the rf pulse, he takes the product state $\rho = \otimes_{j\in V} \phi_j$, where

$$\phi_j(\cdot) = \mathrm{Tr}_j(\cdot \exp[-\gamma\sigma_x^j])/\mathrm{Tr}_j(\exp[-\gamma\sigma_x^j]).$$

This choice is justified by an entropy argument. With the interaction H_V and initial state ρ, he then calculates the time development of the magnetization in the x direction, $S_x = [1/N(V)] \sum_{j\in V} \sigma_x^j$, and obtains without approximation

$$\rho(\alpha_t^V[S_x]) = \rho(S_x)\left(\prod_{j\in V} \cos^2 2\epsilon(|j|)t\right) \cos 2Bt. \tag{6}$$

By taking an infinite volume limit at this point, Emch shows that interactions with a cutoff give recurrences (with calculable frequency) and that the infinite range interaction of the form $\epsilon(|j|) = 1/2^{|j|}$ gives, with Vieta's identity,[29]

$$\prod_{n=1}^{\infty} \cos\left(\frac{t}{2^n}\right) = \frac{(\sin t)}{t},$$

the nonrecurrent damping exhibited by experiment.

In this section, we generalize this work in the following respects. We consider arbitrary observables and show what range of behavior is possible with different choices of the function ϵ. Furthermore, the class of initial states considered is extended, and the approach to equilibrium is exhibited in the stronger form of an initial state decaying into an equilibrium state, rather than just considering individual expectation values. To simplify calculations, we assume throughout that there is no external field. Inspection of (6) shows that the damping we are looking for comes solely from the spin–spin interaction of the lattice, not the external field.

The general result for finite range models, hereafter denoted F_L, is then:

Lemma 2: In the GIM, if $\epsilon(|j|) = 0$ for all $j \in \mathbf{Z}^\nu$ such that $|j| > L$, $0 < L < \infty$, then $\rho[\alpha_t(A)]$ is almost periodic for all $A \in \mathfrak{A}$, $\rho \in \mathfrak{S}$.

Proof: From the proof of Proposition 2 with the above hypothesis added, we see that the observable S is only a finite linear sum of σ_z's. This implies that the spectrum of S, and hence of $\Pi_\rho(S)$, consists of a finite number of isolated points. Using the spectral theorem in $\mathcal{H}_\rho$, we get $[\alpha_t(A')]$ in the form

$$\sum_{j=1}^{N} a_j e^{ib_j t}, \quad b_j \in \mathbf{R},$$

which is almost periodic. Since $AP(\mathbf{R})$ is also a closed linear subspace of $CB(\mathbf{R})$, we get the result for all $A \in \mathfrak{A}$ as in Proposition 3. QED

It will become apparent later that the full range of behavior of the GIM due to different choices of the function ϵ is already predictable from (6). Because of its importance, therefore, we derive a convenient generalization of (6).

Let ϕ be any state which satisfies $\phi(\sigma^A) = 0$ for all A such that $A_3 \neq \varnothing$. Consider $\phi[\alpha_t(\sigma_x^0)]$ in the GIM. As shown in Proposition 1,

$$\alpha_t(\sigma_x^0) = \operatorname*{norm\text{-}lim}_{V\to\infty} \alpha_t^V(\sigma_x^0), \quad \text{where} \quad V \in F,$$

$$\alpha_t^V(\sigma_x^0) = \sigma_0^x \cos\left(2t\sum_{j\in V} R_j\right) - \sigma_y^0 \sin\left(2t\sum_{j\in V} R_j\right),$$

and $R_j = \epsilon(|j|)\sigma_z^j$. Using the conditions on ϕ, we have

$$\phi[\alpha_t^V(\sigma_x^0)] = \phi\left[\sigma_x^0 \cos\left(2t\sum_{j\in V} R_j\right)\right]. \tag{7}$$

We show by induction on the number of sites in V that

$$\phi[\alpha_t^V(\sigma_x^0)] = \phi(\sigma_x^0)\prod_{i\in V} \cos\,[2\epsilon(|j|)t]. \tag{8}$$

For $V = \{j\}$, $\cos(2tR_j) = \cos[2t\epsilon(|j|)]$, since $(\sigma_z^j)^{2n} = 1$ for all $n \in \mathbf{N}$ and (8) follows from (7). Now assume (8) for V having N sites with site $l \notin V$. Then

$$\phi\left\{\sigma_x^0 \cos\left[2t\left(\sum_{j\in V} R_j + R_l\right)\right]\right\}$$
$$= \phi\left[\sigma_x^0 \cos\left(2t\sum_{j\in V} R_j\right)\cos(tR_l) - \sigma_x^0 \sin\left(2t\sum_{j\in V} R_j\right)\sin(2tR_l)\right].$$

The second term on the rhs vanishes since, in the series expansion of the sines, every term has at least one "unmatched" σ_z in it which is annihilated by ϕ. Again, $\cos(2tR_l) = \cos[2t\epsilon(|l|)]$, so that

$$\phi[\alpha_t^{V\cup l}(\sigma_x^0)] = \phi(\sigma_x^0)\cdot\prod_{j\in V\cup l} \cos\,[2\epsilon(|j|)t].$$

By induction, we have (8) for all $V \in F$, $V \neq \varnothing$. To take the volume limit, we first define $\prod_{j\in\mathbf{Z}^\nu} a_j$, where $a_j \in \mathbf{C}$, as the limit, if it exists, of the net $\prod_{j\in V} a_j$. We make no exceptions for zero factors or convergence to zero. Since $\sum_{j\in\mathbf{Z}^\nu} |\epsilon(|j|)| < \infty$, we must have $\epsilon(|j|) \xrightarrow[j\to\infty]{} 0$, so that it is clear that the limit exists[30] for the net $\prod_{j\in V} \cos\,[2t\epsilon(|j|)]$. Hence,

$$\phi[\alpha_t(\sigma_x^0)] = \phi(\sigma_x^0)\prod_{j\in\mathbf{Z}^\nu} \cos\,[2\epsilon(|j|)t]. \tag{6'}$$

A natural means of investigating the influence of a particular choice of ϵ is thus determining the resulting behavior of $\prod_{n=1}^{\infty} \cos\,[\epsilon(n)t]$. As mentioned above, Vieta's formula shows that the choice $\epsilon(|j|) = 1/2^{|j|}$ produces nonrecurrent behavior. More generally, one might inquire into the time behavior resulting from $\epsilon(|j|) = 1/\xi^{|j|}$, $\xi > 1$. The model with $\epsilon(|j|) = (\xi^{|j|})^{-1}$ for $j \neq 0$, where $\xi > 1$, for stability, is called the exponential model E_ξ. We have[31] that

$$\prod_{n=1}^{\infty} \cos\left(\frac{t}{\xi^n}\right) \xrightarrow[t\to\infty]{} 0, \quad \text{for} \quad \xi > 1,$$

$$\text{if and only if } \xi \notin S/\{2\}, \tag{9}$$

where S is the countable set of all algebraic integers over the rationals with conjugates having moduli strictly less than one. From this, we see that the qualitative behavior of the model is discontinuous in ξ. For this reason, and because of certain results concerning phase transitions by Dyson,[32] we also consider the following form for ϵ: $\epsilon(|j|) = 1/j^{|\alpha|}$ for $j \neq 0$, where α is assumed greater than the dimension ν for stability. We call this the Dyson model D_α. The nonrecurrent time behavior of the Dyson models is shown by the following lemma.

Lemma 3:

$$\prod_{j=1}^{\infty} \cos^2\left(\frac{t}{j^\alpha}\right) \xrightarrow[t\to\infty]{} 0, \quad \text{for all} \quad \alpha > 1.$$

Proof: Define

$$f(t) = \prod_{j=1}^{\infty} \cos^2\left(\frac{t}{j^\alpha}\right) = \prod_{j=1}^{\infty} \left[1 - \sin^2\left(\frac{t}{j^\alpha}\right)\right].$$

Assume that $f(t)$ does not have limit zero as $t \to \infty$. Then there exists a $\delta > 0$ and a sequence $t_n > 0$ such that $t_n \to \infty$ and $f(t_n) > \delta$ for all $n \in \mathbf{N}$. Now $e^x \geq 1 + x$ for all $x \in \mathbf{R}$. Therefore, if $1 + x_j \geq 0$,

$$\prod_{j=1}^{N} e^{x_j} \geq \prod_{j=1}^{N} (1 + x_j).$$

Therefore,

$$\prod_{j=1}^{\infty} \exp\left[-\sin^2\left(\frac{t_n}{j^\alpha}\right)\right] \geq \delta \quad \text{for all } n.$$

Taking logarithms, we obtain

$$-\sum_{j=1}^{\infty} \sin^2\left(\frac{t_n}{j^\alpha}\right) \geq \ln \delta.$$

Hence,

$$\sum_{j=1}^{\infty} \sin^2\left(\frac{t_n}{j^\alpha}\right) \leq -\ln \delta, \quad \text{for all} \quad n \in \mathbf{N}. \tag{10}$$

Let N_n be the number of solutions m in N of the expression $\sin^2(t_n/m^\alpha) \geq \frac{1}{4}$. Clearly, N_n is greater than or equal to the number of solutions m in N of

$$\tfrac{1}{6}\pi \leq t_n/m^\alpha \leq \tfrac{5}{6}\pi$$

or

$$m(\tfrac{1}{6}\pi)^{1/\alpha} \leq (t_n)^{1/\alpha} \leq m(\tfrac{5}{6}\pi)^{1/\alpha}. \tag{11}$$

Therefore, $N_n \to \infty$ as $t_n \to \infty$. This contradicts (10). Hence, $f(t) \to 0$ as $t \to \infty$. QED

We now combine the above facts to prove the following proposition.

Proposition 5: With dimension $\nu = 1$, let the interaction be that of any exponential model E_ξ, where ξ is transcendental, or any Dyson model. Let ϕ be any state which satisfies $\phi(\sigma^A) = 0$ for all A such that $A_3 \neq \varnothing$. Then $M\phi = \otimes_{j\in\mathbf{Z}} \mathrm{Tr}_j$.

Proof: We show that $M\phi$ and $\otimes_{j\in\mathbf{Z}} \mathrm{Tr}_j$ coincide on the set of all σ^A, which by linearity and continuity will give the full result. Note that $\otimes_{j\in\mathbf{Z}} \mathrm{Tr}_j(\sigma^A) = 0$ for all $A \neq (\varnothing, \varnothing, \varnothing)$. For $A = (\varnothing, \varnothing, A_3)$, $\alpha_t(\sigma^A) = \sigma^A$ so that $\phi[\alpha_t(\sigma^A)] = \phi(\sigma^A)$, and the coincidence is obvious. Hence, for the rest of the proof, we assume that $A_1 \cup A_2 \neq \varnothing$. As in the proof of Proposition 2, $\alpha_t(\sigma^A)$ can be put in the form of a finite linear sum of terms, such as

$$\sigma_z^{j_1} \cdots \sigma_z^{j_k} \sigma_{l_{k+1}}^{j_{k+1}} \cdots \sigma_{l_{k+n}}^{j_{k+n}} \exp(iSt),$$

where $S = \sum_{j\in\mathbf{Z}} a_j \sigma_z^j$, and for $j \in \mathbf{Z}$ such that $|j| > W \equiv \max_{m=1,\cdots,n}\{|j_{k+m}|\}$ we have

$$\tfrac{1}{2}a_j = \pm\epsilon(|j - j_{k+1}|) \pm \cdots \pm \epsilon(|j - j_{k+n}|).$$

To show that $M\phi(\sigma^A) = 0$, we first notice that $M(\sigma^A)$ is a finite linear sum of terms of the form

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (\Pi_\phi[\sigma_{l_{k+n}}^{j_{k+n}} \cdots \sigma_z^{j_1}]\Phi_\phi, \exp[i\Pi_\phi(S)t]\Phi_\phi)\, dt.$$

By von Neumann's ergodic theorem,[33] the above expression equals $(\Psi'_\phi, P\Phi_\phi)$, where

$$\Psi'_\phi \equiv \Pi_\phi[\sigma_{l_{k+n}}^{j_{k+n}} \cdots \sigma_z^{j_1}]\Phi_\phi$$

and P is the projection defined by the strong operator limit,

$$\lim_{\epsilon\to 0-} [E(0) - E(\epsilon)],$$

where $\{E(\lambda) \mid \lambda \in \mathbf{R}\}$ is the resolution of the identity corresponding to $\Pi_\phi(S)$. We will show that $P\Phi_\phi = 0$, and this will complete the proof:

$$\begin{aligned} \|P\Phi_\phi\| &= (\Phi_\phi, P\Phi_\phi) \\ &= \lim_{T\to\infty} T^{-1} \int_0^T (\Phi_\phi, \exp[i\Pi_\phi(S)t]\Phi_\phi)\, dt. \end{aligned}$$

By arguing as in the proof of (6'), we get

$$(\Phi_\phi, \exp[i\Pi_\phi(S)t]\Phi_\phi) = \prod_{j\in\mathbf{Z}} \cos(a_j t).$$

All we need to do now is show that

$$\prod_{j\in\mathbf{Z}} \cos(a_j t) \xrightarrow[t\to\infty]{} 0. \tag{12}$$

For the exponential models, we have for all large enough $|j|$ that

$$\tfrac{1}{2}a_j = \pm \frac{1}{\xi^{|j-j_1|}} \pm \cdots \pm \frac{1}{\xi^{|j-j_n|}}.$$

Therefore, if ξ is transcendental, it is clear that a_j does not vanish. Then, since a_j can be factored,

$$\begin{aligned} \tfrac{1}{2}a_j &= \frac{1}{\xi^j}[\pm\xi^{j_1} \pm \cdots \pm \xi^{j_n}], \qquad \text{for} \quad j \gg 0, \\ &= \frac{1}{\xi^{-j}}[\pm\xi^{-j_1} \pm \cdots \pm \xi^{-j_n}], \quad \text{for} \quad j \ll 0, \end{aligned}$$

a simple change of variable in (9) gives (12). For the Dyson models, we use a different argument. First, we need to show that a_j does not vanish for all sufficiently large $|j|$. There are two cases to be treated separately: (1) a_j does not have an equal number of $+2$ and -2 coefficients for the $1/|j - j_1|^\alpha$; (2) it does have an

equal number. In case 1, it is clear that the sign which appears more often gives a_j that same sign for all large enough $|j|$. Case 2 follows from the fact that the functions of a complex variable defined by

$$a_1(z) = \pm \frac{1}{(z - j_1)^\alpha} \pm \cdots \pm \frac{1}{(z - j_n)^\alpha}$$

and

$$a_2(z) = \pm \frac{1}{(j_1 - z)^\alpha} \pm \cdots \pm \frac{1}{(j_n - z)^\alpha}$$

are both analytic at infinity, so that there is a compact set K which contains the zeros of both functions. An argument similar to Lemma 3 then yields (12).

QED

In addition to the result stated in the proposition, we point out that the above proof not only determined the equilibrium value of all local observables, but also showed that this equilibrium value is actually approached for large t. In fact, we have proven the following:

Corollary: With dimension $\nu = 1$, let the interaction be that of any exponential model E_ξ with ξ transcendental or of any Dyson model. Let ϕ be any state which satisfies $\phi(\sigma^A) = 0$ for all A such that $A_3 \neq \varnothing$. Then

$$\rho[\alpha_t(\theta)] \xrightarrow[t\to\infty]{} M\rho[\theta]$$

for every local observable θ.

We conclude this section with an example which shows that one cannot expect the models to be so well behaved on all initial states. In particular, we exhibit a state which shows recurrences for all ferromagnetic Ising-type models.

Define $\phi \in \mathfrak{S}$ as $\otimes_{j\in\mathbf{Z}^\nu} \hat{f}_j$, where $\sigma_z^j f_j = f_j$ for $j \neq 0$, and $\sigma_x^0 f_0 = f_0$. Then, in the GIM, we have from (1) that

$$\phi[\alpha_t(\sigma_x^0)] = \phi[\sigma_x^0 \cos(2P_0 t) - \sigma_y^0 \sin(2P_0 t)].$$

But $\hat{f}_0(\sigma_y^0) = 0$, and so the second term on the rhs vanishes. Hence,

$$\phi[\alpha_t(\sigma_x^0)] = \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} \phi(P_0^{2n}) \cdot 2^{2n} = \cos(2pt),$$

where

$$p = \sum_{j\in\mathbf{Z}^\nu} \epsilon(|j|).$$

6. RATE OF DECAY TO EQUILIBRIUM

Most of the proofs in previous sections depended on properties of functions of the form

$$f(t) = \prod_{n=1}^{\infty} \cos(a_n t).$$

We now want to comment on the essential connection of this function with our problem. We show, in particular, that f is the Fourier transform of a certain measure μ of physical origin and that investigation of the structure of this measure can give detailed information about the behavior of the system. Before we can discuss this further, we need some definitions and facts.

Let μ be a Borel probability measure, hereafter abbreviated Bpm. We denote Borel sets with Lebesgue measure zero by Z, and countable sets by C. Then μ is called

(1) absolutely continuous if $\mu(Z) = 0$ for all Z,
(2) singular continuous if $\mu(Z) = 1$ for some Z and $\mu(C) = 0$ for all C,
(3) discontinuous if $\mu(C) = 1$ for some C.

An equivalent classification is obtained by using the function $\mu(I_x)$ of $x \in \mathbf{R}$, called the distribution function of μ, where $I_x = \{y \in \mathbf{R} \mid y \leq x\}$. Then μ is

(1) absolutely continuous if and only if $\mu(I_x)$ is an absolutely continuous point function,
(2) singular continuous if and only if $\mu(I_x)$ is continuous and $d\mu(I_x)/dx = 0$ for almost all x,
(3) discontinuous if and only if the range of $\mu(I_x)$ is a countable set.

If μ_1 and μ_2 are Bpm's, the set function defined by

$$\mu_1 * \mu_2(A) = \int_{\mathbf{R}} \mu_1(A - x)\, d\mu_2(x)$$

is a Bpm called the convolution of μ_1 and μ_2. If μ is a Bpm, the point function defined by

$$\hat{\mu}(t) = \int_{\mathbf{R}} e^{itx}\, d\mu(x)$$

is called the Fourier transform of μ. We then have the following connection: If μ_1 and μ_2 are Bpm's, then $\widehat{\mu_1 * \mu_2}(t) = \hat{\mu}_1(t)\hat{\mu}_2(t)$. Furthermore, the Fourier transform is a means of determining continuity properties of Bpm's since[34] if γ is greater than the positive integer p, then

$$\hat{\mu}(t) = O_\pm(|t|^{-\gamma}) \quad \text{implies that} \quad \mu(I_x) \in C^p. \tag{13}$$

Here, we use the notation that a function $g(t)$ satisfies $g(t) = O_\pm[h(t)]$ if there exist positive constants c and d such that $|g(t)| \leq ch(t)$ for all $t > d$. Also, $g \in C^p$ means that g has continuous derivatives through order p.

We now come back to $\rho[\alpha_t(\sigma_x^0)]$. From Proposition 1, we have

$$\alpha_t(\sigma_x^0) = \exp[i\tilde{H}_0 t]\sigma_x^0 \exp[-i\tilde{H}_0 t],$$

where $\tilde{H}_0 = \sigma_z^0 \sum_{k \in Z^\nu} \epsilon(|k|)\sigma_z^k$. It is then easy to obtain

$$\rho[\alpha_t(\sigma_x^0)] = \rho(\sigma_x^0 \exp[i2\tilde{H}_0 t]).$$

Defining μ as the spectral measure of $\Pi_\rho(2\tilde{H}_0)$ corresponding to the form $\rho(\sigma_x^0 \cdot) = (\Pi_\rho(\sigma_x^0)\Phi_\rho, \cdot\, \Phi_\rho)$, we have from Stone's theorem that

$$\rho(\alpha_t[\sigma_x^0]) = \hat{\mu}(t). \tag{14}$$

By a simple generalization of (13) to complex measures, we have

Proposition 6: In the GIM, if $\rho[\alpha_{|t|}(\sigma_x^0)] = O_\pm(|t|^{-\gamma})$ for any γ greater than the positive integer p, then $\mu(I_x) \in C^p$, where μ is the spectral measure of $\Pi_\rho(\tilde{H}_0)$ corresponding to the form $\rho(\sigma_x^0 \cdot)$.

To show that the available range for the rates of decay into equilibrium is wide enough to be of interest, we investigate the situation discussed in Sec. 5. Therefore, restricting ourselves to dimension $\nu = 1$ and states ρ which satisfy $\rho(\sigma^A) = 0$ for all A such that $A_3 \neq 0$, we have, as in (6′),

$$\begin{aligned}\rho[\alpha_t(\sigma_x^0)] &= \rho(\sigma_x^0)\rho(e^{-i2\tilde{H}_0 t}) \\ &= \rho(\sigma_x^0) \prod_{n=1}^{\infty} \cos^2[2\epsilon(n)t].\end{aligned}$$

We therefore need to classify the measures μ_L, μ_ξ, and μ_α which have Fourier transform

$$\prod_{n=1}^{\infty} \cos[2\epsilon(n)t],$$

with ϵ coming from the finite-range, exponential, and Dyson models, respectively:

(A) For the finite-range models, it is easy to see that μ_L is discontinuous, producing recurrent behavior.

(B) For the exponential model E_ξ with $\xi > 2$, it is known[35] that μ_ξ is singular continuous so that $\hat{\mu}_\xi \notin C^1$ and, therefore, from (13) that

$$\prod_{n=1}^{\infty} \cos^2\left(\frac{2t}{\xi^n}\right) \neq O_\pm(|t|^{-\gamma}), \quad \text{for any } \gamma > 2 \text{ if } \xi > 2.$$

In fact, it is further known[36] that $\mu_\xi * \mu_\xi$ is singular continuous for $\xi > 3$ so that

$$\prod_{n=1}^{\infty} \cos^2\left(\frac{2t}{\xi^n}\right) \neq O_\pm(|t|^{-\gamma}), \quad \text{for any } \gamma > 1 \text{ if } \xi > 3.$$

(C) For the Dyson model D_α, the following lemma shows that, for some $c > 0$,

$$\prod_{n=1}^{\infty} \cos^2\left(\frac{2t}{n^\alpha}\right) < \exp[-c\,|t|^{1/\alpha}]$$

so that $\mu_\alpha \in C^\infty$.

Lemma 4: If $\alpha > 1$, there exists a $c > 0$ such that

$$\left|\prod_{j=1}^{\infty} \cos\left(\frac{t}{j^\alpha}\right)\right| \leq \exp[-c\,|t|^{1/\alpha}].$$

Proof: For $0 < x < 1$, we have $0 < \cos x < 1 - cx^2$ for some $c > 0$, and $1 - x \leq e^{-x}$. Therefore, for $t > 0$ we have

$$\begin{aligned}\left|\prod_{j=1}^{\infty} \cos\left(\frac{t}{j^\alpha}\right)\right| &\leq \prod_{j > t^{1/\alpha}} \left|\cos\frac{t}{j^\alpha}\right| \\ &\leq \prod_{j > t^{1/\alpha}} \left(1 - c\frac{t^2}{j^{2\alpha}}\right) \\ &\leq \exp\left(-c \prod_{j > t^{1/\alpha}} \frac{t^2}{j^{2\alpha}}\right).\end{aligned}$$

By integral approximation,

$$\sum_{j > t^{1/\alpha}} \frac{t^2}{j^{2\alpha}} \geq t^{1/\alpha}.$$

The transition to negative t then gives the full result. QED

The above classification shows by example how wide a range of rates of decay is attainable. To complete the picture, we note[37] that for no form of interaction in the GIM is there a $c > 0$ such that

$$\prod_{n=1}^{\infty} \cos^2[2\epsilon(n)t] = O_\pm(e^{-c|t|}).$$

7. CONCLUSIONS

The analysis presented in this paper leads to an explicit statement on the relation, in the thermodynamical limit, between the spectrum of the "Hamiltonian" and the time behavior of the expectation values for local observables. In particular, Proposition 6 shows that, for generalized Ising interactions, the degree of continuity of the spectrum of local Hamiltonians, considered in the Hilbert space generated by any initial state, limits the rate at which that initial state can approach equilibrium.

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to my thesis advisor, Dr. G. G. Emch, for his patience and guidance at all stages of this paper. I would also like to thank several members of the Mathematics Department of the University of Rochester for helping to orient me in the mathematical literature.

[1] F. J. Murray and J. von Neumann, Ann. Math. **37**, 116 (1936).
[2] Z. Takeda, Tôhoku Math. J., Ser. 2, **7**, 67 (1955).
[3] This framework was introduced in D. Robinson, Commun. Math. Phys. **6**, 151 (1967).
[4] For more general interactions see Theorem 7.6.2 of D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[5] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[6] G. G. Emch, H. J. F. Knops, and E. J. Verboven, Commun. Math. Phys. **7**, 164 (1968).
[7] The average can be put in the more understandable form of an integral as follows. By the Riesz representation theorem, the means $\{M\}$ on $CB(\mathbf{R})$ are in 1-to-1 correspondence with the probability measures $\{\mu\}$ on the Stone-Čech compactification $\beta\mathbf{R}$ of $\mathbf{R}$, and the correspondence is $Mf = \int_{\beta\mathbf{R}} f(t)\, d(t)$ for all f in $CB(\mathbf{R})$.
[8] A review of the proof shows that a discrete amenable group (e.g., any Abelian group) will give a left invariant mean on the set of all bound functions on the group, and this mean also gives rise to an invariant state. Here, we have the desired continuity anyway.
[9] See Ref. 6.
[10] Since the group is Abelian, we drop "left."
[11] See Appendix 1 of F. P. Greenleaf, *Invariant Means on Topological Groups* (Van Nostrand-Reinhold , New York, 1969). Also note that an ergodic average of the form $\lim 1/T \int_0^T dt$, as $T \to \infty$, does not exist on $CB(\mathbf{R})$.
[12] W. E. Eberlein, Trans. Am. Math. Soc. **67**, 217 (1949). The part of this work which we are using here has recently been generalized to non-Abelian groups. See, for example, Ref. 11, p. 38.
[13] See Ref. 12.
[14] See Ref. 12.
[15] As in Footnote 7, we can give a setting which gives a simple interpretation of our average but also exhibits its uniqueness. Since all the functions of time that we have are in $W(\mathbf{R})$, Theorem 5.3 of K. Deleeuw and I. Glicksberg [Acta Math. **105**, 63 (1961)] implies that the means on $W(\mathbf{R})$ are in one-to-one correspondence with the probability measures on a compactification $\mathbf{R}^\omega$ of $\mathbf{R}$ which is smaller than $\beta\mathbf{R}$. Therefore, given the orbit $\{\alpha_t^* \rho\}$ of an initial state ρ, we are using as our average the unique integral

$$M\rho = \int_{\mathbf{R}^\omega} \alpha_t^* \rho \, d\mu(t),$$

which is independent of the definition of $t = 0$.
[16] Theorem V.3.15 in N. Dunford and J. T. Schwartz, *Linear Operators*, (Interscience, New York, 1957), Part I.
[17] Lemma 11.24 and Prop. 3.4.2 of J. Dixmier, *Les C*-algèbres et leurs représentations* (Gauthier-Villars, Paris, 1964).
[18] The fact that our algebra is simple and antiliminal is proven in Prop. 3.1 of A. Guichardet, Ann. Sci. Ecole Normale Supér., 3rd Ser. **83**, 1 (1966).
[19] H. J. F. Knops, thesis, University of Nijmegen, 1969.
[20] G. G. Emch, H. J. F. Knops and E. J. Verboven, J. Math. Phys. **11**, 1656 (1970).
[21] H. Araki, Commun. Math. Phys. **14**, 120 (1969).
[22] R. V. Kadison, Top.,Supp. 2, **3**, 177 (1965).
[23] An example of what can be used as a substitute when the obvious limit cannot be proven is contained in Sec. 7.3 of Ref. 4.
[24] See Ref. 21.
[25] The essence of this proof is already contained on page 412 of J. L. Lebowitz, Rev. Phys. Chem. **19**, 389 (1968).
[26] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).
[27] G. G. Emch, J. Math. Phys. **7**, 1198 (1966).
[28] I. J. Lowe and R. E. Norberg, Phys. Rev. **107**, 46 (1957).
[29] See, for example, M. Kac, *Statistical Independence in Probability, Analysis and Number Theory* (Wiley, New York, 1959).
[30] In fact, the product is absolutely convergent so that we may rearrange factors arbitrarily.
[31] Theorem II in R. Salam, *Algebraic Numbers and Fourier Series* (Heath, Boston, 1963).
[32] F. J. Dyson, Comm. Math. Phys. **12**, 91 (1969).
[33] J. von Neumann, Proc. Natl. Acad. Sci.(U.S.)**18**, 70 (1932).
[34] See p. 117 of A. Wintner, *The Fourier Transforms of Probability Distributions* (Edwards Brothers, Ann Arbor, Michigan, 1947).
[35] J. P. Kahane and R. Salem, Colloq. Math. **6**, 193 (1958).
[36] See Ref. 35.
[37] C. Radin, to be published.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Classical Thermodynamics Simplified

PETER RASTALL
Physics Department, University of British Columbia, Vancouver 8, B.C.

(Received 5 January 1970)

Classical thermodynamics is developed in a rigorous and quite general form. The approach is similar to Carathéodory's in that entropy and temperature are defined in terms of quantities which are more directly measurable, but Pfaffian forms and quasistatic processes do not appear. The mathematics used is elementary, apart from a small amount of symbolic logic and a very little topology.

1. INTRODUCTION

The second law of thermodynamics is still often stated in the manner of Kelvin: *It is impossible to construct a system that, operating in a cycle, will produce no effect other than the extraction of heat from a reservoir and the performance of work on a mechanical system.* Such formulations have a comfortingly operational sound, but they are unsatisfactory as a basis for a physical theory. Their most serious defect is that they are incomplete. For example, they give no indication of what processes are possible for a physical system. One is forced to rely on intuitive judgements, which makes it impossible to construct a logically sound theory. To make matters worse, the processes required in traditional applications of the second law are often "quasistatic" or "reversible," and can be defined only by subtle limiting procedures.

Carathéodory[1] was the first to attempt an axiomatic formulation of thermodynamics. Although his theory is not completely general, it does apply to a large class of systems. Heat, entropy, and temperature are defined in terms of measurable quantities, and the assumptions of the older theory are made more explicit and simplified. Despite these considerable

[1] F. J. Murray and J. von Neumann, Ann. Math. **37**, 116 (1936).
[2] Z. Takeda, Tôhoku Math. J., Ser. 2, **7**, 67 (1955).
[3] This framework was introduced in D. Robinson, Commun. Math. Phys. **6**, 151 (1967).
[4] For more general interactions see Theorem 7.6.2 of D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[5] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[6] G. G. Emch, H. J. F. Knops, and E. J. Verboven, Commun. Math. Phys. **7**, 164 (1968).
[7] The average can be put in the more understandable form of an integral as follows. By the Riesz representation theorem, the means $\{M\}$ on $CB(\mathbf{R})$ are in 1-to-1 correspondence with the probability measures $\{\mu\}$ on the Stone-Čech compactification $\beta\mathbf{R}$ of $\mathbf{R}$, and the correspondence is $Mf = \int_{\beta\mathbf{R}} f(t)\, d(t)$ for all f in $CB(\mathbf{R})$.
[8] A review of the proof shows that a discrete amenable group (e.g., any Abelian group) will give a left invariant mean on the set of all bound functions on the group, and this mean also gives rise to an invariant state. Here, we have the desired continuity anyway.
[9] See Ref. 6.
[10] Since the group is Abelian, we drop "left."
[11] See Appendix 1 of F. P. Greenleaf, *Invariant Means on Topological Groups* (Van Nostrand-Reinhold , New York, 1969). Also note that an ergodic average of the form $\lim 1/T \int_0^T dt$, as $T \to \infty$, does not exist on $CB(\mathbf{R})$.
[12] W. E. Eberlein, Trans. Am. Math. Soc. **67**, 217 (1949). The part of this work which we are using here has recently been generalized to non-Abelian groups. See, for example, Ref. 11, p. 38.
[13] See Ref. 12.
[14] See Ref. 12.
[15] As in Footnote 7, we can give a setting which gives a simple interpretation of our average but also exhibits its uniqueness. Since all the functions of time that we have are in $W(\mathbf{R})$, Theorem 5.3 of K. Deleeuw and I. Glicksberg [Acta Math. **105**, 63 (1961)] implies that the means on $W(\mathbf{R})$ are in one-to-one correspondence with the probability measures on a compactification $\mathbf{R}^\omega$ of $\mathbf{R}$ which is smaller than $\beta\mathbf{R}$. Therefore, given the orbit $\{\alpha_t^* \rho\}$ of an initial state ρ, we are using as our average the unique integral

$$M\rho = \int_{\mathbf{R}^\omega} \alpha_t^* \rho \, d\mu(t),$$

which is independent of the definition of $t = 0$.
[16] Theorem V.3.15 in N. Dunford and J. T. Schwartz, *Linear Operators*, (Interscience, New York, 1957), Part I.
[17] Lemma 11.24 and Prop. 3.4.2 of J. Dixmier, *Les C*-algèbres et leurs représentations* (Gauthier-Villars, Paris, 1964).
[18] The fact that our algebra is simple and antiliminal is proven in Prop. 3.1 of A. Guichardet, Ann. Sci. Ecole Normale Supér., 3rd Ser. **83**, 1 (1966).
[19] H. J. F. Knops, thesis, University of Nijmegen, 1969.
[20] G. G. Emch, H. J. F. Knops and E. J. Verboven, J. Math. Phys. **11**, 1656 (1970).
[21] H. Araki, Commun. Math. Phys. **14**, 120 (1969).
[22] R. V. Kadison, Top.,Supp. 2, **3**, 177 (1965).
[23] An example of what can be used as a substitute when the obvious limit cannot be proven is contained in Sec. 7.3 of Ref. 4.
[24] See Ref. 21.
[25] The essence of this proof is already contained on page 412 of J. L. Lebowitz, Rev. Phys. Chem. **19**, 389 (1968).
[26] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).
[27] G. G. Emch, J. Math. Phys. **7**, 1198 (1966).
[28] I. J. Lowe and R. E. Norberg, Phys. Rev. **107**, 46 (1957).
[29] See, for example, M. Kac, *Statistical Independence in Probability, Analysis and Number Theory* (Wiley, New York, 1959).
[30] In fact, the product is absolutely convergent so that we may rearrange factors arbitrarily.
[31] Theorem II in R. Salam, *Algebraic Numbers and Fourier Series* (Heath, Boston, 1963).
[32] F. J. Dyson, Comm. Math. Phys. **12**, 91 (1969).
[33] J. von Neumann, Proc. Natl. Acad. Sci.(U.S.)**18**, 70 (1932).
[34] See p. 117 of A. Wintner, *The Fourier Transforms of Probability Distributions* (Edwards Brothers, Ann Arbor, Michigan, 1947).
[35] J. P. Kahane and R. Salem, Colloq. Math. **6**, 193 (1958).
[36] See Ref. 35.
[37] C. Radin, to be published.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Classical Thermodynamics Simplified

PETER RASTALL
Physics Department, University of British Columbia, Vancouver 8, B.C.

(Received 5 January 1970)

Classical thermodynamics is developed in a rigorous and quite general form. The approach is similar to Carathéodory's in that entropy and temperature are defined in terms of quantities which are more directly measurable, but Pfaffian forms and quasistatic processes do not appear. The mathematics used is elementary, apart from a small amount of symbolic logic and a very little topology.

1. INTRODUCTION

The second law of thermodynamics is still often stated in the manner of Kelvin: *It is impossible to construct a system that, operating in a cycle, will produce no effect other than the extraction of heat from a reservoir and the performance of work on a mechanical system.* Such formulations have a comfortingly operational sound, but they are unsatisfactory as a basis for a physical theory. Their most serious defect is that they are incomplete. For example, they give no indication of what processes are possible for a physical system. One is forced to rely on intuitive judgements, which makes it impossible to construct a logically sound theory. To make matters worse, the processes required in traditional applications of the second law are often "quasistatic" or "reversible," and can be defined only by subtle limiting procedures.

Carathéodory[1] was the first to attempt an axiomatic formulation of thermodynamics. Although his theory is not completely general, it does apply to a large class of systems. Heat, entropy, and temperature are defined in terms of measurable quantities, and the assumptions of the older theory are made more explicit and simplified. Despite these considerable

advances, his theory has been ignored by most physicists, probably because they have been unwilling to master the necessary mathematics of Pfaffian forms.[2]

In recent years a number of authors have tried to simplify or generalize Carathéodory's work. The common feature of the new theories is that they make assumptions about the possibility of processes with arbitrary initial and final states, whereas Carathéodory is mainly concerned with neighboring states. The simplest of the theories is that of Buchdahl.[3] Its generality and rigor are similar to Carathéodory's, but it uses only elementary mathematics. The most general and rigorous theory is due to Giles[4]; that of Falk and Jung[5] resembles it, but is less fully worked out. Both theories require nonelementary mathematics—different from Carathéodory's, but no simpler. The same is true to some extent of the theory of Buchdahl and Greve.[6]

We shall attempt to construct a theory of the same type as those just described, but simple, general, and rigorous. We would, however, point out that this is by no means the only sensible approach to thermodynamics. If one wishes to apply or generalize the theory, it is often convenient to postulate the existence of entropy and temperature functions with the desired properties. This procedure, which derives from Gibbs, is well exemplified in the books of Guggenheim,[7] Callen,[8] and Tisza.[9] As Tisza emphasizes, the two points of view are complementary. The advantage of ours is that it gives a more immediate insight into the meaning of entropy and temperature.

We shall be mainly concerned with the formal structure of thermodynamics—the statement of axioms and the proof of theorems—which is of course only the skeleton of the subject. To give the bare bones life, one must interpret the undefined elements of the formal theory in terms of measurable quantities. Such interpretations, and the informal arguments by which one makes the axioms plausible, are not part of the formal theory. We shall always try to keep this distinction clear, since failure to do so can cause great confusion. We shall, for example, speak of the *definition* of elements in the formal theory, but of *interpretation* or *characterization* in the informal discussion. (E.g., an adiabatic process is an undefined element of the formal theory, but it is characterized informally as a process in which a system is thermally insulated—where thermal insulation is something that can be demonstrated in the laboratory.)

2. INTERNAL ENERGY

We begin by giving an account of the first law of thermodynamics and the definition of internal energy. The axioms and theorems of this section are needed later, but there is nothing really new; so we will be brief.

Informally, thermodynamics is concerned with *systems* that can sometimes be described in terms of *macroscopic variables*. A macroscopic variable is a measurable quantity whose values are not subject to significant random fluctuations and are independent of any sufficiently careful measurements made on the system. (Thus, the mass of gas in a very small volume and the momentum of an electron are not macroscopic variables). The number of macroscopic variables of a given system may differ from time to time. When the number is maximal, one says that the system is in a *state*.[10] Any state is determined by the values of a maximal set of macroscopic variables. A system in a state may evolve in time, either spontaneously or by interaction with its surroundings, into another state or into no state at all. If it evolves from a state to a state, we say that it undergoes a *process*.

As axiomatists, we are not concerned with the many difficulties associated with the ideas of the last paragraph. We do not have to talk even about systems. All we need say is that there exists a set Σ whose elements are called *states*. Time does not appear in the formal theory. There is a set Π, whose elements are called *processes*, and there are mappings $\mathfrak{I}:\Pi \to \Sigma$ and $\mathfrak{F}:\Pi \to \Sigma$. If $\mathfrak{I}(\pi) = a$, $\mathfrak{F}(\pi) = b$, then a is the *initial state* of π, b is the *final state* of π, and π is a process *from a to b*. If $\mathfrak{I}(\pi) = \mathfrak{F}(\pi)$, then π is said to be *cyclic*. We assume that, if there is a process from a to b and a process from b to c, then there is a process from a to c:

$$\forall\, \pi, \sigma \in \Pi\ \exists\, \tau \in \Pi : \mathfrak{F}(\pi) = \mathfrak{I}(\sigma) \Rightarrow \mathfrak{I}(\tau) = \mathfrak{I}(\pi)\ \&\ \mathfrak{F}(\tau) = \mathfrak{F}(\sigma).$$

We use conventional logical notation: read $\forall$ as *for all*, $\in$ as *belonging to*, $\exists$ as *there exists*, $\Rightarrow$ as *implies* or *only if*, and & as *and*. We shall also use $\exists!$ for *there exists a unique*, $\Leftrightarrow$ for *iff* or *if and only if*, and $\vee$ for *or*.

If a system undergoes a process, work may be done on it. We assume that this work is measurable. Formally speaking, there is a mapping $W:\Pi \to R^1$, where R^1 is the set of real numbers, and $W(\pi)$ is called the *work done in π*. When interpreting the theory, one chooses a system of units and takes $W(\pi)$ to be the work done on the system as measured in those units.

We shall be concerned for the rest of this section with a special class of processes called *adiabatic processes*. These are characterized as processes that a system undergoes when it is thermally insulated. In the formal theory we assume that there exists a set $\Delta \subset \Pi$, whose elements are called *adiabatic processes*,

and we assume that, if there is an adiabatic process from a to b and one from b to c, then there is one from a to c:

$$\forall\, \pi, \sigma \in \Delta\ \exists\, \tau \in \Delta : \mathcal{F}(\pi) = \mathcal{I}(\sigma) \Rightarrow \mathcal{I}(\tau) = \mathcal{I}(\pi)\ \&\ \mathcal{F}(\tau) = \mathcal{F}(\sigma).$$

There is not always an adiabatic process from a state a to a state b. When there is, we say that b is *adiabatically accessible* from a and write $a \rightharpoonup b$. The formal definition is $a \rightharpoonup b \Leftrightarrow \exists\, \pi \in \Delta : \mathcal{I}(\pi) = a\ \&\ \mathcal{F}(\pi) = b$. We assume the following axioms for all $a, b, c \in \Sigma$:

I. $a \rightharpoonup a$;

II. $a \rightharpoonup b\ \&\ b \rightharpoonup c \Rightarrow a \rightharpoonup c$;

III. $a \rightharpoonup b \vee b \rightharpoonup a$.

An example of an adiabatic process that satisfies I is one that begins and ends at the same instant. II is simply a reformulation of our previous assumption about adiabatic processes. Axiom III, which says that one can always go from a to b or from b to a by an adiabatic process, entails a restriction on the type of system the theory can deal with. It is easy to see, for example, that it is not valid for many systems in which friction occurs. It might be possible to replace III by something weaker: One could try $\forall\, a, b \in \Sigma\ \exists\, c \in \Sigma : (a \rightharpoonup c \vee c \rightharpoonup a)\ \&\ (b \rightharpoonup c \vee c \rightharpoonup b)$. This would considerably complicate the theory, however, and we shall therefore be content with III.

The *first law of thermodynamics* states that the work done in an adiabatic process from a to b is independent of the adiabatic process chosen:

IV. $\forall\, \lambda, \pi \in \Delta : \mathcal{I}(\lambda) = \mathcal{I}(\pi)\ \&\ \mathcal{F}(\lambda) = \mathcal{F}(\pi) \Rightarrow W(\lambda) = W(\pi)$.

It follows that one can speak without ambiguity of the work done in an adiabatic process from a to b. The next axiom says that this work plus the work done in an adiabatic process from b to c is the work done in an adiabatic process from a to c.

V. $\forall\, \lambda, \pi, \mu \in \Delta : \mathcal{I}(\lambda) = \mathcal{I}(\mu)\ \&\ \mathcal{F}(\lambda) = \mathcal{I}(\pi)\ \&\ \mathcal{F}(\pi) = \mathcal{F}(\mu) \Rightarrow W(\lambda) + W(\pi) = W(\mu)$.

Note that the processes λ, π, and μ need not be distinct. This fact is used in the proof of our first theorem.

Theorem 1: (i) If μ is a cyclic, adiabatic process, then $W(\mu) = 0$. (ii) If λ is an adiabatic process from a to b and π is an adiabatic process from b to a, then $W(\lambda) = -W(\pi)$.

Proof: (i) Consider $\lambda = \pi = \mu$ in V. (ii) From I, there is an adiabatic process μ from a to a, and, from V and (i), $W(\lambda) + W(\pi) = W(\mu) = 0$.

For any state $a \in \Sigma$, a function $U_a : \Sigma \to R^1$, called the *internal energy with respect to a*, is defined as follows. For all $b \in \Sigma$, III implies that there is an adiabatic process λ from a to b or an adiabatic process π from b to a. In the first case we define $U_a(b) = W(\lambda)$, and in the second case $U_a(b) = -W(\pi)$. From IV, it does not matter how λ and π are chosen, and, from Theorem 1, the definition is unambiguous when $a \rightharpoonup b$ and $b \rightharpoonup a$.

The internal energies with respect to different states are very simply related.

Theorem 2: $U_c = U_a + K_{ca}$ for all $a, c \in \Sigma$, where K_{ca} is a constant function.

Proof (outline): From III, $a \rightharpoonup c$ or $c \rightharpoonup a$. Without loss of generality one may assume $a \rightharpoonup c$. Given $b \in \Sigma$, it follows from III that there are four possibilities. These are: (i) $a \rightharpoonup b\ \&\ c \rightharpoonup b$; (ii) $a \rightharpoonup b\ \&\ b \rightharpoonup c$; (iii) $b \rightharpoonup a\ \&\ c \rightharpoonup b$; and (iv) $b \rightharpoonup a\ \&\ b \rightharpoonup c$. The theorem is now proved using V, the definitions of U_a and U_c, and, in case (iii), Theorem 1. In case (i), for example, let λ, π, and μ be adiabatic processes from a to c, c to b, and a to b, respectively. Then $W(\lambda) + W(\pi) = W(\mu)$ from V, and hence $U_a(c) + U_c(b) = U_a(b)$ and $K_{ca}(b) = -U_a(c)$. The other cases are similar.

Any function of the form $U_a + K$, where $a \in \Sigma$ and K is a constant, is called an *internal energy*. It follows from Theorem 2 that any pair of internal energies differ by a constant. Since it rarely matters which internal energy is used, we shall normally speak of *the* internal energy and denote it by U. (One may say that U is *arbitrary to the extent of an additive constant*.)

Theorem 3: The work done in an adiabatic process is equal to the increase in the internal energy.

Proof: Let $\pi \in \Delta$, $\mathcal{I}(\pi) = a$, $\mathcal{F}(\pi) = b$. From the definition of U_a, $U_a(b) = W(\pi)$. From I and Theorem 1, $U_a(a) = 0$, and hence $W(\pi) = U_a(b) - U_a(a)$. Since any pair of internal energies differ by a constant function, one has $W(\pi) = U(b) - U(a)$ for any internal energy U.

For nonadiabatic processes it is, in general, untrue that the work done is equal to the increase in internal energy. We define a mapping $Q : \Pi \to R^1$ by $Q(\pi) = U(b) - U(a) - W(\pi)$, where $a = \mathcal{I}(\pi)$ and $b = \mathcal{F}(\pi)$

and we call $Q(\pi)$ the *heat absorbed in* π. It follows that the increase in internal energy in any process is the sum of the work done and the heat absorbed in the process and that the heat absorbed in an adiabatic process is zero.

3. ADIABATIC ACCESSIBILITY

The axioms introduced so far are compatible with any state being adiabatically accessible from any other. In this section, we impose restrictions on adiabatic accessibility, and, in particular, we require that in any neighborhood of a state there be adiabatically inaccessible states. Before we can make this precise, we must define what is meant by a *neighborhood*, which means that we must impose on the set of states the structure of a topological space.

In this section we shall not be concerned with the whole set of states Σ, but with a subset $\mathcal{S}$ of Σ. (It will turn out that $\mathcal{S}$ is the set of equilibrium states of the system, see Secs. 4, 6.) We assume that there exists a set Λ whose elements are subsets of $\mathcal{S}$, and such that (i) $\mathcal{S} \in \Lambda$, (ii) $\varnothing \in \Lambda$, where $\varnothing$ is the null set, (iii) $\forall B_1, B_2, \cdots, B_n \in \Lambda: B_1 \cap B_2 \cap \cdots \cap B_n \in \Lambda$, (iv) $\forall K \subset \Lambda: \bigcup_{B \in K} B \in \Lambda$. Here $\cap$ denotes the intersection of sets, $\bigcup$ their union, and $A \subset B$ means that A is a subset of B. In words, (iii) says that any intersection of a finite number of elements of Λ is an element of Λ, and (iv) says that any union, finite or not, of elements of Λ is an element of Λ. The pair $(\mathcal{S}, \Lambda)$ is called a *topological space*, and the elements of Λ are called *open sets*. By abuse of language, we speak of the *topological space* $\mathcal{S}$ and the *open sets of* $\mathcal{S}$. In most applications of the theory, one can label the states of a system by sets of n coordinates, that is, by points of R^n. The open sets are then defined in the obvious way in terms of the usual open sets of R^n. Although coordinates are probably necessary to specify the topology in applications of the theory, they need not be introduced into the formal structure.

We need very little topology. We recall that, for any sets A and B, one defines $B - A$ to be the set of elements of B which are not elements of A; that is, $B - A = \{a \in B \mid a \notin A\}$. If $(\mathcal{S}, \Lambda)$ is a topological space and $A \subset \mathcal{S}$, then $\operatorname{Fr} A$, the *frontier of* A, is defined to be the set of points $b \in \mathcal{S}$ such that any open set that contains b also contains points of A and $\mathcal{S} - A$; i.e.,

$$b \in \operatorname{Fr} A \Leftrightarrow b \in \mathcal{S} \;\&\; \forall B \in \Lambda (b \in B \Rightarrow \exists a, c \in B: a \in A \;\&\; c \in \mathcal{S} - A).$$

If $\operatorname{Fr} A \subset A$, then A is said to be *closed*. One proves easily that A is closed iff $\mathcal{S} - A$ is open.

For any state $a \in \mathcal{S}$, we consider the set $A(a)$ of all states adiabatically accessible from a; that is,

$$A(a) = \{b \in \mathcal{S} \mid a \Rightarrow b\}.$$

We assume that $A(a)$ is closed for all $a \in \mathcal{S}$. Note that this assumption is made purely for mathematical convenience, and usually has no physical content (there is no way to test it experimentally). We call $\operatorname{Fr} A(a)$ the *frontier set of* a.

We assume the following axioms for all $a, b \in \mathcal{S}$:

VI. $a \in \operatorname{Fr} A(a)$;

VII. $b \in \operatorname{Fr} A(a) \Rightarrow a \in \operatorname{Fr} A(b)$.

Axiom VI is called the *principle of Carathéodory*. It implies that in any neighborhood of a state a, i.e., in any open set that contains a, there are states not adiabatically accessible from a. Axiom VII is a slight extension of III. If $b \in \operatorname{Fr} A(a)$, then in any neighborhood B of b there is a state c_B which is not adiabatically accessible from a. From III, a is adiabatically accessible from c_B, or $a \in A(c_B)$. It does not follow from III that a is adiabatically accessible from b, the *limit point* of the states c_B, but this is implied by VII, since $a \in \operatorname{Fr} A(b) \Rightarrow a \in A(b)$.

We now prove that two states are mutually adiabatically accessible iff all states adiabatically accessible from one are adiabatically accessible from the other and iff one belongs to the frontier set of the other.

Theorem 4:

$$\forall a, b \in \mathcal{S}: a \rightharpoonup b \;\&\; b \rightharpoonup a \Leftrightarrow A(a) = A(b) \Leftrightarrow b \in \operatorname{Fr} A(a).$$

Proof: (i) If $a \rightharpoonup b$ and $c \in A(b)$, then $a \rightharpoonup b \;\&\; b \rightharpoonup c$, and $a \rightharpoonup c$ from II. Thus $c \in A(a)$ and $A(b) \subset A(a)$. Similarly, $b \rightharpoonup a \Rightarrow A(a) \subset A(b)$, and hence $a \rightharpoonup b \;\&\; b \rightharpoonup a \Rightarrow A(a) = A(b)$. (ii) If $A(a) = A(b)$, then $\operatorname{Fr} A(a) = \operatorname{Fr} A(b)$, and $b \in \operatorname{Fr} A(a)$ from VI. (iii) If $b \in \operatorname{Fr} A(a)$, then $b \in A(a)$ since $A(a)$ is closed, and it follows that $a \rightharpoonup b$. From VII, $b \in \operatorname{Fr} A(a) \Rightarrow a \in \operatorname{Fr} A(b)$, and hence $b \rightharpoonup a$. Thus $b \in \operatorname{Fr} A(a) \Rightarrow a \rightharpoonup b \;\&\; b \rightharpoonup a$. The proof is completed by using the transitive property of $\Rightarrow$.

If $\mathfrak{Q}$ is a set of sets and A is a set, then one says that $\mathfrak{Q}$ *partitions* A if every element of A is an element of exactly one of the elements of $\mathfrak{Q}$. (As an example, take $A = R^2$ and $\mathfrak{Q}$ to be the set of all lines in R^2 parallel to a given line.)

Theorem 5: The set $\mathfrak{F} = \{\operatorname{Fr} A(a) \mid a \in \mathcal{S}\}$ partitions $\mathcal{S}$.

Proof: From VI, $a \in \operatorname{Fr} A(a)$, and so every element of $\mathcal{S}$ is an element of one of the elements of $\mathfrak{F}$. If $a \in \operatorname{Fr} A(b)$, then $A(a) = A(b)$ from Theorem 4, and

hence Fr $A(b)$ = Fr $A(a)$. We call $\mathfrak{F}$ the *set of frontier sets.*

4. ANERGIC PROCESSES

It is not obvious how many frontier sets there are, nor whether one can label them with one or more real parameters. We prove in the next section that, in fact, one real parameter is sufficient. But, first, we must introduce a new set of processes called *anergic processes* (they are a slight generalization of what Buchdahl[3] calls *isometric processes*).

One can partially characterize an anergic process as a process in which no external force does work on the system in any time interval. (Very roughly, one can say that anergic processes are "no work" processes, just as adiabatic processes are "no heat" processes.) This characterization is sufficient for some systems, but not for all. To be more precise, we assume that the states of Σ can be labeled by a set of coordinates. That is, there are functions $x_i: \Sigma \to R^1$, $i = 1, 2, \cdots$, such that the mapping $a \to (x_1(a), x_2(a), \cdots)$ is an injection (*or one-to-one*). We assume that the cooordinates can be chosen so that: (i) $x_1 = U$, the internal energy; (ii) there is a subset $I \subset \{2, 3, \cdots\}$ such that the x_i with $i \in I$ correspond to physical quantities that can be measured at all times (not only when the system is in one of the states of Σ); (iii) in any process in which the x_i are constant at all times for all $i \in I$, the work done by each external force is zero in every time interval.

We can now characterize an anergic process from a to b as a process in which x_i has the same value whenever it can be measured, for all $i > 1$. In particular, the x_i are constant at all times for $i \in I$, and $x_i(a) = x_i(b)$ for all $i > 1$. It follows from (iii) that this is compatible with the previous, partial characterization.

The coordinates x_i are useful in characterizing anergic processes, but they play no part in the formal theory, where anergic processes are undefined elements. Formally, we assume that there exists a set $\mathrm{N} \subset \Pi$ whose elements are called *anergic processes* and that the following axioms are satisfied:

VIII. $\forall\, a \in \Sigma\ \exists\, \pi \in \mathrm{N}: \mathfrak{I}(\pi) = \mathfrak{F}(\pi) = a$;

IX. $\forall\, \pi, \sigma \in \mathrm{N}\ \exists\, \tau \in \mathrm{N}: \mathfrak{F}(\pi) = \mathfrak{I}(\sigma) \Rightarrow \mathfrak{I}(\tau) = \mathfrak{I}(\pi)\ \&\ \mathfrak{F}(\tau) = \mathfrak{F}(\sigma)$;

X. $\forall\, \pi \in \mathrm{N}: W(\pi) = 0$.

Axioms VIII and IX are similar to I and II for adiabatic processes. Note that X is not a definition of anergic processes; i.e., it is possible to have $W(\pi) = 0$, $\pi \notin \mathrm{N}$. If π is an anergic process from a to b, then the heat absorbed in π is $Q(\pi) = U(b) - U(a) - W(\pi) = U(b) - U(a)$ by X. Thus, the heat absorbed is the same in all anergic processes from a to b.

In what follows, we restrict ourselves once more to the subset $\mathcal{S}$ of Σ.

Let dom q be the set of all $(a, b) \in \mathcal{S} \times \mathcal{S}$ such that there is an anergic process from a to b [that is, $(a, b) \in \mathrm{dom}\, q \Leftrightarrow a, b \in \mathcal{S}\ \&\ \exists\, \pi \in N: \mathfrak{I}(\pi) = a\ \&\ \mathfrak{F}(\pi) = b$].

Define a mapping q; dom $q \to R^1$ by $q(a, b) = U(b) - U(a)$. From the remarks above, $q(a, b)$ is the heat absorbed in any anergic process from a to b. Axiom VIII implies that $(a, a) \in \mathrm{dom}\, q$ for all $a \in \mathcal{S}$. Axiom IX implies that if (a, b), $(b, c) \in \mathrm{dom}\, q$, then $(a, c) \in \mathrm{dom}\, q$, and, from the definition of q, one has $q(a, c) = q(a, b) + q(b, c)$.

If there is an anergic process from a to b in which the heat absorbed in nonnegative, then the next axiom asserts that there is also an adiabatic process from a to b. If the heat absorbed is negative, then there is no such adiabatic process.

XI. $\forall\, (a, b) \in \mathrm{dom}\, q: q(a, b) \geq 0 \Leftrightarrow a \rightharpoonup b$.

If π is an energic process from a to b in which the heat absorbed is k, then, in terms of our previous informal characterization, we have $x_1(b) = U(b) = U(a) + k = x_1(a) + k$. Since $x_i(b) = x_i(a)$ for $i > 1$, it follows that the final state b of the anergic process is uniquely[11] determined by the initial state a and by the heat absorbed k. It also seems plausible that, by suitably choosing k, one should be able to go from a to some state in any frontier set. Fortified by these heuristic arguments, we assume the following axiom:

XII. $\forall\, a \in \mathcal{S}, \forall\, \beta \in \mathfrak{F}\ \exists!\, b \in \beta: (a, b) \in \mathrm{dom}\, q$.

It is easy to prove that there is no anergic process from a state b to a distinct state c belonging to the same frontier set and to show that, if there is an anergic process from a to b, then there is one from b to a.

Theorem 6:

$\forall\, \beta \in \mathfrak{F}, \forall\, a, b \in \beta: (a, b) \in \mathrm{dom}\, q \Leftrightarrow a = b$.

Proof: (i) If $a = b$, then $(a, b) \in \mathrm{dom}\, q$ by VIII. (ii) Assume $b \in \beta$, $(a, b) \in \mathrm{dom}\, q$. From XII, b is unique. But $(a, a) \in \mathrm{dom}\, q$ from (i), and $a \in \beta$. Hence, $a = b$.

Theorem 7:

$\forall\, a, b \in \mathcal{S}: (a, b) \in \mathrm{dom}\, q \Rightarrow (b, a) \in \mathrm{dom}\, q$.

Proof: Let $a \in \alpha \in \mathfrak{F}$. From XII, $\exists!\, c \in \alpha: (b, c) \in \mathrm{dom}\, q$. Since $(a, b) \in \mathrm{dom}\, q$, IX implies that $(a, c) \in \mathrm{dom}\, q$, and $c = a$ by Theorem 6. Since $q(a, b) = U(b) - U(a)$, one has $q(a, b) = -q(b, a)$.

The set of anergic *and* adiabatic processes is very restricted. In fact, if the initial state of such a process belongs to $\mathcal{S}$, the final state is the same as the initial state (the process is cyclic).

Theorem 8:

$$\forall a \in \mathcal{S}, \forall \lambda \in \Delta \cap \mathrm{N}: \mathfrak{I}(\lambda) = a \Rightarrow \mathcal{F}(\lambda) = a.$$

Proof: Suppose that $a \in \mathcal{S}$ and λ is an anergic and adiabatic process such that $\mathfrak{I}(\lambda) = a$ and $\mathcal{F}(\lambda) = c$. Since λ is anergic and adiabatic, $(a, c) \in \mathrm{dom}\, q$ and $q(a, c) = 0$. From Theorem 7, $(c, a) \in \mathrm{dom}\, q$, and $q(c, a) = 0$. Hence, $a \rightharpoonup c$ & $c \rightharpoonup a$ from XI, and $c \in \mathrm{Fr}\, A(a)$ from Theorem 4. It follows from Theorem 6 that $c = a$.

Physically speaking, we say that a system is *isolated* if it is thermally insulated and no force does work on it in any time interval. We characterize an *equilibrium state* as a state in which an isolated system can remain indefinitely. The processes which an isolated system undergoes must be adiabatic, but they need not be anergic. In our characterization of anergic processes at the beginning of this section, we required both that the coordinates x_i for $i \in I$ should be constant (which is all that is needed to ensure that no force does work) and that each of the x_i with $i > 1$ should have the same value at the beginning and at the end of the process.[12] The latter condition need not always be satisfied by an isolated system. However, it is very often satisfied for the states of $\mathcal{S}$ (in the simplest cases there are no x_i such that $i > 1$, $i \notin I$). Theorem 8 implies that in such cases all the states of $\mathcal{S}$ are equilibrium states.

If there are anergic processes from a to b and from c to d, where a and c belong to the frontier set α and b and d to the frontier set β, then the heat absorbed $q(a, b)$ need not be the same as $q(c, d)$, but it does have the same sign.

Theorem 9:

$$\forall \alpha, \beta \in \mathcal{F}, \forall (a, b), (c, d) \in \mathrm{dom}\, q: a, c \in \alpha \;\&\; b, d \in \beta \Rightarrow (q(a, b) \geq 0 \Leftrightarrow q(c, d) \geq 0).$$

Proof: Suppose not: Suppose $q(a, b) \geq 0$ and $q(c, d) < 0$. From Theorem 7, $(d, c) \in \mathrm{dom}\, q$, and $q(d, c) = -q(c, d) > 0$. It follows from XI that $a \rightharpoonup b$ and $d \rightharpoonup c$. Since $c \in \alpha = \mathrm{Fr}\, A(a)$, one has $c \rightharpoonup a$ by Theorem 4, and similarly $b \rightharpoonup d$. From II, $a \rightharpoonup d$ & $d \rightharpoonup a$, and $d \in \mathrm{Fr}\, A(a) = \alpha$ from Theorem 4. Since c, $d \in \alpha$, and $(c, d) \in \mathrm{dom}\, q$, Theorem 6 implies that $c = d$ and $q(c, d) = U(d) - U(c) = 0$. This contradicts the assumption $q(c, d) < 0$. For the case $q(c, d) \geq 0, q(a, b) < 0$, one has only to exchange the roles of (a, b) and (c, d).

5. EMPIRICAL ENTROPY

We now label the frontier sets with a real parameter; that is, we define a real function on $\mathcal{F}$. One can, in fact, do this in many ways. For example, given any state $a \in \mathcal{S}$, one may take the value of the function at the frontier set β to be the heat absorbed in an anergic process from a to a state of β. If b is any state of β, then we call the value of the function at β an *empirical entropy of b*. Different choices of function give different empirical entropies, but we prove that in all cases the *principle of increase of entropy* is satisfied (Theorem 12).

Given any $a \in \mathcal{S}$, it follows from XII that $\forall \beta \in \mathcal{F}\, \exists!\, b \in \beta : (a, b) \in \mathrm{dom}\, q$. Hence one can define a mapping $f_a: \mathcal{F} \to R^1$ by $f_a(\beta) = q(a, b)$. We prove that f_a is an injection.

Theorem 10:

$$\forall a \in \mathcal{S}, \forall \beta, \gamma \in \mathcal{F}: f_a(\beta) = f_a(\gamma) \Rightarrow \beta = \gamma.$$

Proof: From XII, there exist unique states $b \in \beta$, $c, c' \in \gamma$ such that (a, b), (a, c), $(b, c') \in \mathrm{dom}\, q$. From IX, $(a, c') \in \mathrm{dom}\, q$, and hence $c = c'$ by XII. Since $q(a, c) = q(a, b) + q(b, c)$, $q(a, c) = f_a(\gamma)$, $q(a, b) = f_a(\beta)$, and $f_a(\beta) = f_a(\gamma)$, one has $q(b, c) = 0$, and $b \rightharpoonup c$ from XI. From Theorem 7 and XI, $(c, b) \in \mathrm{dom}\, q$, $q(c, b) = 0$, and $c \rightharpoonup b$. Theorem 4 implies $b \in \mathrm{Fr}\, A(c)$, and $\beta = \gamma$.

Corollary: From the proof of the theorem and the equation $q(b, c) = f_b(\gamma)$, it follows that

$$\forall a \in \mathcal{S}, \forall \beta, \gamma \in \mathcal{F}\, \exists\, b \in \beta: f_a(\gamma) = f_a(\beta) + f_b(\gamma).$$

If $a \in \alpha \in \mathcal{F}$, then, from Theorem 6, $f_a(\alpha) = q(a, a) = 0$. Conversely, from Theorem 10, if $f_a(\alpha) = 0$, then $a \in \alpha$.

From Theorem 9 one can deduce conditions that must be satisfied by any pair of functions f_a, f_b.

Theorem 11:

$$\forall a, b \in \mathcal{S}, \forall \beta, \gamma \in \mathcal{F}: f_a(\gamma) - f_a(\beta) \geq 0 \Leftrightarrow f_b(\gamma) - f_b(\beta) \geq 0.$$

Proof: From the Corollary to Theorem 10, there exist c, $d \in \beta$ such that $f_a(\gamma) = f_a(\beta) + f_c(\gamma)$, $f_b(\gamma) = f_b(\beta) + f_d(\gamma)$. From XII there exist c', $d' \in \gamma$ such that $f_c(\gamma) = q(c, c')$, $f_d(\gamma) = q(d, d')$. Theorem 9 implies that $f_c(\gamma) \geq 0 \Leftrightarrow f_d(\gamma) \geq 0$, and the theorem follows.

Since f_a is an injection, the inverse f_a^{-1} exists for all $a \in \mathcal{S}$. The transformation functions F_{ba} can therefore

be defined by $F_{ba} = f_b \circ f_a^{-1}$ for all $a, b \in \mathcal{S}$. We assume that the F_{ba} are differentiable. That is, the derivative $F'_{ba}(x)$ exists for all $x \in f_a(\mathcal{T})$. One proves from Theorem 11 that F'_{ba} is everywhere nonnegative. Since F'_{ab} exists everywhere, it follows that F'_{ba} has no zeros, and is therefore strictly positive for all $a, b \in \mathcal{S}$.

It is convenient to consider a set of functions which includes the f_a as a subset. We define $f: \mathcal{T} \to R^1$ to be a *parametrization* of $\mathcal{T}$ if for some $a \in \mathcal{S}$, one has $f = F_a^{-1} \circ f_a$, where the functions F_a and F_a^{-1} are differentiable and F'_a is strictly positive. It follows at once that f is an injection. From the results of the last paragraph we see that parameterizations can be defined in terms of any of the functions f_a and that f_a is itself a parametrization of $\mathcal{T}$ for all $a \in \mathcal{S}$.

If f is a parametrization of $\mathcal{T}$ and $a \in \mathcal{S}$, there is a unique $\alpha \in \mathcal{T}$ such that $a \in \alpha$. One can therefore define a function $Z: \mathcal{S} \to R^1$ by $Z(a) = f(\alpha)$. We call Z an *empirical entropy*, and $Z(a) = f(\alpha)$ is an *empirical entropy of a*. Note that different parametrizations of $\mathcal{T}$ give rise to different empirical entropies. If we speak of *the* empirical entropy, it is to be understood that we are considering a definite parametrization. The physical significance of empirical entropy is shown by the following theorem.

Theorem 12 (Principle of Increase of Entropy): For all $a, b \in \mathcal{S}$, b is adiabatically accessible from a iff the empirical entropy of b is not less than that of a.

Proof: (i) If $a, b \in \alpha \in \mathcal{T}$, then $a \rightharpoonup b$ by Theorem 4. Since the empirical entropy of a is the same as that of b, the theorem is trivially satisfied. (ii) (a) Suppose $a \in \alpha$, $b \in \beta$, where $\alpha, \beta \in \mathcal{T}$, $\alpha \neq \beta$. From XII and Theorem 7, $\exists!\, c \in \alpha: (b, c), (c, b) \in \operatorname{dom} q$. Since $\alpha \neq \beta, f_b(\alpha) = q(b, c) \neq 0$ and $f_c(\beta) = q(c, b) \neq 0$. Hence by XI, $f_b(\alpha) > 0 \Leftrightarrow b \rightharpoonup c$, and $f_c(\beta) > 0 \Leftrightarrow c \rightharpoonup b$. Now let f be any parameterization of $\mathcal{T}$. One has $f_b = F_b \circ f$, where F'_b is strictly positive, and it follows from the mean value theorem that $f_b(\alpha) - f_b(\beta) = K(f(\alpha) - f(\beta))$, where $K > 0$. Since $f_b(\beta) = 0$, it follows that $f(\alpha) - f(\beta) > 0 \Leftrightarrow b \rightharpoonup c$. Similarly, or from III, one shows $f(\alpha) - f(\beta) < 0 \Leftrightarrow c \rightharpoonup b$. (ii) (b) Assume that $a \rightharpoonup b$ and $f(\alpha) - f(\beta) > 0$. From (a), $b \rightharpoonup c$ and, from Theorem 4, $c \rightharpoonup a$. Hence $a \rightharpoonup b \,\&\, b \rightharpoonup a$, and $\alpha = \beta$, from II and Theorem 4. This contradicts the assumption $\alpha \neq \beta$, and it follows that $a \rightharpoonup b \Rightarrow f(\alpha) - f(\beta) < 0$. Conversely, assume that $f(\alpha) - f(\beta) < 0$. From (a), $c \rightharpoonup b$, and, since $a \rightharpoonup c$ by Theorem 4, one has $a \rightharpoonup b$ from II.

To understand the physical significance of the F'_a, let f be a parameterization of $\mathcal{T}$ and write $f(\alpha) = z$ and $f(\beta) = z + h$, so that $f_a(\alpha) = F_a(z)$ and $f_a(\beta) = F_a(z + h)$. From the Corollary to Theorem 10, there exists $b \in \alpha$ such that $f_a(\beta) = f_a(\alpha) + f_b(\beta)$, and, from XII, there exists $c \in \beta$ such that $f_b(\beta) = q(b, c) = U(c) - U(b)$. It follows that

$$f_a(\beta) - f_a(\alpha) = F_a(z + h) - F_a(z) = U(c) - U(b)$$

and

$$hF'_a(z) = U(c) - U(b) + o(h) = q(b, c) + o(h)$$

as $h \to 0$. Thus $F'_a(z)$ is the rate of change of internal energy with respect to z in an anergic process or the heat absorbed per unit change of z in an anergic process.

6. STATES HAVING THE SAME TEMPERATURE

One can often give empirical meaning to the statement "state a has the same temperature as state b." The operations by which one verifies such a statement usually involve some kind of thermometer, but they do not require any particular temperature scale. (We can graduate the thermometer arbitrarily.) One can imagine a system for which two states have the same temperature iff they have the same empirical entropy. This we exclude. We assume, in fact, that there are states which have the same temperature and *any* values of the empirical entropy.

The above ideas are formalized in terms of an *equivalence relation* on the set of states $\mathcal{S}$. We recall that an equivalence relation on a set B is a subset R of $B \times B$ such that, for all $a, b, c \in B$, $(a, a) \in R$, $(a, b) \in R \Rightarrow (b, a) \in R$, and $(a, b) \in R \,\&\, (b, c) \in R \Rightarrow (a, c) \in R$. The axiom is the following.

XIII. There exists an equivalence relation $\mathcal{C}$ on $\mathcal{S}$ and a state $a_0 \in \mathcal{S}$ such that $\forall\, \beta \in \mathcal{T}\ \exists\, b \in \beta: a_0 \mathcal{C} b$.

If $(a, b) \in \mathcal{C}$, then we write $a \mathcal{C} b$ and say that *a has the same temperature as b*. It follows that for all $a, b, c \in \mathcal{S}$, $a \mathcal{C} a$, $a \mathcal{C} b \Rightarrow b \mathcal{C} a$, and $a \mathcal{C} b \,\&\, b \mathcal{C} c \Rightarrow a \mathcal{C} c$. One can, of course, replace a_0 in XIII by any state $b_0 \in \mathcal{S}$ such that $b_0 \mathcal{C} a_0$. The existence of the equivalence relation $\mathcal{C}$ is roughly equivalent to what is called the *zeroth law of thermodynamics* in other treatments of the subject.

For the sake of simplicity, we shall use Axiom XIII, but one can develop the theory on the basis of a more complicated but much weaker assumption (see XIII′ below). Roughly speaking, the idea is that one should demand that XIII hold, not for all sets of $\mathcal{T}$, but only for neighboring sets. To make this precise, let f be a parameterization of $\mathcal{T}$ and J be a finite or infinite set of consecutive integers; e.g., one might have $J = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$. For all $n \in J$ let $I_n \subset R^1$ be an open interval such that $\bigcup_{n \in J} I_n = f(\mathcal{T})$, $I_n \cap I_{n+1} \neq \phi$, and $I_n \cap I_{n+2} = \varnothing$ for all n,

$n+1$, $n+2 \in J$. Define $\mathfrak{I}_n = \{\alpha \in \mathfrak{I} \mid f(\alpha) = I_n\}$. We then have $\bigcup_{n\in J} \mathfrak{I}_n = \mathfrak{I}$, $\mathfrak{I}_n \cap \mathfrak{I}_{n+1} \neq \varnothing$, and $\mathfrak{I}_n \cap \mathfrak{I}_{n+2} = \varnothing$ for all n, $n+1$, $n+2 \in J$. A set $\mathfrak{Q} = \{\mathfrak{I}_n \mid n \in J\}$, where J and $\mathfrak{I}_n$ satisfy the above conditions, is called a *countable, open covering of* $\mathfrak{I}$. We can now state the modified form of Axiom XIII.

XIII′. There exists an equivalence relation $\mathcal{C}$ on $\mathcal{S}$ and a countable open covering $\mathfrak{Q}$ of $\mathfrak{I}$ such that $\forall\, \mathfrak{I}_n \in \mathfrak{Q}\ \exists\, a_n \in \alpha \in \mathfrak{I}_n, \forall\, \beta \in \mathfrak{I}_n\ \exists\, b \in \beta : a_n \mathcal{C} b$.

7. TEMPERATURE AND ENTROPY

The parameterizations f_a of $\mathfrak{I}$ and the corresponding empirical entropies depend on an arbitrary choice of state a, and they cannot be expected to have any deep physical significance. In order to find a nonarbitrary entropy function and also a nonarbitrary temperature function, we use another of Carathéodory's ideas. We consider a compound system that consists of two identical subsystems which have the same temperature (a more precise statement is given below). We assume that all the previous theory applies both to the compound system and to the subsystems. We assume too that certain specified processes of the subsystems correspond to adiabatic processes of the compound system. One can then show that the empirical entropy of the compound system is a function only of the empirical entropies of the subsystems and that the parametrizations f_a must satisfy a fairly restrictive condition (Axiom XIV below). Using this result, one easily proves the existence of nonarbitrary temperature and entropy functions.

In carrying out the program just outlined, one has a choice of tactics: Either one can construct a completely rigorous theory of compound systems, or one can give a heuristic argument to make it plausible that the parametrizations f_a should satisfy a certain condition, and can then adopt this condition as an axiom of the formal theory. We choose the second alternative because it is simpler. (A brief, formal account of compound systems is given in the Appendix.)

More formally, one considers a set $\Sigma_C = \Sigma \times \Sigma$ and a set $\mathcal{S}_C \subset \mathcal{S} \times \mathcal{S} \subset \Sigma_C$ such that $(a, b) \in \mathcal{S}_C \Leftrightarrow a \mathcal{C} b$, where Σ and $\mathcal{S}$ are sets of states of a subsystem and $\mathcal{C}$ is the relation on $\mathcal{S}$ that satisfies Axiom XIII. We assume that all the theory that has been developed up to now is valid for Σ, $\mathcal{S}$, and the sets of processes Π, Δ, and N and that it is also valid for Σ_C, $\mathcal{S}_C$, and the corresponding sets of processes Π_C, Δ_C, and N_C. (If $\lambda_C \in \Delta_C$, we say that λ_C is an *adiabatic process of the compound system*, etc.) The relation $\mathcal{C}$ on $\mathcal{S}$ and the corresponding relation $\mathcal{C}_C$ on $\mathcal{S}_C$ are assumed to satisfy $(a, b)\mathcal{C}_C(c, d) \Leftrightarrow a\mathcal{C}c$ for all (a, b), $(c, d) \in \mathcal{S}_C$.

Let us denote adiabatic accessibility in the compound system by $\rightarrow_C$. We assume that adiabatic processes in the subsystems induce an adiabatic process in the compound system: $a \rightarrow b\ \&\ c \rightarrow d \Rightarrow (a, c) \rightarrow_C (b, d)$ for all (a, c), $(b, d) \in \mathcal{S}_C$. If in addition $a, b \in \alpha$ and $c, d \in \beta$, where $\alpha, \beta \in \mathfrak{I}$, then Theorem 4 implies that $a \rightarrow b\ \&\ b \rightarrow a$ and $c \rightarrow d\ \&\ d \rightarrow c$. It follows that $(a, c) \rightarrow_C (b, d)$ and $(b, d) \rightarrow_C (a, c)$, and Theorem 4 for the compound system implies that (a, c) and (b, d) belong to the same frontier set: (a, c), $(b, d) \in \gamma_C \in \mathfrak{I}_C$. We have therefore proved that to each $(\alpha, \beta) \in \mathfrak{I} \times \mathfrak{I}$ there corresponds a unique $\gamma_C \in \mathfrak{I}_C$ such that $a \in \alpha\ \&\ b \in \beta \Rightarrow (a, b) \in \gamma_C$ for all $(a, b) \in \mathcal{S}_C$, and we can define a mapping $\kappa : P \times P \rightarrow P_C$, $\kappa(\alpha, \beta) = \gamma_C$. In terms of parametrizations f of $\mathfrak{I}$ and f_C of $\mathfrak{I}_C$, one writes $z_C = f_C(\gamma_C)$, $z_1 = f(\alpha)$, $z_2 = f(\beta)$, and

$$g = f \circ \kappa \circ (f^{-1}, f^{-1}),$$

and one has $z_C = g(z_1, z_2)$. Since z_1, z_2, and z_C are the empirical entropies of states $a \in \alpha$, $b \in \beta$, $(a, b) \in \gamma_C$, respectively, we have shown that the empirical entropy of any state $(a, b) \in \mathcal{S}_C$ is determined by the empirical entropies of a and b.

We do not assume that the only adiabatic processes of the compound system are those induced by adiabatic processes of the subsystems. Since the compound system absorbs no heat in an adiabatic process, it may seem reasonable that the amount of heat absorbed by one subsystem should be minus the amount absorbed by the other. However, an implicit assumption here is that the heat absorbed by the compound system is the sum of the heats absorbed by the subsystems, which is not always true. We shall assume it to be true in the special case when the subsystems undergo anergic processes and are always at the same, or almost the same, temperature.

To be specific, we consider an adiabatic process of the compound system from $(a, b) \in \mathcal{S}_C$ to $(c, d) \in \mathcal{S}_C$, where both states have the same empirical entropy z_C. The empirical entropies of the states a, b, c, and d of the subsystems are z_1, z_2, $z_1 + \delta z_1$, and $z_2 + \delta z_2$, respectively. One subsystem undergoes an anergic process from the state a to a state a' with empirical entropy $z_1 + \delta z_1$ and then an adiabatic process from a' to c. The other subsystem undergoes an anergic process from b to a state b', with empirical entropy $z_2 + \delta z_2$, and then an adiabatic process from b' to d. We assume that, for small δz_1, and δz_2, the heat absorbed by the compound system is the sum of the heats absorbed by the subsystems. Since the compound system undergoes an adiabatic process, this

implies that $F_a'(z_1)\delta z_1 + F_b'(z_2)\delta z \approx 0$ (cf. the end of Sec. 5). Since $z_C = g(z_1, z_2)$ and the states (a, b) and (c, d) have the same empirical entropy, we have

$$\delta z_C = 0 \approx \frac{\partial g}{\partial z_1}(z_1, z_2)\delta z_1 + \frac{\partial g}{\partial z_2}(z_1, z_2)\delta z_2.$$

Provided that $\partial g/\partial z_1$ and $\partial g/\partial z_2$ do not vanish, we can solve these equations, and find

$$\frac{F_a'(z_1)}{F_b'(z_2)} = \frac{\partial g}{\partial z_1}(z_1, z_2) \Big/ \frac{\partial g}{\partial z_2}(z_1, z_2).$$

The right-hand side of this equation depends only on z_1 and z_2. On the left-hand side, a and b are any states of $\mathcal{S}$ with empirical entropies z_1 and z_2, respectively, subject only to the condition that a has the same temperature as b [since $(a, b) \in \mathcal{S}_C$]. It follows that, if $i, j \in \mathcal{S}$ have empirical entropies z_1 and z_2, respectively, and i has the same temperature as j, then $F_a'(z_1)/F_b'(z_2) = F_i'(z_1)/F_j'(z_2)$.

We emphasize that the argument of the last paragraphs is not rigorous and is not part of our axiomatic structure. Its purpose is only to make the next axiom plausible.

XIV. $\forall\, z, w \in f(\mathcal{F}), \forall\, a, c \in f^{-1}(z), \forall\, b, d \in f^{-1}(w)$: $a\mathcal{C}b \;\&\; c\mathcal{C}d \Rightarrow F_a'(z)/F_b'(w) = F_c'(z)/F_d'(w)$.

We have stated XIV in terms of a parameterization f of $\mathcal{F}$ and the functions $F_a = f_a \circ f^{-1}$, but one shows easily that it is, in fact, independent of the choice of f and represents a condition imposed on the f_a. To see the physical meaning of XIV, use the interpretation of F_a' at the end of Sec. 5.

As a special case of XIV, we put $z = w$ and $a = b$. One then has $c\mathcal{C}d \Rightarrow F_c'(z) = F_d'(z)$ for all $z \in f(\mathcal{F})$ and for all $c, d \in f^{-1}(z)$.

A *temperature function* is defined to be a mapping $T: \mathcal{S} \to R^1$ that satisfies the conditions (i) $\forall\, a, b \in \mathcal{S}: a\mathcal{C}b \Rightarrow T(a) = T(b)$, (ii) $\forall\, \alpha \in \mathcal{F}, \forall\, a, b \in \alpha: z = f(\alpha) \Rightarrow T(b) = T(a)F_b'(z)/F_a'(z)$, and (iii) $\exists\, a \in \mathcal{S}$: $T(a) > 0$. We note that (ii) is independent of the choice of parametrization f.

To prove the existence of a temperature function, we first choose $T(a_0) > 0$, where a_0 is the state that appears in XIII. If $d \in \mathcal{S}$, then there exists a unique $\beta \in \mathcal{F}$ such that $d \in \beta$. From XIII, there exists $b \in \beta$ such that $a_0\mathcal{C}b$. Define $T(d) = T(a_0)F_d'(w)/F_b'(w)$, where $w = f(\beta)$. (Recall that F_b' is strictly positive.) This definition is independent of the choice of the state b, since, if $c \in \beta$, and $a_0\mathcal{C}c$, then $b\mathcal{C}c$ and $F_b'(w) = F_c'(w)$ from XIV. Again using XIV, one verifies that (i), (ii), and (iii) are satisfied. Since F_a' is strictly positive for all $a \in \mathcal{S}$, it follows that T is strictly positive.

If $\bar{T}$ is any temperature function, then one can define a function $r = \bar{T}/T$ (this makes sense because T is strictly positive). The conditions (i) and (ii) imply that $a\mathcal{C}b \Rightarrow r(a) = r(b)$ for all $a, b \in \mathcal{S}$ and that $r(a) = r(b)$ for all $\alpha \in \mathcal{F}$ and for all $a, b \in \alpha$. It follows from XIII that $r(a) = \lambda$, a constant, for all $a \in \mathcal{S}$, and hence $\bar{T} = \lambda T$. From (iii), one has $\lambda > 0$. We have therefore proved that a temperature function is uniquely determined by conditions (i), (ii), (iii), and by its value (necessarily positive) at a single state.

If one uses Axiom XIII′ instead of XIII, the proof of the existence and uniqueness of a temperature function is only slightly more complicated. The proof already given establishes the existence and uniqueness properties of temperature functions restricted to the sets $\mathcal{F}_n$, and one has only to show that these functions can be chosen so that they coincide on the intersections of their domains (proof by induction on n).

From condition (ii), one sees that for any temperature function T and for any $\alpha = f^{-1}(z) \in \mathcal{F}$ the ratio $F_b'(z)/T(b)$ is constant for all $b \in \alpha$. One can therefore define a function $s': f(\mathcal{F}) \to R^1$ by $s'(z) = F_b'(z)/T(b)$, where $b \in f^{-1}(z)$. We assume that s' is continuous on $f(\mathcal{F})$, and it follows that there exists a function s on $f(\mathcal{F})$ whose derivative is s'. Since the functions F_b' and T are strictly positive, so is s', and this implies that $s \circ f$ is a parametrization of $\mathcal{F}$. For a given temperature function T, $s \circ f$ is arbitrary to the extent of an additive constant, but it is independent of the parametrization f (i.e., if $\bar{f}$ is another parametrization of $\mathcal{F}$ and $\bar{F}_b = f_b \circ \bar{f}^{-1}$, $\bar{s}'(z) = \bar{F}_b'(z)/T(b)$, then $\bar{s} \circ \bar{f} = s \circ f$). If T is replaced by λT, where $\lambda > 0$, then s is replaced by s/λ.

The function $S: \mathcal{S} \to R^1$ defined by $S(a) = s \circ f(\alpha)$, where $a \in \alpha$, is called the *entropy*. Like $s \circ f$, it is arbitrary to the extent of an additive constant for a given temperature function T, and it is replaced by S/λ if T is replaced by λT. Since S is an empirical entropy, it satisfies Theorem 12, the principle of increase of entropy.

It was shown in Sec. 5 that the heat absorbed in an anergic process π from $a \in f^{-1}(z)$ to $b \in f^{-1}(z + h)$ is $Q(\pi) = U(b) - U(a) = hF_a'(z) + o(h)$ as $h \to 0$. Since $hF_a'(z) = T(a)hs'(z) + o(h)$, one has $Q(\pi) = U(b) - U(a) = T(a)[S(b) - S(a)] + o[S(b) - S(a)]$ as $S(b) - S(a) \to 0$.

The existence of temperature and entropy functions is the essential content of the traditional second law of thermodynamics. For completeness, we note that the third law states that the internal energy and the entropy are bounded below, and that, as the internal energy approaches its lower bound, so does the entropy, and the temperature approaches zero.

XV. The internal energy U has a greatest lower bound glb U, the entropy S has a greatest lower bound glb S, and for all sequences

$$\{a_n \in \mathcal{S} \mid n = 1, 2, 3, \cdots\}:$$

$$\lim_{n\to\infty} U(a_n) = \text{glb } U \Rightarrow \lim_{n\to\infty} S(a_n) = \text{glb } S \;\&\; \lim_{n\to\infty} T(a_n) = 0,$$

where T is any temperature function.

One can, of course, choose the arbitrary constant in S so that glb $S = 0$.

We note finally that the theory can be easily generalized to include states of negative temperature. Instead of a single set $\mathcal{S}$, one assumes the existence of two disjoint sets $\mathcal{S}$, $\overline{\mathcal{S}}$, such that $\mathcal{S}$, $\overline{\mathcal{S}} \subset \Sigma$. The axioms satisfied by $\overline{\mathcal{S}}$ exactly parallel those satisfied by $\mathcal{S}$. In particular, XIII (or XIII') holds for $\overline{\mathcal{S}}$. However, no state of $\overline{\mathcal{S}}$ has the same temperature as any state of $\mathcal{S}$. A temperature function T is defined on $\mathcal{S}$ as before. A function $\bar{T}: \overline{\mathcal{S}} \to R^1$ is defined which satisfied the conditions (i) and (ii) for a temperature function. Instead of condition (iii), we require that there exists $a \in \overline{\mathcal{S}}$ such that $\bar{T}(a) < 0$. One can then define a generalized temperature T^* on $\mathcal{S} \cup \overline{\mathcal{S}}$ such that T^* restricted to $\mathcal{S}$ is T and T^* restricted to $\overline{\mathcal{S}}$ is $\bar{T}$.

8. CONCLUSION

We have shown that one can develop thermodynamics in a manner which is rigorous, and also quite simple and general. The distinguishing features of our treatment are the use of anergic processes, rather than quasistatic or reversible processes, and the very sparing use of arguments that involve compound systems.

The main purpose of axiomatizations of the present kind is to make clearer the physical significance of such abstract concepts as entropy and temperature. It is also possible that axiomatization may lead to generalizations of the theory (cf. the remarks following Axiom III). From a broader point of view, one can argue that no physical theory is complete until it has been axiomatized. Certainly, one cannot understand the nature of physical theories until one has performed a few such dissections.

APPENDIX

We here give an account of compound systems which is a little more complete and general than that in Sec. 7. One should note that, following Carathéodory, we use *compound system* in a very special sense. The subsystems of a compound system are not thermally insulated from one another, and the sum of their internal energies is equal to the internal energy of the compound system. Other kinds of compound system are sometimes discussed in thermodynamics.

We consider a compound system composed of two subsystems (the generalization to any number is trivial). Quantities referring to the subsystems are labeled by suffixes 1 and 2, and those referring to the compound system, by a suffix C. We introduce sets of states Σ_n, $\mathcal{S}_n$, sets of processes Π_n, Δ_n, N_n, etc., where $n = 1, 2, \mathrm{C}$, and we assume that all our previous assumptions about Σ, $\mathcal{S}$, etc., also hold for Σ_n, $\mathcal{S}_n$, etc.

We assume that there exists an equivalence relation $\overline{\mathcal{C}}$ on $\mathcal{S}_1 \cup \mathcal{S}_2$ such that the restriction of $\overline{\mathcal{C}}$ to $\mathcal{S}_P$ is $\mathcal{C}_P$, for $P = 1, 2$. The set of states Σ_C is defined by $\Sigma_\mathrm{C} = \Sigma_1 \times \Sigma_2$, and $\mathcal{S}_\mathrm{C}$ is defined by

$$\mathcal{S}_\mathrm{C} \subset \mathcal{S}_1 \times \mathcal{S}_2, \quad (a_1, a_2) \in \mathcal{S}_\mathrm{C} \Leftrightarrow a_1 \overline{\mathcal{C}} a_2.$$

The equivalence relation $\mathcal{C}_\mathrm{C}$ on $\mathcal{S}_\mathrm{C}$ is defined by

$$(a_1, a_2)\mathcal{C}_\mathrm{C}(b_1, b_2) \Leftrightarrow a_1 \overline{\mathcal{C}} b_1.$$

In accordance with the general assumption of the last paragraph, the $\mathcal{C}_n$ satisfy Axioms XIII and XIV, for $n = 1, 2, \mathrm{C}$.

If $a_n, b_n \in \Sigma_n$ and b_n is adiabatically accessible from a_n, we write $a_n \rightharpoonup_n b_n$. We assume that for all (a_1, a_2), $(b_1, b_2) \in \mathcal{S}_\mathrm{C}$

$$a_1 \rightharpoonup b_1 \;\&\; a_2 \rightharpoonup b_2 \Rightarrow (a_1, a_2) \rightharpoonup_\mathrm{C} (b_1, b_2).$$

By an easy modification of the proof of Sec. 7, one shows that there exists a mapping $\bar{\kappa}: \mathcal{F}_1 \times \mathcal{F}_2 \to \mathcal{F}_\mathrm{C}$ such that

$$a_1 \in \alpha_1 \;\&\; a_2 \in \alpha_2 \Rightarrow (a_1, a_2) \in \bar{\kappa}(\alpha_1, \alpha_2) \quad \text{for all } (a_1, a_2) \in \mathcal{S}_\mathrm{C}.$$

Further, if f_n is a parametrization of the set $\mathcal{F}_n$ of frontier sets, then there exists a mapping

$$\bar{g}: f_1(\mathcal{F}_1) \times f_2(\mathcal{F}_2) \to f_\mathrm{C}(\mathcal{F}_\mathrm{C})$$

which determines the empirical entropy of any state $(a_1, a_2) \in \mathcal{S}_\mathrm{C}$ in terms of the empirical entropies of a_1 and a_2.

We assume that anergic processes in the subsystems induce an anergic process in the compound system. That is, for all (a_1, a_2), $(b_1, b_2) \in \mathcal{S}_\mathrm{C}$,

$$(a_1, b_1) \in \text{dom } q_1 \;\&\; (a_2, b_2) \in \text{dom } q_2 \Rightarrow ((a_1, a_2), (b_1, b_2)) \in \text{dom } q_\mathrm{C},$$

where the functions q_n are defined in the manner of Sec. 4. We also assume that, with a suitable choice of the arbitrary constants, the internal energy of the compound system is the sum of the internal energies of the subsystems; i.e.,

$$U_\mathrm{C}(a_1, a_2) = U_1(a_1) + U_2(a_2)$$

for all $(a_1, a_2) \in \mathcal{S}_C$. It follows from the definitions of the parametrizations f_{na} of $\mathcal{T}_n$ that

$$f_{Ca_C}(\beta_C) = f_{1a_1}(\beta_1) + f_{2a_2}(\beta_2)$$

for all $a_C = (a_1, a_2) \in \mathcal{S}_C$, where

$$\beta_n \in \mathcal{T}_n \quad \text{and} \quad \beta_C = \bar{\kappa}(\beta_1, \beta_2).$$

In terms of any parametrizations f_n of $\mathcal{T}_n$, we write $f_n(\beta_n) = z_n$, $F_{na} = f_{na} \circ f_n^{-1}$, and we have

$$F_{Ca_C}(z_C) = F_{1a_1}(z_1) + F_{2a_2}(z_2)$$

for all $z_1 \in f_1(\mathcal{T}_1)$ and for all $z_2 \in f_2(\mathcal{T}_2)$, where $z_C = \bar{g}(z_1, z_2)$. It is assumed that $\bar{g}$ and the F_{na} are differentiable, and it follows that

$$F'_{Ca_C}(z_C) \frac{\partial g}{\partial z_P}(z_1, z_2) = F'_{Pa_P}(z_P)$$

for $P = 1, 2$. Using this result, one shows that Axiom XIV for the compound system is compatible with Axiom XIV for the subsystems and that

$$\forall\, z_1 \in f_1(\mathcal{T}_1), \forall\, z_2 \in f_2(\mathcal{T}_2), \forall\, a_1, b_1 \in f_1^{-1}(z_1),$$
$$\forall\, a_2, b_2 \in f_2^{-1}(z_2): a_1 \overline{\mathcal{C}} a_2 \;\&\; b_1 \overline{\mathcal{C}} b_2 \Rightarrow F'_{1a_1}(z_1)/F'_{2a_2}(z_2) = F'_{1b_1}(z_1)/F'_{2b_2}(z_2).$$

As in Sec. 7, temperature functions T_P are defined on $\mathcal{S}_P$ for $P = 1, 2$. Given $(a_1, a_2) \in \mathcal{S}_C$, one can determine the T_P uniquely by choosing

$$T_1(a_1) = T_2(a_2) = T_0,$$

where $T_0 > 0$. If $a_P \in \alpha_P \in \mathcal{T}_P$, $b_P \in \alpha_P$, and $b_1 \overline{\mathcal{C}} b_2$, then, from the equation at the end of the last paragraph and property (ii) of temperature functions, we have $T_1(b_1) = T_2(b_2)$. If $d_P \in \beta_P \in \mathcal{T}_P$ and $d_1 \overline{\mathcal{C}} d_2$, then, from XIII, there exist $c_P \in \beta_P$, $e_P \in \alpha_P$ such that $c_P \mathcal{C}_P e_P$ and hence $c_P \overline{\mathcal{C}} e_P$. We assume that the e_P may be chosen so that $e_1 \overline{\mathcal{C}} e_2$. It then follows from our previous result that $T_1(e_1) = T_2(e_2)$, while from the transitive property of $\overline{\mathcal{C}}$ we have $c_1 \overline{\mathcal{C}} c_2$. From property (i) of temperature functions, $T_P(c_P) = T_P(e_P)$, and hence $T_1(c_1) = T_2(c_2)$. Again using the equation at the end of the last paragraph and property (ii), one finds that $T_1(d_1) = T_2(d_2)$. We have therefore proved that

$$\forall\, d_P \in \mathcal{S}_P: d_1 \overline{\mathcal{C}} d_2 \Rightarrow T_1(d_1) = T_2(d_2).$$

A temperature function $T_C: \mathcal{S}_C \to R^1$ is defined by $T_C(a_1, a_2) = T_1(a_1) = T_2(a_2)$ for all $(a_1, a_2) \in \mathcal{S}_C$. Conditions (i) and (iii) of Sec. 7 are trivially satisfied. To show that (ii) is satisfied, we must prove that

$$\forall\, \alpha_C \in \mathcal{T}_C, \forall\, (a_1, a_2), (d_1, d_2) \in \alpha_C: z_C = f_C(\alpha_C)$$
$$\Rightarrow T_C(d_1, d_2) = T_C(a_1, a_2) F'_{C(d_1,d_2)}(z_C)/F'_{C(a_1,a_2)}(z_C).$$

We take

$$a_P \in \alpha_P = f_P^{-1}(z_P), \; d_P \in \beta_P = f_P^{-1}(w_P),$$

where

$$\bar{g}(z_1, z_2) = \bar{g}(w_1, w_2) = z_C.$$

It follows from Axiom XIII that there exist $e_P \in \alpha_P$, $c_P \in \beta_P$ such that $e_P \overline{\mathcal{C}} c_P$. We again assume that the e_P may be chosen so that $e_1 \overline{\mathcal{C}} e_2$, and it follows that $c_1 \overline{\mathcal{C}} c_2$ and $(c_1, c_2) \in \mathcal{S}_C$. Since $(c_1, c_2) \overline{\mathcal{C}}_C (e_1, e_2)$ and $(c_1, c_2), (e_1, e_2) \in \alpha_C$, Axiom XIV implies that

$$F'_{C(e_1,e_2)}(z_C) = F'_{C(c_1,c_2)}(z_C).$$

Substituting $F'_{C(d_1,d_2)}$ in terms of F'_{1d_1}, etc., one finds

$$\begin{aligned} T_C(a_1, a_2) F'_{C(d_1,d_2)}(z_C)/F'_{C(a_1,a_2)}(z_C) &= T_1(a_1) F'_{1d_1}(w_1) F'_{1e_1}(z_1)/F'_{1c_1}(w_1) F'_{1a_1}(z_1) \\ &= T_1(a_1) T_1(d_1) T_1(e_1)/T_1(c_1) T_1(a_1) \\ &= T_1(d_1) = T_C(d_1, d_2). \end{aligned}$$

The functions $s'_n: f_n(\mathcal{T}_n) \to R^1$ are defined by

$$s'_n(z) = F'_{na}(z)/T_n(a),$$

where $a \in f_n^{-1}(z)$. Because

$$T_C(a_1, a_2) = T_P(a_P)$$

and

$$F'_{C(a_1,a_2)}(z_C)(\partial \bar{g}/\partial z_P)(z_1, z_2) = F'_{Pa_P}(z_P)$$

for $P = 1, 2$, $z_C = \bar{g}(z_1, z_2)$, and for all $(a_1, a_2) \in f_C^{-1}(z_C)$, we have

$$s'_C(z_C) \frac{\partial \bar{g}}{\partial z_P}(z_1, z_2) = s'_P(z_P).$$

Integrating these equations gives

$$s_C(z_C) = s_1(z_1) + s_2(z_2) + K,$$

where K is an arbitrary constant. It follows that the entropy functions S_n satisfy

$$S_C(a_1, a_2) = S_1(a_1) + S_2(a_2) + K$$

for all $(a_1, a_2) \in \mathcal{S}_C$.

[1] C. Carathéodory, Math. Ann. **67**, 355 (1909); Sitzber. Preuss. Akad. Wiss. Physik-Math. Kl. **39**, (1925).

[2] An extended version of the Carathéodory theory is given in P. T. Landsberg, *Thermodynamics* (Interscience, New York, 1961).

[3] H. A. Buchdahl, Z. Phys. **168**, 316 (1962).

[4] R. Giles, *Mathematical Foundations of Thermodynamics* (Pergamon, Oxford, 1964).

[5] G. Falk and H. Jung, *Handbuch der Physik* III/2 (Springer, Berlin, 1959).

[6] H. A. Buchdahl and W. Greve, Z. Phys. **168**, 386 (1962).

[7] E. A. Guggenheim, *Thermodynamics* (North-Holland, Amsterdam, 1957), 3rd ed.

[8] H. B. Callen, *Thermodynamics* (Wiley, New York, 1960).

[9] L. Tisza, *Generalized Thermodynamics* (M.I.T. Press, Cambridge, Mass., 1966).

[10] One could say *macroscopic state*, but we will be concerned with no other kind.

[11] It is to ensure this uniqueness that we introduce the coordinates x_i with $i > 1$, $i \notin I$. Consider, for example, a system that consists of a solid block. A given amount of heat absorbed in a process might warm the whole block without changing its phase, or it might melt a corner off. By including among the coordinates some that determine the geometry of the block, we can specify which of these alternatives is to occur in an anergic process.

[12] A similar restriction on anergic processes is implicit in Axiom X of the formal theory.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Coulomb Green's Function in f-Dimensional Space*

LEVERE C. HOSTLER
Department of Physics, Wilkes College, Wilkes-Barre, Pennsylvania 18703

(Received 3 April 1970)

It is shown that the f-dimensional nonrelativistic Coulomb Green's function and the associated reduced Green's functions can be obtained by differentiation of the corresponding functions in the 1-dimensional (f odd) or 2-dimensional (f even) case. A new expansion of the 3-dimensional coordinate-space Coulomb Green's function and a new sum formula for a product of two Laguerre polynomials with different arguments are derived.

Recently there has been some interest in the nonrelativistic Coulomb Green's function in f-dimensional space.[1] We will here show that the f-dimensional Coulomb Green's function can be obtained by successive differentiation of the Coulomb Green's function in the 1-dimensional case (f odd) or 2-dimensional case (f even). The proof of this is based on an integral representation of the f-dimensional Green's function analogous to a previous representation[2] of the 3-dimensional Green's function. The mathematical techniques that were used to obtain the integral representation in the 3-dimensional case can be applied also to the f-dimensional problem.

The f-dimensional Coulomb Green's function G_f will be defined as the solution of the differential equation

$$[\nabla_2^2 + (2k\nu/r_2) + k^2]G_f(\mathbf{r}_2, \mathbf{r}_1) = \delta^f(\mathbf{r}_2 - \mathbf{r}_1), \quad \mathrm{Im}\,(k) > 0, \quad (1)$$

subject to suitable regularity conditions at the origin and at infinity. Here

$$\nabla^2 \equiv \sum_{j=1}^{f} \left(\frac{\partial}{\partial x_j}\right)^2 \quad (2)$$

denotes the Laplacian operator of the f-dimensional space. The f-dimensional Dirac δ function occurs on the right-hand side of Eq. (1). The parameters k and ν are taken to be independent complex parameters, arbitrary except for the condition $\mathrm{Im}\,(k) > 0$. Our calculation begins with the partial wave expansion of the solution to Eq. (1). This is[3]

$$\begin{aligned} G_f(\mathbf{r}_2, \mathbf{r}_1) &= \sum_{l=0}^{\infty} \tfrac{1}{2}\pi^{-\frac{1}{2}f}\Gamma(\tfrac{1}{2}f)(2l + f - 2) \\ &\quad \times (f-2)^{-1}C_l^{\frac{1}{2}(f-2)}(\cos\theta)\mathfrak{G}_l(r_2, r_1), \\ \mathfrak{G}_l(r_2, r_1) &= (2ik)^{-1}(r_2r_1)^{-\frac{1}{2}(f-1)}\Gamma(l - i\nu + \tfrac{1}{2}(f-1)) \\ &\quad \times W_{i\nu;\frac{1}{2}(2l+f-2)}(-2ikr_>) \\ &\quad \times \mathcal{M}_{i\nu;\frac{1}{2}(2l+f-2)}(-2ikr_<), \end{aligned} \quad (3)$$

$$\cos\theta \equiv (\mathbf{r}_2 \cdot \mathbf{r}_1)/r_2r_1, \quad f = 3, 4, 5, 6, \cdots.$$

The function $\mathfrak{G}_l(r_2, r_1)$ is the radial Green's function.[4] The 2-dimensional Green's function is treated separately. Its partial wave expansion is[5]

$$G_2(\mathbf{r}_2, \mathbf{r}_1) = \sum_{m=-\infty}^{+\infty} \mathfrak{G}_m(r_2, r_1)(2\pi)^{-1}e^{im\theta}, \quad \theta \equiv \theta_2 - \theta_1, \quad (4)$$

$$\mathfrak{G}_m(r_2, r_1) = (2ik)^{-1}(r_2r_1)^{-\frac{1}{2}}\Gamma(\tfrac{1}{2} + |m| - i\nu) \times W_{i\nu;|m|}(-2ikr_>)\mathcal{M}_{i\nu;|m|}(-2ikr_<).$$

One now uses an integral representation for a product of two Whittaker functions with different arguments and the Neumann's series for[6] $J_\nu(kz)$ to obtain the integral representation

$$\begin{aligned} G_f(x, y) &= \frac{(-ik)^{f-2}}{(4\pi)^{\frac{1}{2}(f-1)}} e^{\pi i[i\nu - \frac{1}{2}(f-1)]} \\ &\quad \times \frac{\pi}{\sin\pi[i\nu - \tfrac{1}{2}(f-1)]} \frac{1}{2\pi i} \\ &\quad \times \int_{+\infty \arg(\zeta\pm1)=0}^{(1+)} d\zeta(\zeta+1)^{i\nu+\frac{1}{2}(f-3)} \\ &\quad \times (\zeta-1)^{-i\nu+\frac{1}{2}(f-3)}e^{ikx\zeta} \\ &\quad \times \frac{I_{\frac{1}{2}(f-3)}[-ik(x^2-y^2)^{\frac{1}{2}}(\zeta^2-1)^{\frac{1}{2}}]}{[-\frac{1}{2}ik(x^2-y^2)^{\frac{1}{2}}(\zeta^2-1)^{\frac{1}{2}}]^{\frac{1}{2}(f-3)}}, \end{aligned}$$

$$x = r_2 + r_1, \quad y = |\mathbf{r}_2 - \mathbf{r}_1|, \quad f = 1, 2, 3, 4, \cdots, \quad (5)$$

for the f-dimensional nonrelativistic Coulomb Green's function. We learn here that the Green's function depends upon $\mathbf{r}_2$ and $\mathbf{r}_1$ only through the two variables x and y. It is quite remarkable that Eq. (5) holds for all f, despite the necessity of treating the 2-dimensional case separately in the derivation. One can even substitute $f = 1$ in (5), and one will obtain the correct answer for the 1-dimensional Coulomb Green's function of Meixner[7,8]

$$G_1(x, y) = (2ik)^{-1}\Gamma(1 - i\nu) \times W_{i\nu;\frac{1}{2}}(-ik(x+y))\mathcal{M}_{i\nu;\frac{1}{2}}(-ik(x-y)), \quad (6)$$

provided that one interprets y as $y = |r_2 - r_1|$. (Note that this is consistent with the general definition $y = |\mathbf{r}_2 - \mathbf{r}_1|$. The magnitude of the difference between two 1-dimensional vectors having components r_2 and r_1 is $|r_2 - r_1|$.) The identity[9]

$$C_n^{\nu}(z) = (-1)^n\Gamma(n + 2\nu)[n!\,\Gamma(2\nu)]^{-1} \times {}_2F_1(-n, n + 2\nu, \nu + \tfrac{1}{2}; \tfrac{1}{2}(1 + z)), \quad \nu \neq 0, \quad (7)$$

was used to identify the Neumann's series occurring in the derivation of the integral representation in the case $f \geq 3$. The derivation in the case $f = 2$ involves a rather tricky limiting case of the Neumann's series.[10]

We are now in a position to establish the general relation

$$G_{f+2}(x, y) = -\frac{\partial}{2\pi y \partial y} G_f(x, y), \quad (8)$$

which exists between the nonrelativistic Coulomb Green's functions in spaces of different dimensionality. By successive applications of this recurrence relation, we find the identities

$$G_f(x, y) = \left(\frac{-\partial}{2\pi y \partial y}\right)^{\frac{1}{2}(f-1)} G_1(x, y), \quad f = 1, 3, 5, \cdots, \quad (9)$$

$$G_f(x, y) = \left(\frac{-\partial}{2\pi y \partial y}\right)^{\frac{1}{2}(f-2)} G_2(x, y), \quad f = 2, 4, 6, \cdots, \quad (10)$$

expressing the fact that the Coulomb Green's function in a space of arbitrary dimensionality can be obtained from either the 1-dimensional Green's function (f odd) or the 2-dimensional Green's function (f even) by successive differentiation. The closed expression (6) for $G_1(x, y)$ has already been discussed, but no closed expression for $G_2(x, y)$ seems to be known. The recurrence relation (8) can be obtained quite simply by applying the operator $(2\pi)^{-1} \cdot (-\partial/y\partial y)$ to both sides of Eq. (5). On the right-hand side of this equation we take the operator under the integral sign, where it acts on the Bessel function, to give

$$\frac{1}{2\pi}\left(-\frac{\partial}{y\partial y}\right)\left(\frac{I_{\frac{1}{2}(f-3)}(z)}{(\frac{1}{2}z)^{\frac{1}{2}(f-3)}}\right) = \frac{(-ik)^2}{4\pi}(\zeta^2 - 1)\left(\frac{I_{\frac{1}{2}(f+2-3)}(z)}{(\frac{1}{2}z)^{\frac{1}{2}(f+2-3)}}\right),$$
$$z = -ik(x^2 - y^2)^{\frac{1}{2}}(\zeta^2 - 1)^{\frac{1}{2}}, \quad (11)$$

which follows from the identity[11]

$$\left(\frac{d}{z\,dz}\right)^m\left(\frac{I_\nu(z)}{z^\nu}\right) = \frac{I_{\nu+m}(z)}{z^{\nu+m}}. \quad (12)$$

The integrand for $G_f(x, y)$ is thereby converted into the integrand for $G_{f+2}(x, y)$, and Eq. (8) is seen to be true.

Reduced Green's functions, defined by

$$\mathcal{G}(\mathbf{r}_2, \mathbf{r}_1, E_n) = -\frac{\hbar^2}{2m}\sum_{k\neq n}\sum_{l}\frac{\varphi_{kl}(\mathbf{r}_2)\varphi_{kl}(\mathbf{r}_1)^*}{E_k - E_n}, \quad (13)$$

play an essential role in Rayleigh–Schrödinger bound-state perturbation theory.[12] In the Coulomb case we now assume that k and ν satisfy a relation of the form

$$\nu = \lambda/k, \quad \lambda > 0. \quad (14)$$

A general relation between a reduced Green's function and the corresponding full Green's function $G(\mathbf{r}_2, \mathbf{r}_1, E)$ is[13]

$$\mathcal{G}(\mathbf{r}_2, \mathbf{r}_1, E_n) = \frac{d}{dE}[(E - E_n)G(\mathbf{r}_2, \mathbf{r}_1, E)]\big|_{E=E_n}. \quad (15)$$

In view of this relation, the identities (8)–(10), connecting Coulomb Green's functions in spaces of different dimensionality, are seen to apply also to the reduced Coulomb Green's functions. By using (8) one finds that

$$\frac{\partial}{\partial E}[(E - E_n)G_{f+2}(x, y, E)] = \frac{\partial}{\partial E}\left((E - E_n)\frac{-\partial}{2\pi y\partial y}G_f(x, y, E)\right) = \frac{-\partial}{2\pi y\partial y}\frac{\partial}{\partial E}[(E - E_n)G_f(x, y, E)],$$

where E_n is the nth energy eigenvalue of the $(f + 2)$-dimensional problem. One observes that each energy eigenvalue of G_{f+2} is also an energy eigenvalue of G_f.[14] Evaluating the above at $E = E_n$ and using Eq. (15) gives

$$\mathcal{G}_{f+2}(x, y, E_n) = \frac{-\partial}{2\pi y\partial y}\mathcal{G}_f(x, y, E_n). \quad (16)$$

This is the analog of Eq. (8). That the analogs of Eqs. (9) and (10) are true follows immediately. One can therefore compute all reduced Coulomb Green's functions for f odd (even) once the corresponding 1-dimensional (2-dimensional) function is known.

Now the 3-dimensional reduced ground-state Coulomb Green's function has been obtained in closed form earlier.[15] The connection with the 1-dimensional problem was not known at that time. For

the sake of completeness, we here give the corresponding 1-dimensional result:

$$\mathcal{G}_1(x, y, E_1) = \tfrac{1}{2}(x + y)e^{-\lambda y} - xe^{-\lambda x} - \tfrac{1}{2}\lambda(x^2 - y^2)e^{-\lambda x} \times \{\tfrac{7}{2} - \gamma - \lambda x - \ln[\lambda(x + y)] + g(\lambda(x - y))\},$$

$$g(z) \equiv \int_0^z dt\,\frac{e^t - 1}{t} = -ze^z \int_0^1 dt e^{-tz} \ln(1 - t). \quad (17)$$

Equation (17) can be obtained quite readily using the work of Ref. 13. The expression (17) has been checked by verifying explicitly that the Green's function equation and the boundary conditions are satisfied. [It was also verified that the previous 3-dimensional result is obtained upon application of the differential operator $(2\pi)^{-1}(-\partial/y\partial y)$, but this checks (17) only to within an arbitrary additive function of x.]

It is known that the nonrelativistic Coulomb Green's function possesses an eigenfunction expansion which involves a sum over a discrete set of states only.[16] This can be obtained by writing the Green's function in the form

$$G_f(\mathbf{r}_2, \mathbf{r}_1, E) = -\frac{\hbar^2}{2m}\langle \mathbf{r}_2|\,(r)^{\frac{1}{2}}\frac{1}{\Lambda - (Ze^2/4\pi)}(r)^{\frac{1}{2}}|\mathbf{r}_1\rangle,$$

$$\Lambda = (r)^{\frac{1}{2}}\frac{p^2}{2m}(r)^{\frac{1}{2}} - Er, \quad E < 0, \quad (18)$$

and inserting a complete set of eigenfunctions of the Hermitian operator Λ. For this purpose we limit our consideration to E real and negative. The Green's function for general complex E is obtained from this special case by analytic continuation. The operator Λ has no continuous spectrum. Its eigenvalues λ_n are just the values of the coupling constant $Ze^2/4\pi$ that would be required to produce an nth bound-state eigenfunction of the Coulomb Hamiltonian at the preassigned negative real energy E. When this method is applied to the 1- and 3-dimensional Green's functions, we obtain the two expansions

$$G_1(x, y) = \sum_{n=1}^{\infty}(2n)^{-1}(n - i\nu)^{-1}ik(x^2 - y^2)e^{ikx} \times L_{n-1}^1[-ik(x + y)]L_{n-1}^1[-ik(x - y)] \quad (19)$$

and

$$G_3(x, y) = 2ike^{ik(r_2+r_1)}\sum_{n=1}^{\infty}(n - i\nu)^{-1} \times \sum_{l=0}^{n-1}\frac{(2l + 1)P_l}{4\pi}\frac{(n - l - 1)!}{(n + l)!} \times (-2ikr_2)^l(-2ikr_1)^l \times L_{n-l-1}^{2l+1}(-2ikr_2)L_{n-l-1}^{2l+1}(-2ikr_1), \quad (20)$$

respectively. On the other hand, a more compact form for $G_3(x, y)$ could be obtained by using (8) in conjunction with the 1-dimensional result (19). We thus obtain the new expansion

$$G_3(x, y) = -\sum_{n=1}^{\infty}\frac{ik}{4\pi}n^{-1}(n - i\nu)^{-1}e^{ikx}\frac{\partial}{y\partial y}(x^2 - y^2) \times L_{n-1}^1(-ik(x + y))L_{n-1}^1(-ik(x - y)) \quad (21)$$

of the 3-dimensional coordinate-space Coulomb Green's function. The individual terms of this expansion have the same general structure as the closed form expression previously obtained for the nonrelativistic Coulomb Green's function, but are free of hypergeometric functions. The nuclear charge appears in Eq. (21) only through the denominator $(n - i\nu)^{-1}$. Substituting $i\nu = 0$ in Eq. (21) leads to the expansion

$$-(4\pi y)^{-1}e^{iky} = -\sum_{n=1}^{\infty}\frac{ik}{4\pi n^2}e^{ikx}\frac{\partial}{y\partial y}(x^2 - y^2) \times L_{n-1}^1(-ik(x + y))L_{n-1}^1(-ik(x - y))$$

of the free-particle Green's function. Another form of Eq. (21) can be obtained by separating out the free-particle part. This is achieved by means of the identity

$$(n - i\nu)^{-1} = n^{-1} + i\nu n^{-1}(n - i\nu)^{-1}.$$

Thus,

$$G_3(x, y) = G_0(x, y) + \mathcal{G}_3(x, y),$$

where $G_0(x, y)$ is the free-particle Green's function and

$$\mathcal{G}_3(x, y) = -\sum_{n=1}^{\infty}i\nu\frac{ik}{4\pi}n^{-2}(n - i\nu)^{-1}e^{ikx}\frac{\partial}{y\partial y}(x^2 - y^2) \times L_{n-1}^1(-ik(x + y))L_{n-1}^1(-ik(x - y)). \quad (22)$$

Using the fact that $i\nu$ and k in the expansions (20) and (21) can be independent variables, one concludes that the equality of (20) and (21) must hold term-by-term. [The pole terms, proportional to $(n - i\nu)^{-1}$, must agree in both expansions.] We thus obtain a new sum formula

$$\sum_{l=0}^{n-1}\frac{2l + 1}{4\pi}P_l\frac{(n - l - 1)!}{(n + l)!}(2tr_2)^l(2tr_1)^l \times L_{n-l-1}^{2l+1}(2tr_2)L_{n-l-1}^{2l+1}(2tr_1) = -\frac{1}{8\pi n}\frac{\partial}{y\partial y}(x^2 - y^2)L_{n-1}^1(t(x + y))L_{n-1}^1(t(x - y)),$$

$$x = r_1 + r_2, \quad y = |\mathbf{r}_2 - \mathbf{r}_1|, \quad t \text{ arbitrary}, \quad (23)$$

for a product of two Laguerre polynomials with different arguments. This sum formula was originally derived by another method, reported in the Appendix. As mentioned in Ref. 16, the momentum space counterpart of Eq. (20) has been given by Schwinger, who exploits the 4-dimensional rotational invariance of the Coulomb problem. His expansion of the momentum-space Green's function involves a sum of products of 4-dimensional spherical harmonics. The identity (23), which enables one to express the sum over the nth energy eigenspace in Eq. (20) in closed form, corresponds in momentum space to the addition theorem for the 4-dimensional spherical harmonics.

APPENDIX[17]

The sum formula (23) was originally obtained from the identity[18]

$$\begin{aligned} G_3(\mathbf{r}_2, \mathbf{r}_1, E) &= -\frac{\hbar^2}{2m}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\int_0^{\infty} dk\,\frac{\varphi_{lm}(k;\mathbf{r}_2)\varphi_{lm}(k;\mathbf{r}_1)^*}{(\hbar^2k^2/2m)-E} \\ &\quad -\frac{\hbar^2}{2m}\sum_{n=1}^{\infty}\sum_{l=0}^{n-1}\sum_{m=-l}^{l}\frac{\varphi_{nlm}(\mathbf{r}_2)\varphi_{nlm}(\mathbf{r}_1)^*}{E_n-E} \\ &= -\frac{\Gamma(1-iv)}{4\pi ik}\frac{\partial}{y\partial y}W_{iv;\frac{1}{2}}(-ik(x+y)) \\ &\quad \times \mathcal{M}_{iv;\frac{1}{2}}(-ik(x-y)) \qquad \text{(A1)} \end{aligned}$$

by computing the residue of each side of the equation at $E =$ the nth Bohr energy level E_n. On the left-hand side of the equation, only terms from the discrete sum contribute to this residue, the integral over the continuous spectrum being an analytic function of E in the neighborhood of $E = E_n < 0$. This contribution is

$$\frac{\hbar^2}{2m}\sum_{l=0}^{n-1}\sum_{m=-l}^{l}\varphi_{nlm}(\mathbf{r}_2)\varphi_{nlm}(\mathbf{r}_1)^* \qquad \text{(A2)}$$

and has the interpretation of (the coordinate space representative of) the perpendicular projector onto the nth energy eigensubspace of the Coulomb Hamiltonian, times $\hbar^2/2m$. The evaluation of the residue of the right-hand side is simplified by the fact that the (simple) pole of the right-hand side is the pole of the gamma-function factor $\Gamma(1-iv)$. The residue is found to be

$$\frac{(-1)^n}{(n-1)!}\frac{\hbar^2}{4\pi mn^2a_1}\frac{\partial}{y\partial y}W_{n;\frac{1}{2}}\left(\frac{x+y}{na_1}\right)\mathcal{M}_{n;\frac{1}{2}}\left(\frac{x-y}{na_1}\right), \qquad a_1 = \frac{4\pi\hbar^2}{mZe^2}. \qquad \text{(A3)}$$

The identity (23) emerges when one equates (A2) and (A3) and simplifies. The simplifications include expressing the Whittaker functions in (A3) in terms of Laguerre polynomials.[19] Also, in the resulting identity, $(na_1)^{-1}$ can be replaced by a new parameter t since a_1 is arbitrary.

It has been known for some time that the infinite sum of terms in Eq. (A1) is a function of only the two variables x and y and not three (as is allowed by 3-dimensional rotational invariance). This fact can be understood in terms of the dynamical symmetry of the Coulomb problem. Specifically, it has been shown[20] that, as a consequence of this symmetry, the coordinate space representative of any function of the Coulomb Hamiltonian will be a function of only the two variables x and y. That we obtain a function of x and y only in (A2) after just a finite sum of terms is a consequence of this theorem, for the perpendicular projector onto the nth energy eigensubspace of the Coulomb Hamiltonian is a function of the Coulomb Hamiltonian, viz.,

$$P_n = \frac{1}{2\pi i}\oint_{E_n^+}\frac{dE}{E-H}, \qquad \text{(A4)}$$

and is therefore subject to the above theorem. It follows that the coordinate space representative of (A4), which (aside from the factor $\hbar^2/2m$) is just the finite sum of terms (A2), is a function of x and y only.

* Supported in part by Wilkes College Research Fund Grant No. 7.

[1] R. J. White and F. H. Stillinger, Jr., "An Analytic Approach to Electron Correlation in Atoms," Bell Telephone Laboratories Preprint, Murray Hill, New Jersey.

[2] L. Hostler, J. Math. Phys. **5**, 591 (1964), Eq. (1.13).

[3] To obtain this result, the addition theorem [H. Bateman, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II, Chap. XI]

$$\sum_{\alpha=1}^{h(l,f)} Y_l^{\alpha}(\mathbf{u}_2)\cdot Y_l^{\alpha}(\mathbf{u}_1)^* = 2^{-1}\pi^{-\frac{1}{2}f}\Gamma(\tfrac{1}{2}f)(2l+f-2)(f-2)^{-1}\cdot C_l^{\frac{1}{2}(f-2)}(\mathbf{u}_2\cdot\mathbf{u}_1),$$

$$h(l,f) = (2l+f-2)(l+f-3)!/l!\,(f-2)!, \qquad f = 3, 4, 5, 6, \cdots,$$

for the f-dimensional spherical harmonics was used. Here α is a degeneracy quantum number and distinguishes between the $h(l,f)$ independent eigenfunctions to the same eigenvalue of the angular part of the f-dimensional Laplace operator. The normalization is

$$\oint\!\!\oint d\Omega\, Y_l^{\alpha}(\mathbf{u})\, Y_{l'}^{\beta}(\mathbf{u})^* = \delta_{ll'}\delta_{\alpha\beta},$$

the integration being over the unit hypersphere.

[4] It satisfies the differential equation

$$\left(\frac{\partial^2}{\partial r_2^2}+(f-1)\frac{\partial}{r_2\partial r_2}-l(l+f-2)r_2^{-2}+\frac{2kv}{r_2}+k^2\right)g_l(r_2,r_1) = \delta(r_2-r_1)(r_2r_1)^{-\frac{1}{2}(f-1)}.$$

[5] See Ref. 1 for a similar calculation.

[6] Reference 2, Eqs. (1.9) and (1.11).

[7] J. Meixner, Math. Z. **36**, 677 (1933).

[8] In the 1-dimensional case the arguments of the Whittaker functions in (6) become $-ik(x+y) = -2ikr_>$ and $-ik(x-y) = -2ikr_<$. We avoid the use of the arguments $-2ikr_>$ and $-2ikr_<$ since we want to focus on the x and y dependence of G_1.

[9] Reference 3, Eq. (20), p. 176.

[10] This limiting case, obtained by evaluating the limit as μ approaches zero in Eq. (1.11) of Ref. 2, is

$$\Gamma(\nu+1)(\tfrac{1}{2}kz)^{-\nu}I_\nu(kz) = I_0(z) + 2\sum_{l=1}^{\infty} {}_2F_1(-l, l, \nu+1, k^2)(-1)^l I_{2l}(z).$$

Also, the identity [Ref. 3, Vol. I, Eq. (11), p. 101] $\cos(mz) = (-1)^m {}_2F_1(m, -m, \frac{1}{2}, \frac{1}{2}(1+\cos z))$ was required.

[11] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1962), 2nd ed., Eq. (6), p. 79.

[12] Here the E_k denote the distinct energy eigenvalues of the system, including continuum eigenvalues, if they occur, and l is a degeneracy quantum number. The energy E_n is one of the bound-state eigenvalues. Quantization in a large sphere would be required to give the sum in Eq. (13) a meaning for the continuum states.

[13] L. Hostler, Phys. Rev. **178**, 126 (1969), Eq. (2.4).

[14] The converse, however, is not necessarily true. The eigenvalues correspond to $i\nu = 1, 2, 3, \cdots$ in the 1-dimensional case and to $i\nu = \frac{1}{2}(f-3)+n$, $n = 1, 2, 3, \cdots$, in case $f \geq 2$. The operation $(2\pi)^{-1}(-\partial/y\,\partial y)$ must remove pole terms in G_f which are not also in G_{f+2}.

[15] Reference 13, Eq. (1.6). The S-wave part only had been obtained in closed form by H. F. Hameka [J. Chem. Phys. **47**, 2728 (1967); **48**, 4810 (1968)].

[16] In momentum space this expansion is due to J. Schwinger, J. Math. Phys. **5**, 1606 (1964). A specific literature citation for the coordinate space expansion (Eq. 20) does not seem to exist.

[17] The following work was done with H.-L. Hwang and is part of a master's thesis submitted by Hwang to Wilkes College, Wilkes-Barre, Pennsylvania, May, 1970.

[18] Reference 2, Eqs. (1.3) and (1.18).

[19] Herbert Buchholz, *The Confluent Hypergeometric Function*, translated by H. Lichtblau and K. Wetzel (Springer, New York, 1969), Eq. (28a), p. 23.

[20] L. Hostler, J. Math. Phys. **8**, 642 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

An Axiomatic Formulation of the Theory of Coinciding Simple Poles and Multiple Poles*

NOBORU NAKANISHI†
Applied Mathematics Department, Brookhaven National Laboratory, Upton, New York 11973

(Received 11 February 1970)

In the Bethe–Salpeter formalism, the scattering Green's function is known to have multiple poles synthesized out of coinciding simple poles. The present paper proposes an axiomatic approach to the problem of finding the residues of the multiple poles in terms of those of M coinciding simple poles. The latter residues are regarded as finite-dimensional, mutually orthogonal projection operators on a reflexive Banach space and its dual. Then various properties of the residues of the multiple poles are derived without recourse to the original Bethe–Salpeter equation, and especially it is shown mathematically that they can be decomposed into a direct sum of operators which commute with the Bethe–Salpeter operator. The residues of multiple poles are explicitly determined in two particular cases, $M = N + 1$ and $M = 2$, where N denotes the highest order of the singularities (in a parameter) of the residues of the coinciding simple poles.

1. INTRODUCTION

As is well known, quantum-theoretical states are represented by vectors in a Hilbert space. Every state has positive norm, which is usually normalized to unity. In the field theory, however, one sometimes needs vectors whose squared norm is negative. Since those negative-norm states are not probabilistically interpretable if they do not disappear asymptotically, they are usually called ghost states. When the difference between a normal state vector and a ghost one tends to zero, one has a zero-norm state and a state which cannot be an eigenstate of the Hamiltonian. The latter is called a dipole ghost; it was first proposed by Heisenberg[1] in the Lee model. Subsequently, several authors[2] investigated physical implications of the dipole ghost and made various extensions to other models from theoretical points of view. As a more practical application, the present author[3] made use of dipole ghosts in order to quantize the electromagnetic field in the Landau gauge in a manifestly covariant way.

In the scattering amplitude or the Green's function, 1-particle intermediate states correspond to simple poles in the invariant energy. If it has double poles, then dipole ghosts must exist in the intermediate states. In general, higher-order or multiple poles correspond to multipole ghosts. The present author[4] found in the Bethe–Salpeter formalism that the scattering Green's function generally has multiple poles at some particular values of invariant energy squared s. For example, at $s = 0$, M bound-state simple poles coincide to produce multiple poles whose highest order is M.

A general theory of coinciding simple poles and multiple poles was formulated by the present author.[5] From the consistency, one obtains the cancellation conditions, which are closely related to the unequal-mass conspiracy conditions of Freedman and Wang.[6]

[10] This limiting case, obtained by evaluating the limit as μ approaches zero in Eq. (1.11) of Ref. 2, is

$$\Gamma(\nu+1)(\tfrac{1}{2}kz)^{-\nu}I_\nu(kz) = I_0(z) + 2\sum_{l=1}^{\infty} {}_2F_1(-l, l, \nu+1, k^2)(-1)^l I_{2l}(z).$$

Also, the identity [Ref. 3, Vol. I, Eq. (11), p. 101] $\cos(mz) = (-1)^m {}_2F_1(m, -m, \frac{1}{2}, \frac{1}{2}(1+\cos z))$ was required.

[11] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1962), 2nd ed., Eq. (6), p. 79.

[12] Here the E_k denote the distinct energy eigenvalues of the system, including continuum eigenvalues, if they occur, and l is a degeneracy quantum number. The energy E_n is one of the bound-state eigenvalues. Quantization in a large sphere would be required to give the sum in Eq. (13) a meaning for the continuum states.

[13] L. Hostler, Phys. Rev. **178**, 126 (1969), Eq. (2.4).

[14] The converse, however, is not necessarily true. The eigenvalues correspond to $i\nu = 1, 2, 3, \cdots$ in the 1-dimensional case and to $i\nu = \frac{1}{2}(f-3)+n$, $n = 1, 2, 3, \cdots$, in case $f \geq 2$. The operation $(2\pi)^{-1}(-\partial/y\,\partial y)$ must remove pole terms in G_f which are not also in G_{f+2}.

[15] Reference 13, Eq. (1.6). The S-wave part only had been obtained in closed form by H. F. Hameka [J. Chem. Phys. **47**, 2728 (1967); **48**, 4810 (1968)].

[16] In momentum space this expansion is due to J. Schwinger, J. Math. Phys. **5**, 1606 (1964). A specific literature citation for the coordinate space expansion (Eq. 20) does not seem to exist.

[17] The following work was done with H.-L. Hwang and is part of a master's thesis submitted by Hwang to Wilkes College, Wilkes-Barre, Pennsylvania, May, 1970.

[18] Reference 2, Eqs. (1.3) and (1.18).

[19] Herbert Buchholz, *The Confluent Hypergeometric Function*, translated by H. Lichtblau and K. Wetzel (Springer, New York, 1969), Eq. (28a), p. 23.

[20] L. Hostler, J. Math. Phys. **8**, 642 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

An Axiomatic Formulation of the Theory of Coinciding Simple Poles and Multiple Poles*

NOBORU NAKANISHI†
Applied Mathematics Department, Brookhaven National Laboratory, Upton, New York 11973

(Received 11 February 1970)

In the Bethe–Salpeter formalism, the scattering Green's function is known to have multiple poles synthesized out of coinciding simple poles. The present paper proposes an axiomatic approach to the problem of finding the residues of the multiple poles in terms of those of M coinciding simple poles. The latter residues are regarded as finite-dimensional, mutually orthogonal projection operators on a reflexive Banach space and its dual. Then various properties of the residues of the multiple poles are derived without recourse to the original Bethe–Salpeter equation, and especially it is shown mathematically that they can be decomposed into a direct sum of operators which commute with the Bethe–Salpeter operator. The residues of multiple poles are explicitly determined in two particular cases, $M = N + 1$ and $M = 2$, where N denotes the highest order of the singularities (in a parameter) of the residues of the coinciding simple poles.

1. INTRODUCTION

As is well known, quantum-theoretical states are represented by vectors in a Hilbert space. Every state has positive norm, which is usually normalized to unity. In the field theory, however, one sometimes needs vectors whose squared norm is negative. Since those negative-norm states are not probabilistically interpretable if they do not disappear asymptotically, they are usually called ghost states. When the difference between a normal state vector and a ghost one tends to zero, one has a zero-norm state and a state which cannot be an eigenstate of the Hamiltonian. The latter is called a dipole ghost; it was first proposed by Heisenberg[1] in the Lee model. Subsequently, several authors[2] investigated physical implications of the dipole ghost and made various extensions to other models from theoretical points of view. As a more practical application, the present author[3] made use of dipole ghosts in order to quantize the electromagnetic field in the Landau gauge in a manifestly covariant way.

In the scattering amplitude or the Green's function, 1-particle intermediate states correspond to simple poles in the invariant energy. If it has double poles, then dipole ghosts must exist in the intermediate states. In general, higher-order or multiple poles correspond to multipole ghosts. The present author[4] found in the Bethe–Salpeter formalism that the scattering Green's function generally has multiple poles at some particular values of invariant energy squared s. For example, at $s = 0$, M bound-state simple poles coincide to produce multiple poles whose highest order is M.

A general theory of coinciding simple poles and multiple poles was formulated by the present author.[5] From the consistency, one obtains the cancellation conditions, which are closely related to the unequal-mass conspiracy conditions of Freedman and Wang.[6]

Indeed, the Khuri scattering amplitude, instead of the Regge one, has multiple poles at $s = 0$ (see Ref. 7), according to the prescription given in the above general theory.[5]

Recently, Ida[8] has made some interesting mathematical considerations on the theory of multiple poles. His approach is based on the Riesz–Schauder theory of compact, nonnormal operators. The present paper is, in some sense, an extension of his work. Our emphasis, however, is on a more axiomatic formulation of the theory. We show that we can construct the theory of coinciding simple poles and multiple poles without using the properties of compact operators.

In the next section, we briefly review our problem in an intuitive way. Section 3 is devoted to mathematical preliminaries. In the subsequent two sections, we present an axiomatic formulation of the theory of coinciding simple poles and multiple poles. Two interesting special cases are investigated in detail in Sec. 6. In the final section, we make some additional remarks.

2. INTUITIVE CONSIDERATION

The scattering Green's function $G(s, \lambda)$ satisfies the inhomogeneous Bethe–Salpeter equation

$$[K(s) - \lambda I(s)]G(s, \lambda) = 1 \tag{2.1}$$

in the operator notation. Here s is the invariant energy squared and λ is a parameter which can be identified with the coupling constant squared in the ladder approximation; $[K(s)]^{-1}$ denotes a product of the propagators of the two particles and $I(s)$ stands for the integral kernel.[9]

Suppose that $G(s, \lambda)$ has M simple poles in λ (with s fixed)[10] whose locations tend to a common value $\lambda = \lambda_0 \neq 0$ as $s \to s_0$. Then it is convenient to introduce new variables $w \equiv \lambda^{-1} - \lambda_0^{-1}$ and $z \equiv s - s_0$. We set

$$A(z) \equiv I(s)[K(s)]^{-1} - \lambda_0^{-1}, \tag{2.2}$$

$$F(z, w) \equiv \lambda K(s)G(s, \lambda). \tag{2.3}$$

Then (2.1) is rewritten as

$$[w - A(z)]F(z, w) = 1. \tag{2.4}$$

We also have

$$F(z, w)[w - A(z)] = 1. \tag{2.5}$$

We assume that $A(z)$ is an operator-valued analytic function holomorphic at $z = 0$.

By assumption, $F(z, w)$ has M simple poles in w whose locations tend to zero as $z \to 0$:

$$F(z, w) = S(z, w) + \hat{F}(z, w), \tag{2.6}$$

$$S(z, w) \equiv \sum_{m=1}^{M} \frac{R_m(z)}{w - w_m(z)}, \quad z \neq 0, \tag{2.7}$$

where $\hat{F}(z, w)$ is holomorphic near $z = 0$ and $w = 0$, and the functions $w_1(z), \cdots, w_M(z)$ are holomorphic at $z = 0$ and vanish there.

The residues $R_m(z)$ have the following important properties:

$$A(z)R_m(z) = R_m(z)A(z) = w_m(z)R_m(z), \tag{2.8}$$

$$R_l(z)R_m(z) = \delta_{lm}R_m(z), \tag{2.9}$$

where the latter follows from

$$-\partial F/\partial w = F^2. \tag{2.10}$$

We assume that the residues $R_m(z)$ have multiple poles at $z = 0$, whose highest order is denoted by N. Then, as $z \to 0$, $S(z, w)$ tends to

$$S(0, w) \equiv \sum_{n=0}^{N} \frac{S^{[n]}}{w^{n+1}}. \tag{2.11}$$

Since

$$[w - A(z)]S(z, w) = S(z, w)[w - A(z)] = \sum_{m=1}^{M} R_m(z), \tag{2.12}$$

we find

$$[w - A(0)]S(0, w) = S(0, w)[w - A(0)] = \lim_{z \to 0} \sum_{m=1}^{M} R_m(z). \tag{2.13}$$

On substituting (2.11) in (2.13), we obtain

$$A(0)S^{[N]} = S^{[N]}A(0) = 0,$$
$$A(0)S^{[n]} = S^{[n]}A(0) = S^{[n+1]}, \quad n = 0, 1, \cdots, N - 1, \tag{2.14}$$

whence

$$[A(0)]^{N-n+1}S^{[n]} = S^{[n]}[A(0)]^{N-n+1} = 0. \tag{2.15}$$

Now we rewrite (2.7) as

$$S(z, w) = \sum_{m=1}^{M} \left(\sum_{n=0}^{N} \frac{[w_m(z)]^n}{w^{n+1}} R_m(z) + \frac{[w_m(z)]^{N+1}R_m(z)}{w^{N+1}[w - w_m(z)]} \right). \tag{2.16}$$

The last term of (2.16) vanishes at $z = 0$. Since (2.16) should be holomorphic in z for any nonzero value of w near $w = 0$, the quantities

$$S^{[n]}(z) \equiv \sum_{m=1}^{M} [w_m(z)]^n R_m(z), \quad n = 0, 1, \cdots, N - 1, \tag{2.17}$$

have to be holomorphic at $z = 0$,[8] that is, the singularities in the right-hand side have to cancel out. Those constraints are called the cancellation conditions. Their explicit expressions can be written down easily.[5] From (2.16), the residues $S^{[n]}$ of the multiple poles

are given by $S^{[n]}(0)$. (This fact is important for the considerations of Sec. 4.)

In the Bethe–Salpeter formalism, it is convenient to introduce the Bethe–Salpeter amplitudes, which satisfy the homogeneous Bethe–Salpeter equation. When we have no degeneracy for $z \neq 0$ near $z = 0$, we can write

$$R_m(z) = \eta_m \phi_m(z) \bar{\phi}_m(z), \tag{2.18}$$

$$A(z)\phi_m(z) = w_m(z)\phi_m(z),$$

$$\bar{\phi}_m(z)A(z) = w_m(z)\bar{\phi}_m(z), \tag{2.19}$$

$$\bar{\phi}_l(z)\phi_m(z) = \eta_l \delta_{lm}, \tag{2.20}$$

where $\phi_m(z)$ is a Bethe–Salpeter amplitude somewhat modified, and $\bar{\phi}_m(z)$ is a conjugate amplitude; the sign factor η_m is related to the positive or negative norm of the relevant bound state.[9] Let $\phi_m^{(0)}$ be the coefficient of the leading term for $z \to 0$. From (2.19) we evidently have

$$A(0)\phi_m^{(0)} = 0, \tag{2.21}$$

but $\phi_1^{(0)}, \cdots, \phi_M^{(0)}$ are *not* linearly independent unless $N = 0$. The number M_0 of independent solutions to

$$A(0)\varphi = 0 \tag{2.22}$$

is in general less than M. There exist $M - M_0$ linearly independent supplementary amplitudes, which satisfy

$$[A(0)]^n \varphi = 0, \quad 2 \leq n \leq N + 1. \tag{2.23}$$

The latter amplitudes correspond to multipole ghosts. A similar consideration also applies to the conjugate amplitudes $\bar{\psi}$. Our problem is to find M independent amplitudes φ_k and M independent conjugate ones $\bar{\psi}_k$ so as to express $S^{[0]}$ in the "standard" form

$$S^{[0]} = \sum_{k=1}^{M} c_k \varphi_k \bar{\psi}_k \tag{2.24}$$

and $S^{[n]}$ in terms of only those which satisfy

$$[A(0)]^{N-n+1} \varphi_k = 0,$$

$$\bar{\psi}_k [A(0)]^{N-n+1} = 0 \tag{2.25}$$

because of (2.15).

3. MATHEMATICAL PRELIMINARIES

We consider a complex Banach space, i.e., a complete, normed, complex linear space X. Elements of X are called vectors. For any $\varphi \in X$, its norm $\|\varphi\|$ is defined in the mathematical sense (distinguish it from the physical norm). Given a linear operator T on X, the norm of T is defined by

$$\|T\| \equiv \sup_{\|\varphi\| \leq 1} \|T\varphi\|. \tag{3.1}$$

If $\|T\|$ is finite, then T is called bounded. A bounded linear operator is continuous in strong topology.

Let $\tilde{X}$ be the dual space of X, i.e., the totality of continuous linear functionals over X. For any $\varphi \in X$ and any $\tilde{\psi} \in \tilde{X}$, $\tilde{\psi}[\varphi]$ is a complex number. If the norm of $\tilde{\psi} \in \tilde{X}$ is defined by

$$\|\tilde{\psi}\| \equiv \sup_{\|\varphi\| \leq 1} |\tilde{\psi}[\varphi]|, \tag{3.2}$$

$\tilde{X}$ is also a Banach space. We assume that the natural embedding of X in the dual of $\tilde{X}$ is onto, that is, X is a reflexive Banach space.

Let T be a bounded linear operator. If there exists an operator $\tilde{T}$ acting on $\tilde{\psi}$ such that

$$(\tilde{T}\tilde{\psi})[\varphi] = \tilde{\psi}[T\varphi] \tag{3.3}$$

for any $\tilde{\psi} \in \tilde{X}$ and $\varphi \in X$, then $\tilde{T}$ is unique, and $\|\tilde{T}\| = \|T\|$. Hereafter we identify $\tilde{T}$ with T and make it operate on $\tilde{\psi}$ from the right:

$$(\tilde{\psi}T)\varphi \equiv (\tilde{T}\tilde{\psi})[\varphi]. \tag{3.4}$$

We always consider such operators.

Definition 3.1: Two kinds of the *ranges* of an operator T are defined by

$$V[T] \equiv \{T\varphi \mid \varphi \in X\},$$

$$\tilde{V}[T] \equiv \{\tilde{\psi}T \mid \tilde{\psi} \in \tilde{X}\}. \tag{3.5}$$

Duality Principle: If a proposition Γ is true, then so is its dual proposition $\tilde{\Gamma}$, which is obtained from Γ by replacing all vectors and spaces by their duals and by reversing the order of the operation of all operators.

Therefore, we need not describe dual propositions separately.

Definition 3.2: A projection operator P is a bounded linear operator such that

$$P^2 = P. \tag{3.6}$$

The following theorem is a direct consequence of (3.6).

Theorem 1: If P is a projection operator, $\varphi \in V[P]$ if and only if $P\varphi = \varphi$.

It should be remarked that, given a space $\Omega \subset X$, projection operators P such that $V[P] = \Omega$ are *not* unique. When a subspace Ω is of a finite dimension, we denote it by $\dim \Omega$.

Theorem 2: Let P be a projection operator such that $\dim V[P] = m$. If $\varphi_1, \cdots, \varphi_m \in V[P]$ are linearly independent, that is, if they are *basis vectors* of $V[P]$,

then there exist uniquely m linearly independent vectors $\tilde{\varphi}_1, \cdots, \tilde{\varphi}_m \in \tilde{V}[P]$ such that

$$P = \sum_{k=1}^{m} \varphi_k \tilde{\varphi}_k \tag{3.7}$$

with

$$\tilde{\varphi}_k \varphi_l = \delta_{kl}, \quad k, l = 1, \cdots, m. \tag{3.8}$$

In particular, $\dim \tilde{V}[P] = m$.

Proof: We define m linear functionals $\tilde{\psi}_1, \cdots, \tilde{\psi}_m$ over $V[P]$ by

$$\tilde{\psi}_k \varphi_k = \delta_{kl}. \tag{3.9}$$

Then, according to the Hahn–Banach theorem,[11] we can extend $\tilde{\psi}_1, \cdots, \tilde{\psi}_m$ to vectors of $\tilde{X}$ (not uniquely). By setting

$$\tilde{\varphi}_k \equiv \tilde{\psi}_k P, \quad k = 1, \cdots, m, \tag{3.10}$$

we find

$$\tilde{\varphi}_k \varphi_l = \tilde{\varphi}_k (P\varphi_l) = \tilde{\psi}_k \varphi_l = \delta_{kl}. \tag{3.11}$$

Hence $\tilde{\varphi}_1, \cdots, \tilde{\varphi}_m$ are linearly independent. For any $\chi \in X$, we can write

$$P\chi = \sum_{k=1}^{m} \alpha_k \varphi_k, \tag{3.12}$$

α_k being complex numbers. Then

$$\sum_{k=1}^{m} \varphi_k \tilde{\varphi}_k \chi = \sum_k \varphi_k \tilde{\varphi}_k P\chi = \sum_k \sum_l \varphi_k \tilde{\varphi}_k \alpha_l \varphi_l = \sum_k \alpha_k \varphi_k. \tag{3.13}$$

Thus we obtain (3.7). Finally, if we can also write P as

$$P = \sum_k \varphi_k \tilde{\varphi}'_k, \tag{3.14}$$

then

$$0 = \sum_k \varphi_k (\tilde{\varphi}_k - \tilde{\varphi}'_k). \tag{3.15}$$

Hence, $\tilde{\varphi}_l - \tilde{\varphi}'_l = 0$ with the aid of (3.8). QED

Theorem 3: Let P and P' be finite-dimensional projection operators. Then $V[P'] \subset V[P]$ if and only if $PP' = P'$.

Proof: Sufficiency is evident from $P'\varphi = P(P'\varphi)$ for any $\varphi \in X$. Necessity is shown as follows. For any $\chi \in X$, $P'\chi$ belongs to $V[P']$, and hence $P'\chi \in V[P]$, because $V[P'] \subset V[P]$. Therefore, by Theorem 1, we have $P(P'\chi) = P'\chi$. Since χ is arbitrary, we have $PP' = P'$. QED

Theorem 4: Let P and P' be projection operators, and $\dim V[P] = m$. Then

$$PP' = P'P = P', \tag{3.16}$$

if and only if there exist φ_k and $\tilde{\varphi}_k$, $k = 1, \cdots, m$, satisfying (3.8) such that

$$P = \sum_{k=1}^{m} \varphi_k \tilde{\varphi}_k, \tag{3.17}$$

$$P' = \sum_{k=1}^{m'} \varphi_k \tilde{\varphi}_k, \quad m' \leq m. \tag{3.18}$$

In particular,

$$\dim V[P] = \dim V[P'] + \dim V[P - P']. \tag{3.19}$$

Proof: We have only to prove necessity. Let $\dim V[P'] = m' \leq m$. From (3.16), $P - P'$ is a projection operator with $\dim V[P - P'] \geq m - m'$. We take m' linearly independent vectors $\varphi_1, \cdots, \varphi_{m'}$ from $V[P']$ and $m - m'$ linearly independent vectors $\varphi_{m'+1}, \cdots, \varphi_m$ from $V[P - P']$. Then they altogether are basis vectors of $V[P]$. We write (3.17) according to Theorem 2. Then (3.18) follows from $P' = P'P$. QED

Now, we consider vectors and operators depending on a parameter z belonging to a domain D bounded by a smooth Jordan curve in the z plane.

Definition 3.3: A vector $\varphi(z)$ and an operator $T(z)$ are *continuous in* z at $z = z_0$ if for any $\epsilon > 0$ there exists $\delta > 0$ such that, for any z satisfying $|z - z_0| < \delta$, we have

$$\|\varphi(z) - \varphi(z_0)\| < \epsilon, \tag{3.20}$$

$$\|T(z) - T(z_0)\| < \epsilon, \tag{3.21}$$

respectively.

If $T_1(z)$ and $T_2(z)$ are bounded operators continuous in z at $z = z_0$, then so is $T_1(z)T_2(z)$. Let $\Omega(z)$ be a subspace of X depending on z.

Definition 3.4: The space $\Omega(z)$ is *continuous in* z at $z = z_0$ if for any $\epsilon > 0$ there exists $\delta > 0$ such that, for any z satisfying $|z - z_0| < \delta$ and for any $\varphi \in \Omega(z_0)$, we can always find $\psi \in \Omega(z)$ satisfying

$$\|\psi - \varphi\| < \epsilon \|\varphi\|. \tag{3.22}$$

Let $F(z)$ be any of $\varphi(z)$, $T(z)$, and $\Omega(z)$. If $F(z)$ is continuous in z at every point of D, then $F(z)$ is said to be continuous in D. If δ can be chosen independently of z_0 in D, then $F(z)$ is *uniformly continuous* in D. If D is compact, uniform continuity follows from continuity in D.

Theorem 5: If $P(z)$ is a projection operator continuous in D, then $V[P(z)]$ is continuous in D.

Proof: For any $z_0 \in D$ and any $\epsilon > 0$, there exists $\delta > 0$ such that for $|z - z_0| < \delta$ we have

$$\|P(z) - P(z_0)\| < \epsilon. \tag{3.23}$$

For any $\varphi \in V[P(z_0)]$, we set $\psi \equiv P(z)\varphi$; then $\psi \in V[P(z)]$ and

$$\begin{aligned}\|\psi - \varphi\| &= \|P(z)\varphi - P(z_0)\varphi\| \\ &\leq \|P(z) - P(z_0)\| \cdot \|\varphi\| < \epsilon \|\varphi\|. \end{aligned} \tag{3.24}$$

QED

Theorem 6: If $\Omega(z)$ is of a finite dimension and continuous in D, then for any $z_0 \in D$ we have

$$\dim \Omega(z_0) \leq \dim \Omega(z) \tag{3.25}$$

for all z in some neighborhood of z_0.

Proof: Let $\dim \Omega(z_0) = m$. Then there exist m linearly independent vectors $\varphi_1, \cdots, \varphi_m \in \Omega(z_0)$ such that $\|\varphi_k\| = 1$. For $\varphi_1, \cdots, \varphi_m$ fixed, we consider the totality of the quantities $\|\sum_{k=1}^m \alpha_k \varphi_k\|$ under the condition $\sum_k |\alpha_k| = 1$. Since it is closed and bounded below, it has a minimum κ. From the linear independence of $\varphi_1, \cdots, \varphi_m$, we have $\kappa > 0$. We choose ϵ in such a way that $0 < \epsilon < \frac{1}{2}\kappa$. Then there exist some $\delta > 0$ and $\psi_1, \cdots, \psi_m \in \Omega(z)$ such that $\|\psi_k - \varphi_k\| < \epsilon$ for $|z - z_0| < \delta$. Hence, for any $\alpha_1, \cdots, \alpha_m$ satisfying $\sum_k |\alpha_k| = 1$, we have

$$\begin{aligned}\left\| \sum_k \alpha_k \psi_k \right\| &= \left\| \sum_k \alpha_k \varphi_k + \sum_k \alpha_k (\psi_k - \varphi_k) \right\| \\ &> \left\| \sum_k \alpha_k \varphi_k \right\| - \sum_k |\alpha_k| \epsilon \\ &\geq \kappa - \epsilon > \tfrac{1}{2}\kappa. \end{aligned} \tag{3.26}$$

Thus $\sum_k \alpha_k \psi_k \neq 0$, that is, $\psi_1, \cdots, \psi_m$ are linearly independent. QED

From Theorem 6 alone, we cannot exclude the possibility that $\dim \Omega(z)$ becomes lower in a closed subset of D, because κ can become arbitrarily small depending on z_0. In order to avoid this possibility, we have to introduce a somewhat *ad hoc* assumption.

Assumption 1: Let Ω be an arbitrary m-dimensional subspace of X. Then there exist m vectors $\varphi_1, \cdots, \varphi_m \in \Omega$ ($\|\varphi_k\| = 1$) such that for any $\alpha_1, \cdots, \alpha_m$ ($\sum_k |\alpha_k| = 1$) we have

$$\left\| \sum_k \alpha_k \varphi_k \right\| \geq \rho > 0, \tag{3.27}$$

where the constant ρ may depend on m but it is *independent of* Ω.

If X is a Hilbert space, then this assumption is satisfied, because if $\{\varphi_1, \cdots, \varphi_m\}$ is an orthonormal system of vectors in Ω, then

$$\begin{aligned}\left\| \sum_k \alpha_k \varphi_k \right\|^2 &= \sum_k |\alpha_k|^2 \|\varphi_k\|^2 \\ &= \sum_k |\alpha_k|^2 \geq 1/m. \end{aligned} \tag{3.28}$$

Under Assumption 1, we can use ρ instead of κ, and therefore the proof of Theorem 6 remains valid when z and z_0 are interchanged, provided that uniform continuity is assumed. Accordingly, we obtain

$$\dim \Omega(z) \leq \dim \Omega(z_0),$$

and hence $\dim \Omega(z) = \dim \Omega(z_0)$ if D is compact. Thus, $\dim \Omega(z)$ ($z \in D$) is constant since D is connected. This property remains valid even if D is not compact, because D can be approximated by a sequence of its compact subdomains. Combining this result with Theorem 5, we have the following theorem.

Theorem 7: Under Assumption 1, if $P(z)$ is a finite-dimensional projection operator continuous in a domain D, then $\dim V[P(z)]$ is constant in D.

4. GENERAL FRAMEWORK

In order to formulate the theory axiomatically, it is not convenient to start with the resolvent $F(z, w)$ of the operator $A(z)$. We postulate the following four assumptions, which are related to the residues $R_m(z)$ only.

Let D be an open neighborhood of $z = 0$. We introduce M finite-dimensional projection operators $R_m(z)$, $m = 1, \cdots, M$, defined for $z \in D' \equiv D - \{0\}$.

Assumption 2:

$$R_l(z)R_m(z) = \delta_{lm}R_m(z), \quad z \in D'. \tag{4.1}$$

Assumption 3: There exists an operator

$$\hat{R}_m(z) \equiv \sum_{n=0}^{N_m} z^{-N_m+n} R_m^{(n)}, \tag{4.2}$$

where

$$R_m^{(0)} \neq 0, R_m^{(1)}, \cdots, R_m^{(N_m)}, \quad N_m \geq 0,$$

are bounded linear operators independent of z, such that $R_m(z) - \hat{R}_m(z)$ is continuous in D and tends to zero as $z \to 0$, that is,

$$R_m(z) = \hat{R}_m(z) + o(1). \tag{4.3}$$

In particular, $R_m(z)$ is continuous in D'.

Assumption 4: The operator

$$S(z) \equiv \sum_{m=1}^{M} R_m(z), \quad z \in D', \tag{4.4}$$

can be extended to an operator continuous in D.

Let $w_1(z), \cdots, w_M(z)$ be certain given functions of z which vanish at $z = 0$ and can be expanded into

$$w_m(z) = \sum_{n=1}^{N_m} z^n w_m^{(n)} + o(z^{N_m}). \tag{4.5}$$

Assumption 5: The operator

$$H(z) \equiv \sum_{m=1}^{M} w_m(z) R_m(z), \quad z \in D', \tag{4.6}$$

can be extended to an operator continuous in D.

Let

$$S^{[n]}(z) \equiv \sum_{m=1}^{M} [w_m(z)]^n R_m(z), \quad n = 0, 1, \cdots; \tag{4.7}$$

then of course $S^{[0]}(z) \equiv S(z)$, $S^{[1]}(z) \equiv H(z)$. From (4.1) we have

$$S^{[l]}(z) S^{[n]}(z) = S^{[l+n]}(z). \tag{4.8}$$

In particular,

$$[S(z)]^2 = S(z), \tag{4.9}$$

$$H(z)S(z) = S(z)H(z) = H(z), \tag{4.10}$$

that is, $S(z)$ is a projection operator commuting with $H(z)$. Furthermore, since

$$S^{[n]}(z) = [H(z)]^n S(z), \tag{4.11}$$

$S^{[n]}(z)$ is also an operator continuous in D. The cancellation conditions which follow from $S^{[n]}(z)$ for $n \geq 1$ are somewhat complicated,[5] but those which follow from $S(z)$ are simple; they read

$$\sum_{m=1}^{M} R_m^{(N_m - N + n)} = 0, \quad n = 0, 1, \cdots, N - 1, \tag{4.12}$$

where $R_m^{(l)} \equiv 0$ for $l < 0$ and $N \equiv \max N_m$.

The multiple-pole residues $S^{[n]}$ are defined by

$$S^{[n]} \equiv \lim_{z \to 0} S^{[n]}(z). \tag{4.13}$$

From (4.7)–(4.11), they have the following properties:

$$S^{[0]} = S \equiv \lim_{z \to 0} S(z) = \sum_{m=1}^{M} R_m^{(N_m)}, \tag{4.14}$$

$$S^{[1]} = H \equiv \lim_{z \to 0} H(z), \tag{4.15}$$

$$S^{[n]} = H^n S \quad (= H^n \text{ for } n \geq 1), \tag{4.16}$$

$$S^{[l]} S^{[n]} = S^{[l+n]}, \tag{4.17}$$

$$S^2 = S, \tag{4.18}$$

$$SH = HS = H. \tag{4.19}$$

Thus S is a projection operator which commutes with H. Furthermore, Assumption 3, together with $w_m(z) = O(z)$, implies that

$$S^{[n]} = 0 \quad \text{for} \quad n \geq N + 1, \tag{4.20}$$

that is, H is nilpotent. Hence the eigenvalue of H is zero alone.

In some cases, a partial sum of $R_m(z)$, say,

$$S'(z) \equiv \sum_{m=1}^{M'} R_m(z), \quad M' < M,$$

is continuous in z at $z = 0$. We call this case *trivially reducible*. Indeed, then the operators

$$S^{[n]'}(z) \equiv \sum_{m=1}^{M'} [w_m(z)]^n R_m(z), \quad n = 1, \cdots, \tag{4.21}$$

are continuous in z at $z = 0$ because

$$S^{[n]'}(z) = S^{[n]}(z) S'(z) = S'(z) S^{[n]}(z). \tag{4.22}$$

Thus, we may discuss $S^{[n]'}(z)$ and $S^{[n]}(z) - S^{[n]'}(z)$ completely separately.

The relation between the formalisms presented in Sec. 2 and here is as follows. Given a bounded linear operator $A(z)$ satisfying (2.8), we have

$$A(z)S(z) = S(z)A(z) = H(z), \tag{4.23}$$

from which an important relation

$$A(0)S = SA(0) = SA(0)S = H \tag{4.24}$$

follows, and therefore (2.14) is derived from (4.16) and (4.19) *without recourse to* (2.11). By subtracting (4.10) from (4.23), we have

$$B(z)S(z) = S(z)B(z) = 0, \tag{4.25}$$

where $B(z) \equiv A(z) - H(z)$. Thus

$$A(z) = H(z) + [1 - S(z)]B(z)[1 - S(z)]. \tag{4.26}$$

Conversely, given an arbitrary bounded linear operator $B(z)$, we can define $A(z)$ by (4.26). Then it is easy to show, by means of (4.4), (4.6), and (4.1), that $A(z)$ satisfies (2.8). Defining $F(z, w)$ by $[w - A(z)]^{-1}$, therefore, we have

$$R_m(z)F(z, w) = F(z, w)R_m(z) = [w - w_m(z)]^{-1} R_m(z), \tag{4.27}$$

so that $F(z, w)$ contains $S(z, w)$ defined by (2.7).[12] Thus $S^{[N]}, \cdots, S^{[0]}$ are really the residues of multiple poles of $F(z, w)$.

Now, on account of (4.1) and Theorem 4, we have

$$\dim V[S(z)] = \sum_{m=1}^{M} d_m$$

for $z \in D'$, where $d_m \equiv \dim V[R_m(z)]$ is independent of z in D' because of Theorem 7. Applying Theorems 2 and 7 to $S(z)$, we find

$$\dim V[S] = \dim \tilde{V}[S] = \sum_{m=1}^{M} d_m. \tag{4.28}$$

Furthermore, (4.16) with (4.19) implies that

$$V[S^{[n]}] \subset V[S], \quad \tilde{V}[S^{[n]}] \subset \tilde{V}[S]. \tag{4.29}$$

For simplicity of description, we hereafter assume $d_m = 1$; then $\dim V[S] = M$.

Let ψ be a vector such that $R_m^{(0)}\psi \neq 0$. If we set $\phi_m(z) \equiv R_m(z)\psi$, then Theorem 2 implies the existence of $\tilde{\phi}_m(z)$ such that

$$R_m(z) = \phi_m(z)\tilde{\phi}_m(z) \tag{4.30}$$

together with $\tilde{\phi}_m(z)\phi_m(z) = 1$. From (4.3) with (4.2), we can write

$$\phi_m(z) = \sum_{n=0}^{N_m} z^{-N_m+n}\phi_m^{(n)} + o(1), \quad \phi_m^{(0)} \neq 0,$$

$$\tilde{\phi}_m(z) = \sum_{n=0}^{N_m} z^n\tilde{\phi}_m^{(n)} + o(z^{N_m}), \quad \tilde{\phi}_m^{(0)} \neq 0. \tag{4.31}$$

Furthermore, (4.1) is rewritten as

$$\tilde{\phi}_l(z)\phi_m(z) = \delta_{lm}. \tag{4.32}$$

Therefore,

$$\sum_{k=0}^{n} \tilde{\phi}_l^{(k)}\phi_m^{(n-k)} = \delta_{lm}\delta_{nN_m}. \tag{4.33}$$

Theorem 8:

$$SR_m^{(0)} = R_m^{(0)}S = R_m^{(0)}, \tag{4.34}$$

$$HR_m^{(0)} = R_m^{(0)}H = 0, \tag{4.35}$$

$$R_l^{(0)}R_m^{(0)} = 0 \quad \text{unless} \quad l = m, \quad N_m = 0. \tag{4.36}$$

Proof:

$$\begin{aligned} S^{[n]}R_m^{(0)} &= \lim_{z\to 0} \sum_l [w_l(z)]^n R_l(z) \cdot \lim_{z\to 0} z^{N_m} R_m(z) \\ &= \lim_{z\to 0} z^{N_m} \sum_l [w_l(z)]^n R_l(z) R_m(z) \\ &= \lim_{z\to 0} [w_m(z)]^n \cdot z^{N_m} R_m(z) \\ &= \delta_{n0} R_m^{(0)}, \end{aligned} \tag{4.37}$$

$$\begin{aligned} R_l^{(0)}R_m^{(0)} &= \lim_{z\to 0} z^{N_l} R_l(z) \cdot \lim_{z\to 0} z^{N_m} R_m(z) \\ &= \delta_{lm} \lim_{z\to 0} z^{2N_m} R_m(z) \\ &= \delta_{lm}\delta_{N_m 0} R_m^{(0)}. \end{aligned} \tag{4.38}$$

QED

Theorem 9:

$$H^{n+1}R_m^{(n)} = R_m^{(n)}H^{n+1} = 0. \tag{4.39}$$

Proof: For $n = 0$, (4.39) reduces to (4.35). Hence we employ mathematical induction with respect to n:

$$\begin{aligned} H^{n+1}R_m^{(n)} &= \lim_{z\to 0} z^{N_m-n} H^{n+1}\left[R_m(z) - \sum_{j=0}^{n-1} z^{-N_m+j} R_m^{(j)} \right] \\ &= H^{n+1} \lim_{z\to 0} z^{N_m-n} R_m(z) \\ &= \lim_{z\to 0} z^{N_m-n} [w_m(z)]^{n+1} R_m(z) \\ &= 0. \end{aligned} \tag{4.40}$$

QED

Theorems 8 and 9 are rewritten as

$$S\phi_m^{(0)} = \phi_m^{(0)}, \qquad \tilde{\phi}_m^{(0)}S = \tilde{\phi}_m^{(0)}, \tag{4.41}$$

$$H\phi_m^{(0)} = 0, \qquad \tilde{\phi}_m^{(0)}H = 0, \tag{4.42}$$

$$\tilde{\phi}_l^{(0)}\phi_m^{(0)} = 0 \quad \text{unless} \quad l = m, \quad N_m = 0, \tag{4.43}$$

$$H^{n+1}\phi_m^{(n)} = 0, \qquad \tilde{\phi}_m^{(n)}H^{n+1} = 0. \tag{4.44}$$

Thus $\phi_m^{[0]} \in V[S]$ and $\tilde{\phi}_m^{[0]} \in \tilde{V}[S]$ are mutually orthogonal eigenvectors of H. As pointed out in Sec. 2, $\phi_1^{[0]}, \cdots, \phi_M^{[0]}$ are not linearly independent except for the case $N = 0$. Their linear dependence comes out from the cancellation conditions. Since $S^{[0]}(z), \cdots, S^{[N-1]}(z)$ are continuous in z at $z = 0$, $\lim_{z\to 0} z^{N-n}S^{[n]}(z)$ for $n < N$ has to vanish, that is,[5]

$$\sum_{m\in I} [w_m^{(1)}]^n R_m^{(0)} = 0, \quad n = 0, 1, \cdots, N-1, \tag{4.45}$$

where $I \equiv \{m \mid N_m = N\}$. Let M_I be the number of the elements of I. If the first derivatives $w_m^{(1)}$ for all $m \in I$ are different from each other, then (4.45) gives us N independent constraints on $R_m^{(0)}$. Since $R_m^{(0)} \neq 0$, we have $M_I \geq N + 1$. Especially, if $M_I = N + 1$, then all $R_m^{(0)}$, $m \in I$, are proportional. More precisely, we have

$$R_m^{(0)} = \xi_m S^{[N]} \tag{4.46}$$

together with

$$S^{[N]} = \sum_{m\in I} [w_m^{(1)}]^N R_m^{(0)} \neq 0, \tag{4.47}$$

$$\xi_m \equiv 1 \Big/ \prod_{l\in I, l\neq m} (w_m^{(1)} - w_l^{(1)}). \tag{4.48}$$

If $M_I > N + 1$, the number of independent $R_m^{(0)}$ is $M_I - N$. If some of $w_m^{(1)}$ are equal, it is possible to have $M_I < N + 1$. We decompose I into classes $I_1, \cdots, I_k$, in each of which $w_m^{(1)}$ is common. Then

$$\sum_{m\in I_j} R_m^{(0)} = 0, \quad j = 1, \cdots, k. \tag{4.49}$$

Therefore, the number of the elements of I_j is greater than one.

Theorem 10: If there is a condition like

$$\sum_{m=1}^{L} \alpha_m R_m^{(0)} = 0, \quad \alpha_m \neq 0, \tag{4.50}$$

then the sum of the number of independent $\phi_m^{(0)}$ and that of independent $\tilde{\phi}_m^{(0)}$, $1 \leq m \leq L$, cannot exceed L.

Proof: If $\phi_1^{(0)}, \cdots, \phi_K^{(0)}$, $K \leq L$, are linearly independent, then there exist $\tilde{\psi}_1, \cdots, \tilde{\psi}_K$ such that

$$\tilde{\psi}_l \phi_m^{(0)} = \delta_{lm}, \quad l, m = 1, \cdots, K. \tag{4.51}$$

Hence (4.50) yields

$$\alpha_l \tilde{\phi}_l^{(0)} + \sum_{m=K+1}^{L} \alpha_m (\tilde{\psi}_l \phi_m^{(0)}) \tilde{\phi}_m^{(0)} = 0, \quad l = 1, \cdots, K. \tag{4.52}$$

Thus $\tilde{\phi}_1^{(0)}, \cdots, \tilde{\phi}_K^{(0)}$ are expressed by $\tilde{\phi}_{K+1}^{(0)}, \cdots, \tilde{\phi}_L^{(0)}$. QED

Theorem 11: If for some l and m we have

$$R_l^{(0)} = \alpha R_m^{(0)}, \quad \alpha \neq 0, \tag{4.53}$$

then (i) $N_l \geq 1$ if $N_m \geq 1$, and (ii) $N_l \geq 2$ if $N_m \geq 2$.

Proof: Since $N_m \neq 0$, we have $(R_m^{(0)})^2 = 0$, that is,

$$(R_l^{(0)})^2 = 0.$$

(i) If $N_l = 0$, then Assumption 2 implies

$$(R_l^{(0)})^2 = R_l^{(0)} \neq 0.$$

This is a contradiction. (ii) For $N_m \geq 2$, Assumption 2 and (4.53) yield

$$\begin{aligned} R_m^{(0)} R_m^{(1)} + R_m^{(1)} R_m^{(0)} &= 0, \\ R_m^{(0)} R_l^{(1)} + R_m^{(1)} \alpha R_m^{(0)} &= 0, \\ \alpha R_m^{(0)} R_m^{(1)} + R_l^{(1)} R_m^{(0)} &= 0. \end{aligned} \tag{4.54}$$

By eliminating $R_m^{(1)}$ from (4.54), we find

$$R_l^{(0)} R_l^{(1)} + R_l^{(1)} R_l^{(0)} = 0. \tag{4.55}$$

If $N_l = 1$, however, the left-hand side of (4.55) has to equal $R_l^{(0)} \neq 0$. This is a contradiction. QED

It does not seem possible to prove $N_l = N_m$ under (4.53) alone.

5. DECOMPOSITION OF S

We first state some definitions.

Definition 5.1: An operator T is *admissible* if T commutes with H, i.e., $HT = TH$.

Definition 5.2: Let P be an admissible projection operator. If there exists an admissible projection operator $P' \neq 0, P$ such that $PP' = P'P = P'$, then P is *reducible*; otherwise P is *irreducible*.

If P is reducible, P is a direct sum of two admissible projection operators P' and $P - P'$. The problem explored in this section is to decompose S into admissible irreducible projection operators.[13] More generally, we consider the decomposition of an *arbitrary finite-dimensional admissible projection operator* P instead of S.[14]

Definition 5.3: If $H^\nu T \neq 0$ but $H^{\nu+1} T = 0$, then ν is called the *rank* of T. If $H^\nu \varphi \neq 0$ but $H^{\nu+1}\varphi = 0$, then ν is the *rank* of φ.

Since H is nilpotent, any operator and any vector have a finite rank $\nu \leq N$. Evidently, T has a rank ν if and only if the maximum rank of the vectors belonging to $V[T]$ is ν.

Theorem 12: Let $\varphi_1, \cdots, \varphi_k \in V[P]$ and the rank of φ_j be ν_j. If k vectors $H^{\nu_j}\varphi_j$, $j = 1, \cdots, k$, are linearly independent, then $\sum_{j=1}^{k} (\nu_j + 1)$ vectors $H^n \varphi_j$ ($n = 0, 1, \cdots, \nu_j$; $j = 1, \cdots, k$) belong to $V[P]$ and are linearly independent.

Proof: The first statement is obvious because $H^n \varphi_j = H^n P \varphi_j = P(H^n \varphi_j)$. To show the linear independence, we assume the contrary. Suppose that

$$\psi \equiv \sum_{j=1}^{k} \sum_{n=0}^{\nu_j} c_{jn} H^n \varphi_j = 0, \tag{5.1}$$

where complex numbers c_{jn} are not all zero. Let c_{jm_j} be the first nonzero one of $c_{j0}, c_{j1}, \cdots, c_{j\nu_j}$ (if all of them are zero, let $m_j = \nu_j + 1$). We set

$$s \equiv \max_j (\nu_j - m_j) \geq 0. \tag{5.2}$$

Since $H^n \varphi_j = 0$ for $n \geq \nu_j + 1$, we have

$$\begin{aligned} 0 = H^s \psi &= \sum_{j=1}^{k} \sum_{n=m_j}^{\nu_j - s} c_{jn} H^{n+s} \varphi_j \\ &= \sum_{j=1}^{k} c_{jm_j} \delta_{m_j, \nu_j - s} H^{\nu_j} \varphi_j . \end{aligned} \tag{5.3}$$

Since $H^{\nu_j}\varphi_j$, $j = 1, \cdots, k$, are linearly independent, we must have

$$c_{jm_j} = 0 \quad \text{for} \quad m_j = \nu_j - s. \tag{5.4}$$

This contradicts the definition of c_{jm_j}. QED

Hereafter, let ν be the rank of P. We construct an appropriate set of basis vectors of $V[P]$ in the following.

Let $H^\nu P \psi_{0k}$, $k = 1, \cdots, m_0$, be basis vectors of $V_0 \equiv V[H^\nu P]$; of course $H^{\nu+1} P \psi_{0k} = 0$. Next we consider a space V_1 defined by

$$V_1 \equiv \{\psi \in V[H^{\nu-1}P] \mid H\psi = 0\}. \tag{5.5}$$

Since $H^\nu P \psi_{0k} = H^{\nu-1} P (H \psi_{0k})$, we see $H^\nu P \psi_{0k} \in V_1$.

If there exist any other independent vectors in V_1, then we denote them by

$$H^{v-1}P\psi_{1k}, \quad k = 1, \cdots, m_1.$$

Likewise, we consider

$$V_2 \equiv \{\psi \in V[H^{v-2}P] \,|\, H\psi = 0\}.$$

If there exist any independent vectors other than $H^vP\psi_{0k}$ and $H^{v-1}P\psi_{1k}$, we denote them by

$$H^{v-2}P\psi_{2k}, \quad k = 1, \cdots, m_2.$$

We continue this procedure until we reach $P\psi_{vk}$. We thus obtain $\sum_{l=0}^{v} m_l$ vectors $\varphi_{lk} \equiv P\psi_{lk} \in V[P]$ ($l = 0, 1, \cdots, v$; $k = 1, \cdots, m_l$). By definition, the rank of φ_{lk} is $v - l$. Furthermore, since $H^{v-l}\varphi_{lk}$ ($l = 0, 1, \cdots, v$; $k = 1, \cdots, m_l$) are linearly independent, Theorem 12 implies that the vectors

$$H^j\varphi_{lk}, \quad l = 0, 1, \cdots, v, \quad k = 1, \cdots, m_l, \quad j = 0, \cdots, v - l, \tag{5.6}$$

belong to $V[P]$ and are linearly independent.

Theorem 13: Any vector $\psi \in V[P]$ can be written as a linear combination of the vectors listed in (5.6).

Proof: Let n be the rank of ψ. If $n = 0$, ψ is expressed in terms of $H^{v-l}\varphi_{lk}$ ($l = 0, 1, \cdots, v$; $k = 1, \cdots, m_l$) because they are basis vectors of the space of the eigenvectors of H by construction. We employ mathematical induction with respect to n. We assume that any $\psi' \in V[P]$ of a rank $n \leq p$ is written as a linear combination of $H^{v-l-j}\varphi_{lk}$ ($j = 0, 1, \cdots, n$; $l = 0, 1, \cdots, v - j$; $k_l = 1, \cdots, m_l$). Then, for any $\psi \in V[P]$ of a rank $p + 1$, since the rank of $H\psi$ is p, we have

$$H\psi = \sum_{j=0}^{p} \sum_{l=0}^{v-j} c_{jl} H^{v-l-j}\varphi_l \tag{5.7}$$

together with

$$\varphi_l \equiv \sum_{k=1}^{m_l} b_k \varphi_{lk}, \quad \sum_k |b_k| = 1; \tag{5.8}$$

that is,

$$H\left(\psi - \sum_{j=1}^{p+1} \sum_{l=0}^{v-j} c_{j-1,l} H^{v-l-j}\varphi_l\right) = \sum_{j=0}^{p} c_{j,v-j}\varphi_{v-j}. \tag{5.9}$$

We prove that the right-hand side of (5.9) vanishes. If not, let q ($\leq p$) be the smallest j such that $c_{j,v-j} \neq 0$. Let

$$c_{q,v-q}\chi \equiv \sum_{j=q}^{p} c_{j,v-j}\varphi_{v-j}. \tag{5.10}$$

Then of course $H^q\chi = H^q\varphi_{v-q}$. On the other hand, (5.9) implies that we can write $\chi = H\chi'$ ($\chi' \in V[P]$). Hence $H^q\varphi_{v-q}$ is an eigenvector of H such that it can be written as $H^{q+1}P\chi'$. By construction, such a vector has to be expressed in terms of $H^{v-l}\varphi_{lk}$ ($l = 0, \cdots, v - q - 1$; $k_l = 1, \cdots, m_l$), but this contradicts the linear independence of (5.6). Thus the vector in the parentheses of (5.9) is an eigenvector of H. Therefore it can be written as

$$\sum_{l=0}^{v} c_l H^{v-l}\varphi_l. \tag{5.11}$$

QED

Thus the vectors listed in (5.6) are basis vectors of $V[P]$.[15] Hence

$$\dim V[P] = \sum_{l=0}^{v} m_l(v - l + 1), \tag{5.12}$$

while the number of independent eigenvectors of H in $V[P]$ is $\sum_{l=0}^{v} m_l$.

Since (5.6) gives basis vectors of $V[P]$, Theorem 2 implies that there exist uniquely linearly independent vectors $\tilde{\varphi}_{lkj} \in \tilde{V}[P]$ ($l = 0, \cdots, v$; $k = 1, \cdots, m_l$; $j = 0, \cdots, v - l$) such that

$$P = \sum_{l=0}^{v} \sum_{k=1}^{m_l} \sum_{j=0}^{v-l} H^j\varphi_{lk}\tilde{\varphi}_{lkj} \tag{5.13}$$

together with

$$\tilde{\varphi}_{lkj}H^{j'}\varphi_{l'k'} = \delta_{ll'}\delta_{kk'}\delta_{jj'}, \quad 0 \leq j, j' \leq v - l. \tag{5.14}$$

We set $\tilde{\varphi}_{lk} \equiv \tilde{\varphi}_{kl,v-l}$. Then, with the aid of (5.14) and $H^{v-l+1}\varphi_{lk} = 0$, we find

$$\begin{aligned}\tilde{\varphi}_{lk}H^j &= \tilde{\varphi}_{lk}H^jP \\ &= \sum_{l'=0}^{v} \sum_{k'=0}^{m_l} \sum_{j'=0}^{v-l} (\tilde{\varphi}_{lk,v-l}H^{j+j'}\varphi_{l'k'})\tilde{\varphi}_{l'k'j'} \\ &= \tilde{\varphi}_{lk,v-l-j}.\end{aligned} \tag{5.15}$$

Thus

$$P = \sum_{l=0}^{v} \sum_{k=1}^{m_l} \sum_{j=0}^{v-l} H^j\varphi_{lk}\tilde{\varphi}_{lk}H^{v-l-j} \tag{5.16}$$

together with

$$\tilde{\varphi}_{lk}H^n\varphi_{l'k'} = \delta_{ll'}\delta_{kk'}\delta_{n,v-l}, \quad n = 0, 1, \cdots. \tag{5.17}$$

Theorem 14: The projection operator P can be decomposed (not uniquely) into

$$P = \sum_{l=0}^{v} \sum_{k=1}^{m_l} P_{lk} \tag{5.18}$$

with

$$P_{lk}P_{l'k'} = P_{l'k'}P_{lk} = \delta_{ll'}\delta_{kk'}P_{lk}, \tag{5.19}$$

where P_{lk} is an admissible irreducible projection operator of a rank $v - l$ such that $\dim V[P_{lk}] = v - l + 1$ and

$$P_{lk} = \sum_{j=0}^{v-l} H^j\varphi_{lk}\tilde{\varphi}_{lk}H^{v-l-j} \tag{5.20}$$

together with (5.17).

Proof: We have only to prove the irreducibility of P_{lk}. Let P'_k be an admissible projection operator such that $P_{lk}P'_{lk} = P'_{lk}P_{lk} = P'_{lk}$. Since P_{lk} has a rank $\nu - l$, P'_{lk} [or $P_{lk} - P'_{lk}$] has the rank $\nu - l$. Then there exists a vector $\varphi \in V[P'_{lk}]$ of the rank $\nu - l$. Since $H^{\nu-l}\varphi \neq 0$, Theorem 12 implies that dim $V[P'_{lk}] \geq \nu - l + 1$. According to Theorem 4, however, we have

$$\nu - l + 1 = \dim V[P'_{lk}] + \dim V[P_{lk} - P'_{lk}]. \quad (5.21)$$

Therefore, we have dim $V[P_{lk} - P'_{lk}] = 0$, that is, $P'_{lk} = P_{lk}$. QED

Theorem 15: When P is written as a direct sum of admissible irreducible projection operators, we can always express it as (5.18) together with (5.20); and if there are two expressions

$$P = \sum_{l=0}^{\nu}\sum_{k=1}^{m_l} P_{lk} = \sum_{l=0}^{\nu}\sum_{k=1}^{m_l'} P'_{lk}, \quad (5.22)$$

then we have $m_l = m'_l$, $l = 0, 1, \cdots, \nu$.

Proof: Any irreducible component of P of a rank $\nu - l$ is written as (5.20) because, otherwise supposing it as the P of Theorem 14, we could decompose it into P_{lk}. Next, for (5.22), we have

$$H^nP = \sum_{l=0}^{\nu-n}\sum_{k=1}^{m_l} H^nP_{lk} = \sum_{l=0}^{\nu-n}\sum_{k=1}^{m_l'} H^nP'_{lk}. \quad (5.23)$$

Hence

$$\dim V[H^nP] = \sum_{l=0}^{\nu-n} m_l(\nu - l + 1) = \sum_{l=0}^{\nu-n} m'_l(\nu - l + 1),$$
$$n = \nu, \nu - 1, \cdots, 0, \quad (5.24)$$

from which $m_l = m'_l$ follows. QED

By applying the above results to S, we obtain

$$S = \sum_{l=0}^{N}\sum_{k=1}^{m_l} S_{lk}, \quad (5.25)$$

$$S_{lk} = \sum_{j=0}^{N-l} H^j\varphi_{lk}\tilde{\varphi}_{lk}H^{N-l-j}, \quad (5.26)$$

$$\tilde{\varphi}_{lk}H^n\varphi_{l'k'} = \delta_{ll'}\delta_{kk'}\delta_{n,N-l}, \quad (5.27)$$

$$M = \sum_{l=0}^{N} m_l(N - l + 1), \quad (5.28)$$

$$S^{[n]} = \sum_{l=0}^{N-n}\sum_{k=1}^{m_l} H^nS_{lk}, \quad (5.29)$$

$$H^nS_{lk} = \sum_{j=0}^{N-n-l} H^{n+j}\varphi_{lk}\tilde{\varphi}_{lk}H^{N-l-j}, \quad (5.30)$$

$$\dim V[S^{[n]}] = \dim \tilde{V}[S^{[n]}] = \sum_{l=0}^{N-n} m_l(N - n - l + 1). \quad (5.31)$$

Since

$$S = \sum_{m=1}^{M} R_m^{(N_m)} = \sum_{m=1}^{M}\sum_{n=0}^{N_m} \phi_m^{(n)}\tilde{\phi}_m^{(N_m-n)}, \quad (5.32)$$

φ_{lk} is expressed in terms of $\phi_m^{(n)}$ ($m = 1, \cdots, M$; $n = 0, 1, \cdots, N_m$) as

$$\varphi_{lk} = \sum_{m=1}^{M}\sum_{n=0}^{N_m} (\tilde{\phi}_m^{(N_m-n)}\varphi_{lk})\phi_m^{(n)}. \quad (5.33)$$

Since

$$[A(0)]^{N-n+1}(H^{n+j}\varphi_{lk}) = 0$$

and

$$(\tilde{\varphi}_{lk}H^{N-l-j})[A(0)]^{N-n+1} = 0 \quad (5.34)$$

for $0 \leq j \leq N - n - l$ because of (4.24), the expressions (5.25)–(5.30) show the existence of the solutions to the problem posed at the end of Sec. 2. In general it is very difficult, however, to find φ_{lk} and $\tilde{\varphi}_{lk}$ explicitly. In the next section, we explicitly construct them in terms of $\phi_m^{(n)}$ and $\tilde{\phi}_m^{(n)}$ in two special cases $M = N + 1$ and $M = 2$.

6. SPECIAL CASES

A. Case $M = N + 1$ ($N \geq 1$)

Let us take the case of $M = N + 1$ ($N \geq 1$), where all of $w_1^{(1)}, \cdots, w_M^{(1)}$ are different from each other. This case is the most important case in the application to the Bethe–Salpeter formalism.[5]

From the consideration of Sec. 4, we have $N_m = N$ for all m and $M = M_I = N + 1$. The rank of S is N according to (4.47), and dim $V[S] = N + 1$. Hence S is irreducible. More precisely, according to Theorem 14, if we find a vector $\varphi \in V[S]$ of the rank N, we can write

$$S = \sum_{l=0}^{N} H^l\varphi\tilde{\varphi}H^{N-l}, \quad (6.1)$$

where $\tilde{\varphi} \in \tilde{V}[S]$ is a vector satisfying

$$\tilde{\varphi}H^n\varphi = \delta_{nN}, \quad n = 0, 1, \cdots \quad (6.2)$$

Hence

$$S^{[n]} = \sum_{l=0}^{N-n} H^{n+l}\varphi\tilde{\varphi}H^{N-l}. \quad (6.3)$$

From (4.46), we have

$$\phi_m^{(0)}\tilde{\phi}_m^{(0)} = \xi_m S^{[N]}. \quad (6.4)$$

On setting

$$\phi \equiv H^N\varphi, \quad \tilde{\phi} \equiv \tilde{\varphi}H^N, \quad (6.5)$$

therefore, we should have

$$\phi_m^{(0)} = a_m\phi, \quad \tilde{\phi}_m^{(0)} = b_m\tilde{\phi}, \quad (6.6)$$

where a_m and b_m are certain complex numbers satisfying $a_mb_m = \xi_m$.

From (5.32), we write

$$S = \psi\tilde{\phi} + U, \quad (6.7)$$

where

$$U \equiv \sum_{m=1}^{N+1} \sum_{n=0}^{N-1} \phi_m^{(n)} \tilde{\phi}_m^{(N-n)}, \tag{6.8}$$

$$\psi \equiv \sum_{m=1}^{N+1} b_m \phi_m^{(N)}. \tag{6.9}$$

Because of (4.44), the rank of U is *less* than N. Since the rank of S is N, therefore, the rank of ψ has to be N. Hence, with the aid of $H^N S = H^N$, we find that $S\psi$ belongs to $V[S]$ and has the rank N. We can therefore set

$$\varphi \equiv S\psi = (\tilde{\phi}\psi)\psi + U\psi. \tag{6.10}$$

Thus we have obtained $N+1$ independent vectors $H^n S\psi$, $n = 0, 1, \cdots, N$, in $V[S]$ in terms of $\phi_m^{(n)}$ $(m = 1, \cdots, N+1;\ n = 0, 1, \cdots, N)$.[16] We note that, since the rank of U is less than N, we have

$$\tilde{\phi}\psi \neq 0. \tag{6.11}$$

We can make a similar consideration in the dual space. With

$$\tilde{\psi} \equiv \sum_{m=1}^{N+1} a_m \tilde{\phi}_m^{(N)}, \tag{6.12}$$

$\tilde{\psi}S \in \tilde{V}[S]$ has the rank N. According to the dual proposition of Theorem 12, therefore, we can write

$$\tilde{\varphi} = \sum_{n=0}^{N} c_n \tilde{\psi} S H^n, \quad c_0 \neq 0. \tag{6.13}$$

The coefficients c_n can be successively determined by (6.2); they are expressed in terms of $h_n \equiv \tilde{\psi} H^{N-n} S\psi$, $n = 0, 1, \cdots, N$. Indeed, rewriting (6.2) as

$$\sum_{n=0}^{N} x^n \sum_{l=0}^{n} c_l h_{n-l} = 1, \tag{6.14}$$

we obtain

$$c_n = \frac{1}{n!}\left(\frac{d}{dx}\right)^n \left(\sum_{l=0}^{N} h_l x^l\right)^{-1}\Bigg|_{x=0}. \tag{6.15}$$

In particular,

$$c_0^{-1} = h_0 = \tilde{\psi} H^N \psi = \tilde{\phi}\psi = \tilde{\psi}\phi. \tag{6.16}$$

B. Case $M = 2$ $(N \geq 1)$

From the continuity in z of

$$S(z) = R_1(z) + R_2(z), \tag{6.17}$$

we obtain $N = N_1 = N_2$ and the cancellation conditions

$$R_1^{(n)} + R_2^{(n)} = 0, \quad n = 0, 1, \cdots, N-1, \tag{6.18}$$

with $R_m^{(0)} \neq 0$. Because of (6.17), we can write

$$S^{[n]}(z) = \{[w_1(z)]^n - [w_2(z)]^n\} R_1(z) + [w_2(z)]^n S(z) \tag{6.19}$$

for $n \geq 1$, and (6.19) has to be continuous in z at $z = 0$. From the case $n = 1$, we find

$$\sum_{j=1}^{N} z^j (w_1^{(j)} - w_2^{(j)}) \cdot \sum_{k=0}^{N} z^{-N+k} R_1^{(k)} = O(1), \tag{6.20}$$

and therefore[8]

$$w_1^{(j)} = w_2^{(j)}, \quad j = 1, 2, \cdots, N-1. \tag{6.21}$$

Then the continuity in z of $S^{[n]}$ for $n \geq 2$ is automatically satisfied. We also have

$$S = R_1^{(N)} + R_2^{(N)}, \tag{6.22}$$

$$H \equiv S^{[1]} = (w_1^{(N)} - w_2^{(N)}) R_1^{(0)}. \tag{6.23}$$

From (6.23), we find that there is a double pole if and only if $w_1^{(N)} \neq w_2^{(N)}$.[8]

Now, our main task is to solve (6.18). For the moment, we do not consider (4.32) or (4.33). Then $\phi_1^{(n)}$ and $\tilde{\phi}_1^{(n)}$, $n = 0, 1, \cdots, N$, are completely arbitrary. Given them, $\phi_2^{(n)}$ and $\tilde{\phi}_2^{(n)}$, $n = 0, 1, \cdots, N-1$, satisfy (6.18) if and only if

$$\phi_2^{(n)} = a \sum_{j=0}^{n} (-\alpha_j + \beta_j) \phi_1^{(n-j)}, \tag{6.24}$$

$$\tilde{\phi}_2^{(n)} = -a^{-1} \sum_{j=0}^{n} (\alpha_j + \beta_j) \tilde{\phi}_1^{(n-j)}, \tag{6.25}$$

where $a \neq 0$, $\alpha_0 = 0$, and $\alpha_1, \cdots, \alpha_{N-1}$ are complex numbers. Let

$$f(x) \equiv \sum_{j=0}^{\infty} \alpha_j x^j \tag{6.26}$$

with $\alpha_n = 0$ for $n \geq N$. Then a formal power series of a generating function,

$$\{1 + [f(x)]^2\}^{\frac{1}{2}} = \sum_{j=0}^{\infty} \beta_j x^j, \tag{6.27}$$

defines β_j; for example, $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = \alpha_1^2/2$, $\beta_3 = \alpha_1 \alpha_2, \cdots$. The proof of (6.24) and (6.25) is presented in the Appendix, in which we also show that

$$S = \chi \tilde{\phi}_1^{(0)} + \phi_1^{(0)} \tilde{\chi}, \tag{6.28}$$

where

$$\chi \equiv \phi_1^{(N)} - a^{-1} \phi_2^{(N)} + \sum_{j=1}^{N} (-\alpha_j + \beta_j) \phi_1^{(N-j)}, \tag{6.29}$$

$$\tilde{\chi} \equiv \tilde{\phi}_1^{(N)} + a \tilde{\phi}_2^{(N)} + \sum_{j=1}^{N} (\alpha_j + \beta_j) \tilde{\phi}_1^{(N-j)}. \tag{6.30}$$

Next, we take (4.33) into account. For $l = m = 1$, we have $N+1$ independent constraints

$$\sum_{k=0}^{n} \tilde{\phi}_1^{(k)} \phi_1^{(n-k)} = \delta_{nN}, \quad n = 0, 1, \cdots, N. \tag{6.31}$$

Then the remainders of (4.33) are automatically satisfied as long as $n \leq N-1$, because $\phi_2^{(n)}$ and $\tilde{\phi}_2^{(n)}$

are linear combinations of $\phi_1^{(0)}, \cdots, \phi_1^{(n)}$ and of $\tilde{\phi}_1^{(0)}, \cdots, \tilde{\phi}_1^{(n)}$, respectively. For $l = 1$, $m = 2$, and $n = N$, Eqs. (6.24), (6.31), and (6.29) yield

$$0 = \sum_{k=0}^{N} \tilde{\phi}_1^{(k)} \phi_2^{(N-k)} = a(1 - \tilde{\phi}_1^{(0)}\chi), \tag{6.32}$$

so that

$$\tilde{\phi}_1^{(0)}\chi = 1. \tag{6.33}$$

Likewise, we find

$$\tilde{\chi}\phi_1^{(0)} = 1 \tag{6.34}$$

from the case $l = 2$, $m = 1$, and $n = N$ of (4.33). The final case $l = m = 2$ and $n = N$ is automatically satisfied under (6.33) and (6.34). From (6.33) and (6.34) together with $\tilde{\phi}_1^{(0)}\phi_1^{(0)} = 0$, we find that χ and $\phi_1^{(0)}$ ($\tilde{\chi}$ and $\tilde{\phi}_1^{(0)}$) are linearly independent. Thus, dim $V[S] = 2 = M$, as it should be. Furthermore, from $S^2 = S$ we have

$$\tilde{\chi}\chi = 0. \tag{6.35}$$

Since (6.23) implies that

$$H\chi = (w_1^{(N)} - w_2^{(N)})\phi_1^{(0)}, \tag{6.36}$$

$$\tilde{\chi}H = (w_1^{(N)} - w_2^{(N)})\tilde{\phi}_1^{(0)}, \tag{6.37}$$

we can rewrite (6.28) as

$$S = \varphi\tilde{\varphi}H + H\varphi\tilde{\varphi}, \tag{6.38}$$

for $w_1^{(N)} \neq w_2^{(N)}$ by setting $\varphi \equiv \chi$ and $\tilde{\varphi} \equiv (w_1^{(N)} - w_2^{(N)})^{-1}\tilde{\chi}$, where

$$\tilde{\varphi}H^n\varphi = \delta_{n1}. \tag{6.39}$$

For $w_1^{(N)} = w_2^{(N)}$, S is reducible into $\chi\tilde{\phi}_1^{(0)}$ and $\phi_1^{(0)}\tilde{\chi}$.

7. DISCUSSION

In the present paper, we have developed a general theory of the residues of multiple poles synthesized out of coinciding simple poles. Though we have introduced an infinite-dimensional space X, our problem is essentially of finite-dimensional nature. Indeed, $V[S]$ is included in a finite-dimensional space Y spanned by $\phi_m^{(n)}$ ($m = 1, \cdots, M$; $n = 0, 1, \cdots, N_m$). In order to prove dim $V[S] = M$, we have introduced an *ad hoc* assumption (Assumption 1), but it might be possible to avoid it by using the finite dimensionality of Y. The special cases discussed in Sec. 6 and some other simple examples suggest that dim $V[S] = M$ would be a property which can be proven without using any topological concepts.

In the Bethe–Salpeter formalism, given $\varphi \in X$, its conjugate $\bar{\varphi} \in \bar{X}$ can be constructed immediately according to a general rule.[9] It should be noted, however, that $\bar{\varphi}$ is not the adjoint vector of φ in the sense of a Hilbert space. We have to introduce a norm factor η_m ($= \pm 1$) associated with $R_m(z)$.[9] Then we may define the conjugate $\bar{\varphi}$ of every vector

$$\varphi \equiv \sum_{m,n} \alpha_{mn}\phi_m^{(n)} \in Y \tag{7.1}$$

by

$$\bar{\varphi} \equiv \sum_{m,n} \alpha_{mn}^*\eta_m\bar{\phi}_m^{(n)} \in \bar{Y}, \tag{7.2}$$

provided that any identity of the form

$$\sum_{m,n} \beta_{mn}\phi_m^{(n)} = 0 \tag{7.3}$$

can hold if and only if

$$\sum_{m,n} \beta_{mn}^*\eta_m\bar{\phi}_m^{(n)} = 0. \tag{7.4}$$

In particular, we have $\bar{\phi}_m^{(n)} = \eta_m\tilde{\phi}_m^{(n)}$. We can also define the conjugate of a linear operator T if $V[T] \subset Y$ and $\tilde{V}[T] \subset \tilde{Y}$. Then S becomes self-conjugate. Thus, it is possible to introduce the conjugate operation in an abstract way, but the self-conjugate property of S does not seem to be very powerful.[17]

Finally, we consider multiple poles on the z plane. Let $z_m(w)$ be the inverse function of $w_m(z)$, and let

$$T(z, w) \equiv \sum_m \frac{Q_m(w)}{z - z_m(w)} \tag{7.5}$$

be the singular part of $F(z, w)$ on the z plane. Since

$$\frac{\partial F}{\partial z} = F\frac{dA}{dz}F \tag{7.6}$$

instead of (2.10), we have

$$Q_l(w)A'Q_m(w) = \delta_{lm}Q_m(w) \tag{7.7}$$

with $A' \equiv -dA/dz|_{z=0}$. In the same way as in (2.16), we find that the residues of multiple poles are given by

$$T^{[n]} \equiv \lim_{w\to 0} T^{[n]}(w) \tag{7.8}$$

with

$$T^{[n]}(w) \equiv \sum_m [z_m(w)]^n Q_m(w). \tag{7.9}$$

Therefore, on setting $R_m(w) \equiv Q_m(w)A'$ and $S^{[n]}(w) \equiv T^{[n]}(w)A'$, we find that $R_m(w)$ and $S^{[n]}(w)$ satisfy the same assumptions stated in Sec. 4. Thus all results presented in the previous sections are applicable to $T^{[n]}A'$. This fact is a merit of the axiomatic approach. But the formulas obtained in this way are not directly related to the generalized Bethe–Salpeter equations[4,5,9] for $T^{[n]}$, which follow from

$$A(z)\sum_{n=0}^{N} T^{[n]}/z^{n+1} = O(1), \tag{7.10}$$

in contrast with the theory on the w plane.

APPENDIX: PROOF OF (6.24)–(6.30)

We prove (6.24) and (6.25) by mathematical induction. For $n = 0$, we have $\phi_2^{(0)} = a\phi_1^{(0)}$ and $\tilde{\phi}_2^{(0)} = -a^{-1}\tilde{\phi}_1^{(0)}$, $a \neq 0$, from $\phi_1^{(0)}\tilde{\phi}_1^{(0)} + \phi_2^{(0)}\tilde{\phi}_2^{(0)} = 0$.

Hence we assume that the cancellation conditions (6.18) for $n = 0, 1, \cdots, k-1$, $1 \leq k \leq N$, are equivalent to (6.24) and (6.25) for $n = 0, 1, \cdots, k-1$. For $n = k$, we have

$$\begin{aligned}\sum_{m=1}^{2} R_m^{(k)} &= \sum_{j=0}^{k} \phi_1^{(j)} \tilde{\phi}_1^{(k-j)} + \sum_{j=0}^{k} \phi_2^{(j)} \tilde{\phi}_2^{(k-j)} \\ &= \sum_{j=0}^{k} \phi_1^{(j)} \tilde{\phi}_1^{(k-j)} + (\phi_2^{(k)} - \psi^{(k)}) \tilde{\phi}_2^{(0)} \\ &\quad + \phi_2^{(0)} (\tilde{\phi}_2^{(k)} - \tilde{\psi}^{(k)}) + K^{(k)},\end{aligned}$$

where

$$\psi^{(k)} \equiv a \sum_{j=0}^{k} (-\alpha_{kj} + \beta_j) \phi_1^{(k-j)},$$

$$\tilde{\psi}^{(k)} \equiv -a^{-1} \sum_{j=0}^{k} (\alpha_{kj} + \beta_j) \tilde{\phi}_1^{(k-j)},$$

$$K^{(k)} \equiv \psi^{(k)} \tilde{\phi}_2^{(0)} + \phi_2^{(0)} \tilde{\psi}^{(k)} + \sum_{j=1}^{k-1} \phi_2^{(j)} \tilde{\phi}_2^{(k-j)}$$

with $\alpha_{kj} = \alpha_j$ for $j = 0, 1, \cdots, k-1$ and $\alpha_{kj} = 0$ for $j = k$.[18] The induction assumption implies that

$$\begin{aligned}K^{(k)} &= -\sum_{l=0}^{k} \Big[\sum_{i=0}^{l} (-\alpha_{k,l-i} + \beta_{l-i}) \phi_1^{(j)} \\ &\quad \times \sum_{j=0}^{k-l} (\alpha_{k,k-l-j} + \beta_{k-l-j}) \tilde{\phi}_1^{(j)} \Big] \\ &= -\sum_{i=0}^{k} \sum_{j=0}^{k-i} \phi_1^{(i)} \tilde{\phi}_1^{(j)} \\ &\quad \times \Big[\sum_{l=i}^{k-j} (-\alpha_{k,l-i} + \beta_{l-i})(\alpha_{k,k-l-j} + \beta_{k-l-j}) \Big].\end{aligned}$$

Lemma 1:

$$\sum_{p=0}^{n} (-\alpha_p + \beta_p)(\alpha_{n-p} + \beta_{n-p}) = \delta_{n0}.$$

Proof: From (6.26) and (6.27),

$$\begin{aligned}\sum_{n=0}^{\infty} \sum_{p=0}^{n} &(-\alpha_p + \beta_p)(\alpha_{n-p} + \beta_{n-p}) x^n \\ &= \sum_{p=0}^{\infty} (-\alpha_p + \beta_p) x^p \sum_{q=0}^{\infty} (\alpha_q + \beta_q) x^q \\ &= \{-f(x) + [1 + (f(x))^2]^{\frac{1}{2}}\} \\ &\quad \times \{f(x) + [1 + (f(x))^2]^{\frac{1}{2}}\} \\ &= 1.\end{aligned}$$

The formula of Lemma 1 can be rewritten as

$$\sum_{p=0}^{n} (-\alpha_{np} + \beta_p)(\alpha_{n,n-p} + \beta_p) = \delta_{n0},$$

because the coefficient of α_n is $-2\alpha_0 = 0$. Therefore,

$$K^{(k)} = -\sum_{i=0}^{k} \phi_1^{(i)} \tilde{\phi}_1^{(k-i)}.$$

For $k \leq N-1$, since $\sum_m R_m^{(k)} = 0$, we have

$$-a^{-1}(\phi_2^{(k)} - \psi^{(k)}) \tilde{\phi}_1^{(0)} + a \phi_1^{(0)} (\tilde{\phi}_2^{(k)} - \tilde{\psi}^{(k)}) = 0.$$

Accordingly, there is a complex number α_k such that

$$\begin{aligned}-a^{-1}(\phi_2^{(k)} - \psi^{(k)}) &= \alpha_k \phi_1^{(0)}, \\ a(\tilde{\phi}_2^{(k)} - \tilde{\psi}^{(k)}) &= -\alpha_k \tilde{\phi}_1^{(0)},\end{aligned}$$

from which we obtain (6.24) and (6.25) for $n = k$. If $k = N$, since $S = \sum_m R_m^{(N)}$, we obtain (6.28) together with (6.29) and (6.30) by setting

$$\begin{aligned}\chi &\equiv -a^{-1}(\phi_2^{(N)} - \psi^{(N)}), \\ \tilde{\chi} &\equiv a(\tilde{\phi}_2^{(N)} - \tilde{\psi}^{(N)}).\end{aligned}$$

* Work performed under the auspices of the U.S. Atomic Energy Commission.

† On leave of absence from Research Institute for Mathematical Science, Kyoto University, Kyoto 606, Japan.

[1] W. Heisenberg, Nucl. Phys. **4**, 532 (1957).

[2] References on multipole ghosts are found in Nagy's review articles: K. Nagy, Nuovo Cimento Suppl. **17**, 92 (1960); *State Vector Spaces with Indefinite Metric in Quantum Field Theory* (P. Noordhoff, Groningen, 1966). See also J. Lukierski, Acta Phys. Polon. **32**, 551 (1967); **32**, 771 (1967); K. Yokoyama and R. Kubo, Progr. Theoret. Phys. (Kyoto) **41**, 542 (1969).

[3] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **35**, 1111 (1966); **38**, 881 (1967). See also, K. Yokoyama, *ibid.* **40**, 160 (1968); **40**, 421 (1968); **41**, 1384 (1969).

[4] N. Nakanishi, Phys. Rev. **140**, B947 (1965). See also N. Nakanishi, *ibid.* **147**, 1153 (1966); Progr. Theoret. Phys. (Kyoto) **39**, 1585 (1968); J. Arafune, *ibid.* **40**, 620 (1968); S. Naito, *ibid.* **41**, 500 (1969).

[5] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **41**, 233 (1969). See also N. Nakanishi, *ibid.* **41**, 780 (1969); M. Minami, *ibid.* **41**, 1328 (1969).

[6] D. Z. Freedman and J.-M. Wang, Phys. Rev. **153**, 1596 (1967). There are a number of papers on the unequal-mass conspiracy.

[7] N. Nakanishi, Phys. Rev. **136**, B1830 (1964); D. Z. Freedman, C. E. Jones, and J.-M. Wang, *ibid.* **155**, 1645 (1967); N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **41**, 516 (1969); N. Nakanishi and N. Seto, *ibid.* **41**, 1094 (1969).

[8] M. Ida, Progr. Theoret. Phys. (Kyoto) **43**, 808 (1970).

[9] For details, see N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **43**, 1 (1969).

[10] Though it is natural to consider the poles on the s plane from the physical point of view, it is more convenient mathematically to work on the λ plane. Both approaches are equivalent if $w_m'(0) \neq 0$. See also Sec. 7.

[11] R. G. Cooke, *Linear Operators* (Macmillan, London, 1953), p. 350. Extension to a complex functional was made by Bohnenblust and Sobczyk.

[12] Here $F(z, w) - S(z, w)$ may be singular near $z = 0$ and $w = 0$, but this causes no difficulty because $S(z)$ is continuous in z.

[13] If we consider the trivially reducible case, S is of course reducible.

[14] We do not use the property $HS = H$ of S in this section.

[15] A similar construction of possible basis vectors of $V[P]$ was made by Ida (Ref. 8), but unfortunately he did not prove their linear independence. Indeed, in his notation, given linear independent vectors f_p, $p = 1, \cdots, d_\nu$, belonging to $M_{i,\nu}$ but not to $M_{i,\nu-1}$, $D^n f_p$, $p = 1, \cdots, d_\nu$, for $n \geq 1$ are *not* necessarily linearly independent.

[16] In this case, basis vectors were obtained also by Ida (Ref. 8) in a different, more complicated way.

[17] For example, the self-conjugateness of S and H yields only the following results in the two cases considered in Sec. 6: In Sec. 6A, all coefficients c_n are real if $b_m = a_m^* \eta_m$, and in Sec. 6B, $|a| = 1$ and $\mathrm{Re}\, \alpha_j = 0$, if $\eta_1 = -\eta_2 = 1$.

[18] Note that β_k does not depend on α_k.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Lorentz Invariance in a Gravitational Field

ROBERT L. KIRKWOOD*
The Rand Corporation, Santa Monica, California

(Received 23 July 1969; Revised Manuscript Received 28 January 1970)

In any theory of gravity in which free particles move along the geodesics of a 4-dimensional metric tensor, a particular class of metrics can be defined which correspond to the fields of Newton's theory of gravity. In these Newtonian fields the metric coefficients which describe intrinsic properties of space and time are clearly separated from those that describe the gravitational field. This separation suggests an invariance in the gravitational field which is quite similar to the usual Lorentz invariance of electromagnetism. The infinitesimal form of the generalized Lorentz transformation is determined by the fact that the 3-dimensional geometry remains Euclidean under the transformation. The finite form is determined so that the transformations form a group, and the group is found to be the usual Lorentz group. The transformation is then applied to fields that are not necessarily Newtonian.

I. INTRODUCTION

If gravity and electromagnetism are just different aspects of a unified field whose laws are in agreement with the principle of special relativity, then it seems reasonable to expect that there is a large-scale invariance in the gravitational field comparable to the known invariance of the electromagnetic field in the special theory of relativity. In the special theory, the intrinsic properties of space and time are described in a Lorentz frame by the flat-space metric $\eta_{\alpha\beta}$, where

$$\eta_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{pmatrix}. \tag{1}$$

The Lorentz transformation is defined such that it leaves these metric coefficients invariant. The electromagnetic field components are not invariant under the transformation, but instead they obey transformation relations which are chosen to make the form of the electromagnetic field equations invariant, and in this way they are treated very differently than are the intrinsic properties of space and time.

In contrast to this, most of the metric theories of gravity, including Einstein's theory, make no sharp distinction between quantities that describe intrinsic properties of space and time and those that describe the gravitational field. Instead, all of these quantities are combined in the metric tensor $g_{\alpha\beta}$, which is defined such that the local time $d\tau$ measured by a moving clock is determined by the expression

$$-c^2\, d\tau^2 = g_{\alpha\beta}\, dx_\alpha\, dx_\beta \tag{2}$$

and which has the property that its geodesics are the paths of free particles. In general, there is no system of coordinates in which $g_{\alpha\beta} = \eta_{\alpha\beta}$, and the usual Lorentz transformation does not leave the coefficients $g_{\alpha\beta}$ invariant. Most of the attempts that have been made to find some type of generalized Lorentz transformation which leaves the coefficients $g_{\alpha\beta}$ invariant have led to isometries, which are transformations that preserve the functional form of all the functions $g_{\alpha\beta}$. However, isometries are far too restrictive to play a role in the gravitational field similar to the one played by the Lorentz transformation in the electromagnetic field. For example, an isometry with as many parameters as the usual Lorentz transformation exists only in a space of constant curvature, and therefore does not exist even in a field as simple as the Schwarzschild field. Clearly, it is too much to ask that the functional form of *all* the metric coefficients be preserved by the transformation.

However, it may still be possible to preserve the form of *some* of the metric coefficients, provided that the remaining coefficients are allowed to transform in any way that is convenient. If the resulting invariance is to be similar to the usual Lorentz invariance of the electromagnetic field, then the invariant metric coefficients must be those that describe intrinsic properties of space and time, and the remaining metric coefficients will be those that describe the gravitational field. This suggests that the first step in investigating the possible existence of such an invariance must be to find a representation of the metric coefficients which separates the ones that describe intrinsic properties of space and time from those that describe the gravitational field. This will be done in this paper by considering the simplest and most important metric tensors, namely, those which describe the fields considered in Newtonian gravitational theory. It will be shown that for such metrics it is possible to introduce an infinitesimal transformation which keeps the intrinsic properties of space and time invariant, following the pattern of special theory. It

will then be shown that this infinitesimal transformation defines a finite transformation group, and finally this group will be applied to a more general class of fields.

II. NEWTONIAN FIELDS

In Newtonian gravitational theory it is assumed that there is a single universal time variable and that 3-space is Euclidean. In the context of a metric theory of gravity, this means that $g_{\alpha\beta}$ is such that there exists a time function, say x_4, such that the 3-dimensional geometry defined by $g_{\alpha\beta}$ on the surface $x_4 = \text{const}$ is Euclidean. It will be convenient throughout this paper to adopt the notation that Greek indices run from one to four and repeated Greek indices are summed from one to four, while Roman indices run from one to three and repeated Roman indices are summed from one to three. With this notation, the metrics which describe Newtonian fields are those for which there exist coordinates x_α in which

$$g_{ij} = \delta_{ij}, \tag{3}$$

where δ_{ij} is the identity. If g is the determinant of $g_{\alpha\beta}$, then in these coordinates it is seen that

$$g = g_{44} - g_{i3}g_{i4}.$$

Let $g^{\alpha\beta}$ be defined by the relations

$$\begin{aligned} g^{ij} &= \delta_{ij} + (1/g)g_{i4}g_{j4}, \\ g^{i4} &= -(1/g)g_{i4}, \\ g^{44} &= 1/g, \end{aligned} \tag{4}$$

and it is readily verified that

$$g^{\alpha\beta}g_{\beta\gamma} = \delta^\alpha_\gamma,$$

so that $g^{\alpha\beta}$ is the reciprocal of $g_{\alpha\beta}$.

In addition to the requirement that the coefficients g_{ij} satisfy Eq. (3), a restriction must be put on the remaining coefficients $g_{\alpha 4}$ if the laws of mechanics are to be those of Newton's theory. Since the path of a test particle is a geodesic of $g_{\alpha\beta}$, its equation of motion is

$$\frac{d^2x_\alpha}{d\tau^2} = -g^{\alpha\beta}(\gamma\delta, \beta)\frac{dx_\gamma}{d\tau}\frac{dx_\delta}{d\tau}.$$

Expressing $g^{\alpha\beta}$ in the form of Eqs. (4), we have that this equation of motion for $\alpha = 4$ is

$$\frac{d^2x_4}{d\tau^2} = \frac{1}{g}[g_{i4}(\gamma\delta, i) - (\gamma\delta, 4)]\frac{dx_\gamma}{d\tau}\frac{dx_\delta}{d\tau}.$$

Using this relation, we can write the equations for $\alpha = 1, 2, 3$ as

$$\frac{d^2x_i}{d\tau^2} = -(\gamma\delta, i)\frac{dx_\gamma}{d\tau}\frac{dx_\delta}{d\tau} - g_{i4}\frac{d^2x_4}{d\tau^2}. \tag{5}$$

In Newton's theory it is usually assumed that a moving clock will measure the time coordinate, so that $d\tau = dx_4$. To the extent that this nonrelativistic approximation is correct, $dx_4/d\tau = 1$, $d^2x_4/d\tau^2 = 0$, and $d^2x_i/d\tau^2$ is the acceleration of a moving particle relative to the coordinates x_i. To this approximation, then, Eq. (5) shows that the acceleration of a moving particle is

$$-(\gamma\delta, i)\frac{dx_\gamma}{dx_4}\frac{dx_\delta}{dx_4}.$$

Evaluating the Christoffel symbols $(\gamma\delta, i)$ for a metric which satisfies Eq. (3) gives this acceleration in the 3-dimensional form:

$$\left(\frac{\partial g_{j4}}{\partial x_i} - \frac{\partial g_{i4}}{\partial x_j}\right)\frac{dx_j}{dx_4} - \frac{\partial g_{i4}}{\partial x_4} + \frac{1}{2}\frac{\partial g_{44}}{\partial x_i}. \tag{6}$$

In Newton's theory the acceleration does not depend on the particle velocity dx_j/dx_4 and can be expressed as the negative gradient of the Newtonian potential function V. If the above expression is not to depend on dx_j/dx_4, it must be that

$$\frac{\partial g_{j4}}{\partial x_i} - \frac{\partial g_{i4}}{\partial x_j} = 0,$$

which is equivalent to requiring that a function β exists such that

$$g_{i4} = -\frac{\partial\beta}{\partial x_i}. \tag{7}$$

Then the acceleration given by expression (6) becomes

$$\frac{\partial}{\partial x_i}\left(\frac{\partial\beta}{\partial x_4} + \tfrac{1}{2}g_{44}\right),$$

which is of the form $-\partial V/\partial x_i$ if

$$V = -\frac{\partial\beta}{\partial x_4} - \tfrac{1}{2}g_{44} - \tfrac{1}{2}c^2. \tag{8}$$

Here the constant $\frac{1}{2}c^2$ has been added to make $V \to 0$ at great distances from any masses, where the field is static, so that $\partial\beta/\partial x_4 \to 0$, and the metric approaches the flat-space metric, so that $g_{44} \to -c^2$. Solving this for g_{44} and combining the result with Eqs. (3) and (7) shows that a Newtonian field is one in which there exist coordinates such that $g_{\alpha\beta}$ takes the form

$$\begin{aligned} g_{ij} &= \delta_{ij}, \\ g_{i4} &= -\frac{\partial\beta}{\partial x_i}, \\ g_{44} &= -2V - 2\frac{\partial\beta}{\partial x_4} - c^2, \end{aligned} \tag{9}$$

where β and V are arbitrary functions. Throughout this paper a metric field will be said to be Newtonian if such coordinates exist. In Newtonian gravitational theory it is further assumed that V satisfies Poisson's equation, but neither this equation nor Einstein's field equations will be assumed here. The results of this paper will depend on the assumption, made above, that the field is described by a metric whose geodesics are the possible paths of a free particle, but they will be independent of any particular system of field equations.

One important case of a Newtonian field is the one in which $V = -KM/r$ and $\beta = -(8KMr)^{\frac{1}{2}}$, where r is $(x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}}$. When these values are used in the metric coefficients of Eqs. (9), the associated quadratic form of Eq. (2) can be written in polar coordinates r, θ, and ϕ in the form

$$-c^2\,d\tau^2 = dr^2 + r^2(d\phi^2 + \sin^2\phi\,d\theta^2) + 2(2KM/r)^{\frac{1}{2}} \times dr\,dx_4 - [c^2 - (2KM/r)](dx_4)^2.$$

If the time coordinate x_4 is replaced by a variable l defined by

$$l = x_4 - \frac{1}{c^2}(8KMr)^{\frac{1}{2}} + \frac{4KM}{c^3}\tanh^{-1}\left(\frac{2KM}{c^2r}\right)^{\frac{1}{2}},$$

the quadratic form above becomes

$$-c^2\,d\tau^2 = \frac{dr^2}{[1 - (2KM/c^2r)]} + r^2(d\phi^2 + \sin^2\phi\,d\theta^2) - c^2\left(1 - \frac{2KM}{c^2r}\right)dl^2,$$

which is the Schwarzschild line element.

As shown above, the Newtonian fields are sufficient to describe all known fields to the accuracy of the nonrelativistic approximation that $d\tau = dx_4$. The only field in which it has been possible to verify the relativistic corrections to Newton's theory experimentally is the Schwarzschild field, where verifiable corrections have been found by using the Schwarzschild line element in Eq. (2), which determines the time measured by a moving clock, and in the geodesic equations that determine the motion of a particle or a light ray. Since it is now seen that the Schwarzschild field itself is Newtonian, it is clear that both Newtonian gravitational theory and all of the verifiable relativistic corrections to Newton's theory can be explained in terms of Newtonian fields if the geodesic equations and Eq. (2) are assumed to hold. Thus it is at least possible that the Newtonian fields are the only fields that occur in nature.

This observation provides the motivation for the description of gravity that will be used in this paper. As is customary today, gravity will be described in terms of a curved metric 4-space. The metric coefficients determine the time measured by a moving clock by means of Eq. (2), and they determine the paths of moving particles and light rays by means of the geodesic equations. However, in most descriptions of gravity it is further assumed that the metric coefficients are found as solutions of a set of field equations such as Einstein's equations. The weakness of this procedure is that the field equations are derived as much from philosophical considerations as from the observed facts, and this lends some unnecessary uncertainty to any results obtained from them. As far as the observational evidence is concerned, about all that can be said for any set of field equations is that they yield one important class of solutions which are closely approximated by Newtonian fields in which V is at least very nearly a solution of Poisson's equation in the coordinates x_α. If it is assumed that gravitational waves exist, even this last qualification is questionable, since Poisson's equation does not lead to wavelike solutions. Because of this uncertainty concerning the field equations, the analysis of this paper will be based only on the assumption that many physically important gravitational fields can be closely approximated by Newtonian fields. In the next two sections it will be assumed that the field is exactly Newtonian. Then it will be shown that the results of these two sections can be applied to any set of metric coefficients, although the application will be physically meaningful only in fields which closely approximate Newtonian fields.

A discussion of the physical meaning of a class of fields which includes the Newtonian fields has been given previously.[1]

III. THE INFINITESIMAL LORENTZ TRANSFORMATION IN A NEWTONIAN FIELD

The desired separation of the intrinsic properties of space and time from the properties of the gravitational field is very clear in the case of Newtonian fields. Here the intrinsic properties of space and time are described by the time variable x_4 and the Euclidean geometry in the 3-space defined by a constant value of x_4. If the three spatial coordinates x_i are defined to be Cartesian coordinates in this Euclidean 3-space, as has been done in Eqs. (9), then the four coordinates x_α will be called a Lorentz frame. In this Lorentz frame, the gravitational field is described by the quantities $g_{\alpha 4}$ given in Eqs. (9). If a second Lorentz frame is to be defined in a manner that will keep the intrinsic properties of space and time invariant, following the pattern of the special theory, then the coordinates x'_α in the new frame must be chosen such

that the metric coefficients g'_{ij} in the new frame are still equal to δ_{ij}. The coefficients $g'_{\alpha 4}$ may have any desired values and will not be specified in advance. The transformation can be determined from the known law of transformation of $g_{\alpha\beta}$, namely,

$$g_{\alpha\beta} = g'_{\gamma\delta}\frac{\partial x'_\gamma}{\partial x_\alpha}\frac{\partial x'_\delta}{\partial x_\beta}. \tag{10}$$

In this section, Eq. (10) will be solved for an infinitesimal transformation, that is, for a transformation in which $x'_\alpha = x_\alpha + \delta x_\alpha$ and $g'_{\alpha\beta} = g_{\alpha\beta} + \delta g_{\alpha\beta}$, where δx_α and $\delta g_{\alpha\beta}$ are so small that only terms of first order in δx_α and $\delta g_{\alpha\beta}$ need to be considered. For such a transformation, Eq. (10) becomes

$$\delta g_{\alpha\beta} = -g_{\alpha\gamma}\frac{\partial \delta x_\gamma}{\partial x_\beta} - g_{\beta\gamma}\frac{\partial \delta x_\gamma}{\partial x_\alpha}.$$

Since $g_{ij} = \delta_{ij}$, the condition that $g'_{ij} = \delta_{ij}$ is just that $\delta g_{ij} = 0$. Noting that $g_{\alpha\beta}$ has the form given by Eqs. (9), this condition is found to lead to the following differential equation for the infinitesimal transformation:

$$\frac{\partial \delta x_i}{\partial x_j} + \frac{\partial \delta x_j}{\partial x_i} = \frac{\partial \beta}{\partial x_i}\frac{\partial \delta x_4}{\partial x_j} + \frac{\partial \beta}{\partial x_j}\frac{\partial \delta x_4}{\partial x_i}. \tag{11}$$

Differentiating this with respect to x_k and solving for $\partial^2 \delta x_k/\partial x_i \partial x_j$ gives

$$\frac{\partial^2 \delta x_k}{\partial x_i \partial x_j} = \frac{\partial^2 \beta}{\partial x_i \partial x_j}\frac{\partial \delta x_4}{\partial x_k} + \frac{\partial^2 \delta x_4}{\partial x_i \partial x_j}\frac{\partial \beta}{\partial x_k}.$$

The condition that this is integrable for δx_k is that

$$\frac{\partial^2 \beta}{\partial x_i \partial x_j}\frac{\partial^2 \delta x_4}{\partial x_k \partial x_l} - \frac{\partial^2 \beta}{\partial x_i \partial x_l}\frac{\partial^2 \delta x_4}{\partial x_j \partial x_k} + \frac{\partial^2 \delta x_4}{\partial x_i \partial x_j}\frac{\partial^2 \beta}{\partial x_k \partial x_l} - \frac{\partial^2 \delta x_4}{\partial x_i \partial x_l}\frac{\partial^2 \beta}{\partial x_j \partial x_k} = 0.$$

Obviously, one solution of this equation is

$$\frac{\partial^2 \delta x_4}{\partial x_i \partial x_j} = 0. \tag{12}$$

If it is assumed that the determinant of $\partial^2\beta/\partial x_i \partial x_j$ does not vanish, this solution can be shown to be unique. In the Schwarzschild field, $\beta = -(8KMr)^{\frac{1}{2}}$, and the determinant of $\partial^2\beta/\partial x_i \partial x_j$ does not vanish anywhere. Since many fields of physical interest are small perturbations of this field, it follows that Eq. (12) *must* be satisfied by many physical fields. In the following analysis it will be assumed that Eq. (12) is satisfied.

From Eq. (12) it follows immediately that

$$\delta x_4 = m_i x_i + m, \tag{13}$$

where m_i and m are arbitrary functions of x_4. Then Eq. (11) can be rewritten as

$$\frac{\partial}{\partial x_i}(\delta x_j - \beta m_j) + \frac{\partial}{\partial x_j}(\delta x_i - \beta m_i) = 0.$$

The general solution of this equation is

$$\delta x_i = \beta m_i + n_{ij}x_j + n_i, \tag{14}$$

where n_{ij} and n_i are arbitrary functions of x_4 and where $n_{ij} = -n_{ji}$. The functions m_i, m, n_{ij}, and n_i can be evaluated if it is assumed that, in the region far away from all masses, $g_{\alpha\beta}$ approaches $\eta_{\alpha\beta}$ and the relation between dx'_α and dx approaches the usual Lorentz transformation. In this region the field is static and $g_{i4} \to 0$, so that $\partial\beta/\partial x_\alpha$ vanishes and the value of $d\beta$ associated with any dx_α vanishes. If the derivatives of m_i, m, n_{ij}, and n_i with respect to x_4 are denoted by a dot, the differentials of Eqs. (13) and (14) in the region far from all masses become

$$\begin{aligned} d\delta x_4 &= \dot{m}_i x_i\, dx_4 + m_i\, dx_i + \dot{m}\, dx_4, \\ d\delta x_i &= \beta \dot{m}_i\, dx_4 + \dot{n}_{ij}x_j\, dx_4 + n_{ij}\, dx_j + \dot{n}_i\, dx_4. \end{aligned} \tag{15}$$

The infinitesimal form of the usual Lorentz transformation of the special theory is

$$\begin{aligned} \delta x_4 &= -c^{-2}w_i x_i, \\ \delta x_i &= c_{ij}x_j - w_i x_4, \end{aligned} \tag{16}$$

where $c_{ij} = -c_{ji}$ are infinitesimal constants and w_i are the components of the infinitesimal transformation velocity. In order that Eqs. (15) be the same as the differentials of Eqs. (16) for all dx_α, it must be that $m_i = -w_i/c^2$, $n_{ij} = c_{ij}$, $\dot{m} = 0$, and $\dot{n}_i = -w_i$. The last two of these relations can be integrated to give $m = 0$ and $n_i = -w_i x_4$, where the possible additive constants have been dropped for simplicity. Then Eqs. (13) and (14) become

$$\begin{aligned} \delta x_4 &= -c^{-2}w_i x_i, \\ \delta x_i &= c_{ij}x_j - w_i(x_4 + \beta/c^2). \end{aligned} \tag{17}$$

This is the most general infinitesimal transformation which carries metric coefficients of the form of Eqs. (9) into new metric coefficients in such a way that $g'_{ij} = \delta_{ij}$ and which reduces to the usual Lorentz transformation between the coordinate differentials in the region far from all masses.

It is worth noting that, since Eq. (11) is a set of six equations in the four unknowns δx_α, it might generally be expected that the δx_α would be overdetermined. However, Eq. (11) is soluble because of the special form of the coefficients g_{i4}, which indicates that the fields that admit such a transformation are closely related to the fields that actually exist in nature.

IV. THE TRANSFORMATION GROUP

To complete the analogy to the Lorentz invariance of the special theory, the infinitesimal transformation of Eqs. (17) must be extended to apply to finite values of the transformation parameters. It is desirable that the finite transformations form a group, since this ensures that physical phenomena predicted in any one Lorentz frame do not depend on the particular sequence of transformations that may have been used to reach that frame, which is surely necessary if the theory is to correspond to reality. It will be assumed that the coordinate differentials are still related by the usual Lorentz transformation in the region far from any masses, which implies that the structure of such a group, if one exists, must be that of the usual Lorentz group. Therefore, if such a group exists, it must be possible to express the infinitesimal generalized Lorentz transformation, given by Eqs. (17), in the form of the infinitesimal Lorentz group, given by Eqs. (16). This can be done by defining a new set of coordinates X_α by the relations

$$X_i = x_i, \quad X_4 = x_4 + \beta/c^2 \tag{18}$$

and by similarly defining a new set of coordinates X'_α in the new Lorentz frame by the relations

$$X'_i = x'_i, \quad X'_4 = x'_4 + \beta'/c^2,$$

where β' is the transform of β and is an arbitrary function. Then, if $\delta X_\alpha \equiv X'_\alpha - X_\alpha$ and $\delta\beta \equiv \beta' - \beta$, Eqs. (17) can be written as

$$\delta X_4 - c^{-2}\delta\beta = -c^{-2}w_i X_i,$$
$$\delta X_i = c_{ij}X_j - w_i X_4.$$

Comparing these with Eqs. (16), we see that the transformation from X_α to X'_α has exactly the form of Eqs. (16) if $\delta\beta$ is defined to be zero, that is, if β' is chosen to equal β, so that *β is an invariant*. Then the coordinates X'_α are given by

$$X'_i = x'_i, \quad X'_4 = x'_4 + \beta/c^2. \tag{19}$$

Since any finite group is determined by its infinitesimal transformation, it is clear that, if the finite transformation relating X'_α to X_α is assumed to form a group, the group must be the usual Lorentz group of the special theory. *The transformation group in which the X'_α of Eqs. (19) are related to the X_α of Eqs. (18) by the usual Lorentz transformation will be taken to be the generalized Lorentz transformation in a Newtonian gravitational field.*

Ignoring a 3-dimensional rotation, we can write the usual Lorentz transformation as

$$X'_i = X_i + \left(\frac{\gamma - 1}{w^2} w_j X_j - \gamma X_4\right) w_i,$$
$$X'_4 = \gamma(X_4 - c^{-2}w_j X_j),$$

where the w_i are the components of the velocity associated with the transformation and are the parameters of the transformation, $w^2 = w_i w_i$, and $\gamma = (1 - w^2/c^2)^{-\frac{1}{2}}$. Writing this in terms of x_α and x'_α by the use of Eqs. (18) and (19) gives the relations

$$x'_i = x_i + \left(\frac{\gamma - 1}{w^2} w_j x_j - \gamma x_4 - \frac{\gamma}{c^2}\beta\right) w_i,$$
$$x'_4 = \gamma\left(x_4 - \frac{1}{c^2} w_j x_j + \frac{\gamma - 1}{\gamma c^2}\beta\right). \tag{20}$$

The most general form of the generalized Lorentz transformation in a Newtonian field is given by Eqs. (20) with the addition of an arbitrary 3-dimensional rotation. The coordinates x'_α given by Eqs. (20) are those of the new Lorentz frame.

From this it follows immediately that the defining equation of the generalized Lorentz transformation, which must hold for all dX_α, is

$$\eta_{\alpha\beta}\, dX'_\alpha\, dX'_\beta = \eta_{\alpha\beta}\, dX_\alpha\, dX_\beta$$

or, from Eqs. (18) and (19),

$$dx'_i\, dx'_i - c^2(dx'_4 + c^{-2}\, d\beta)^2 = dx_i\, dx_i - c^2(dx_4 + c^{-2}\, d\beta)^2.$$

Expanding, cancelling the two terms containing $(d\beta)^2$, and requiring that this relation hold for all dx_α give the relation

$$\sigma'_{\alpha\beta}\frac{\partial x'_\alpha}{\partial x_\gamma}\frac{\partial x'_\beta}{\partial x_\delta} = \sigma_{\gamma\delta},$$

where

$$\sigma_{ij} = \delta_{ij}, \quad \sigma_{i4} = -\frac{\partial\beta}{\partial x_i}, \quad \sigma_{44} = -c^2 - 2\frac{\partial\beta}{\partial x_4},$$
$$\sigma'_{ij} = \delta_{ij}, \quad \sigma'_{i4} = -\frac{\partial\beta}{\partial x'_i}, \quad \sigma'_{44} = -c^2 - 2\frac{\partial\beta}{\partial x'_4}. \tag{21}$$

It is seen that *the defining equation of the generalized Lorentz transformation is the condition that the quantities $\sigma_{\alpha\beta}$ transform as a covariant tensor with two indices.*

Comparing the first three of Eqs. (21) with Eqs. (9) shows that $g_{ij} = \sigma_{ij}$, $g_{i4} = \sigma_{i4}$, and $g_{44} = \sigma_{44} - 2V$, so that the metric coefficients of Eqs. (9) can be written as

$$g_{\alpha\beta} = \sigma_{\alpha\beta} - 2V\frac{\partial x_4}{\partial x_\alpha}\frac{\partial x_4}{\partial x_\beta}. \tag{22}$$

If this equation is multiplied by $(\partial x_\alpha/\partial x'_\gamma)(\partial x_\beta/\partial x'_\delta)$ and it is noted that this process transforms $g_{\alpha\beta}$ into $g'_{\gamma\delta}$ and $(\partial x_4/\partial x_\alpha)(\partial x_4/\partial x_\beta)$ into $(\partial x_4/\partial x'_\gamma)(\partial x_4/\partial x'_\delta)$ when x'_α is any system of coordinates and also transforms $\sigma_{\alpha\beta}$ into $\sigma'_{\gamma\delta}$ when the coordinates x'_α are those of another Lorentz frame, it is seen that the metric coefficients in an arbitrary Lorentz frame can be written as

$$g'_{\gamma\delta} = \sigma'_{\gamma\delta} - 2V \frac{\partial x_4}{\partial x'_\gamma} \frac{\partial x_4}{\partial x'_\delta}. \tag{23}$$

Using the fourth of Eqs. (21), we further see that the 3-dimensional geometry in the new Lorentz frame is described by the metric tensor

$$g'_{ij} = \delta_{ij} - 2V \frac{\partial x_4}{\partial x'_i} \frac{\partial x_4}{\partial x'_j}. \tag{24}$$

The nature of this geometry becomes clear if the terms $\partial x_4/\partial x'_i$ are evaluated by differentiating the reciprocal of the transformation given by Eqs. (20). This reciprocal can be obtained from Eqs. (20) by interchanging x_α and x'_α and replacing w_i by $-w_i$, with the result that

$$\frac{\partial x_4}{\partial x'_i} = \frac{\partial}{\partial x'_i}\left[\gamma\left(x'_4 + \frac{1}{c^2} w_j x'_j + \frac{\gamma - 1}{\gamma c^2}\beta\right)\right]$$
$$= \frac{\gamma}{c^2}\left(w_i + \frac{\gamma - 1}{\gamma}\frac{\partial \beta}{\partial x'_i}\right).$$

For transformation velocities much less than c, $\gamma \approx 1$, and this becomes $\partial x_4/\partial x'_i \approx w_i/c^2$, so that Eq. (24) is approximately

$$g'_{ij} \approx \delta_{ij} - \frac{2V}{c^2}\frac{w_i w_j}{c^2}. \tag{25}$$

It is clear that the new geometry will be *exactly* Euclidean (that is, $g'_{ij} = \delta_{ij}$) only if $w_i = 0$, in which case the transformation reduces to the identity, or if $V = 0$, in which case the gravitational field vanishes. However, it is easily seen that the geometry in the new coordinates is not likely to differ from Euclidean geometry in any measurable way. First, the factor $-2V/c^2$ has the value of about 4×10^{-6} at the surface of the sun and is much less than this throughout most of the solar system. It would approach unity at the Schwarzschild singularity, which is not known to be approached anywhere in nature. Second, if we consider the transformation to the rest frame of an observer who is in orbit about the sun, the maximum value of $|w_i|/c$ is that associated with an observer who moves in a circular orbit very close to the sun, and is about 1.4×10^{-3}. Thus, the maximum achievable value of the last term in Eq. (25) is about 8×10^{-12}, which must be compared with terms of order unity in δ_{ij}. Finally, it must be remembered that nature does not provide us with the coordinates x'_α, so that it is not possible to measure the magnitude of the last term of Eq. (25) directly in these coordinates. Instead, it is necessary to measure the curvature of g'_{ij}, a process which requires the determination of the difference in the values of g'_{ij} at points which are separated by a distance appreciable compared with the size of the solar system. All of these facts considered, it seems unlikely that the last term of Eq. (25) is large enough to be detected by any measurements in the foreseeable future.

V. THE GENERALIZED LORENTZ TRANSFORMATION IN NON-NEWTONIAN FIELDS

It has been shown that, when a Newtonian field is described in the coordinates in which $g_{\alpha\beta}$ takes the form of Eqs. (9), it is always possible to introduce an infinitesimal transformation which preserves the Euclidean nature of the 3-dimensional geometry. The finite transformation group associated with this infinitesimal transformation preserves the Euclidean nature of 3-space only to terms of first order in the parameters w_i, but the non-Euclidicity arising from higher-order terms appears to be too small to be measurable. Applying the generalized Lorentz transformation to Eqs. (9) shows that the form of the metric coefficients in an arbitrary Lorentz frame is given in terms of the three functions β, V, and x_4 by Eq. (23) and the last three of Eqs. (21). The functions β, V, and x_4 are treated as invariants under the generalized Lorentz transformation, but there is one Lorentz frame in which the function x_4 plays the role of the time variable. In this frame the metric coefficients take the particularly simple form of Eqs. (9).

The preferential treatment of this one Lorentz frame can be removed, at least in principle, by considering a slightly more general class of fields, namely, those in which there exists a coordinate system in which the metric coefficients can be expressed in the form

$$g_{\alpha\beta} = \sigma_{\alpha\beta} - 2V \frac{\partial t}{\partial x_\alpha}\frac{\partial t}{\partial x_\beta}, \tag{26}$$

where $\sigma_{\alpha\beta}$ is given by the first three of Eqs. (21) and where V, β, and t are three arbitrary functions of the coordinates, all of which are invariant under the generalized Lorentz transformation. The functional form of the right-hand side of Eq. (26) is the same in all Lorentz frames because α and β transform as covariant indices, and this implies that the term $-2V(\partial t/\partial x_\alpha) \times (\partial t/\partial x_\beta)$ will take the same form in any system of

coordinates and that $\sigma_{\alpha\beta}$ will take the same form in any Lorentz frame. Therefore, in *any* Lorentz frame the metric coefficients of Eq. (26) can be written in the 3-dimensional form

$$\begin{aligned} g_{ij} &= \delta_{ij} - 2V \frac{\partial t}{\partial x_i}\frac{\partial t}{\partial x_j}, \\ g_{i4} &= -\frac{\partial \beta}{\partial x_i} - 2V \frac{\partial t}{\partial x_4}\frac{\partial t}{\partial x_i}, \\ g_{44} &= -c^2 - 2\frac{\partial \beta}{\partial x_4} - 2V\left(\frac{\partial t}{\partial x_4}\right)^2. \end{aligned} \tag{27}$$

If the components of this metric in the coordinates X_α of Eqs. (18) are denoted by $G_{\alpha\beta}$, it is found that the coefficients $G_{\alpha\beta}$ have the convenient 4-dimensional form

$$G_{\alpha\beta} = \eta_{\alpha\beta} + \frac{1}{c^2}\frac{\partial \beta}{\partial X_\alpha}\frac{\partial \beta}{\partial X_\beta} - 2V\frac{\partial t}{\partial X_\alpha}\frac{\partial t}{\partial X_\beta}. \tag{28}$$

If it should happen that there is one particular Lorentz frame x_α in which $x_4 = t$, then, in that frame, $g_{\alpha\beta}$ will be given by Eqs. (9), and the field will be Newtonian. However, if there is no frame in which $x_4 = t$, then no one Lorentz frame is preferred in principle to any other. As a result, the preferred position of one Lorentz frame has been removed by considering a class of fields that is slightly more general than the Newtonian fields.

It is not difficult to consider even more general forms of $g_{\alpha\beta}$ by adding to the right-hand side of Eq. (26) additional terms of the form $\lambda_n(\partial\mu_n/\partial x_\alpha)(\partial\mu_n/\partial x_\beta)$, where λ_n and μ_n are arbitrary invariant functions and n runs from one to as many terms as may be desired. In this way, any metric tensor can be put in a functional form which is invariant under the generalized Lorentz transformation. However, the transformation group has been defined such that it maintains the Euclidean nature of 3-space as closely as possible, and this physical motivation for the transformation will be meaningful only if the field closely approximates a Newtonian field. For this to be the case, the terms involving λ_n and μ_n must be very small, and at present there appears to be no observational evidence which indicates that these terms do not vanish entirely.

VI. CONCLUSIONS

If there is a close fundamental relation between gravitation and electromagnetism, it seems likely that there is an invariance in the gravitational field similar to the known invariance in the electromagnetic field, as given in the special theory of relativity. In the special theory the intrinsic properties of space and time are treated very differently than are the electromagnetic field quantities, which suggests that it is desirable to separate these two types of quantities and to treat them very differently in the gravitational field also. A natural way to do this is suggested by the Newtonian gravitational fields, in which coordinates x_α exist such that $g_{\alpha\beta}$ takes the form of Eqs. (9). The intrinsic properties of space and time are the Newtonian time variable x_4 and the Euclidean 3-dimensional geometry described by the metric coefficients g_{ij}. The remaining metric coefficients $g_{\alpha 4}$ are the quantities that describe the gravitational field.

With this interpretation, the infinitesimal form of the generalized Lorentz transformation is defined such that it keeps the 3-dimensional geometry Euclidean. The resulting infinitesimal transformation is given by Eqs. (17). The finite form of the transformation is determined through the assumption that the transformations form a group and is given by Eqs. (20). The structure of the group is found to be that of the Lorentz group of the special theory. The finite group does not keep the 3-dimensional geometry exactly Euclidean, but it seems unlikely that the non-Euclidicity will be detectable in the foreseeable future.

Of all of the Lorentz frames defined by the generalized Lorentz transformation, there is one in which the metric coefficients take a particularly simple form, namely, the one in which the time variable is x_4 and the metric coefficients have the values given by Eqs. (9). This preferential position of one Lorentz frame can be removed by considering a class of fields slightly more general than the Newtonian fields. These fields are described in a Lorentz frame by Eqs. (27), where β, V, and t are three arbitrary invariant functions. Other fields can be written in a form invariant under the transformation, but they do not appear to be needed to describe any known fields.

All of the results obtained here are independent of Einstein's field equations.

ACKNOWLEDGMENT

The author is very indebted to Dr. K. S. Thorne of the California Institute of Technology for reading the manuscript of this paper and offering several valuable comments concerning it.

* Present address: 1355 Berea Place, Pacific Palisades, California 90272.

[1] R. L. Kirkwood, Phys. Rev. **92**, 1557 (1953); **95**, 1051 (1954).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Time Translations in the Algebraic Formulation of Statistical Mechanics

D. A. DUBIN AND G. L. SEWELL
Department of Physics, Queen Mary College, London E1, England

(Received 26 March 1970)

We present a new description of time translations in the C^*-algebraic formulation of statistical mechanics. This description is based on weaker assumptions than the hitherto accepted ones, due to Haag, Hugenholtz, and Winnink (HHW) [Commun. Math. Phys. **5**, 215 (1967)]. It is shown that these weaker assumptions still lead to the principal results of HHW for Gibbs states and, further, that our assumptions, unlike those of HHW, are valid for the ideal Bose gas and strong-coupling BCS models.

1. INTRODUCTION

In the algebraic formulation of statistical mechanics, time translations of a system have hitherto been described in terms of a 1-parameter group of automorphisms of the appropriate C^*-algebra of quasilocal bounded observables. Such a description has been introduced by Haag, Hugenholtz, and Winnink[1] (HHW) on the basis of assumptions concerning the existence of certain thermodynamical limits, which will be specified below. These assumptions have been shown by Robinson[2] to be valid for a large class of lattice systems. On the other hand, they have not been substantiated for any class of continuous systems and, indeed, as we show in the appendices to this article, they are invalid in the cases of the ideal Bose gas and BCS models. Thus, it is desirable to weaken the assumptions of HHW.

In the present paper we formulate a theory of time translations in statistical mechanics, on the basis of assumptions weaker than those of HHW. It is shown that these weaker assumptions have the dual merit of yielding the principal results of HHW for Gibbs states and of being valid for both the ideal Bose gas and strong-coupling BCS models.

As a preliminary to formulating our theory, we first recall that the algebraic formulation of statistical mechanics has been formulated as follows (cf. HHW). One defines a ν-dimensional Euclidean space Γ to be the physical space of the system under consideration and L to be the set $\{\Lambda\}$ of all bounded, measurable open subsets of Γ. One then constructs a Fock–Hilbert space $\mathcal{H}_F$ corresponding to CCR (for bosons) or CAR (for fermions). For each $\Lambda \in L$, one constructs a closed subspace $\mathcal{H}_{F\Lambda}$ of $\mathcal{H}_F$ and a von Neumann algebra $\mathfrak{A}_\Lambda$, in $\mathcal{H}_{F\Lambda}$, such that $\mathfrak{A}_\Lambda$ and $\mathcal{H}_{F\Lambda}$ are isotonic with respect to Λ, i.e., that $\Lambda \subset \Lambda'$ implies $\mathfrak{A}_\Lambda \subset \mathfrak{A}_{\Lambda'}$ and $\mathcal{H}_{F\Lambda} \subset \mathcal{H}_{F\Lambda'}$. The algebras $\{\mathfrak{A}_\Lambda\}$ are termed local algebras and are constructed so that $\mathfrak{A}_\Lambda$ corresponds to the algebra of bounded observables for the region Λ. One then defines $\mathfrak{A}_L$ to be $\bigcup_{\Lambda\in L} \mathfrak{A}_\Lambda$, and $\mathfrak{A}$ to be the norm closure of $\mathfrak{A}_L$. Thus $\mathfrak{A}$ is a C^*-algebra, possessing a unit element I and is usually termed the algebra of quasilocal bounded observables for the system.

The states of the system are represented by positive normalized linear functionals on $\mathfrak{A}$. Of particular importance in statistical mechanics are the Gibbs states, which are constructed as follows. One assumes that there exists an increasing sequence $M = \{\Lambda_n\}$ of elements of L, with $\bigcup_n \Lambda_n = \Gamma$, such that, for each $\Lambda_n \in M$, there exist self-adjoint operators H_n and N_n in $\mathcal{H}_{F\Lambda_n}$ corresponding to the Hamiltonian and particle number for a system of the specified particles occupying Λ_n, subject to prescribed boundary conditions. It is also assumed that, for real μ less than some fixed μ_c ($> -\infty$), and real, positive β, the operator $(H_n - \mu N_n)$ is self-adjoint and lower bounded, and $\exp[-\beta(H_n - \mu N_n)]$ is of trace class in $\mathcal{H}_{F\Lambda_n}$, for all $\Lambda_n \in M$. Thus one may define a normal state $\phi^{(n)}$ on $\mathfrak{A}_{\Lambda_n}$ by

$$\phi^{(n)}(A) = \frac{\mathrm{Tr}_n\,[A \exp -\beta(H_n - \mu N_n)]}{\mathrm{Tr}_n\,[\exp -\beta(H_n - \mu N_n)]}, \qquad \forall\, A \in \mathfrak{A}_{\Lambda_n}, \quad (1.1)$$

where Tr_n denotes trace over $\mathcal{H}_{F\Lambda_n}$. It follows from this definition, together with the isotony of $\mathfrak{A}_\Lambda$, $\mathcal{H}_{F\Lambda}$ with respect to Λ, that, if $A \in \mathfrak{A}_L$, then $\phi^{(n)}(A)$ is defined for sufficiently large n. It is now assumed that

(I) $\lim_{n\to\infty} \phi^{(n)}(A)$ exists for each $A \in \mathfrak{A}_L$.

Since $\mathfrak{A}_L$ is norm dense in $\mathfrak{A}$, it follows from this assumption that there exists a state ϕ on $\mathfrak{A}$ that is uniquely defined by

$$\phi(A) = \lim_{n\to\infty} \phi^{(n)}(A), \quad \forall\, A \in \mathfrak{A}_L. \qquad (1.2)$$

This state ϕ is usually referred to as the Gibbs state of the system, for chemical potential μ and inverse

temperature β. Since $\mathfrak{A}_\Lambda$ is a von Neumann algebra, ϕ is locally normal.[3]

In order to formulate time translations, one starts by defining $\tau_t^{(n)}: \mathfrak{A}_{\Lambda_n} \to \mathfrak{B}(\mathcal{H}_{F\Lambda_n})$ by

$$\tau_t^{(n)} A = U_t^{(n)} A (U_t^{(n)})^{-1} \equiv A^{(n)}(t), \qquad \forall\, A \in \mathfrak{A}_{\Lambda_n}, \quad t \in R, \tag{1.3}$$

with

$$U_t^{(n)} = \exp\,[i(H_n - \mu N_n)t]. \tag{1.4}$$

It is assumed that $\mathfrak{A}_{\Lambda_n}$, H_n, and N_n are defined so that $\{\tau_t^{(n)}\} \subset \operatorname{Aut} \mathfrak{A}_{\Lambda_n}$.

We now come to the crucial assumption made by HHW in their treatment of time translations. This assumption is that

(II) for each $A \in \mathfrak{A}_L$ and $t \in R$, $\tau_t^{(n)} A$ converge normwise as $n \to \infty$.

A direct consequence of this assumption is that, since $\mathfrak{A}_L$ is norm dense in $\mathfrak{A}$, there exists a 1-parameter group $\{\tau_t\}$ of automorphisms of $\mathfrak{A}$, uniquely defined by

$$\|\tau_t^{(n)} A - \tau_t A\| \to 0 \quad \text{as} \quad n \to \infty, \qquad \forall\, A \in \mathfrak{A}_L, \quad t \in R.$$

The theory of time translations in the present article is based on the replacement of (II) by the following weaker[4] assumptions:

(III) $\lim_{n\to\infty} \phi^{(n)}(A_1^{(n)}(t_1) \cdots A_k^{(n)}(t_k))$ exists for all $A_1, \cdots, A_k \in \mathfrak{A}_L$, $t_1, \cdots, t_k \in R$, $k < \infty$;

(IV)

$$\lim_{m\to\infty} \lim_{n\to\infty} \phi^{(n)}(A_1^{(n)}(t_1) \cdots A_k^{(n)}(t_k) A_1'^{(m)}(t_1') \cdots A_l'^{(m)}(t_l'))$$

exists and is equal to

$$\lim_{n\to\infty} \phi^{(n)}(A_1^{(n)}(t_1) \cdots A_k^{(n)}(t_k) A_1'^{(n)}(t_2) \cdots A_l'^{(n)}(t_l'))$$

for all

$$A_1, \cdots, A_k, A_1', \cdots, A_l' \in \mathfrak{A}_L,$$
$$t_1, \cdots, t_k, t_1', \cdots, t_l' \in R, \quad k, l < \infty.$$

Since (III) trivially implies (I), one should regard (III) and (IV) as the basic assumptions of our theory. The principal results that we derive from these assumptions may be summarized as follows. Let the GNS representation of $\mathfrak{A}$ corresponding to the Gibbs state ϕ be a $*$-homomorphism π of $\mathfrak{A}$ into the bounded operators in a Hilbert space $\mathcal{H}$, with cylindrical vector Ω, such that

$$\phi(A) = (\Omega, \pi(A)\Omega), \quad \forall\, A \in \mathfrak{A}.$$

Then:

(i) Time translations are represented in $\mathcal{H}$ by a 1-parameter group of automorphisms of $\pi(\mathfrak{A})''$, implemented by a unitary group $\{\hat{U}_t\}$ of transformations of transformations of $\mathcal{H}$, for which $\hat{U}_t$ is strongly continuous with respect to t and $\hat{U}_t\Omega = \Omega$, $\forall\, t \in R$; on the other hand, time translations do not necessarily correspond to automorphisms of $\mathfrak{A}$ in our description.

(ii) There is a conjugation operator J, in $\mathcal{H}$, such that

$$J\Omega = \Omega, \quad [J, \hat{U}_t]_- = 0, \quad \forall\, t \in R,$$

and

$$J\pi(\mathfrak{A})''J = \pi(\mathfrak{A})'.$$

(iii) The infinite-volume two-time correlation functions, appropriately defined, satisfy the Kubo–Martin–Schwinger (KMS) boundary conditions.

Thus we recover the main results of HHW concerning Gibbs states, despite weakening the basic assumptions from (I) and (II) to the weaker pair (III) and (IV).

Our theory will be set out as follows. In Sec. 2, we shall present a mathematical formalism that enables us to derive, from postulate (III), a representation π of $\mathfrak{A}$ and a unitary representation of R in a Hilbert space $\mathcal{H}$. It will be seen that this derivation is similar to Wightman's field-theoretical reconstruction theorem.[5] In Sec. 3, we shall show that, in view of postulate (IV), the representation π of $\mathfrak{A}$ in $\mathcal{H}$ corresponds precisely to the GNS representation of $\mathfrak{A}$ for the Gibbs state ϕ; we shall then derive the results (i)–(iii) summarized above. In Sec. 4 we shall make some concluding comments on these results and their possible ramifications. In Appendix A, we shall show that the ideal Bose gas model satisfies postulates (III) and (IV), but violates (II). In Appendix B, we shall show that the BCS model violates (II), but that, at least in the strong coupling case, it satisfies (III) and (IV).

2. MATHEMATICAL CONSTRUCTIONS

We now present our formalism. It will be seen that the contents of the section depend partly on postulate (III), but are independent of (IV).

We refer to the set of real numbers as D or R, according to whether it is equipped with the discrete or usual topology. The set $\mathfrak{A}$, when considered as a topological space, will always have the norm topology. The set of all positive integers and the field of complex numbers will be denoted by the usual symbols Z_+ and C, respectively.

A. The Space S

For each $k \in Z_+$, we define $S^{(k)}$ to be the topological product $\mathfrak{A}^k \times D^k$ and S to be the topological sum $\sum_{k\in Z_+} S^{(k)}$. For each $\Lambda \in L$, we define $S_\Lambda^{(k)}$ to be the

subset $\mathfrak{A}_\Lambda^k \times D^k$ of $S^{(k)}$. We then define

$$S_\Lambda = \bigcup_{k\in Z_+} S_\Lambda^{(k)} \quad \text{and} \quad S_L = \bigcup_{\Lambda\in L} S_\Lambda.$$

Thus, S_L is dense in S.

Elements of S will usually be denoted by σ, sometimes by s. Thus, each σ ($\in S$) will correspond to an ordered set $(A_1, \cdots, A_k; t_1, \cdots, t_k)$, with the $A_j \in \mathfrak{A}$, the $t_j \in D$, and $k < \infty$.

We define the following operations in S:

(i) a binary multiplication, such that if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k)$$

and

$$\sigma' = (A_1', \cdots, A_l'; t_1', \cdots, t_l'), \tag{2.1}$$

then

$$\sigma\sigma' = (A_1, \cdots, A_k, A_1', \cdots, A_l'; t_1, \cdots, t_k, t_1', \cdots, t_l');$$

(ii) an involution ($\sigma \to \sigma^*$), such that if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$\sigma^* = (A_k^*, \cdots, A_1^*; t_k, \cdots, t_1), \tag{2.2}$$

and thus

$$(\sigma^*\sigma')^* = \sigma'^*\sigma, \quad \forall\ \sigma, \sigma' \in S; \tag{2.3}$$

(iii) a set $\{V_t\} \in \text{Aut } S$, with the index set $\{t\} = D$, such that if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$V_t\sigma = (A_1, \cdots, A_k; t_1 + t, \cdots, t_k + t), \tag{2.4}$$

where the sums $(t_j + t)$ are defined in the usual sense of addition of real numbers.

It follows from Eq. (2.4) that, for each $\Lambda \in L$, $V_t \in \text{Aut } S_\Lambda$ and also that $\{V_t\}$ is a 1-parameter group, with

$$V_tV_{t'} = V_{t+t'}, \quad \forall\ t \in D, \quad \text{and } V_0 = I_S, \tag{2.5}$$

where I_S is the unit operator on S. Further, by Eqs. (2.1) and (2.4),

$$V_t(\sigma\sigma') = (V_t\sigma)(V_t\sigma'), \quad \forall\ \sigma, \sigma' \in S,\ t \in D. \tag{2.6}$$

For $\Lambda_n \in M$ (defined in Sec. I), we define $\gamma^{(n)}: S_{\Lambda_n} \to \mathfrak{A}_{\Lambda_n}$ as follows: if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$\gamma^{(n)}(\sigma) = A_1^{(n)}(t_1) \cdots A_k^{(n)}(t_k). \tag{2.7}$$

It follows from this definition and Eqs. (1.1), (2.1), and (2.2) that

$$\gamma^{(n)}(\sigma\sigma') = \gamma^{(n)}(\sigma)\gamma^{(n)}(\sigma'), \quad \forall\ \sigma, \sigma' \in S_{\Lambda_n}, \tag{2.8}$$

$$\gamma^{(n)}(\sigma^*) = (\gamma^{(n)}(\sigma))^*, \quad \forall\ \sigma \in S_{\Lambda_n}, \tag{2.9}$$

and

$$\phi^{(n)}(\gamma^{(n)}(V_t\sigma)) = \phi^{(n)}(\gamma^{(n)}(\sigma)), \quad \forall\ t \in D, \quad \sigma \in S_{\Lambda_n}. \tag{2.10}$$

We define the map $\eta: \mathfrak{A} \to S$ by

$$\eta(A) = (A; 0)\ (\in S^{(1)}), \quad \forall\ A \in \mathfrak{A}. \tag{2.11}$$

It follows from this definition, together with Eq. (2.2), that

$$(\eta(A))^* = \eta(A^*), \quad \forall\ A \in \mathfrak{A}. \tag{2.12}$$

We define $\eta_t: \mathfrak{A} \to S$ by

$$\eta_t(A) = V_t\eta(A), \quad \forall\ A \in \mathfrak{A},\ t \in D. \tag{2.13}$$

Thus, it follows from Eqs. (2.1), (2.4), and (2.13) that

$$(A_1, A_2, \cdots, A_k; t_1, t_2, \cdots, t_k) \equiv \eta_{t_1}(A_1) \cdots \eta_{t_k}(A_k). \tag{2.14}$$

B. The Map $W: S \to C$

It follows from postulate (III) and our definition of the topology of S that we may define a continuous map $W: S_L \to C$ by

$$W(\sigma) = \lim_{n\to\infty} \phi^{(n)}(\gamma^{(n)}(\sigma)), \quad \forall\ \sigma \in S_L. \tag{2.15}$$

Since S_L is dense in S, we may extend W to be a continuous linear function from S to C, uniquely defined by Eq. (2.15).

It follows from Eqs. (1.1), (1.3), (1.4), (2.2), (2.4), and (2.15) that W possesses the following properties:

$$W(V_t\sigma) = W(\sigma), \quad \forall\ t \in D, \quad \sigma \in S, \tag{2.16}$$

$$W(\sigma^*) = \overline{W(\sigma)}, \quad \forall\ \sigma \in S, \tag{2.17}$$

$$W(\sigma^*\eta(AB)\sigma') = W(\sigma^*\eta(A)\eta(B)\sigma'), \quad \forall\ \sigma, \sigma' \in S,\ A, B \in \mathfrak{A}, \tag{2.18}$$

$$\left.\begin{aligned} &W(\sigma\eta_t(I)) = W(\sigma\eta(I)) = W(\eta(I)\sigma\eta(I)) = W(\sigma), \\ &\text{and} \\ &W(\sigma^*\eta(I)\sigma') = W(\sigma^*\sigma'), \end{aligned}\right\} \quad \forall\ \sigma, \sigma' \in S, \quad t \in D. \tag{2.19}$$

Further, if $a_1, \cdots, a_N \in C$ and $\sigma_1, \cdots, \sigma_N \in S_L$, then it follows from Eqs. (2.1), (2.2), and (2.15) and from the positivity of the states $\{\phi^{(n)}\}$ that

$$\sum_{j,k=1}^N \bar{a}_j a_k W(\sigma_j^*\sigma_k) = \lim_{n\to\infty} \phi^{(n)}\left[\left(\sum_{j=1}^N s_j\gamma^{(n)}(\sigma_j)\right)^*\left(\sum_{j=1}^N a_j\gamma^{(n)}(\sigma_j)\right)\right] \geq 0.$$

Since S_L is dense in S, it follows by continuity that

$$\sum_{j=1}^{N} \bar{a}_j a_k W(\sigma_j^* \sigma_k) \geq 0, \quad \forall\, a_1, \cdots, a_N \in C, \quad \sigma_1, \cdots, \sigma_N \in S, \quad N < \infty. \tag{2.20}$$

The restriction of W to $S^{(k)}$ will be denoted by $W^{(k)}$. Thus, if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$W(\sigma) = W^{(k)}(A_1, \cdots, A_k; t_1, \cdots, t_k). \tag{2.21}$$

The functions $W^{(k)}$ will play a role in our theory which parallels that of the Wightman functions in quantum field theory, with our equations (2.16)–(2.20) serving the same purpose as the Wightman axioms. Specifically, we shall use the above properties of W to construct a Hilbert space representation of $(\mathfrak{A}, D)$ by a method similar to that by which Wightman derived his reconstruction theorem.[5] In fact, the essential reason why one cannot directly apply Wightman's theorem here is that it was designed for cases where, for each $k \in Z_+$, there are at most a countable number of Wightman functions (on R^{4k}), whereas, in the present situation, $\mathfrak{A}$ can be non-separable (as in the case of CCR), and thus there could be a nondenumerable set of functions $W^{(k)}(A_1, \cdots, A_k; \cdot)$ on D^k.

C. The Space K

We define K to be the set of functions, from S to C, whose supports are finite point sets in S. Thus, with the usual definition of linearity, K is a linear vector space over C.

Denoting the set of all linear transformations of K by $\mathfrak{L}(K)$, we define $\theta: S \to \mathfrak{L}(K)$ by

$$\operatorname{supp}(\theta(\sigma)f) = \sigma \operatorname{supp} f \equiv \{\sigma\sigma' \mid \sigma' \in \operatorname{supp} f\}, \quad \forall\, \sigma \in S, f \in K,$$

and

$$(\theta(\sigma)f)(\sigma\sigma') = f(\sigma'), \quad \sigma, \sigma \in S, f \in K. \tag{2.22}$$

Thus,

$$\theta(\sigma\sigma') = \theta(\sigma)\theta(\sigma'), \quad \forall\, \sigma, \sigma' \in S. \tag{2.23}$$

For $t \in D$, we define $U_t \in \mathfrak{L}(K)$ by

$$(U_t f)(\sigma) = f(V_{-t}\sigma), \quad \forall\, f \in K, \sigma \in S. \tag{2.24}$$

Thus, by Eqs. (2.5) and (2.24), $\{U_t\}$ is a 1-parameter group, with

$$U_{t_1} U_{t_2} = U_{t_1 t_2}, \quad \forall\, t_1, t_2 \in D, \quad \text{and} \quad U_0 = I_K, \tag{2.25}$$

where I_K is the unit operator in K. Further, it follows from Eqs. (2.22) and (2.24) that

$$U_t \theta(\sigma) U_{-t} = \theta(V_t \sigma), \quad \forall\, \sigma \in S, t \in D. \tag{2.26}$$

We define functions ρ and ρ_t from $\mathfrak{A}$ to $\mathfrak{L}(K)$ by

$$\rho = \theta \circ \eta, \quad \rho_t = \theta \circ \eta_t, \quad \forall\, t \in D. \tag{2.27}$$

Hence, by Eqs. (2.11), (2.13), (2.14), (2.22), (2.26), and (2.27),

$$\rho_t(A) = U_t \rho(A) U_{-t}, \quad \forall\, A \in \mathfrak{A}, t \in D, \tag{2.28}$$

and, if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$\theta(\sigma) = \rho_{t_1}(A_1) \cdots \rho_{t_k}(A_k). \tag{2.29}$$

We equip K with a sesquilinear form

$$\langle f, g \rangle = \sum_{\sigma, \sigma'} \overline{f(\sigma)} g(\sigma) W(\sigma^* \sigma'), \quad \forall\, f, g \in K, \tag{2.30}$$

the summations with respect to σ and σ' being taken over $\operatorname{supp} f$ and $\operatorname{supp} g$, respectively. Thus, defining ω to be the element of K given by

$$\operatorname{supp} \omega = \eta(I) \quad \text{and} \quad \omega(\eta(I)) = 1, \tag{2.31}$$

it follows from Eqs. (2.19), (2.22), (2.27), and (2.30) that

$$\langle \omega, \theta(\sigma)\omega \rangle = W(\sigma) \quad \text{and}$$

$$\langle \theta(\sigma)\omega, \theta(\sigma')\omega \rangle = \langle \theta(\sigma)\omega, \rho(I)\theta(\sigma')\omega \rangle = W(\sigma^* \sigma'), \quad \forall\, \sigma, \sigma' \in S. \tag{2.32}$$

Further, it follows from Eqs. (2.3), (2.12), (2.16)–(2.20), (2.22), (2.30), (2.31) and (2.32) that if f, $g \in K$, $t \in D$, $\sigma \in S$, and A, $B \in \mathfrak{A}$, then

$$\langle f, U_t g \rangle = \langle U_{-t} f, g \rangle, \tag{2.33}$$

$$\langle f, U_t \omega \rangle = \langle f, \omega \rangle, \tag{2.34}$$

$$\langle g, f \rangle = \sum_s \langle g, \theta(s)\omega \rangle f(s) = \langle g, \rho(I) f \rangle, \tag{2.35}$$

$$\langle f, \theta(\sigma) g \rangle = \langle \theta(\sigma^*) f, g \rangle, \tag{2.36}$$

$$\langle f, \rho(AB) g \rangle = \langle f, \rho(A)\rho(B) g \rangle, \tag{2.37}$$

$$\langle f, g \rangle = \overline{\langle g, f \rangle}, \tag{2.38}$$

and

$$\langle f, f \rangle \geq 0. \tag{2.39}$$

In view of the linearity of K, Eqs. (2.38) and (2.39) imply the Schwartz inequality

$$\langle f, f \rangle \geq |\langle f, g \rangle|^2. \tag{2.40}$$

D. The Subspace K_0

Let $K_0 = \{h \mid h \in K; \langle h, h \rangle = 0\}$. Then it follows from Eq. (2.40) that, if $h \in K_0$, then $\langle g, h \rangle = 0$, $\forall\, g \in K$. Conversely, if this latter condition holds, then $\langle h, h \rangle = 0$, i.e., $h \in K_0$. It follows that K_0 is the linear subspace of K given by $\{h \mid h \in K; \langle g, h \rangle = 0, \forall\, g \in K\}$.

Using this latter specification of K_0 and defining the binary relation $\sim$ in K by $(f \sim f') \equiv (f - f' \in K_0)$, it follows that $\sim$ is an equivalence relationship. Further, it follows from Eqs. (2.30), (2.33)–(2.36), and (2.38) and our definition of $\sim$ that

$$\text{if } f \sim f' \text{ and } g \sim g', \quad \text{then } \langle f', g' \rangle = \langle f, g \rangle, \tag{2.41}$$

$$\text{if } f \sim f' \text{ and } \sigma \in S, \quad \text{then } \theta(\sigma) f' \sim \theta(\sigma) f, \tag{2.42}$$

$$\text{if } f \sim f' \text{ and } t \in D, \quad \text{then } U_t f' \sim U_t f, \tag{2.43}$$

$$U_t \omega \sim \omega, \quad \forall\, t \in D, \tag{2.44}$$

$$f \sim \rho(I) f \sim \sum_s f(s)\theta(s)\omega, \quad \forall\, f \in K. \tag{2.45}$$

E. The Hilbert Space $\mathcal{H}$

Let K/K_0 be the quotient space corresponding to the equivalence relation $\sim$. We denote by $\hat{f}$ the element of K/K_0 corresponding to the equivalence class of f in K. It follows from Eq. (2.43) that we may unambiguously define a sesquilinear form on K/K_0 by

$$(\hat{f}, \hat{g}) = \langle f, g \rangle, \quad \forall f, g \in K. \tag{2.46}$$

It follows from this definition, together with Eqs. (2.38) and (2.39), that K/K_0 is a pre-Hilbert space, with inner product $(\hat{f}, \hat{g})$. The completion of this space will be denoted by $\mathcal{H}$.

For $t \in D$, we define $\hat{U}_t : \mathcal{H} \to \mathcal{H}$ by

$$\hat{U}_t \hat{f} = \widehat{U_t f}, \quad \forall f \in K, \tag{2.47}$$

this definition being unambiguous in view of Eq. (2.43). Since $\mathcal{H}$ is the completion of K/K_0, it follows from Eqs. (2.25), (2.33), (2.46), and (2.47) that

$$\hat{U}_{t_1} \hat{U}_{t_2} = \hat{U}_{t_1 + t_2}, \quad \forall\, t_1, t_2 \in D,$$
$$\hat{U}_0 = \hat{I}, \tag{2.48}$$

where $\hat{I}$ is the unit operator in $\mathcal{H}$; and

$$(\psi_1, \hat{U}_t \psi_2) = (\hat{U}_{-t} \psi_1, \psi_2), \quad \forall\, t \in D, \psi_1, \psi_2 \in \mathcal{H}. \tag{2.49}$$

Thus, by Eqs. (2.48) and (2.49), $\{\hat{U}_t\}$ is a unitary representation of D in $\mathcal{H}$. Further, it follows from Eqs. (2.34) and (2.47) that

$$\hat{U}_t \hat{\omega} = \hat{\omega}, \quad \forall\, t \in D. \tag{2.50}$$

We define $\hat{\theta} : S \to \mathfrak{B}(\mathcal{H})$ by

$$\hat{\theta}(\sigma) \hat{f} = \widehat{\theta(\sigma) f}, \quad \forall\, \sigma \in S, f \in K, \tag{2.51}$$

this definition being unambiguous in view of Eq. (2.42). It follows from this definition, together with Eqs. (2.32), (2.36), and (2.45), that

$$W(\sigma) = (\hat{\omega}, \hat{\theta}(\sigma)\hat{\omega}), \quad \forall\, \sigma \in S, \tag{2.52}$$

$$W(\sigma^* \sigma') = (\hat{\theta}(\sigma)\hat{\omega}, \hat{\theta}(\sigma')\hat{\omega}), \quad \forall\, \sigma, \sigma' \in S, \tag{2.53}$$

$$(\hat{\theta}(\sigma))^* = \hat{\theta}(\sigma^*), \quad \forall\, \sigma \in S, \tag{2.54}$$

and

$$\hat{f} = \sum_s f(s)\hat{\theta}(s)\hat{\omega}, \quad \forall\, f \in K. \tag{2.55}$$

We define maps $\hat{\rho}$ and $\hat{\rho}_t$, from $\mathfrak{A}$ to $\mathfrak{B}(\mathcal{H})$, by

$$\hat{\rho} = \hat{\theta} \circ \eta, \quad \hat{\rho}_t = \hat{\theta} \circ \eta_t. \tag{2.56}$$

It follows from this definition, together with Eqs. (2.12), (2.18), (2.28), (2.29), (2.45), (2.47), (2.48), and (2.51)–(2.55), that

$$\hat{\rho}(A^*) = (\hat{\rho}(A))^*, \quad \forall\, A \in \mathfrak{A}, \tag{2.57}$$

$$\hat{\rho}(AB) = \hat{\rho}(A)\hat{\rho}(B), \quad \forall\, A, B \in \mathfrak{A}, \tag{2.58}$$

$$\hat{\rho}_t(A) = \hat{U}_t \hat{\rho}(A) \hat{U}_t^{-1}, \quad \forall\, A \in \mathfrak{A}, t \in D, \tag{2.59}$$

$$\hat{\rho}(I) = \hat{I}, \tag{2.60}$$

where $\hat{I}$ is the unit operator in $\mathcal{H}$; and if

$$\sigma = (A_1, \cdots, A_k; t_1, \cdots, t_k),$$

then

$$\hat{\theta}(\sigma) = \hat{\rho}_{t_1}(A_1) \cdots \hat{\rho}_{t_k}(A_k). \tag{2.61}$$

3. THE REPRESENTATION

We shall now enunciate our results in the form of six propositions, whose proofs will be given below. It will be seen that these propositions embody the results (i)—(iii), stated in Sec. 1, and that the substance of the first proposition has been established in Sec. 2.

Proposition 1: There exists a $*$-representation $\hat{\rho}$ of $\mathfrak{A}$ and a unitary representation $\{\hat{U}_t\}$ of R in a Hilbert space $\mathcal{H}$, possessing a vector $\hat{\omega}$, such that:

(i) $\hat{\rho}(I) = \hat{I}$, the unit operator in $\mathcal{H}$;

(ii) $\hat{U}_t \hat{\omega} = \hat{\omega}$, $\forall\, t \in R$;

(iii) $W^{(k)}(A_1, \cdots, A_k; t_1, \cdots, t_k) = (\hat{\omega}, \hat{\rho}_{t_1}(A_1) \cdots \hat{\rho}_{t_k}(A_k)\hat{\omega})$,

with

$$\hat{\rho}_t(A) \equiv \hat{U}_t \hat{\rho}(A) \hat{U} \forall, \quad \forall\, A_1, \cdots, A_k \in \mathfrak{A}, \quad t_1, \cdots, t_k \in R;$$

(iv) the set

$$\left\{ \sum_{j=1}^{N} a_j \hat{\theta}(\sigma_j)\hat{\omega} \;\middle|\; \{a_j\} \in C, \{\sigma_j\} \in S, N < \infty \right\}$$

is dense in $\mathcal{H}$; equivalently, the space $\mathcal{H}$ may be generated by application to $\hat{\omega}$ of linear combinations of all products $\hat{\rho}_{t_1}(A_1) \cdots \hat{\rho}_{t_k}(A_k)$, with $A_1, \cdots, A_k \in \mathfrak{A}$, $t_1, \cdots, t_k \in R$, and $k \in Z_+$.

Note: This proposition is not concerned with any topological properties of the real numbers $\{t\}$ and thus does not involve any distinction between D and R. In fact, it is not established until Proposition 5 that $\{\hat{U}_t\}$ is a continuous representation of R in $\mathcal{H}$.

Proposition 2: $\hat{\rho}(\mathfrak{A}_L)\hat{\omega}$ [and thus $\hat{\rho}(\mathfrak{A})\hat{\omega}$] is strongly dense in $\mathcal{H}$. Hence, $\hat{\omega}$ is a cyclical vector with respect to $\hat{\rho}(\mathfrak{A})$, and thus $\hat{\rho}$ is the GNS representation of $\mathfrak{A}$ for the state ϕ.

Proposition 3: If, for $t \in R$, $\hat{\tau}_t$ is the automorphism of $\mathfrak{B}(\mathcal{H})$ defined by $\hat{\tau}_t Q = \hat{U}_t Q \hat{U}_t^{-1}$, $\forall\, Q \in \mathfrak{B}(\mathcal{H})$, then $\{\hat{\tau}_t\}$ is a 1-parameter group of automorphisms of $\hat{\rho}(\mathfrak{A})''$.

Proposition 4: The functions $W^{(2)}(A, B; t, 0)$ and $W^{(2)}(B, A; 0, t)$ satisfy the KMS boundary conditions. Thus, if $A, B \in \mathfrak{A}$, then $\exists$ functions f and g on C such that:

(i) $f(z)$ and $g(z)$ are respectively analytic in the strips $\operatorname{Im} z \in (-\beta, 0)$ and $\operatorname{Im} z \in (0, \beta)$ and are continuous on their boundaries;

(ii) $f(t) = W^{(2)}(A, B; t, 0)$ and

$$g(t) = W^{(2)}(A, B; 0, t), \quad \forall\, t \in R;$$

(iii) $f(z + i\beta) = g(z)$.

Proposition 5: $\{\hat{U}_t\}$ is a strongly continuous representation of R.

Proposition 6: There exists a conjugation operator J, in $\mathcal{H}$, such that

$$J^2 = \hat{I}, \quad J\hat{\omega} = \hat{\omega}, \quad [J, \hat{U}_t]_- = 0, \quad \forall\, t \in R,$$

and

$$J\hat{\rho}(\mathfrak{A})''J = \hat{\rho}(\mathfrak{A})'.$$

Proof of Proposition 1: It follows from Eqs. (2.48), (2.49), (2.58), and (2.59) that $\hat{\rho}$ is a $*$-representation of $\mathfrak{A}$ and $\{\hat{U}_t\}$ a unitary representation of R, in $\mathcal{H}$. Further, (i) and (ii) are established in Eqs. (2.60) and (2.50), respectively; (iii) follows from Eqs. (2.21), (2.52), and (2.61); (iv) follows from Eq. (2.55) and the definitions of K and $\mathcal{H}$ $(= \overline{K/K_0})$. QED

Proof of Proposition 2: Since S_L is dense in S and W is a continuous function on S, it follows from Eq. (2.53) that if $\sigma \in S$, then $\exists$ a sequence $\{\sigma_n\} \in S_L$ such that $\hat{\theta}(\sigma_n)\hat{\omega}$ tends strongly to $\hat{\theta}(\sigma)\hat{\omega}$ as $n \to \infty$. Hence, it follows from Proposition 1 (iv) that if Δ is defined as the set

$$\left\{ \sum_{j=1}^{N} a_j \hat{\theta}(\sigma_j)\hat{\omega} \mid \{\sigma_j\} \in S,\ \{a_j\} \in C,\ N < \infty \right\},$$

then Δ is dense in $\mathcal{H}$. Thus, in order to prove Proposition 2, it suffices for us to establish that if $\sigma \in S_L$, then $\exists$ a sequence $\{B_n\} \in \mathfrak{A}$, such that $\hat{\rho}(B_n)\hat{\omega}$ converges strongly to $\hat{\theta}(\sigma)\hat{\omega}$ as $n \to \infty$, i.e., that

$$\|\hat{\rho}(B_n)\hat{\omega}\| \to \|\hat{\theta}(\sigma)\hat{\omega}\| \quad \text{as} \quad n \to \infty \tag{3.1}$$

and

$$(\mathrm{w}, \mathcal{H})\text{-}\lim_{n\to\infty} \hat{\rho}(B_n)\hat{\omega} = \hat{\theta}(\sigma)\hat{\omega}.$$

In view of the definition of the dense set Δ, this last equation may be rewritten as

$$(\hat{\theta}(\sigma')\hat{\omega},\ \hat{\rho}(B_n)\hat{\omega}) \to (\hat{\theta}(\sigma')\hat{\omega},\ \hat{\theta}(\sigma)\hat{\omega}),$$
$$\text{as} \quad n \to \infty,\ \forall\, \sigma \in S_L. \tag{3.2}$$

In order to show that $\exists$ $\{B_n\}$ which satisfies (3.1) and (3.2), we note that, in view of Eqs. (2.7), (2.11), and (2.15), postulate (IV) may be expressed in the form

$$W(s'^*\eta(\gamma^{(n)}(s))) \to W(s'^*s) \quad \text{as} \quad n \to \infty, \forall\, s, s' \in S_L,$$

i.e., by Eqs. (2.53) and (2.56),

$$(\hat{\theta}(s')\hat{\omega},\ \hat{\rho}(\gamma^{(n)}(s))\hat{\omega}) \to (\hat{\theta}(s')\hat{\omega},\ \hat{\theta}(s)\hat{\omega})$$
$$\text{as} \quad n \to \infty,\ \forall\, s, s' \in S_L. \tag{3.3}$$

On putting $s' = \eta(I)$ and $s = \sigma^*\sigma$ in Eq. (3.3) and using Eqs. (2.8), (2.56)–(2.58), and (2.60), we obtain the formula

$$\|\hat{\rho}(\gamma^{(n)}(\sigma))\hat{\omega}\| \to \|\hat{\theta}(\sigma)\hat{\omega}\| \quad \text{as} \quad n \to \infty,\ \forall\, \sigma \in S_L. \tag{3.4}$$

Further, on putting $s' = \sigma'$ and $s = \sigma$ in Eq. (3.3), we obtain

$$(\hat{\theta}(\sigma')\hat{\omega},\ \hat{\rho}(\gamma^{(n)}(\sigma))\hat{\omega}) \to (\hat{\theta}(\sigma')\omega,\ \hat{\theta}(\sigma)\hat{\omega})$$
$$\text{as} \quad n \to \infty,\ \forall\, \sigma, \sigma' \in S_L. \tag{3.5}$$

Equations (3.4) and (3.5) yield the required formulas (3.1) and (3.2) with B_n chosen as $\gamma^{(n)}(\sigma)$. QED

Proof of Proposition 3: The group property of $\{\hat{\tau}_t\}$ follows immediately from the definition of this set (in the statement of Proposition 3), together with Eq. (2.48). It remains for us, therefore, to show that if $t \in R$, then $\hat{\tau}_t\hat{\rho}(\mathfrak{A})'' = \hat{\rho}(\mathfrak{A})''$. For this purpose we note that, by postulate (IV) and Eqs. (2.1), (2.15), and (2.21),

$$W^{(3)}(B^*, A^{(n)}(t), C; 0, 0, 0) \to W^{(3)}(B^*, A, C; 0, t, 0)$$
$$\text{as} \quad n \to \infty,\ \forall\, A, B, C \in \mathfrak{A}_L,\ t \in R.$$

Hence, by Eqs. (2.52), (2.53), and (2.56),

$$(\hat{\rho}(B)\hat{\omega},\ \hat{\rho}(A^{(n)}(t))\hat{\rho}(C)\hat{\omega}) \to (\hat{\rho}(B)\hat{\omega},\ \hat{\rho}_t(A)\hat{\rho}(C)\hat{\omega})$$
$$\text{as} \quad n \to \infty,\ \forall\, A, B, C \in \mathfrak{A}_L,\ t \in R.$$

Thus, since $\hat{\rho}(\mathfrak{A}_L)\hat{\omega}$ is dense in $\mathfrak{K}$ (by Proposition 2), it follows that

$$(\mathrm{w}, \mathfrak{K}) - \lim_{n\to\infty} \hat{\rho}(A^{(n)}(t)) = \hat{\rho}_t(A), \quad \forall\, A \in \mathfrak{A},\ t \in R.$$

This implies that $\hat{\tau}_t\hat{\rho}(\mathfrak{A}_L) \subset \hat{\rho}(\mathfrak{A})''$, since the bicommutant of $\hat{\rho}(\mathfrak{A})$ is also its weak closure. Hence, since $\mathfrak{A}_L$ is norm dense in $\mathfrak{A}$, it follows from the definition of $\hat{\tau}_t$ that $\hat{\tau}_t\hat{\rho}(\mathfrak{A}) \subset \hat{\rho}(\mathfrak{A})''$; consequently,

$$\hat{\tau}_t(\hat{\rho}(\mathfrak{A})'') \equiv (\hat{\tau}_t\hat{\rho}(\mathfrak{A}))'' \subset \hat{\rho}(\mathfrak{A})''.$$

Likewise $\hat{\tau}_{-t}(\hat{\rho}(\mathfrak{A})'') \subset \hat{\rho}(\mathfrak{A})''$, from which it follows that $\hat{\rho}(\mathfrak{A})'' \subset \hat{\tau}_t(\hat{\rho}(\mathfrak{A})'')$. Thus we have shown that $\hat{\rho}(\mathfrak{A})'' \subset \hat{\tau}_t(\hat{\rho}(\mathfrak{A})'') \subset \hat{\rho}(\mathfrak{A})''$, from which it follows that $\hat{\tau}_t(\hat{\rho}(\mathfrak{A})'') = \rho(\mathfrak{A})''$. QED

Proof of Proposition 4: By Eqs. (2.15) and (2.21),

$$W^{(2)}(A, B; t, 0) = \lim_{n\to\infty} \phi^{(n)}(A^{(n)}(t)B) \qquad (3.6)$$

and

$$W^{(2)}(B, A; 0, t) = \lim_{n\to\infty} \phi^{(n)}(BA^{(n)}(t)). \qquad (3.7)$$

Using Proposition 5 (proved below) and Eqs. (1.1), (1.3), and (1.4), one may readily use the method of HHW to derive the KMS conditions for $W^{(2)}(A, B; t, 0)$ and $W^{(2)}(B, A; 0, t)$. QED

Proof of Proposition 5: By Eqs. (1.1), (1.3), (1.4), and (3.7), $W^{(2)}(B, A; 0, t)$ is the pointwise limit of a sequence of continuous functions of t, and is therefore a measurable function of t. Hence, as

$$W^{(2)}(B, A; 0, t) \equiv (\hat{\rho}(B^*)\hat{\omega}, \hat{U}_t\hat{\rho}(A)\hat{\omega}),$$

and as $\rho(\mathfrak{A})\hat{\omega}$ is dense in $\hat{\mathfrak{K}}$, it follows that $\hat{U}_t$ is weakly measurable. Further, as noted in Sec. 1 (following the definition of ϕ), ϕ is locally normal. This implies[6] that $\hat{\mathfrak{K}}$ is separable and hence[7] that $\hat{U}_t$ is strongly continuous. QED

Proof of Proposition 6: $\hat{\rho}(\mathfrak{A})''$ is a W^*-algebra, and therefore a C^*-algebra. We define a state Φ on this algebra by

$$\Phi(Q) = (\hat{\omega}, Q\hat{\omega}), \quad \forall\, Q \in \hat{\rho}(\mathfrak{A})''.$$

Since the algebra $\hat{\rho}(\mathfrak{A})$ contains I [by Proposition 1(i)], it follows that this algebra is strongly dense in $\hat{\rho}(\mathfrak{A})''$ (cf. Dixmier[8]). Hence, it follows from Proposition 4 and our definition of Φ that Φ is a KMS state on $\hat{\rho}(\mathfrak{A})''$, i.e., that $\Phi((\hat{\tau}_t Q)Q')$, $\Phi(Q'\hat{\tau}_t Q)$ satisfy the KMS conditions for all $Q, Q' \in \hat{\rho}(\mathfrak{A})''$.

Hence the analysis of HHW may be directly applied to the KMS state Φ on $\hat{\rho}(\mathfrak{A})''$ to yield the required result. QED

4. CONCLUSION

We have established Propositions 1–6 on the basis of the postulates (III) and (IV). Thus we have shown that the principal results of HHW, which those authors derived from postulates (I) and (II), are also consequences of the weaker postulates (III) and (IV). One may similarly show that Hugenholtz's theorem[9] of factor types from postulates (III) and (IV), together with the assumption that $\hat{\omega}$ is the only vector in $\mathfrak{K}$ that is invariant under $\{\hat{U}_t\}$ and that there is no nonzero element of $\hat{\rho}(\mathfrak{A})''$ that annihilates $\hat{\omega}$.

The advantages, from a physical standpoint, of basing the theory on (III) and (IV), rather than (I) and (II), were discussed in Sec. 1. Clearly, an essential outstanding problem is that of obtaining conditions on the interparticle forces for which an assembly of particles of a given species fulfills (III) and (IV).

Leaving aside this very difficult problem, one may extend our formalism so as to include local unbounded observables by methods already developed by one of us.[10] Alternatively, one might seek to extend the methods of the present article so as to formulate the properties of such observables in terms of Wightman functions, defined as appropriate thermodynamical limits (assuming that these exist) of spatio–temporal correlation functions between the field operators describing the system in second quantization. The construction of such Wightman functions would then be based on the Hamiltonian formulations of specified quantum-mechanical systems of particles enclosed in finite volumes. Thus the theory would be based on the well-defined nonrelativistic quantal laws for finite systems. Consequently, it would have a definite mechanistic basis, and in this respect it would be different from the present form of relativistic field theory, where the assumed properties of the Wightman functions are not derived from any mechanistic model.

ACKNOWLEDGMENT

The authors would like to thank M. Sirugue, R. Streater, and especially M. Winnink for some constructively critical comments on a first draft of this article.

APPENDIX A

We shall now show that the ideal Bose gas model satisfies conditions (III) and (IV), but violates (II). In deriving the properties of this model, we shall make extensive use of the treatment of Araki and Woods[8] (AW).

We use the same notation as in Sec. 1 of the present article for the sets Γ, L, M, $\mathfrak{A}_\Lambda$, and $\mathfrak{A}$, with the additional specification that each Λ_n ($\in M$) is a parallelepiped, of hypervolume V_n, with edges parallel to the chosen Cartesian axis for Γ. We denote by $\mathfrak{L}$ (resp. $\mathfrak{L}_n$) the Hilbert space of square-integrable functions on Γ (resp. Λ_n). The subset of $\mathfrak{L}$ with compact supports in

Γ will be denoted by $\mathfrak{L}_L$. The symbols $\mathfrak{D}$ and $\mathcal{S}$ will be used in the conventional sense to denote the Schwartz spaces of functions on Γ. The Fourier map $\mathcal{F}:\mathfrak{L}\to\mathfrak{L}$ will be defined by

$$(\mathcal{F}f)(k)\equiv\tilde{f}(k)\equiv\int_\Gamma dx f(x)e^{-ik\cdot x},\quad \forall f\in\mathfrak{L}. \tag{A1}$$

Finally, the space $\mathcal{F}(\mathfrak{D})$ will be denoted, in a usual way, by $\mathfrak{Z}$.

We shall be concerned with free particles whose masses, in the chosen units, are $\frac{1}{2}$. The Hamiltonian operator, in $\mathfrak{L}$, for a single such particle will be denoted by h, and the corresponding evolution operator, $\exp(iht)$, by u_t. Thus

$$(u_tf)(x)\equiv[(2\pi)^\nu]^{-1}\int_\Gamma dk\tilde{f}(k)\exp[i(k\cdot x+k^2t)],\quad \forall f\in\mathfrak{L}, \tag{A2}$$

where ν is the dimensionality of Γ. It follows from this equation that

$$(\mathcal{F}u_tf)(k)=e_t(k)(\mathcal{F}f)(k) \tag{A3}$$

with

$$e_t(k)=\exp ik^2t. \tag{A4}$$

It may readily be verified[12] from this last equation that if $t\neq 0$, then $e_t(k)$ is not a multiplier in $\mathfrak{Z}$ in which case it follows from Eq. (A3) that if $f\in\mathfrak{D}\setminus 0$, i.e., if $\mathcal{F}f\in\mathfrak{Z}\setminus 0$, then $u_tf\notin\mathfrak{D}$. Thus, by Eq. (A2),

$$u_t(\mathfrak{D}\setminus 0)\subset\mathcal{S}\setminus\mathfrak{D},\quad \forall t\neq 0. \tag{A5}$$

We denote by h_n the Hamiltonian operator, in $\mathfrak{L}_n$, for a single free particle of the same species, confined to Λ_n, with periodic boundary conditions, and we define $u_t^{(n)}=\exp(ih_nt)$. Thus, defining K_n as the set $\{k\}$ for which $e^{ik\cdot x}$ satisfies the periodicity conditions for Λ_n and denoting the characteristic function on Λ_n by χ_n, we have

$$(u_t^{(n)}f)(x)=\chi_n(x)\sum_{k\in K_n}\tilde{f}(k)\exp[i(k\cdot x+k^2t)],\quad \forall f\in\mathfrak{L}_n. \tag{A6}$$

It follows easily from Eqs. (A2) and (A6) that, since $\mathcal{F}(\mathfrak{D})\subset\mathcal{S}$, then

$$(\mathrm{s},\mathfrak{L})-\lim_{n\to\infty}u_t^{(n)}f=u_tf,\quad \forall f\in\mathfrak{D}.$$

Hence, since $\mathfrak{D}$ is a dense subset of $\mathfrak{L}_L$, in the strong $\mathfrak{L}$-topology, it follows that

$$(\mathrm{s},\mathfrak{L})-\lim_{n\to\infty}u_t^{(n)}f\to u_tf,\quad \forall f\in\mathfrak{L}_L. \tag{A7}$$

In order to formulate the ideal Bose gas model, we construct a CCR representation of $\mathfrak{L}$ in a Fock space $\mathcal{H}_F$. Thus, in a usual way, we construct $\mathcal{H}_F$ so that:

(i) For each $f\in\mathfrak{L}$, $\exists$ operators $\psi(f)$ and $\psi^*(f)=(\psi(\bar{f}))^*$ in $\mathcal{H}_F$ such that

$$[\psi(f),\psi(g)]_-=0,\quad [\psi(f),\psi^*(g)]_-=(g,f)_{\mathfrak{L}},\quad \forall f,g\in\mathfrak{L}; \tag{A8}$$

(ii) $\mathcal{H}_F$ contains a vector Ω_F such that

$$\psi(f)\Omega_F=0,\quad \forall f\in\mathfrak{L};$$

(iii) if $\mathcal{A}$ is the algebra of all polynomials in $\{\psi^*(f)\,|\,f\in\mathfrak{L}\}$, then $\mathcal{A}\Omega_F$ is dense in $\mathcal{H}_F$.

It is useful to define self-adjoint operators $F(f)$ and $G(f)$ in $\mathcal{H}_F$ by the formulas

$$F(f)=\psi(f)+\psi^*(\bar{f}),\quad G(f)=i[\psi(f)-\psi^*(\bar{f})],\quad \forall f\in\mathfrak{L}. \tag{A9}$$

The algebra[13] $\mathfrak{A}$ is then constructed in terms of these operators by defining $\mathfrak{A}_\Lambda$ to be the W^*-algebra generated by

$$\{\exp[iF(f)],\exp[iG(f)]\,|\,f\in\mathfrak{L},\ \mathrm{supp}\,f\in\Lambda\}$$

and then defining $\mathfrak{A}$ to be the norm closure of

$$\mathfrak{A}_L\equiv\bigcup_{\Lambda\in L}\mathfrak{A}_\Lambda.$$

Let H_n and N_n be the operators in $\mathcal{H}_F$ corresponding to the Hamiltonian and particle number for an ideal Bose gas, confined to Λ_n, subject to the prescribed boundary conditions. Then, in a usual way, the evolution operator $U_t^{(n)}$, generated by $(H_n-\mu N_n)$, may be defined in terms of the single-particle operator $u_t^{(n)}$ by the formulas

$$U_t^{(n)}\Omega_F=\Omega_F,\quad U_t^{(n)}\psi(f)(U_t^{(n)})^{-1}=\psi(e^{-i\mu t}u_t^{(n)}f),\quad \forall t\in R, f\in\mathfrak{L}_n. \tag{A10}$$

Hence, by Eqs. (1.3), (A9), and (A10),

$$\tau_t^{(n)}\exp[iF(f)]=\exp[iF(e^{-i\mu t}u_t^{(n)}f)],\quad \forall t\in R, f\in\mathfrak{L}_n, \tag{A11}$$

and

$$\tau_t^{(n)}\exp[iG(f)]=\exp[iG(e^{-i\mu t}u_t^{(n)}f)];$$

thus, by Eqs. (A7) and (A11),

$$(\mathrm{s},\mathcal{H}_F)\lim_{n\to\infty}\tau_t^{(n)}\exp[iF(f)]=\exp[iF(e^{-i\mu t}u_tf)],\quad \forall t\in R, f\in\mathfrak{L}_L, \tag{A12}$$

and

$$\lim_{n\to\infty}\tau_t^{(n)}\exp[iG(f)]=\exp[iG(e^{-i\mu t}u_tf)].$$

On applying the techniques of the Appendix of AW to the model and noting that $\mathfrak{D}$ is a dense subset of $\mathfrak{L}_L$, in the strong $\mathfrak{L}$-topology, it follows readily from Eqs. (A2), (A6), and (A11) that conditions (III) and (IV) are satisfied when restricted to A_j and A_k' in the

set $\{\exp[iF(f)], \exp[iG(f)] \mid f \in \mathfrak{L}\}$. Hence, it follows from the definition of $\mathfrak{A}_L$ that (III) and (IV) are satisfied absolutely.

In order to show that the model violates condition (II), it suffices to show that the following statement is invalid:

$$\tau_t^{(n)} \exp[iF(f)] \text{ is norm convergent as } n \to \infty, \quad \forall f \in \mathfrak{D} \text{ and fixed } t \in R. \quad \text{(A13)}$$

Thus, we note that, in view of Eq. (A12), the statement (A13) implies that, for all $f \in \mathfrak{D}$,

$$\|\tau_t^{(n)} \exp[iF(f)] - \exp[iF(e^{-i\mu t}u_t f)]\| \to 0 \quad \text{as} \quad n \to \infty,$$

i.e., by Eqs. (A8) and (A9),

$$\|\exp[iF(e^{-i\mu t}(u_t f - u_t^{(n)} f))] - I\| \to 0 \quad \text{as} \quad n \to \infty, \quad \forall f \in \mathfrak{D}. \quad \text{(A14)}$$

Further, as is well known, it follows from the definition of F that, if g is a nonnull element of $\mathfrak{L}$, then $F(g)$ is an unbounded, self-adjoint operator with continuous spectrum; thus,

$$\|\exp[iF(g)] - I\| = 2 \quad \text{unless} \quad \|g\|_{\mathfrak{L}} = 0.$$

Hence, Eq. (A14) implies that if $f \in \mathfrak{D}$, then $\exists\, N \in Z_+$, such that $\|u_t^{(n)} f - u_t f\| = 0$, $\forall\ n > N$. This result contradicts Eqs. (A6) and (A7), since they imply that $u_t^{(n)} f \in \mathfrak{L}_L$ and $u_t f \in \mathcal{S} \setminus \mathfrak{D}$ if $f \in \mathfrak{D} \setminus 0$ and $t \neq 0$. This contradiction suffices to establish that the model violates (II).

APPENDIX B

In this appendix we shall show that the BCS model violates condition (II) and that, in the strong coupling case at least, it satisfies (III) and (IV). Here we shall make extensive use of results obtained by Thirring and Wehrl (TW),[14] and by Thirring.[15]

In order to show that the BCS model violates (II), we note that the article of TW contains results which may be summarized as follows. There exists a set of representations, which may be designated by an index set $S = \{\alpha\}$, of the C^*-algebra $\mathfrak{A}$, such that:

(i) For each $\alpha \in S$, $\exists$ a unique faithful representation π_α of $\mathfrak{A}$ in a Hilbert space $\mathcal{H}_\alpha$;

(ii) for each $\alpha \in S$, $\exists$ a 1-parameter group $\{\tau_t^\alpha\}$ of automorphisms of $\mathfrak{A}$, such that τ_t^α varies with α (the group $\{\tau_t^\alpha\}$ is implemented in $\mathcal{H}_F$ by a unitary group, whose generator is the α-dependent Bogoliubov–Haag Hamiltonian);

(iii) Constructing $\{\Lambda_n\}$ and $\{\tau_t^{(n)}\}$ as in Sec. 1 of this paper,

$$(\mathrm{s}, \mathcal{H}_\alpha) \lim_{n\to\infty} \pi_\alpha(\tau_t^{(n)} A) = \pi_\alpha(\tau_t^\alpha A), \quad \forall\ A \in \mathfrak{A}_L. \quad \text{(B1)}$$

Suppose now that condition (II) were fulfilled. Then, in view of (i) and (ii), this would mean that, for $A \in \mathfrak{A}_L$, $\tau_t^{(n)} A$ converges normwise to $\tau_t^\alpha A$ as $n \to \infty$, for each $\alpha \in S$. This implies that τ_t^α is independent of α, in contradiction with (ii). Thus we conclude that the model violates (II).

As regards conditions (III) and (IV), it was shown by Thirring[15] that the model satisfies (III) in the strong coupling case. Further, Thirring's results imply that any Gibbs state ϕ may be expressed as a direct integral

$$\phi = \int_0^{2\pi} d\theta \phi_\theta, \quad \text{(B2)}$$

where the index $\theta \in S$ and where the GNS space and representation $(\mathcal{H}_\theta, \pi_\theta)$ corresponding to ϕ_θ thus satisfy the above properties (i)–(iii). Further, it was shown by Thirring that

$$\lim_{n\to\infty} \phi^{(n)}(A_1^{(n)}(t_1) \cdots A_k^{(n)}(t_k)) = \int_0^{2\pi} d\theta \phi_\theta((\tau_{t_1}^\theta A_1) \cdots (\tau_{t_k}^\theta A_k)), \quad \text{(B3)}$$

where $\phi^{(n)}$ is defined as in Sec. 1. It follows readily from Eqs. (B1)–(B3) that the model satisfies (IV).

[1] R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).

[2] D. W. Robinson, Commun. Math. Phys. **6**, 151 (1967); **7**, 337 (1968).

[3] S. Sakai, Proc. Japan Acad. **33**, 439 (1957), Proposition 1.

[4] In order to verify that (III) and (IV) form a weaker pair of assumptions than (I) and (II), it suffices to note that: (a) By employing the procedure used by HHW at the start of their Sec. IV, one readily verifies that (I) and (II) imply (III) and (IV); and (b) as shown in the Appendices of this article, there are models satisfying (III) and (IV), but not (II).

[5] A. S. Wightman, Phys. Rev. **101**, 860 (1956); see also Chap. 3.4 of R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).

[6] N. M. Hugenholtz and J. D. Wieringa, Commun. Math. Phys. **16**, 81 (1969), Theorem 3.

[7] E. Hille and R. S. Phillips, *Functional Analysis and Semi-groups* (Am. Math. Soc., Providence, R.I., 1957), Vol. 31, p. 73, Corollary 2 and Theorem 10.2.3.

[8] J. Dixmier, *Les algebres d'operateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1957), Chap. I.3, Theorem 2, Corollary 1.

[9] N. M. Hugenholtz, Commun. Math. Phys. **6**, 189 (1967).

[10] G. L. Sewell, J. Math. Phys. **11**, 1868 (1970).

[11] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).

[12] Cf. I. M. Gelfand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, p. 159.

[13] It should be noted that this algebra is not strictly identical to that used in AW. Our choice of $\mathfrak{A}$ enables us to use the local normality of ϕ, as in Proposition 5, and also to take over the calculation of AW for the expectation functional.

[14] W. Thirring and A. Wehrl, Commun. Math. Phys. **4**, 303 (1967).

[15] W. Thirring, Commun. Math. Phys. **7**, 181 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Integral Transformations in Momentum Space and Conformal Invariance

L. CASTELL
Max-Planck Institut für Physik und Astrophysik, 8 München 23, Germany

(Received 19 January 1970)

A certain element Z of the identity component of the conformal group together with the Poincaré subgroup generate the whole conformal group. In order to prove the conformal invariance of an S-matrix, only the invariance under Z has to be checked, once relativistic invariance has been established. The explicit form of Z for certain physically important representations of the different covering groups of the conformal group will be derived. The transformation Z turns out to be an integral transformation.

1. INTRODUCTION

The explicit form of the different irreducible unitary representations of the proper orthochronous Poincaré group and its universal covering group in momentum space is well known.[1] These unitary transformations have the property that an "improper eigenstate" of the 4-momentum is transformed into an improper eigenstate of the same kind. However, we are faced with a completely new situation if we generalize the Poincaré group to the conformal group. The action of the identity component (subgroup whose elements can be continuously connected with the identity) of the conformal group $SO_0(4, 2)/C_2$ and its physically important covering groups $SO_0(4, 2)$ and $SU_0(2, 2)$ is, in general, an *integral transformation* of the square-integrable functions in momentum space. An improper eigenstate of the 4-momentum is smeared out into a superposition of eigenstates if transformed by a special conformal transformation.

We shall give the structure of these integral transformations. The reason for the appearance of integral transformations lies in the fact that the momentum space is no longer an integral over different homogeneous spaces of the conformal group with respect to certain subgroups, which is true for the Poincaré group. More physically expressed, plane waves in Minkowski space, for example, are transformed into certain superpositions of "spherical" waves under a special conformal transformation.

2. AN IMPORTANT DISCRETE SUBGROUP OF THE CONFORMAL GROUP

In order to derive that, for example, an S-matrix theory is invariant under the identity component of the spin-covering group of the conformal group, it is sufficient to show that it is invariant under the identity component of the universal covering group of the Poincaré group[2] and one further discrete transformation Z, as these transformations generate the whole group. In Minkowski space (metric $+++-$) Z is given by

$$y_i' = -y_i/y^2, \quad y_4' = +y_4/y^2.$$

[Expressed in homogeneous coordinates,

$$x_\rho,\ \rho = 1, 2, \cdots, 6; \quad (x) \neq 0;$$
$$(x) \triangleq (\lambda x), \quad \lambda > 0;$$
$$x_1^2 + x_2^2 + x_3^2 - x_4^2 + x_5^2 - x_6^2 = 0;$$
$$y_\mu = \frac{x_\mu}{x_5 + x_6}, \quad \mu = 1, 2, 3, 4;$$

Z is represented by a rotation in the $x_4 - x_6$ plane by an angle π.] The inner automorphism induced by $Z = e^{i\frac{1}{2}\pi(P_4 - K_4)}$ in the Lie algebra of the conformal group

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(g_{\mu\rho}M_{\nu\sigma} - g_{\mu\sigma}M_{\nu\rho} - g_{\nu\rho}M_{\mu\sigma} + g_{\nu\sigma}M_{\mu\rho}),$$
$$[M_{\mu\nu}, P_\rho] = i(g_{\mu\rho}P_\nu - g_{\nu\rho}P_\mu), \quad [P_\mu, P_\nu] = 0,$$
$$[D, M_{\mu\nu}] = 0, \quad [D, P_\mu] = iP_\mu,$$
$$[D, K_\mu] = -iK_\mu,$$
$$[M_{\mu\nu}, K_\rho] = i(g_{\mu\rho}K_\nu - g_{\nu\rho}K_\mu),$$
$$[K_\mu, K_\nu] = 0, \quad [P_\mu, K_\nu] = -2i(g_{\mu\nu}D - M_{\mu\nu})$$

has the following form:

$$M_{ik} \to M_{ik}, \quad M_{i4} \to -M_{i4}, \quad P_i \to K_i, \quad P_4 \to -K_4,$$
$$D \to -D, \quad K_i \to P_i, \quad K_4 \to -P_4, \quad i, k = 1, 2, 3.$$

From this inner automorphism one can see that the kernel $z(p, q)$ of the integral operator Z in momentum space is an improper eigenvector of the operator K_μ with the eigenvalue $k_\mu = g_{\mu\mu'}q_{\mu'}$ (P_μ represented by p_μ). In the following we shall deal with two series of unitary irreducible representations of the group $SO_0(4, 2)$ and $SU_0(2, 2)$ which are of special physical interest. Thus, Z is an unitary operator $Z^+Z = 1$, which has the property $ZZ = \pm 1$, depending whether we are dealing with representations which belong to integer or half-integer spin. A special conformal

transformation, which is represented in Minkowski space, for example, by

$$y'_\mu = \frac{y_\mu - b_\mu y^2}{1 - 2by + b^2y^2},$$

induces on the square-integrable functions ψ in momentum space the transformation

$$\psi' = Z^+(e^{-ip_\mu b_\mu}1)Z\psi.$$

In the following we shall give the mathematical results concerning the transformation Z. Physical consequences will be published later.

3. DISCRETE DEGENERATE REPRESENTATION OF $SO_0(4, 2)$

The reduction of the most degenerate unitary representations of the principal series of $SO_0(4, 2)$ with respect to the Poincaré group has been dealt with.[3] The representation of the Lie algebra for the physically interesting discrete series is given by

$$M_{\mu\nu} = i^{-1}(g_{\nu\nu'}p_\mu\partial_{\nu'} - g_{\mu\mu'}p_\nu\partial_{\mu'}),$$
$$P_\mu = p_\mu, \quad D = i(p_\mu\partial_\mu + 2),$$
$$K_\mu = p_\mu(g_{\rho\rho'}\partial_\rho\partial_{\rho'} + \nu^2/p^2) - 2g_{\mu\mu'}(p_\rho\partial_\rho + 2)\partial_{\mu'},$$
$$\partial_\mu = \frac{\partial}{\partial p_\mu}, \quad \mu, \nu, \rho = 1, 2, 3, 4,$$

in the Hilbert space of square-integrable functions

$$\int_\Delta \psi^*(p)\psi(p)d^4p < \infty.$$

Here

$$\Delta = \{p_\mu : p_4 > 0, p^2 < 0; \text{ or } p_4 < 0, p^2 < 0; \text{ for } \nu = 0, 1, 2, \cdots\},$$

where $\nu = 0$ belongs already to the continuous series. These representations contain only massive particles with spin 0. The kernel of the integral operator Z is given by

$$z(p, q) = \frac{1}{4\pi}\frac{\partial}{\partial a}(J_\nu\{[a + (a^2 - b^2)^{\frac{1}{2}}]^{\frac{1}{2}}\} \times J_\nu\{[a - (a^2 - b^2)^{\frac{1}{2}}]^{\frac{1}{2}}\}),$$
$$a = p_iq_i + p_4q_4, \quad b^2 = p^2q^2,$$
$$p^2 = p_i^2 - p_4^2 < 0, \quad q^2 = q_i^2 - q_4^2 < 0,$$
$$p_4, q_4 > 0 \quad \text{or} \quad p_4, q_4 < 0.$$

The expression given in Ref. 4 as eigensolution of the K_μ (for the special case $\nu = 1$) does not satisfy the eigenvalue equations.

To prove the relation

$$\int z(q', p)z(p, q)d^4p = \delta^4(q' - q),$$

one introduces the variables $p_4 \pm |p|$ and uses the completeness relation for Bessel functions.

4. THE EXCEPTIONAL DISCRETE REPRESENTATIONS OF $SU_0(2, 2)$

The reduction of the exceptional degenerate discrete series of unitary representations[5] of $SU_0(2, 2)$ with respect to the Poincaré group[6,7] shows that the irreducible representations stay irreducible and contain just representations of the Poincaré group, which belong to mass 0 and discrete helicity λ. The generators, which are not difficult to calculate,[8] are given by

$$M_{ik} = i^{-1}(p_i\partial_k - p_k\partial_i) + \lambda\epsilon_{ikl}\frac{p_l + \delta_{l3}p}{p + p_3},$$
$$M_{41} = -ip\partial_1 - \lambda\frac{p_2}{p + p_3},$$
$$M_{42} = -ip\partial_2 + \lambda\frac{p_1}{p + p_3},$$
$$M_{43} = -ip\partial_3, \quad P_i = p_i, \quad P_4 = p,$$
$$D = i(p_k\partial_k + 1),$$
$$K_1 = p_1\nabla^2 - 2(p_k\partial_k + 1)\partial_1 - 2i\lambda\left(\frac{p_2}{p + p_3}\partial_3 - \partial_2\right),$$
$$K_2 = p_2\nabla^2 - 2(p_k\partial_k + 1)\partial_2 - 2i\lambda\left(\partial_1 - \frac{p_2}{p + p_3}\partial_3\right),$$
$$K_3 = p_3\nabla^2 - 2(p_k\partial_k + 1)\partial_3 - \frac{2i\lambda}{p + p_3}(p_1\partial_2 - p_2\partial_1 + i\lambda),$$
$$K_4 = p\nabla^2 + \frac{2i\lambda}{p + p_3}(p_1\partial_2 - p_2\partial_1 + i\lambda),$$
$$i, k = 1, 2, 3, \quad p = +(p_kp_k)^{\frac{1}{2}}, \quad \nabla^2 = \partial_k\partial_k.$$

If we replace M_{4i}, P_4, and K_4 by $-M_{4i}$, $-P_4$, and $-K_4$, respectively, we get the representations which belong to negative energy. The Hilbert space is defined over the square-integrable functions

$$\int \psi^*(p)\psi(p)\frac{d^3p}{p} < \infty.$$

The number $\lambda = 0, \pm\frac{1}{2}, \pm 1, \cdots$ represents the helicity. The kernel of the integral transformation Z is given by

$$z(p, q) = \frac{(i)^{2\lambda}}{4\pi}\left(\frac{(p + p_3)(q + q_3) + (p_1 - ip_2)(q_1 + iq_2)}{(p + p_3)(q + q_3) + (p_1 + ip_2)(q_1 - iq_2)}\right)^\lambda \times J_{2\lambda}\{[2(pq + p_iq_i)]^{\frac{1}{2}}\}.$$

The eigenfunctions of K_μ for $\lambda = 0$ have been given in Ref. 9. [They are not the kernel of a transformation of $SU_0(2, 2)$.]

To prove the relations

$$\int z^*(p, q')z(p, q)\frac{d^3p}{p} = q\delta^3(q - q'),$$

$$\int z(q', p)z(p, q)\frac{d^3p}{p} = (-1)^{2\lambda}q\delta^3(q - q'),$$

one uses Graf's addition theorem, introduces the variables $p \pm p_3$, and applies the completeness relation for Bessel functions.

ACKNOWLEDGMENT

The author would like to thank Dipl. Phys. W. Jass for discussions.

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[2] It is sufficient to ask for translation invariance only.
[3] L. Castell, Nucl. Phys. **B4**, 343 (1967).
[4] H. A. Kastrup, Phys. Rev. **142**, 1060 (1966).
[5] T. Yao, J. Math. Phys. **9**, 1615 (1968).
[6] G. Mack and I. Todorov, J. Math. Phys. **10**, 2078 (1969).
[7] L. Castell, Nucl. Phys. **B13**, 231 (1969).
[8] See, for example, S. K. Bose and R. Parker, J. Math. Phys. **10**, 812 (1969).
[9] H. A. Kastrup, Phys. Rev. **140**, B183 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Relative Position Representation for a Relativistic Two-Particle System*

S. C. McDONALD†
Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa 50010

(Received 6 April 1970)

Two-particle angular momentum states are constructed which are localized with respect to the magnitude of the relative position in the rest system and which have arbitrary 3-momentum dependence. The associated relative position operator is constructed, and a quantum-mechanical analog of the classical impact parameter is identified. Two-particle angular momentum states are constructed, which are also localized with respect to the "mean-position" of the 2-particle system, and the associated "mean-position" operator is seen to be a generalization of the 1-particle Newton–Wigner position operator.

1. INTRODUCTION

In the analysis of 2-particle scattering processes, one frequently employs a potential function to represent the basic interaction mechanism. Based on macroscopic considerations, it is apparent that such a potential should in some way depend upon the relative displacement of the two particles. Relativistically speaking, however, it is not clear how the parametrization of such a potential by a relative displacement may be effected since different inertial observers will not, in general, agree on what is meant by "relative position." Intimately related to this question is the more formal problem of the construction of a covariant relative-position representation of the 2-particle Hilbert space in question.

Several nonequivalent 1-particle position representations have been investigated by Pryce,[1] Møller,[2] and Newton and Wigner,[3] and their manifestly covariant generalizations have been given by Fleming.[4] Since it is based on invariance principles, the Newton–Wigner representation seems to offer the most promise for generalization to the 2-particle case. The natural generalization of the Newton–Wigner formalism to the case of 2-particle states which are localized only with respect to the rest system relative position is suggested by the states

$$|\tilde{P}\rho\rangle = (2\pi)^{-\frac{3}{2}}\int d^3q[4\omega_1\omega_2]^{-\frac{1}{2}}e^{i\rho\cdot q}\,|\tilde{P}\mathbf{q}\rangle, \quad (1.1)$$

where $\tilde{P} = (M, \mathbf{0})$ and q are respectively the total and relative momenta in the rest system and $\omega_i = (m_i^2 + \mathbf{q}^2)^{\frac{1}{2}}$, $i = 1, 2$. The momentum states $|\tilde{P}\mathbf{q}\rangle$ are taken to have the Lorentz-invariant normalization, given in the rest system by

$$\langle\tilde{P}\mathbf{q}'\,|\,\tilde{P}\mathbf{q}\rangle = 4\omega_1\omega_2\delta^3(\mathbf{q}' - \mathbf{q}), \quad (1.2)$$

the over-all momentum Dirac distribution having been factored out. Due to the non-Lorentz covariance of the Newton–Wigner localization criterion,[3] it is evident that the states of Eq. (1.1) do not transform into corresponding localized states of arbitrary over-all momentum. One notes, however, that the magnitude of the rest system relative momentum is a Lorentz invariant, given by

$$q = (2M)^{-1}\{[M^2 - (m_1 + m_2)^2][M^2 - (m_1 - m_2)^2]\}^{\frac{1}{2}}. \quad (1.3)$$

Consequently, states localized only with respect to the magnitude of the rest-system relative position may admit a direct generalization to states of arbitrary momentum.

To prove the relations

$$\int z^*(p, q')z(p, q)\frac{d^3p}{p} = q\delta^3(q - q'),$$

$$\int z(q', p)z(p, q)\frac{d^3p}{p} = (-1)^{2\lambda}q\delta^3(q - q'),$$

one uses Graf's addition theorem, introduces the variables $p \pm p_3$, and applies the completeness relation for Bessel functions.

ACKNOWLEDGMENT

The author would like to thank Dipl. Phys. W. Jass for discussions.

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[2] It is sufficient to ask for translation invariance only.
[3] L. Castell, Nucl. Phys. **B4**, 343 (1967).
[4] H. A. Kastrup, Phys. Rev. **142**, 1060 (1966).
[5] T. Yao, J. Math. Phys. **9**, 1615 (1968).
[6] G. Mack and I. Todorov, J. Math. Phys. **10**, 2078 (1969).
[7] L. Castell, Nucl. Phys. **B13**, 231 (1969).
[8] See, for example, S. K. Bose and R. Parker, J. Math. Phys. **10**, 812 (1969).
[9] H. A. Kastrup, Phys. Rev. **140**, B183 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Relative Position Representation for a Relativistic Two-Particle System*

S. C. McDONALD†
Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa 50010

(Received 6 April 1970)

Two-particle angular momentum states are constructed which are localized with respect to the magnitude of the relative position in the rest system and which have arbitrary 3-momentum dependence. The associated relative position operator is constructed, and a quantum-mechanical analog of the classical impact parameter is identified. Two-particle angular momentum states are constructed, which are also localized with respect to the "mean-position" of the 2-particle system, and the associated "mean-position" operator is seen to be a generalization of the 1-particle Newton–Wigner position operator.

1. INTRODUCTION

In the analysis of 2-particle scattering processes, one frequently employs a potential function to represent the basic interaction mechanism. Based on macroscopic considerations, it is apparent that such a potential should in some way depend upon the relative displacement of the two particles. Relativistically speaking, however, it is not clear how the parametrization of such a potential by a relative displacement may be effected since different inertial observers will not, in general, agree on what is meant by "relative position." Intimately related to this question is the more formal problem of the construction of a covariant relative-position representation of the 2-particle Hilbert space in question.

Several nonequivalent 1-particle position representations have been investigated by Pryce,[1] Møller,[2] and Newton and Wigner,[3] and their manifestly covariant generalizations have been given by Fleming.[4] Since it is based on invariance principles, the Newton–Wigner representation seems to offer the most promise for generalization to the 2-particle case. The natural generalization of the Newton–Wigner formalism to the case of 2-particle states which are localized only with respect to the rest system relative position is suggested by the states

$$|\tilde{P}\rho\rangle = (2\pi)^{-\frac{3}{2}}\int d^3q[4\omega_1\omega_2]^{-\frac{1}{2}}e^{i\rho\cdot q}\,|\tilde{P}\mathbf{q}\rangle, \quad (1.1)$$

where $\tilde{P} = (M, \mathbf{0})$ and q are respectively the total and relative momenta in the rest system and $\omega_i = (m_i^2 + \mathbf{q}^2)^{\frac{1}{2}}$, $i = 1, 2$. The momentum states $|\tilde{P}\mathbf{q}\rangle$ are taken to have the Lorentz-invariant normalization, given in the rest system by

$$\langle\tilde{P}\mathbf{q}'\,|\,\tilde{P}\mathbf{q}\rangle = 4\omega_1\omega_2\delta^3(\mathbf{q}' - \mathbf{q}), \quad (1.2)$$

the over-all momentum Dirac distribution having been factored out. Due to the non-Lorentz covariance of the Newton–Wigner localization criterion,[3] it is evident that the states of Eq. (1.1) do not transform into corresponding localized states of arbitrary over-all momentum. One notes, however, that the magnitude of the rest system relative momentum is a Lorentz invariant, given by

$$q = (2M)^{-1}\{[M^2 - (m_1 + m_2)^2][M^2 - (m_1 - m_2)^2]\}^{\frac{1}{2}}. \quad (1.3)$$

Consequently, states localized only with respect to the magnitude of the rest-system relative position may admit a direct generalization to states of arbitrary momentum.

In this paper such a relative-position representation will be developed. In view of the noncovariance of the localization criterion, it will be necessary to first construct a displacement-dependent momentum-space basis in terms of which the relative position states of arbitrary momentum may be defined. Toward this end, Sec. 2 will consist of a brief review of the canonical formalism of Foldy[5] with regard to the construction of 1-particle canonical states. In Sec. 3, 2-particle canonical states will be constructed by first performing a relative displacement on the standard canonical states of Macfarlane[6] and by then projecting out those states of sharp orbital angular momentum. Such states are then used in Sec. 4 as a basis for the construction of angular momentum states of arbitrary $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$ which are localized with respect to the magnitude of the rest-system relative displacement. The corresponding "relative-position" operator will be constructed, and an "impact parameter" operator identified as the quantum-mechanical generalization of the classical impact parameter. In Sec. 5, 2-particle angular momentum states localized with respect to the over-all "mean position" will be constructed, and the corresponding "mean position" operator seen to be the natural extension of the Newton-Wigner position operator[3] to a 2-particle theory.

In this work, only the cases for which $m_1 > 0$ and $m_2 > 0$ will be considered. As usual, units will be chosen such that $c = \hbar = 1$.

2. THE CANONICAL FORMALISM

The canonical formalism which was first developed by Foldy[5] in 1956 is a prime example of the application of group-theoretical techniques to the formal kinematics of noninteracting particles. As used in this paper, the term "particle" will be synonomous with "elementary system," defined by Newton and Wigner[3] to be a system described completely by states which transform within a given irreducible representation of the Poincaré group. It will be recalled that the Poincaré group is that group of transformations of the form (l, a) which, when acting on vectors x^μ in 4-dimensional Minkowski space, yield

$$(l, a): x^\mu \to x'^\mu = l^\mu{}_\nu x^\nu + a^\mu, \tag{2.1}$$

where

$$l^\mu{}_\nu l_\mu{}^\lambda = \delta^\lambda{}_\nu. \tag{2.2}$$

The quantum-mechanical generators of infinitesimal homogeneous transformations of the form $(\delta l, 0)$ are the well-known angular momentum and boost generators J^i and N^i, $i = 1, 2, 3$, respectively. On the other hand, the generators of infinitesimal translations $(1, \delta a)$ are the 4-momentum operators P^μ, $\mu = 0, 1, 2, 3$. The commutation rules satisfied by these generators are

$$\begin{aligned} &[P^0, J^i] = 0, \qquad [P^0, N^i] = iP^i, \\ &[P^i, N^j] = i\delta^{ij}P^0, \quad [N^i, N^j] = -i\epsilon^{ijk}J^k, \\ &[J^i, A^j] = i\epsilon^{ijk}A^k, \end{aligned} \tag{2.3}$$

where $\mathbf{A} = \mathbf{P}$, $\mathbf{J}$, or $\mathbf{N}$.

Consistent with the above commutation rules, the rest states $|\tilde{p}\, s\, \mu\rangle$ of a particle with mass m and spin s may be defined to be eigenstates of P, $\mathbf{J}^2$, and J_z with respective eigenvalues $\tilde{p} = (m, \mathbf{0})$, $s(s+1)$, and μ. Such states are seen to transform under a rotation $r(\alpha\, \beta\, \gamma)$, with Euler angles α, β, and γ and corresponding operator

$$R[r(\alpha\, \beta\, \gamma)] = \exp(-i\alpha J_z)\exp(-i\beta J_y)\exp(-i\gamma J_z), \tag{2.4}$$

by the transformation

$$R\,|\tilde{p}\, s\, \mu\rangle = \sum_{\mu'} D^s_{\mu'\mu}(R)\,|\tilde{p}\, s\, \mu'\rangle, \tag{2.5}$$

where $D^s(R)$ is the $s(s+1)$-dimensional irreducible representation matrix for the rotation R, given by Rose[7] as

$$D^s_{\mu\nu}(R) = \exp(-i\mu\alpha)d^j_{\mu\nu}(\beta)\exp(-i\nu\gamma). \tag{2.6}$$

The canonical states for general momenta p are then defined by

$$|p\, s\, \mu\rangle_c = L(p)\,|\tilde{p}\, s\, \mu\rangle. \tag{2.7}$$

Here, the Lorentz transformation without rotation $l(p)$ is defined by McKerrell[8] as

$$\begin{aligned} p^\mu &= l^\mu{}_\nu(p)\tilde{p}^\nu, \\ l(p)^0{}_0 = p^0/m, \quad l(p)^0{}_k &= l(p)^k{}_0 = p^k/m, \\ l(p)^i{}_j &= \delta^{ij} + p^i p^j/m(m + p^0) \end{aligned} \tag{2.8}$$

with the corresponding operator

$$L(p) = \exp(-i\zeta\hat{\mathbf{p}}\cdot\mathbf{N}), \tag{2.9}$$

$\hat{\mathbf{p}}$ being a unit vector in the $\mathbf{p}$ direction and ζ defined by[8]

$$p = (m\cosh\zeta, m\sinh\zeta(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)). \tag{2.10}$$

From Eq. (2.5) and the fact that, for an arbitrary Lorentz transformation l, with operator L, the operator

$$\bar{R}(L, p) \equiv L(lp)^{-1}\, L\, L(p) \tag{2.11}$$

represents a rotation, the states of Eq. (2.7) transform in the canonical manner

$$L\,|p\, s\, \mu\rangle_c = \sum_{\mu'} D^s_{\mu'\mu}[\bar{R}(L, p)]\,|lp\, s\, \mu'\rangle_c. \tag{2.12}$$

Furthermore, the canonical states of Eq. (2.7) satisfy

$$P^\alpha\,|p\, s\, \mu\rangle_c = p^\alpha\,|p\, s\, \mu\rangle_c. \tag{2.13}$$

3. TWO-PARTICLE ANGULAR MOMENTUM STATES

In a recent discussion of 1-particle localized states by the present author,[9] a set of intermediate states is implicitly constructed from the general canonical states $|p\,\lambda\rangle_c$ by defining for each translation $t(\mathbf{x})$, with operator $T(\mathbf{x})$, the states

$$|p\,\lambda;\mathbf{x}\rangle_c \equiv T(\mathbf{x})\,|p\,\lambda\rangle_c = \exp(i\mathbf{p}\cdot\mathbf{x})\,|p\,\lambda\rangle_c. \tag{3.1}$$

The states of Eq. (3.1) are elements of the rays in Hilbert space which have as standard representatives the states $|p\,\lambda\rangle_c$, and as such are also canonical states, each of which is specified by a point in the parameter space of the 3-dimensional translation group. In order to insure that the interpretation of the $\mathbf{x}$'s as the points of translation space is preserved in the construction of localized states, so that such states may indeed be taken as a basis for a "position" representation, the states localized at $\mathbf{x}$ are defined as a superposition of the states of Eq. (3.1), with the $\mathbf{x}$ dependence contained entirely in said basis states. The Newton–Wigner localization criterion[3] requires that localized states defined by

$$|\mathbf{x}\lambda\rangle = \int d^3p\,\Psi^s(p)\,|p\lambda;\mathbf{x}\rangle_c \tag{3.2}$$

be normalized, within a constant factor, to

$$\langle\mathbf{x}'\lambda'\,|\,\mathbf{x}\lambda\rangle = \delta_{\lambda'\lambda}\delta^3(\mathbf{x}'-\mathbf{x}). \tag{3.3}$$

If the canonical states are taken to have the non-relativistic normalization

$${}'_c\langle p'\lambda';\mathbf{x}'\,|\,p\lambda;\mathbf{x}\rangle'_c = \delta_{\lambda'\lambda}\delta^3(\mathbf{p}'-\mathbf{p})\exp[-i\mathbf{p}\cdot(\mathbf{x}'-\mathbf{x})], \tag{3.4}$$

it then follows that $\Psi^s(p)$ is a constant, and the corresponding localized states are the 3-dimensional Fourier transforms of the standard ray representatives of the canonical states. If, on the other hand, the canonical states are chosen to have the Lorentz-invariant normalization

$${}_c\langle p'\lambda';\mathbf{x}'\,|\,p\lambda;\mathbf{x}\rangle_c = \delta_{\lambda'\lambda}\tilde{\delta}(p',p)\exp[-i\mathbf{p}\cdot(\mathbf{x}'-\mathbf{x})], \tag{3.5}$$

where $\tilde{\delta}(p',p) \equiv 2\omega(p)\delta^3(\mathbf{p}'-\mathbf{p})$ with $\omega(p) = (\mathbf{p}^2+m^2)^{\frac{1}{2}}$, the localized states are seen to be given by

$$|\mathbf{x}\lambda\rangle = (2\pi)^{-\frac{3}{2}}\int\frac{d^3p}{2\omega}(2\omega)^{\frac{1}{2}}\,|p\lambda;\mathbf{x}\rangle_c = (2\pi)^{-\frac{3}{2}}\int\frac{d^3p}{2\omega}(2\omega)^{\frac{1}{2}}\exp(i\mathbf{p}\cdot\mathbf{x})\,|p\lambda\rangle_c, \tag{3.6}$$

a "modified" Fourier transform of the standard canonical states $|p\lambda\rangle_c$.

However, in the case of 2-particle states localized with respect to the magnitude of the rest system relative displacement, it is not so apparent how the integral transformation corresponding to that of Eq. (3.6) is effected so as to result in a meaningful interpretation of relative position. The problem of finding the "suitable" integral transformation which allows for the interpretation of a position space as the parameter space of a specified translation group may be partially resolved by first finding the ray representatives of the canonical states which are related to the standard canonical states by the appropriate translation. In the 1-particle case the suitable basis is given by Eq. (3.1). The remainder of this section will hence be devoted to the construction of possible sets of canonical states which "carry" the displacement dependence.

Starting with the direct-product state

$$|\tilde{p}_1\lambda_1;\tilde{p}_2\lambda_2\rangle = |\tilde{p}_1\lambda_1\rangle_c\otimes|\tilde{p}_2\lambda_2\rangle_c \tag{3.7}$$

for particles of mass m_1 and m_2 and spin s_1 and s_2, respectively, and momenta $\tilde{\mathbf{p}}_1 = -\tilde{\mathbf{p}}_2 = \tilde{\mathbf{q}} = (0,0,q)$, we may construct rest states of over-all spin s as the Clebsch–Gordan series

$$|\tilde{P}\,s\,\lambda\rangle = \sum_{\lambda_1,\lambda_2} C(s_1s_2s\,|\,\lambda_1\lambda_2\lambda)\,|\tilde{p}_1\lambda_1;\tilde{p}_2\lambda_2\rangle, \tag{3.8}$$

where $\tilde{P} = \tilde{p}_1+\tilde{p}_2 = (M,\mathbf{0})$ and $M = (m_1^2+q^2)^{\frac{1}{2}} + (m_2^2+q^2)^{\frac{1}{2}}$. Two-particle angular momentum states may now be constructed by applying to the states of Eq. (3.8) the Wigner projection operator[10] for orbital angular momentum l and z component m and defining the states

$$|\tilde{P}lms\lambda\rangle = \int d\Omega D^{l*}_{m0}(R)R^L(r)\,|\tilde{P}s\lambda\rangle. \tag{3.9}$$

Here $r = r(\varphi\theta 0)$ is a rotation with Euler angles φ, θ, and 0, with corresponding operator $R^L(r)$ which acts only in momentum space and volume element $d\Omega = d\varphi d(\cos\theta)$.

In his construction of 2-particle canonical states, McKerrell[8] has defined the total angular momentum states in the rest system by

$$|\tilde{P}\alpha[Mj]ls\rangle_c = (\tfrac{1}{2}q)^{\frac{1}{2}}\left(\frac{2l+1}{4\pi}\right)^{\frac{1}{2}}\sum_{m,\lambda}C(lsj\,|\,m\lambda\alpha)\,|\tilde{P}lms\lambda\rangle, \tag{3.10}$$

which are seen to satisfy[8]

$$R\,|\tilde{P}\alpha[Mj]ls\rangle_c = \sum_{\alpha'}D^j_{\alpha'\alpha}(R)\,|\tilde{P}\alpha'[Mj]ls\rangle_c \tag{3.11}$$

for arbitrary rotations r, with operator $R = R^LR^S = R^SR^L$ which acts in both momentum and spin spaces. Moreover, the states of Eq. (3.10) are noted by McKerrell[8] to be eigenstates of P, J_z, $\mathbf{J}^2$, $\mathbf{L}^2$, and $\mathbf{S}^2$.

McKerrell[8] then obtains the states with general total momentum P by applying the operator $L(P)$ of Eq. (2.9) to the rest states of Eq. (3.10):

$$|P\alpha[Mj]ls\rangle_c = L(P)\,|\tilde{P}\alpha[Mj]ls\rangle. \tag{3.12}$$

These states are then shown to transform under an arbitrary Lorentz transformation L in the canonical manner:

$$L\,|P\alpha[Mj]ls\rangle_c = \sum_{\alpha'} D^j_{\alpha'\alpha}[\bar{R}(L,P)]\,|lP\alpha'[Mj]ls\rangle_c, \tag{3.13}$$

where the transformation $\bar{R}(L,P)$ is the natural direct-product generalization of the transformation of Eq. (2.11).

An alternative approach to the construction of 2-particle canonical states is to first apply to the rest states of Eq. (3.9) for which the orbital angular momentum is zero a "relative displacement" of amount $\hat{\boldsymbol{\rho}} = (0, 0, \rho)$. A "relative displacement" may be defined in general as a transformation $\tilde{t}(\mathbf{x})$ on the coordinate pair $(\mathbf{x}_1, \mathbf{x}_2)$ with action

$$\tilde{t}(\mathbf{x}):(\mathbf{x}_1, \mathbf{x}_2) \to (\mathbf{x}_1 + \tfrac{1}{2}\mathbf{x}, \mathbf{x}_2 - \tfrac{1}{2}\mathbf{x}). \tag{3.14}$$

The quantum-mechanical operator corresponding to $\tilde{t}(\mathbf{x})$ is taken to be

$$\tilde{T}(\mathbf{x}) = T^{(1)}(\tfrac{1}{2}\mathbf{x})T^{(2)}(\tfrac{1}{2}\mathbf{x})^{-1} = T^{(1)}(\tfrac{1}{2}\mathbf{x})T^{(2)}(-\tfrac{1}{2}\mathbf{x}), \tag{3.15}$$

where $T^{(1)}$ and $T^{(2)}$ are the respective 1-particle translation operators. Due to the Abelian nature of the translation group, the set of $\tilde{t}(\mathbf{x})$ for all $\mathbf{x}$ forms a representation of the translation group. Such a representation is not, however, a subgroup of the direct-product representation of the Poincaré group defined by

$$\Lambda^{(1\times2)}(l, a) = \Lambda^{(1)}(l, a)\Lambda^{(2)}(l, a), \tag{3.16}$$

for which

$$(l, a):x^\mu \to l^\mu{}_\nu x^\nu + a^\mu. \tag{3.17}$$

The effect of the relative translation $\tilde{t}(\hat{\boldsymbol{\rho}})$ on the state $|\tilde{P}00s\lambda\rangle$ is

$$\begin{aligned}\tilde{T}(\hat{\boldsymbol{\rho}})\,|\tilde{P}00s\lambda\rangle &= \int d\Omega \tilde{T}(\hat{\boldsymbol{\rho}})R^L(r)\,|\tilde{P}s\lambda\rangle \\ &= \int d\Omega \exp(i\hat{\boldsymbol{\rho}}\cdot\mathbf{q})R^L(r)\,|\tilde{P}s\lambda\rangle,\end{aligned} \tag{3.18}$$

where the property

$$R^{(1\times2)}(r)\tilde{T}(\mathbf{x})R^{(1\times2)}(r^{-1}) = \tilde{T}(r\mathbf{x}) \tag{3.19}$$

has been used. Here, $R^{(1\times2)}(r) = R^{(1)}(r)R^{(2)}(r)$ and

$$\mathbf{q} = r(\varphi\theta0)\tilde{\mathbf{q}} = q(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta). \tag{3.20}$$

The rest state of orbital angular momentum l, z component m, and relative displacement ρ is then given by the effect of the Wigner projection operator on the states of Eq. (3.18):

$$\begin{aligned}&|\tilde{P}lms\lambda;\,\rho\rangle \\ &= \int d\Omega D^{l*}_{m0}(R)R^L(r)\cdot\tilde{T}(\hat{\boldsymbol{\rho}})\,|\tilde{P}00s\lambda\rangle \\ &= \int d\Omega\, d\Omega' D^{l*}_{m0}(R)\exp(i\hat{\boldsymbol{\rho}}\cdot\mathbf{q})R^L(rr')\,|\tilde{P}s\lambda\rangle.\end{aligned} \tag{3.21}$$

It then follows that upon making the variable change $r'' = rr'$, using the plane-wave expansion

$$\exp(i\hat{\boldsymbol{\rho}}\cdot\mathbf{q}) = \sum_{l'}(i)^{l'}(2l'+1)j_{l'}(q\rho)D^{l'*}_{00}(r), \tag{3.22}$$

and using the orthogonality of the rotation matrices

$$\int d\Omega D^{j'*}_{m'n'}(r)D^j_{mn}(r) = \frac{4\pi}{2j+1}\delta_{j'j}\delta_{m'm}\delta_{n'n}, \tag{3.23}$$

one may perform the integral over $d\Omega'$ in Eq. (3.21), yielding

$$\begin{aligned}|\tilde{P}lms\lambda;\,\rho\rangle &= 4\pi(i)^l j_l(q\rho)\int d\Omega D^l_{m0}(R)R^L(r)\,|\tilde{P}s\lambda\rangle \\ &= 4\pi(i)^l j_l(q\rho)\,|\tilde{P}lms\lambda\rangle_c\end{aligned} \tag{3.24}$$

by Eq. (3.9). Corresponding to Eq. (3.10), the canonical states of relative displacement ρ may then be defined by

$$\begin{aligned}&|\tilde{P}\alpha[Mj]ls;\,\rho\rangle \\ &= (\tfrac{1}{2}q)^{\frac{1}{2}}\left(\frac{2l+1}{4\pi}\right)^{\frac{1}{2}}\sum_{m\lambda} C(lsj\mid m\lambda\alpha)\,|\tilde{P}lms\lambda;\,\rho\rangle \\ &= 4\pi(i)^l j_l(q\rho)\,|\tilde{P}\alpha[Mj]ls\rangle.\end{aligned} \tag{3.25}$$

Moreover, since the relative momentum q is a Poincaré invariant quantity, it then follows that the canonical states of general momentum may also be defined via the transformation $L(P)$ as are McKerrell's standard canonical states in Eq. (3.12):

$$\begin{aligned}|P\alpha[Mj]ls;\,\rho\rangle &\equiv L(P)\,|\tilde{P}\alpha[Mj]ls;\,\rho\rangle \\ &= 4\pi(i)^l j_l(q\rho)\,|P\alpha[Mj]ls\rangle_c.\end{aligned} \tag{3.26}$$

Such states are clearly eigenstates of P, J_z, $\mathbf{J}^2$, $\mathbf{L}^2$, and $\mathbf{S}^2$ and are proportional to the standard canonical states except on a set of measure zero, that is, the set of zeros of the $j_l(q\rho)$. Furthermore, the states of Eq. (3.26) have the same Lorentz transformation properties as the canonical states:

$$L\,|P\alpha[Mj]ls;\,\rho\rangle = \sum_{\alpha'} D^j_{\alpha'\alpha}[\bar{R}(L,P)]\,|lP\alpha'[Mj]ls;\,\rho\rangle \tag{3.27}$$

for an arbitrary Lorentz transformation L. The normalization of the states of Eq. (3.26) is seen from

McKerrell's results[8] [Eq. (4.21)] to be

$$\langle P'\alpha'[M'j']l's';\ \rho' \mid P\alpha[Mj]ls;\ \rho\rangle = (4\pi)^2 j_l(q\rho')j_l(q\rho)\delta_{\alpha'\alpha}\delta_{j'j}\delta_{l'l}\delta_{s's}\delta(M'-M)\delta(P',P). \tag{3.28}$$

More generally, the relative displacement $\tilde{T}(\boldsymbol{\rho})$ may be applied to over-all rest states of Eq. (3.9) which have nonzero orbital angular momentum and, corresponding to Eq. (3.21), a set of canonical states defined for each $l_0 \geq 0$ by

$$|\tilde{P}lm(l_0m_0)s\lambda;\ \rho\rangle \equiv \int d\Omega D^{l*}_{m0}(R)R^L(r)\cdot\tilde{T}(\hat{\boldsymbol{\rho}})\,|\tilde{P}\ l_0\ m_0\ s\ \lambda\rangle_c. \tag{3.29}$$

By following an argument similar to that leading up to Eq. (3.25), the above equation may be simplified to

$$|\tilde{P}lm(l_00)s\lambda;\ \rho\rangle = 4\pi(i)^l\mathfrak{F}_{l,l_0}(q\rho)\,|\tilde{P}lms\lambda\rangle_c, \tag{3.30}$$

where

$$\mathfrak{F}_{l,l_0}(q\rho) = (-)^{l_0}\sum_{l'}(i)^{l'-l}[C(l_0ll'\mid 000)]^2 j_{l'}(q\rho) \tag{3.31}$$

and where only states for which $m_0 = 0$ occur. In parallel with Eqs. (3.25) and (3.26), the corresponding states of total angular momentum and general momentum are then defined by

$$|P\alpha[Mj]ls;\ l_0,\ \rho\rangle = L(P)\left[(\tfrac{1}{2}q)^{\frac{1}{2}}\left(\frac{2l+1}{4\pi}\right)^{\frac{1}{2}} \times \sum_{m\lambda} C(lsj\mid m\lambda\alpha)\,|\tilde{P}lm(l_00)sl;\ \rho\rangle\right], \tag{3.32}$$

which, by using Eqs. (3.10) and (3.12), becomes

$$|P\alpha[Mj]ls;\ l_0,\ \rho\rangle = 4\pi(i)^l\mathfrak{F}_{l,l_0}(q\rho)\,|P\alpha[Mj]ls\rangle_c. \tag{3.33}$$

The ray representatives of the canonical states thus given by Eq. (3.33) provide for each l_0 a possible basis set for the construction of relative position states, localized with respect to the magnitude ρ of the relative displacement in the rest system. Of particular interest is the set of canonical states given by Eq. (3.26), which is the specialization of Eq. (3.33) to the case for which $l_0 = 0$. Only for this set is the appearance of orbital angular momentum states with $l > 0$ due entirely to the initial relative displacement.

4. RELATIVE POSITION STATES

Angular momentum states localized with respect to the rest-system relative position may now be constructed by imposing the following requirements: (a) A complete set of such states be constructed as a superposition of the states of one of the sets of canonical states of Eq. (3.33), so that the magnitude of the rest-system relative displacement be taken as the relative position parameter; (b) the only possible states localized to zero relative position be those states of zero orbital angular momentum; (c) two states localized to different relative positions be orthogonal.

Conditions (a) and (b) require that relative position states with definite total, orbital, and spin angular momenta be written as a superposition of that set of Eqs. (3.33) for which $l_0 = 0$, that is, the set of which all canonical states are generated by the relative displacement $\tilde{r}(\hat{\boldsymbol{\rho}})$ in the rest frame. Thus the relative position states are defined by

$$|\mathbf{P}\rho(j\mu)ls\rangle = \int_0^\infty dq\Psi^l(q)\,|\mathbf{P}\mu(Mj)ls;\ \rho\rangle = 4\pi(i)^l\int_0^\infty dq\Psi^l(q)j_l(q\rho)\,|P\mu(Mj)ls\rangle_c, \tag{4.1}$$

where $P^0 = (\mathbf{P}^2 + M^2)^{\frac{1}{2}}$ and $M = (m_1^2 + q^2)^{\frac{1}{2}} + (m_2^2 + q^2)^{\frac{1}{2}}$. The function $\Psi^l(q)$ may be determined up to a phase by condition (c), the "localization" criterion for relative position, which requires that

$$\langle \mathbf{P}'\rho'(j'\mu')l's' \mid \mathbf{P}\rho(j\mu)ls\rangle = \delta_{j'j}\delta_{\mu'\mu}\delta_{l'l}\delta_{s's}\delta^3(P'-P)\rho^{-2}\delta(\rho'-\rho). \tag{4.2}$$

From the normalization of the states of Eq. (3.28) and the orthogonality property of the spherical Bessel functions,

$$\int_0^\infty q^2\,dqj_l(q\rho)j_l(q\rho') = \tfrac{1}{2}\pi\rho^{-2}\delta(\rho'-\rho), \tag{4.3}$$

it follows that, in order for the states defined by Eq. (4.1) to satisfy Eq. (4.2), the functions $\Psi^l(q)$ must be, up to a phase,

$$\Psi^l(q) = (i)^{-l}(4\pi^{\frac{3}{2}})^{-1}[q^3/\epsilon(q)E]^{\frac{1}{2}}, \tag{4.4}$$

where $E = (M^2 + \mathbf{P}^2)^{\frac{1}{2}}$ is the total energy and $\epsilon(q)$ is the relativistic analog of the classical reduced mass, defined in the rest system by

$$\epsilon(q)^{-1} = (m_1^2 + q^2)^{-\frac{1}{2}} + (m_2^2 + q^2)^{-\frac{1}{2}} \tag{4.5}$$

and called the invariant "reduced energy."

The relative position states defined by Eq. (4.1) are then given in terms of 2-particle standard canonical states at time $t = 0$ by

$$|\mathbf{P}\rho(j\mu)ls\rangle = \int d\tilde{q}\left(\frac{q\epsilon}{\pi E}\right)^{\frac{1}{2}}j_l(q\rho)\,|P\mu(Mj)ls\rangle_c, \tag{4.6}$$

where $d\tilde{q} \equiv [q/\epsilon(q)]\,dq = dM$ is the Lorentz-invariant volume element, so that the corresponding "wave-function" distribution is

$$_c\langle P'\mu(Mj)ls \mid \mathbf{P}\rho(j\mu)ls\rangle = \langle \mathbf{P}\rho(j\mu)ls \mid P'\mu(Mj)ls\rangle_c = [q\epsilon/\pi E]^{\frac{1}{2}}j_l(q\rho)\delta(P',P). \tag{4.7}$$

The relative position states of Eq. (4.6) are, like the canonical states, eigenstates of $\mathbf{P}$, $\mathbf{J}^2$, J_z, $\mathbf{L}^2$, and $\mathbf{S}^2$ with corresponding eigenvalues $\mathbf{P}$, $j(j+1)$, μ, $l(l+1)$, and $s(s+1)$, but are clearly not eigenstates of the total mass M or energy $E = (M^2 + \mathbf{P}^2)^{\frac{1}{2}}$. The relative position states defined at time t are then given by the time-translated states

$$|\mathbf{P}\rho(j\mu)ls\rangle_t = T(t)\,|\mathbf{P}\rho(j\mu)ls\rangle = \int_0^\infty d\tilde{q}\left(\frac{q\epsilon}{\pi E}\right)^{\frac{1}{2}} j_l(q\rho)\exp(-iEt)\,|P\mu(Mj)ls\rangle_c. \quad (4.8)$$

As is the case with the Newton–Wigner localized states,[3] the relative position states of Eq. (4.6) do not transform covariantly under Lorentz transformations. This is seen from the mass dependence of the transformation matrix $D^j(\bar{R}(L, P))$ given in Eq. (3.13).

By using the orthogonality property of the spherical Bessel functions [Eq. (4.3)], the standard canonical states are seen to be related to the relative position states by

$$|P\,\mu(Mj)l\,s\rangle_c = 2\pi^{-\frac{1}{2}}(q\epsilon E)^{\frac{1}{2}}\int_0^\infty \rho^2\,d\rho j_l(q\rho)\,|\mathbf{P}\;\rho(j\mu)l\;s\rangle. \quad (4.9)$$

The "relative-position" operator for which the states of Eq. (4.6) are eigenstates is now readily found by first constructing the operator R^2 for which

$$R^2\,|\mathbf{P}\rho(j\mu)ls\rangle = \rho^2\,|\mathbf{P}\rho(j\mu)ls\rangle \quad (4.10)$$

for the states of Eq. (4.6) with fixed j, μ, l, and s. This implies that the corresponding operator $\hat{R}^2$ which operates in wavefunction space satisfies, by Eqs. (4.6) and (4.7),

$$\hat{R}^2\,{}_c\langle P'\,\mu(Mj)l\,s\,|\,\mathbf{P}\;\rho(j\mu)ls\rangle = \rho^2\,{}_c\langle P'\mu(Mj)ls\,|\,\mathbf{P}\;\rho(j\mu)l\;s\rangle = \rho^2(q\epsilon/\pi E)^{\frac{1}{2}} j_l(q\rho)\delta(P', P). \quad (4.11)$$

Upon noting that, from Bessel's equation,

$$\rho^2 j_l(q\rho) = \left(-q^{-2}\frac{d}{dq}q^2\frac{d}{dq} + \frac{l(l+1)}{q^2}\right)j_l(q\rho), \quad (4.12)$$

the operator $\hat{R}^2$ is seen to be

$$\hat{R}^2 = -(q\epsilon E)^{\frac{1}{2}}q^{-2}\frac{d}{dq}q^2\frac{d}{dq}(q\epsilon E)^{-\frac{1}{2}} + \frac{\mathbf{L}^2}{q^2}. \quad (4.13)$$

In terms of derivatives with respect to the total mass M, it then follows that

$$\hat{R} = [\hat{R}_0^2 + q^{-2}\mathbf{L}^2]^{\frac{1}{2}}, \quad (4.14)$$

where the operator $\hat{R}_0$ is defined by

$$\hat{R}_0 = -i\left(\frac{Eq}{\epsilon}\right)^{\frac{1}{2}}\frac{d}{dM}\left(\frac{q}{\epsilon E}\right)^{\frac{1}{2}} \quad (4.15)$$

and satisfies

$$[\hat{R}_0, M] = -i(q/\epsilon), \quad [\hat{R}_0, q] = -i. \quad (4.16)$$

The commutation rules of Eq. (4.16) suggest that the operator $\hat{R}_0$ may be taken to be the "component" of the relative position operator $\mathbf{R}$ in the direction of the rest-system relative momentum $\mathbf{q}$. In this case, the operator

$$S = (q^{-2}\mathbf{L}^2)^{\frac{1}{2}} \quad (4.17)$$

represents the magnitude of the component of the relative position operator transverse to $\mathbf{q}$, and so may be interpreted as an "impact parameter" operator in momentum space. This "impact parameter" operator corresponds exactly to the magnitude of the classical-impact-parameter vector $\mathbf{s}$ in the rest system, which satisfies

$$\mathbf{s}\cdot\mathbf{q} = 0, \quad \mathbf{L} = \mathbf{s}\times\mathbf{q}. \quad (4.18)$$

5. TWO-PARTICLE LOCALIZED STATES

Two-particle angular momentum states, which are localized with respect to the parameter space of the direct-product representation of the translation group,

$$t^{(1\times2)}(\mathbf{x}):(\mathbf{x}_1, \mathbf{x}_2) \to (\mathbf{x}_1 + \mathbf{x}, \mathbf{x}_2 + \mathbf{x}), \quad (5.1)$$

may now be constructed from the relative position states of Eq. (4.6). The set of localized states is defined at the origin by

$$|\mathbf{0}\rho(j\mu)ls\rangle_L = \int d^3P\varphi^l(\mathbf{P})\,|\mathbf{P}\rho(j\mu)ls\rangle, \quad (5.2)$$

and such states satisfy the Newton–Wigner[3] postulates of symmetry, localization, and continuity if the functions $\varphi^l(\mathbf{P})$ are chosen such that, if

$$|\mathbf{x}\rho(j\mu)ls\rangle_L = T^{(1\times2)}(\mathbf{x})\,|\mathbf{0}\rho(j\mu)ls\rangle_L = \int d^3P\varphi^l(\mathbf{P})\exp(i\mathbf{P}\cdot\mathbf{x})\,|P\rho(j\mu)ls\rangle, \quad (5.3)$$

then

$${}_L\langle\mathbf{x}'\rho'(j'\mu')l's'\,|\,\mathbf{x}\rho(j\mu)ls\rangle_L = \delta_{j'j}\delta_{\mu'\mu}\delta_{l'l}\delta_{s's}\rho^{-2}\delta(\rho' - \rho)\delta^3(\mathbf{x}' - \mathbf{x}). \quad (5.4)$$

Upon applying this "localization" criterion to the states of Eq. (5.2) and using Eq. (4.2), it follows that the functions φ^l must be constants and up to a phase chosen to be $(2\pi)^{-\frac{3}{2}}$ for correct normalization. Hence, the 2-particle angular momentum states which are localized with respect to both the overall "mean position" $\mathbf{x}$ and the magnitude of the rest-system relative position ρ are given by

$$|\mathbf{x}\rho(j\mu)ls\rangle_L = (2\pi)^{-\frac{3}{2}}\int d^3P\exp(i\mathbf{P}\cdot\mathbf{x})\,|\mathbf{P}\rho(j\mu)ls\rangle. \quad (5.5)$$

In terms of the 2-particle canonical states, it follows from Eq. (4.6) that

$$|\mathbf{x}\rho(j\mu)ls\rangle_{\mathrm{L}} = (2\pi)^{-\frac{3}{2}}\int d^3P\, d\tilde{q}\left(\frac{q\epsilon}{\pi E}\right)^{\frac{1}{2}} j_l(q\rho) \times \exp(i\mathbf{P}\cdot\mathbf{x})\,|P\mu(Mj)ls\rangle_{\mathrm{c}}, \quad (5.6)$$

where

$${}_{\mathrm{c}}\langle P\mu(Mj)ls \mid \mathbf{x}\rho(j\mu)ls\rangle_{\mathrm{L}} = (q\epsilon E/2\pi^4)^{\frac{1}{2}} j_l(q\rho)\exp(i\mathbf{P}\cdot\mathbf{x}). \quad (5.7)$$

The 2-particle localized states of Eq. (5.6) are clearly eigenstates of $\mathbf{J}^2$, J_z, $\mathbf{L}^2$, and $\mathbf{S}^2$ in addition to the relative position operator $R = (R^2)^{\frac{1}{2}}$ of Eq. (4.14). Furthermore, these states are eigenstates of a "mean position" operator X^i, $i = 1, 2, 3$, corresponding to the wavefunction operators for which

$$\hat{X}^i\,{}_{\mathrm{c}}\langle P\mu(Mj)ls \mid \mathbf{x}\rho(j\mu)ls\rangle_{\mathrm{L}} = x^i\,{}_{\mathrm{c}}\langle P\mu(Mj)ls \mid \mathbf{x}\rho(j\mu)ls\rangle_{\mathrm{L}}, \quad (5.8)$$

where, from Eq. (5.7), it follows that

$$\hat{X}^i = -\frac{1}{i}\left(\frac{\partial}{\partial P_i} + \frac{P^i}{2E^2}\right). \quad (5.9)$$

This position operator is of the same form as the 1-particle Newton–Wigner position operator in its canonical representation,[9] and satisfies the commutation rules

$$[\hat{X}^i, \hat{X}^j] = 0, \quad [\hat{X}^i, p^j] = -i\delta^{ij}. \quad (5.10)$$

6. SUMMARY AND CONCLUSIONS

In this paper an infinite number of sets of 2-particle canonical states have been constructed by first considering the effect of a relative displacement of rest-system standard canonical states and by then projecting out of these transformed states their respective orbital angular momentum components. Of particular interest was that set of states which were generated from canonical rest states with zero orbital angular momentum, for which the resultant orbital angular momentum was seen to be due entirely to the initial relative displacement.

By then postulating that a complete set of states localized to a zero relative position contains only zero orbital angular momentum components, a set of states localized with respect to the magnitude of the rest-system relative position were constructed, as was the corresponding relative position operator. Lastly, 2-particle angular momentum states were localized with respect to the "mean position" of the 2-particle system, and the corresponding position operator was seen to be similar in form to the 1-particle Newton–Wigner position operator in its canonical form.

The relative-position states discussed herein are to be distinguished from the relativistic impact-parameter states constructed by Chang and Raman.[11] In their case, impact parameter space is taken to be the parameter space of the 2-dimensional translation group which, in the rest system, acts in a plane which is orthogonal to the direction of the overall 3-momentum. The impact parameter states defined by Chang and Raman[11] in the rest system are

$$|\mathbf{b}\kappa\rangle = (2\pi)^{-1}\int d^2k_t(4\omega_1\omega_2)^{-\frac{1}{2}}\exp(i\mathbf{k}_t\cdot\mathbf{b})\,|\tilde{P}\mathbf{q}\rangle, \quad (6.1)$$

where $\mathbf{k}_t$ is the transverse (relative) momentum, κ is the z component of the relative momentum, and the impact parameter vector $\mathbf{b}$ is orthogonal to the direction of the over-all momentum. Such states may be thought of as the transverse projections of the localized states of Eq. (1.1) in the rest system. In the particular case for which the $\mathbf{P}$ direction is along the z axis, it follows that

$$\langle \mathbf{b}\kappa \mid \tilde{P}\boldsymbol{\rho}\rangle \sim e^{i\kappa\rho_z}\delta^2(\boldsymbol{\rho}_t - \mathbf{b}), \quad (6.2)$$

where ρ_t is the transverse component of the relative position vector $\boldsymbol{\rho}$ in the rest system.

On the other hand, the relative position states given by Eq. (4.6) are, in the rest system, the angular momentum components of those states of Eq. (1.1) upon which an integration of the form

$$\int d\varphi\, d(\cos\theta)\, Y_l{}^m(\theta, \varphi)$$

has been performed. Such states are, as in the case of Fq. (6.2), eigenstates of $\rho = (\mathbf{b}^2 + \rho_z^2)^{\frac{1}{2}}$. Whereas the relative-position states defined in the rest system by Eq. (1.1) do not readily admit a covariant generalization to arbitrary momentum $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$, it is believed that the procedure followed in this work has provided a meaningful relative-position representation for systems of arbitrary momentum.

An interesting feature of this development is that a quantum-mechanical analog of the classical impact parameter has arisen in a natural way as a consequence of localization with respect to the magnitude of the rest-system relative position. As opposed to the impact parameter interpretation given by Chang and Raman,[11] the impact parameter operator given by Eq. (4.17) has as its classical limit the magnitude of that component of the rest-system relative position which is orthogonal to the relative momentum. Indeed, it appears that, if one requires that the correspondence principle be satisfied, it is unlikely that the corresponding impact parameter "vector" operator $\mathbf{S}$ arises directly via a Newton–Wigner[3] type of localization procedure in the rest system. Accordingly, one would expect $\mathbf{S}$ to be a nonlocal operator, in the sense that $[S^i, S^j] \neq 0$.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor D. L. Pursey for many valuable conversations and to Dr. A. McKerrell for a critical reading of the manuscript.

* This work was performed in the Ames Laboratory of the U.S. Atomic Energy Commission, Contribution No. 2680.

† Present address: Department of Applied Mathematics, The University, P.O. Box 147, Liverpool, L69 3BX, United Kingdom.

[1] M. H. L. Pryce, Proc. Roy. Soc. (London) **195A**, 62 (1948).
[2] C. Møller, Dublin Institute for Advanced Studies, **A**, No. 5, 1949 (unpublished).
[3] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[4] G. N. Fleming, Phys. Rev. **137**, B188 (1965).
[5] L. L. Foldy, Phys. Rev. **102**, 568 (1956).
[6] A. J. Macfarlane, J. Math. Phys. **4**, 490 (1963).
[7] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1967).
[8] A. McKerrell, Nuovo Cimento **34**, 1289 (1964).
[9] S. C. McDonald, J. Math. Phys. **11**, 1558 (1970).
[10] E. P. Wigner, *Group Theory* (Academic, New York, 1964), 2nd ed.
[11] N. P. Chang and K. Raman, Phys. Rev. **181**, 2048 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Pure Thermodynamical Phases as Extremal KMS States

GÉRARD G. EMCH AND HUBERT J. F. KNOPS*
Department of Physics and Astronomy, The University of Rochester, Rochester, New York 14627

(Received 9 April 1970)

We compare the dynamical characterization of pure thermodynamical phases as extremal KMS states and their characterization as extremal time- or space-invariant states. We find that, for a class of Weiss–Ising models with periodic potentials, the extremal KMS states coincide exactly with the solutions of the self-consistency equations familiar from molecular field methods. We show that the models considered are not η-asymptotically Abelian in time. We conclude that the characterization of pure thermodynamical phases as extremal KMS states is the only correct one for these models. We pay special attention (in particular, in the decomposition of an arbitrary KMS state into its extremal KMS components) to the fact that the time evolution is not an automorphism of the C^*-algebra of the quasilocal observables.

1. INTRODUCTION

The Kubo–Martin–Schwinger (KMS) boundary condition, first discovered as an analyticity property of thermal Green's functions,[1] was introduced in the C^*-algebraic approach to statistical mechanics by Haag, Hugenholtz, and Winnink.[2] This condition has proved itself to be such a useful and elegant tool that many speculations have since appeared about its interpretation in this formalism.[3]

The aim of the present paper is to confront these speculations—and, more specifically, the dynamical characterization of thermodynamical pure phases as extremal KMS states[4]—with what actually happens in some exactly solvable models which exhibit a phase transition associated with a spontaneous symmetry breaking.

We consider for this purpose a slightly generalized version of the Weiss theory of ferromagnetism and antiferromagnetism, which we define precisely and treat with the conventional methods of statistical mechanics in Sec. 2. Everything which could possibly be said about the didactic value and pitfalls of this type of naive model has been repeatedly expounded in the literature[5]; we shall henceforth not come back to this and just accept these models as concrete examples on which one can test more abstract theories.

Since, on the one hand, we want to check, for the explicit models, whether the thermodynamical pure phases can be characterized as extremal KMS states and since, on the other hand, the KMS condition involves in an essential way the time evolution, we first have to define the time evolution in the limit where infinite systems are considered. Our Sec. 3 is devoted to this preliminary question, which we answer in three successive steps: The first one involves a convergence assumption which we later prove to hold true for the KMS states of our models; to avoid circular reasoning, the second step bypasses this assumption; the third step is concerned with the particular case of time-invariant states. The class of potentials for which these results are obtained contains, in particular, the case discussed in Sec. 2.

The time evolution obtained in Sec. 3 happens to be of a slightly more general type than that allowed by the usual assumptions of the C^*-algebraic approach; specifically, the time evolution in the van der Waals limit is no longer an automorphism of the C^*-algebra of the quasilocal observables. In Sec. 4 we make the necessary accommodations to take this fact into account, and we establish in this context the essential properties of KMS states as well as their decomposition into extremal KMS states. A complete characterization of extremal KMS states is also obtained in this section in the form of self-consistency equations which

ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor D. L. Pursey for many valuable conversations and to Dr. A. McKerrell for a critical reading of the manuscript.

* This work was performed in the Ames Laboratory of the U.S. Atomic Energy Commission, Contribution No. 2680.

† Present address: Department of Applied Mathematics, The University, P.O. Box 147, Liverpool, L69 3BX, United Kingdom.

[1] M. H. L. Pryce, Proc. Roy. Soc. (London) **195A**, 62 (1948).
[2] C. Møller, Dublin Institute for Advanced Studies, **A**, No. 5, 1949 (unpublished).
[3] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[4] G. N. Fleming, Phys. Rev. **137**, B188 (1965).
[5] L. L. Foldy, Phys. Rev. **102**, 568 (1956).
[6] A. J. Macfarlane, J. Math. Phys. **4**, 490 (1963).
[7] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1967).
[8] A. McKerrell, Nuovo Cimento **34**, 1289 (1964).
[9] S. C. McDonald, J. Math. Phys. **11**, 1558 (1970).
[10] E. P. Wigner, *Group Theory* (Academic, New York, 1964), 2nd ed.
[11] N. P. Chang and K. Raman, Phys. Rev. **181**, 2048 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Pure Thermodynamical Phases as Extremal KMS States

GÉRARD G. EMCH AND HUBERT J. F. KNOPS*
Department of Physics and Astronomy, The University of Rochester, Rochester, New York 14627

(Received 9 April 1970)

We compare the dynamical characterization of pure thermodynamical phases as extremal KMS states and their characterization as extremal time- or space-invariant states. We find that, for a class of Weiss–Ising models with periodic potentials, the extremal KMS states coincide exactly with the solutions of the self-consistency equations familiar from molecular field methods. We show that the models considered are not η-asymptotically Abelian in time. We conclude that the characterization of pure thermodynamical phases as extremal KMS states is the only correct one for these models. We pay special attention (in particular, in the decomposition of an arbitrary KMS state into its extremal KMS components) to the fact that the time evolution is not an automorphism of the C^*-algebra of the quasilocal observables.

1. INTRODUCTION

The Kubo–Martin–Schwinger (KMS) boundary condition, first discovered as an analyticity property of thermal Green's functions,[1] was introduced in the C^*-algebraic approach to statistical mechanics by Haag, Hugenholtz, and Winnink.[2] This condition has proved itself to be such a useful and elegant tool that many speculations have since appeared about its interpretation in this formalism.[3]

The aim of the present paper is to confront these speculations—and, more specifically, the dynamical characterization of thermodynamical pure phases as extremal KMS states[4]—with what actually happens in some exactly solvable models which exhibit a phase transition associated with a spontaneous symmetry breaking.

We consider for this purpose a slightly generalized version of the Weiss theory of ferromagnetism and antiferromagnetism, which we define precisely and treat with the conventional methods of statistical mechanics in Sec. 2. Everything which could possibly be said about the didactic value and pitfalls of this type of naive model has been repeatedly expounded in the literature[5]; we shall henceforth not come back to this and just accept these models as concrete examples on which one can test more abstract theories.

Since, on the one hand, we want to check, for the explicit models, whether the thermodynamical pure phases can be characterized as extremal KMS states and since, on the other hand, the KMS condition involves in an essential way the time evolution, we first have to define the time evolution in the limit where infinite systems are considered. Our Sec. 3 is devoted to this preliminary question, which we answer in three successive steps: The first one involves a convergence assumption which we later prove to hold true for the KMS states of our models; to avoid circular reasoning, the second step bypasses this assumption; the third step is concerned with the particular case of time-invariant states. The class of potentials for which these results are obtained contains, in particular, the case discussed in Sec. 2.

The time evolution obtained in Sec. 3 happens to be of a slightly more general type than that allowed by the usual assumptions of the C^*-algebraic approach; specifically, the time evolution in the van der Waals limit is no longer an automorphism of the C^*-algebra of the quasilocal observables. In Sec. 4 we make the necessary accommodations to take this fact into account, and we establish in this context the essential properties of KMS states as well as their decomposition into extremal KMS states. A complete characterization of extremal KMS states is also obtained in this section in the form of self-consistency equations which

we then compare to those obtained in Sec. 2, thus establishing the desired connection between thermodynamical pure phases and extremal KMS states.

Section 5 contains our conclusions and some possible extensions of our methods. Six short appendices deal with the more technical computations and proofs which would otherwise unnecessarily disrupt the main line of the argumentation.

2. VARIATIONS ON A THEME BY KAC

In his Brandeis lectures[6] (see also Refs. 7), Kac showed that the molecular field method gives the correct thermodynamical behavior of the Weiss model. The elegance of his proof is based on the observation of an elementary property of the exponential function with square argument. We show in Appendix A that Kac's observation can actually be generalized quite simply to the case where the argument of the exponential is a positive quadratic form, thus enabling, by a method paralleling closely that of Kac, treatment of the case of a periodic Weiss–Ising interaction. We also compute the higher equilibrium space correlations. These natural generalizations of Kac's idea constitute the material of this section and are thus derived in the most orthodox spirit of traditional statistical mechanics. All results are obtained, *without approximations*, in the thermodynamical limit.

Let $\mathsf{N} = \{0, 1, \cdots, N-1\}$ be a lattice, each point of which is occupied by a spin σ_i. The energy of a configuration $\{\mu\}$ is

$$E(\mathsf{N}, \{\mu\}) = -B\sum_{i\in\mathsf{N}}\mu_i - \sum_{i,j\in\mathsf{N}}\mu_i v_{ij}(\mathsf{N})\mu_j .$$

We assume $v_{ij}(\mathsf{N})$ to be symmetric [$v_{ij}(\mathsf{N}) = v_{ji}(\mathsf{N})$], real, translation-invariant [$v_{i+n,j+n}(\mathsf{N}) = v_{ij}(\mathsf{N})$ with cyclic boundary conditions] so that $v_{ij}(\mathsf{N}) = v_{|i-j|}(\mathsf{N})$, and periodic of period p [$v_{n+p}(\mathsf{N}) = v_n(\mathsf{N})$]; for simplicity, we write $N = pZ$, where Z is an integer. To keep the energy per particle finite in the thermodynamical limit ($Z \to \infty$), we assume further $v_n(\mathsf{N}) = f(n)/Z$. We define $\mathsf{P} = \{0, 1, \cdots, p-1\}$ and the sublattice observable M_j for each j in P as

$$M_j = \sum_m \mu_{j+mp} .$$

We can then write

$$E(\mathsf{N}, \{\mu\}) = -B\sum_{i\in\mathsf{P}} M_i - \sum_{i,j\in\mathsf{P}} M_i v_{ij}(\mathsf{N}) M_j .$$

We assume finally that the matrix (f_{ij}) of rank p, defined above from the interaction $v_{ij}(\mathsf{N})$, is positive.

Using the lemma of Appendix A, we see that the partition function for our system is

$$\begin{aligned} Q(B, \beta, Z) &= \sum_{\{\mu\}} e^{-\beta E(N,\{\mu\})} \\ &= \prod_{i\in\mathsf{P}} (2\beta Z\lambda_i/2\pi)^{\frac{1}{2}} \int_{-\infty}^{+\infty} d\zeta_1 \cdots \int_{-\infty}^{+\infty} d\zeta_p \\ &\quad \times \left\{ \exp\left(-\beta \sum_{i,j} \zeta_i f_{ij} \zeta_j\right) \right. \\ &\quad \left. \times \prod_{i\in\mathsf{P}} 2\cosh\left[\beta\left(B + 2\sum_{j\in\mathsf{P}} f_{ij}\zeta_j\right)\right] \right\}^Z \end{aligned}$$

(where the λ_i are the eigenvalues of the matrix f_{ij}). As the thermodynamical limit ($Z \to \infty$; p, B, and β fixed) is approached, the integrand becomes more and more peaked, and the familiar steepest descent method gives an asymptotic expansion of the partition function, the first term of which is proportional to

$$\begin{aligned} &2^{pZ} Z^{p/2} \max_{(\zeta_1,\cdots,\zeta_p)} \left\{ \exp\left(-\beta \sum_{i,j\in\mathsf{P}} \zeta_i f_{ij} \zeta_j\right) \right. \\ &\qquad \left. \times \prod_{i\in\mathsf{P}} \cosh\left[\beta\left(B + 2\sum_{j\in\mathsf{P}} f_{ij}\zeta_j\right)\right] \right\}^Z . \end{aligned}$$

The position of the stationary points of the integrand is then obtained in an implicit form, as the self-consistency equations

$$B_i = 2\sum_{j\in\mathsf{P}} f_{ij} \tanh\left[\beta(B + B_j)\right],$$

where we write

$$B_i = 2\sum_{j\in\mathsf{P}} f_{ij}\zeta_j .$$

These self-consistency equations, considered by themselves, only determine the stationary points of the integrand, whereas only the maxima of the integrand with greatest magnitude contribute to the leading term of the integral. Due to the transcendental character of the expression at hand, the selection of these maxima from among the multiplicity of the solutions of the self-consistency equations might turn out to be a somewhat cumbersome task in the general case; we shall therefore assume that this selection (always possible in principle when an explicit choice of f_{ij} is made) is achieved, and we shall use the index set $\Lambda(B, \beta)$ to label the maxima corresponding to the largest value of the integrand.

We now notice that the trick used above to compute the partition function can be used all the same for the computation of the expectation value of any local observable A_Ω in the thermodynamical canonical equilibrium state. Since these observables are finite linear combinations of observables of the form $A_\Omega = \prod_{i\in\Omega} A_i$, it is sufficient to compute the expectation value of the latter. For those observables, one finds,

in the thermodynamical limit,

$$\langle A_\Omega \rangle_{B,\beta} = \sum_{\lambda \in \Lambda(B,\beta)} \alpha_\lambda \langle A_\Omega \rangle_\lambda ,$$

where α_λ corresponds to the relative weight of the maximum λ, as determined by the steepest descent method,

$$\langle A_\Omega \rangle_\lambda = \prod_{i \in \Omega} \langle A_i \rangle_\lambda$$

and

$$\langle A_i \rangle_\lambda = \left(\sum_{\mu=\pm 1} e^{\beta(B+B_{i,\lambda})\mu} \right)^{-1} \sum_{\mu=\pm 1} e^{\beta(B+B_{i,\lambda})\mu} A_i(\mu).$$

Consequently, we recognize $(B + B_{i,\lambda})$ as the effective magnetic field acting at site i to give

$$\langle \mu_i \rangle_\lambda = \tanh [\beta(B + B_{i,\lambda})].$$

Since the $B_{i,\lambda}$ only depend on λ and on the sublattice $\{i + mp\}$, we write

$$m_i = \lim_{Z \to \infty} Z^{-1} M_i ,$$

and then $\langle m_i \rangle_\lambda = \langle \mu_i \rangle_\lambda$ is the magnetization of this sublattice, characteristic of $\lambda \in \Lambda(B, \beta)$. We then have

$$B_{i,\lambda} = 2 \sum_{j \in \mathsf{P}} f_{ij} \langle m_j \rangle_\lambda ,$$

and we can rewrite the self-consistency equations in the familiar form of the usual equations of state relating the applied magnetic field B and the sublattice magnetizations $\langle m_i \rangle_\lambda$:

$$\langle m_i \rangle_\lambda = \tanh \left[\beta \left(B + 2 \sum_{j \in \mathsf{P}} f_{ij} \langle m_j \rangle_\lambda \right) \right].$$

From this expression we see that the free energy per site, defined as

$$-\beta f(B, \beta) = \lim_{Z \to \infty} (pZ)^{-1} \ln Q(B, Z),$$

can be written as

$$f = p^{-1} \sum_{j \in \mathsf{P}} f_j ,$$

with

$$-\beta f_j = \ln 2 - \tfrac{1}{2} \beta B_j \langle m_j \rangle - \tfrac{1}{2} \ln (1 - \langle m_j \rangle^2),$$

which is the same for every $\lambda \in \Lambda(B, \beta)$.

The relations obtained in this section show that a natural extension of Kac's trick[6] allows us to establish, in the thermodynamical limit, the exact validity of the molecular field method for a periodic Weiss potential. In particular, we can interpret $\Lambda(B, \beta)$ as labeling the pure thermodynamical phases, the prescribed mixture of which constitutes the canonical equilibrium state. In this respect, we might mention that all the pure thermodynamical phases which are obtained from one another by translation ($B_j \to B_{j+n}$) or, when $B = 0$, by flip-flop ($B_j \to -B_j$) naturally occur in the canonical equilibrium state with the same weight since the integrand of the partition function is invariant under these transformations, thus reflecting in the canonical equilibrium state the symmetry of the original Hamiltonian.

3. TIME EVOLUTION

In the classical treatment presented in Sec. 2, we could ignore the fact that, at each site j of the lattice, sits a full spin $\boldsymbol{\sigma}_j$ and not only a "classical spin" σ_j^z. This reflects the fact that the time evolution plays no role in this classical treatment of Ising-like models. However, in view of the fact that the emphasis in the present paper is on a *dynamical* characterization of pure thermodynamical phases, we should be prepared to allow for situations where the above simplification is not allowed any more.

We first consider the time evolution for a finite system. To define the time evolution in the van der Waals limit (i.e., infinite volume, infinitely long-range and infinitely weak interactions), we proceed in three successive steps. In the first step we consider a particular case where the van der Waals limit can be carried out explicitly, leading to a Hamiltonian, the interpretation of which provides a rigorous basis for the molecular field method; we show, under the assumptions of this particular case, the existence of a mean free field acting at each point of the lattice. This field is due to the collective action of all the other spins on the lattice. To free ourselves from the necessity of assuming the conditions under which this first step can be carried out, we introduce in a second step a mathematically legitimate construct (the generalized mean $\bar{r}_l$) which properly allows us to define a time evolution; this construction, however, suffers from the physical disadvantage of lacking the uniqueness we would expect from a model as simple as that considered here. The third step, carried out after the material of Sec. 4 has been expounded, is contained in Appendix D, where we *prove*, using this intermediary step, that the assumptions of the first step are indeed justified and that, consequently, the ambiguity introduced in the second step disappears in the final result. This section ends with a discussion of the time-invariant states.

The physical system we consider is therefore an infinite lattice, each site of which is occupied by a quantum spin $\boldsymbol{\sigma}_j$. For the sake of simplicity in the notation, we assume this lattice to be 1 dimensional, although this restriction can be trivially lifted; let us henceforth index the sites of our lattice by j in Z, where Z is the set of all integers. With each site j on the lattice, we associate the algebra $\mathcal{A}(j)$ of all complex 2×2 matrices, i.e.,

$$\mathcal{A}(j) = \left\{ \sum_{\mu=0}^{3} a_\mu \sigma_j^\mu \;\middle|\; a_\mu \in \mathsf{C} \right\},$$

where σ^0 is the identity matrix and σ^i, $i = 1, 2, 3$, are the three Pauli matrices. The algebra $\mathcal{A}(\Omega)$ of all local observables associated with a finite region $\Omega \subset \mathbf{Z}$ is then obtained as the set of all finite linear combinations of all finite tensor products of the form $\otimes_{j\in\Omega} A_j$ with A_j in $\mathcal{A}(j)$. A structure of C^*-algebra is imposed on $\mathcal{A}(\Omega)$ by the usual rules of addition, multiplication, Hermitian conjugation, and norm of finite-rank matrices. With the canonical embedding

$$A_j \to \cdots \otimes I \otimes I \otimes A_j \otimes I \otimes I \otimes \cdots,$$

$\mathcal{A}(\Omega_1)$ is clearly a sub-C^*-algebra of $\mathcal{A}(\Omega_2)$ whenever Ω_1 is contained in Ω_2; this property is referred to as "isotony." "Locality" is then expressed by

$$[\mathcal{A}(\Omega_1), \mathcal{A}(\Omega_2)] = 0 \quad \text{whenever} \quad \Omega_1 \cap \Omega_2 = \phi,$$

which makes sense via the canonical embedding of $\mathcal{A}(\Omega_i)$ into $\mathcal{A}(\Omega_1 \cup \Omega_2)$. The algebra $\mathcal{A}$ of all quasilocal observables of our system is finally defined, via the canonical embedding defined above, as the completion with respect to the norm of the union of all $\mathcal{A}(\Omega)$:

$$\mathcal{A} = \overline{\bigcup_{\Omega\subset \mathbf{Z}} \mathcal{A}(\Omega)}.$$

This last step, referred to as the C^*-inductive limit, has been expounded with great mathematical precision by Guichardet[8]; the naive description given above will, however, already suffice for our purpose.

We next define the time evolution via the following limiting process. With every finite region $\Omega \subset \mathbf{Z}$, we associate $H(\Omega) \in \mathcal{A}(\Omega)$:

$$H(\Omega) = -B\sum_{i\in\Omega} \sigma_i^z - \sum_{i,j\in\Omega} \nu_{ij}(\Omega)\sigma_i^z\sigma_j^z,$$

where, for the time being, we only assume that the $\nu_{ij}(\Omega)$ are real couplings depending on Ω, with

$$\nu_{ij}(\Omega) = \nu_{ji}(\Omega), \quad \nu_{ii}(\Omega) = 0.$$

Since the Hamiltonian $H(\Omega)$ is bounded, it defines a time evolution on $\mathcal{A}$, via the formula

$$\alpha^t(\Omega) = \sum_{n=0}^{\infty} \frac{t^n}{n!} L(\Omega)^n,$$

where the Liouville operator $L(\Omega)$, acting on $\mathcal{A}$, is defined by

$$L(\Omega)[A] = -i[H(\Omega), A] \quad \text{for all } A \text{ in } \mathcal{A}.$$

We now observe that the action of $\alpha^t(\Omega)$ on $\mathcal{A}$ is entirely determined by its action on the various $A_l \in \mathcal{A}(l)$, since, first, $\alpha^t(\Omega)$ is an automorphism of $\mathcal{A}$ and, second, $\mathcal{A}$ is generated, as a C^*-algebra, by the A_l. On those particular elements, $\alpha^t(\Omega)$ takes the following simple form (for $l \in \Omega$):

$$\alpha^t(\Omega)[A_l] = \sum_{n=0}^{\infty} \frac{t^n}{n!} L_l(\Omega)^n[A_l],$$

where the local Liouville operator $L_l(\Omega)$, acting on $\mathcal{A}$, is defined by

$$L_l(\Omega)[A] = -i[H_l(\Omega), A], \quad \text{for all } A \text{ in } \mathcal{A},$$

in which enters the local Hamiltonian

$$H_l(\Omega) = -\left(B + 2\sum_{j\in\Omega} \nu_{lj}(\Omega)\sigma_j^z\right)\sigma_l^z;$$

we use in the sequel the abbreviation

$$B_l(\Omega) = 2\sum_{j\in\Omega} \nu_{l,j}(\Omega)\sigma_j^z.$$

We now eliminate the volume dependence Ω in the time evolution by passing to the limit $\Omega \to \infty$; to do this, we need some further restrictions, the strength of which will be discussed later.

We first assume that the couplings $\nu_{ij}(\Omega)$ are such that for each i in $\mathbf{Z}$ there exists a positive constant c_i such that

$$\sum_{j\in\Omega} |\nu_{ij}(\Omega)| \leq \tfrac{1}{2}c_i$$

for all finite subsets Ω of Z containing i; we further assume that

$$\lim_{\Omega\to\infty} \nu_{ij}(\Omega) = 0, \quad \text{for all fixed } i, j \in \mathbf{Z}.$$

These assumptions (which we will refer to, respectively, as the "stability condition" and the "van der Waals condition") are rather benign in the context of mean free field methods; the first of these assumptions implies that, for every i in $\mathbf{Z}$, $\|B_i(\Omega)\| \leq c_i$ for all finite Ω containing i, with c_i *independent* of Ω.

We now restrict our attention to the study of those particular representations π of $\mathcal{A}$ for which $B_{l,\pi}(\Omega) \equiv \pi(B_l(\Omega))$ converges strongly as Ω goes to infinity; this restriction means that we assume the existence of a bounded linear operator $B_{l,\pi}$ acting on the representation space $\mathcal{H}_\pi$, such that for every $\epsilon > 0$ and ψ in $\mathcal{H}_\pi$ there exists a finite integer $N(\epsilon, \psi)$ such that

$$\|(B_{l,\pi}(\Omega) - B_{l,\pi})\psi\| \leq \epsilon$$

for every finite interval $\Omega = (l - N, l + N)$ with $N \geq N(\epsilon, \psi)$.

The fact that $B_{l,\pi}$ is obtained as the strong limit of $B_{l,\pi}(\Omega)$ in $\pi(\mathcal{A})$ implies first of all that $B_{l,\pi}$ belongs to $\pi(\mathcal{A})''$; furthermore, together with the "van der Waals condition," this implies that $B_{l,\pi}$ commutes with all $\pi(A_k)$ $(k \in \mathbf{Z})$ and hence belongs to $\pi(\mathcal{A})'$. Hence, $B_{l,\pi}$ belongs to the center $\pi(\mathcal{A})'' \cap \pi(\mathcal{A})'$ of $\pi(\mathcal{A})''$.

Under these conditions we prove in Appendix B that, for every local quantity $A \in \mathcal{A}(\Omega_0)$ with Ω_0 arbitrary finite subset of Z and every fixed time t, $\pi(\alpha^t(\Omega)[A])$ (with $\Omega \supseteq \Omega_0$) converges in the strong operator topology, as $\Omega \to \infty$, to an element $\alpha_\pi^t[\pi(A)]$ of $\pi(\mathcal{A})''$. We further show that α_π^t preserves the norm of $\pi(A)$ and can hence be extended to $\pi(\mathcal{A})$; finally, we show that, for every finite subset $\Omega \subset Z$, the mapping α_π^t is unitarily implemented from an effective Hamiltonian

$$H_{\Omega,\pi}(\infty) = -\sum_{l\in\Omega}(B + B_{l,\pi})\pi(\sigma_l^z)$$

[belonging to $\pi(\mathcal{A})''$] which therefore generates the actual time evolution (in the van der Waals limit) of any *local* quantity $A \in \mathcal{A}(\Omega)$. Whereas

$$U_{\Omega,\pi}^t \equiv \exp[-iH_{\Omega,\pi}(\infty)t]$$

clearly belongs to $\pi(\mathcal{A})''$, the *global* GNS U_π^t, constructed in Appendix C, does *not* belong to $\pi(\mathcal{A})''$; this is in agreement with the results of Haag, Hugenholtz,and Winnink (Ref. 2; see especially p. 233).

For mathematical convenience we now want to generalize the preceding discussion to the cases where $B_{l,\pi}(\Omega)$ does not necessarily converge strongly as $\Omega \to \infty$. To achieve this, we consider the C^*-algebra $\mathcal{C}(Z)$ of all bounded functions on Z, equipped with the sup-norm; let $\mathcal{C}_v(Z)$ be the linear manifold of all $f \in \mathcal{C}(Z)$ for which

$$\nu_l[f] \equiv \lim_{\substack{\Omega\to\infty \\ \Omega\ni l}} 2\sum_{j\in\Omega}\nu_{lj}(\Omega)f(j)$$

exists; ν_l is then clearly a positive bounded linear functional on $\mathcal{C}_v(Z)$. By the Hahn–Banach theorem, we can then extend it to a positive bounded linear functional $\tilde{\nu}_l$ on $\mathcal{C}(Z)$. This extension is evidently not unique, but we shall see that this nonuniqueness actually neither hampers our discussion nor alters the generality of our final results; indeed, we shall prove the KMS states have a time evolution which is actually determined by ν_l and therefore unique, since all quantities entering its definition belong to $\mathcal{C}_v(Z)$. For every representation π of $\mathcal{A}$ and every fixed observable $A \in B(\mathbb{C}^2)$, we define $\tilde{\nu}_{l,\pi}[A_j]$ in $\pi(\mathcal{A})'' \cap \pi(\mathcal{A})'$ by

$$(\Phi, \tilde{\nu}_{l,\pi}[A_j]\Psi) \equiv \tilde{\nu}_l[(\Phi, \pi(A_j)\Psi], \quad \text{for all } \Phi, \Psi \in \mathcal{H}_\pi .$$

In particular, we then write $B_{l,\pi} = \tilde{\nu}_{l,\pi}[\sigma_j^z]$, and we notice that this new $B_{l,\pi}$ coincides with the operator designated previously by this symbol whenever $B_{l,\pi}(\Omega)$ converges strongly as Ω goes to infinity.

We then generalize our previous local Hamiltonian $H_{\Omega,\pi}(\infty)$ by replacing in it our previous $B_{l,\pi}$ by their generalized versions, just defined. As before, these Hamiltonians generate a time evolution for all local observables A belonging to the corresponding $\pi(\mathcal{A}(\Omega))$. We denote by α_π^t this evolution, which again maps each $\pi(\mathcal{A}(\Omega))$ into $\pi(\mathcal{A})''$.

The mathematical structure of the time evolution, as developed up to this point, differs in several ways from that postulated in the usual algebraic approach. First, the time evolution α_π^t is defined for each representation separately. Second, α_π^t is locally unitarily implemented. Third, whereas the domain of α_π^t is restricted to the (quasi-) local observables, its range cannot be shown, in general, to be restricted to $\pi(\mathcal{A})$ but only to $\pi(\mathcal{A})''$, and it is therefore not an automorphism of $\mathcal{A}$. We shall see later that these differences cannot be due to a faulty or awkward approach to the thermodynamical limit, but are indeed strongly linked to the existence, in the van der Waals limit, of several different thermodynamical phases.

We now want to show that, in the case of time-invariant states, the time evolution can be extended to an automorphism of $\pi(\mathcal{A})''$.

Given any state ϕ on $\mathcal{A}$, we denote by π the GNS representation associated to it and by $\tilde{\phi}$ the canonical extension of ϕ from $\mathcal{A}$ to $\pi(\mathcal{A})''$ defined by

$$\langle\tilde{\phi}; A\rangle = (\Phi, A\Phi), \quad \text{for all } A \text{ in } \pi(\mathcal{A})'',$$

where Φ is the cyclic vector in $\mathcal{H}_\pi$ corresponding to ϕ. We now say that a state ϕ on $\mathcal{A}$ is time invariant whenever $\langle\tilde{\phi}; \alpha_\pi^t[\pi(A)]\rangle$ is a constant in t for every (quasi-) local observable A in $\mathcal{A}$. Although this condition is weaker than the usual one, we can still prove (see Appendix C) that there then exists a unitary operator U_π^t on $\mathcal{H}_\pi$ [not to be confused with the various $U_{\Omega,\pi}^t$ defined by the local Hamiltonians $H_{\Omega,\pi}(\infty)$]: (i) $U_\pi^t\Phi = \Phi$ and (ii) $\tilde{\alpha}_\pi^t[A] \equiv U_\pi^t A U_\pi^{-t}$, for all A in $\pi(\mathcal{A})''$, is an extension of α_π^t to an automorphism of the von Neumann algebra $\pi(\mathcal{A})''$.

4. KMS STATES

We now want to check whether the dynamical characterization of pure thermodynamical phases suggested in a previous paper[4] makes sense in the specific model studied in Sec. 2 by means of traditional methods.

We recall that, in order to get a consistent theory of symmetry breaking (including crystallization), we were led[4] to assume that pure thermodynamical phases are to be identified, in the C^*-algebraic approach to statistical mechanics, as extremal KMS states. In view of the character of the time evolution, as discussed in the previous section, we rephrase the fundamental Kubo–Martin–Schwinger boundary condition familiar in the treatment of thermal Green's functions as

follows: A state ϕ on the algebra $\mathcal{A}$ of quasilocal observables is said to be KMS for the natural temperature $\beta = (1/kT)$ if the two-time correlation functions associated with it satisfy the condition

$$\int dt f_\beta(t)\langle\tilde{\phi};\, \pi(C)\alpha_\pi^t[\pi(A)]\rangle = \int dt f_0(t)\langle\tilde{\phi};\, \alpha_\pi^t[\pi(A)]\pi(C)\rangle \quad (4.1)$$

for all finite subsets Ω and Ω' of Z, all A and C in $\mathcal{A}(\Omega)$ and $\mathcal{A}(\Omega')$, respectively, and all f of the form

$$f_\gamma(t) = \int d\omega \hat{f}(\omega) e^{-i\omega(t+i\gamma)}, \quad \gamma = 0 \text{ or } \beta,$$

where $\hat{f}$ runs over the space of all infinitely differentiable functions with compact support.

Although this version of the KMS boundary condition is slightly weaker than the usual one (see, for instance, Ref. 4 and references quoted therein), it still keeps the essential strength of the latter. We indeed first notice that ϕ KMS on $\mathcal{A}$ (in the present sense) implies by the usual argument (e.g., via Bochner's theorem on measures obtained as Fourier transforms of continuous, bounded functions of positive type) that ϕ is time invariant (in the sense of Sec. 3); in conjunction with the last results of Sec. 3, this is sufficient to show that $\tilde{\phi}$ is KMS (in the usual sense) on $\pi(\mathcal{A})''$ with respect to $\tilde{\alpha}_\pi^t$ and that a KMS state ψ on $\mathcal{A}$ (in the present sense) is extremal KMS if and only if $\pi(\mathcal{A})''$ is a factor.[9]

In view of these results, we can now immediately compute the extremal KMS states for our model. We recall that $B_{l,\pi}$, as defined in Sec. 3, belongs to the center of $\pi(\mathcal{A})''$ and that this center is trivial whenever π is the GNS representation associated with an extremal KMS state ψ on $\mathcal{A}$; let us denote by $B_{l,\psi}$ the value of the c-number $B_{l,\pi}$, thus associated to every extremal KMS state ψ on $\mathcal{A}$. Since $B_{l,\psi}$ is a c-number, α_π^t is not only an automorphism of the von Neumann algebra $\pi(\mathcal{A})''$ associated with ψ, but moreover its restrictions to $\pi(\mathcal{A})$ and to $\pi(\mathcal{A}(\Omega))$ (where Ω is any finite subset of Z) are also automorphisms of these C^*-algebras. We now recall that $\mathcal{A}(\Omega)$ is the algebra of all bounded operators on a finite-dimensional Hilbert space, so that all its representations are faithful; α_π^t hence defines an automorphism of $\mathcal{A}(\Omega)$ which is generated by the Hamiltonian

$$H_{\Omega,\psi}(\infty) = -\sum_{l\in\Omega}(B + B_{l,\psi})\sigma_l^z.$$

Every state on $\mathcal{A}(\Omega)$ is *normal* since $\mathcal{A}(\Omega)$ is finite dimensional; this is in particular true for the restriction ψ_Ω of ψ to $\mathcal{A}(\Omega)$. Furthermore, ψ_Ω is KMS with respect to the time evolution generated by $H_{\Omega,\psi}(\infty)$ on $\mathcal{A}(\Omega)$; finally, $\mathcal{A}(\Omega)$ is irreducible. The density matrix representing ψ_Ω is then uniquely determined, and so is, therefore, ψ_Ω; we can thus conclude that ψ extremal KMS implies the following for each A in $\mathcal{A}(\Omega)$:

$$\langle\psi;\, A\rangle = \mathrm{Tr}\ \rho(\Omega)A, \quad (4.2)$$

where

$$\rho(\Omega) = \{\mathrm{Tr}\exp[-\beta H_{\Omega,\psi}(\infty)]\}^{-1}\exp[-\beta H_{\Omega,\psi}(\infty)].$$

This implies immediately the following four consequences [Eqs. (4.3)–(4.6)]:

$$\psi \text{ is } \textit{not} \text{ extremal time-invariant} \quad (4.3)$$

since we can, for instance, write $\psi = \lambda_1\psi^{(1)} + \lambda_2\psi^{(2)}$, where $\psi^{(1)}$ and $\psi^{(2)}$ are again product states defined respectively by

$$\langle\psi^{(i)};\, A_l\rangle = \begin{cases}\langle\psi;\, A_l\rangle, & l \neq j\\ \mathrm{Tr}\ \rho_j^{(i)}A_j, & l = j\end{cases}$$

j arbitrary, but fixed in Z,

with

$$\rho_j^{(1)} = \tfrac{1}{2}(1 + \sigma_j^z) \quad \text{and} \quad \rho_j^{(2)} = \tfrac{1}{2}(1 - \sigma_j^z)$$

and

$$\lambda_1 = \tfrac{1}{2}\{1 + \tanh[\beta(B + B_{l,\psi})]\},$$
$$\lambda_2 = \tfrac{1}{2}\{1 - \tanh[\beta(B + B_{l,\psi})]\};$$

hence,[10] the time evolution is *not* asymptotically Abelian, as can also be explicitly seen from

$$\eta\langle\psi;\, C^*[\alpha_t[A], B]C\rangle \neq 0, \quad (4.4)$$

with

$$A = \sigma_l^z, \quad B = \sigma_l^x, \quad \text{and} \quad C = 1 + i\sigma_l^x.$$

$$\langle\psi;\, AC\rangle = \langle\psi;\, A\rangle\langle\psi;\, C\rangle \quad (4.5)$$

whenever

$$A \in \mathcal{A}(\Omega), \quad C \in \mathcal{A}(\Omega'), \quad \Omega \cap \Omega' = \phi,$$

so that ψ is a product state and hence satisfies the strongest possible cluster property;

$$\langle\psi;\, \sigma_l^x\rangle = 0 = \langle\psi;\, \sigma_l^y\rangle,$$
$$\langle\psi;\, \sigma_l^z\rangle = \tanh[\beta(B + B_{l,\psi})], \quad (4.6)$$
$$B_{l,\psi} = 2\tilde{v}_l[\langle\psi;\, \sigma_j^z\rangle].$$

The above three equations determine ψ; they appear here as formal generalizations of the self-consistency equations of Sec. 2, and we check in Appendix D that the above relations actually reduce to those of Sec. 2 in the particular case considered there.

The equations determine only the stationary points of our integrand; this implies that *not all* extremal KMS states occur in the decomposition of the Gibbs state ϕ. We illustrate explicitly this feature in Appendix E for periodic potentials of period $p = 2$. In this simple case, one can already exhibit as many as nine solutions

to the self-consistency equations. Among these, at most two occur in the decomposition of ϕ: those corresponding to absolute maxima of our integrand, i.e., the absolute minima of the free energy; the remaining solutions are then either relative minima, saddle points, or maxima of the free energy, so that these extremal KMS states can at best be interpreted as metastable or unstable pure thermodynamical phases.

We now want to examine the question of the decomposition of an arbitrary KMS state ϕ in its extremal KMS components ψ and the related problem of consistency between the definitions of the time evolution in the representations respectively associated with ϕ and the various ψ entering in the decomposition of ϕ.

Many of the results proved above for extremal KMS states carry over to any general KMS state ϕ; we already noticed that ϕ is time invariant and that, as a consequence (see Sec. 3), $\alpha^t_{\pi_\phi}$ is unitarily implemented in the GNS representation π_ϕ associated with ϕ and extends to an automorphism $\tilde{\alpha}^t_\phi$ of $\pi_\phi(\mathcal{A})''$ with respect to which the extension $\tilde{\phi}$ of ϕ to $\pi_\psi(\mathcal{A})''$ is KMS in the ordinary sense. Due to the fact that the time evolution is not an automorphism of $\mathcal{A}$, the analysis of the decomposition of a KMS state into its extremal components, as carried out in our Ref. 4, needs some refinements. Under the assumption[11] that

(i) there exist p's finite such that

$$\nu_l[f] = \nu_{l+p}[f],$$

for all l in Z and all f in $\mathcal{C}_v(\mathrm{Z})$, and

(ii) χ^k, the characteristic function of the sublattices, $\{i = np + k \mid n \in \mathrm{Z}\}$, with k in P, belongs to $\mathcal{C}_v(\mathrm{Z})$, we prove in Appendix F that ϕ KMS can actually be written as a discrete statistical mixture of extremal KMS states satisfying Eqs. (4.2)–(4.6) and that the time evolution defined on $\pi_\phi(\mathcal{A})''$ is consistent (in a sense which will be made more precise in Appendix F) with that defined on each $\pi_\psi(\mathcal{A})''$.

5. CONCLUSIONS

Our main results, for a class of Weiss–Ising models with periodic potentials, are that: (i) The thermodynamical pure phases are extremal KMS states; (ii) the set of extremal KMS states is exactly given as the manifold of solutions of (a slight extension of) the usual self-consistency equations of the molecular field methods; (iii) every KMS state (and, in particular, the Gibbs state in the thermodynamical limit) can be written uniquely as a discrete statistical mixture of extremal KMS states.

The model analyzed here allows a comparison between the various definitions of pure thermodynamical phases which have been suggested in the past[4,14–16] and points to the fact that the definition chosen in Ref. 4 is the most appropriate. First, the model is not asymptotically Abelian in time so that the pure thermodynamical phases which *are* extremal KMS states are *not* extremal time invariant, a fact which we exhibit explicitly in Sec. 4. Hence the characterization of pure thermodynamical phases as extremal time-invariant states is not appropriate in general situations where the time evolution is not asymptotically Abelian. Secondly, since a breaking of the translation invariance Z of the lattice is involved in the phase transition occurring here, a characterization of pure thermodynamical phases as extremal Z-invariant states is awkward. We notice in this connection that our pure thermodynamical phases ψ are invariant with respect to a proper subgroup Z_ψ of Z; since, on the one hand, the local structure of our system implies that the translations act in an asymptotically Abelian manner and since, on the other hand, the representations π_ψ are primary, it follows that ψ are extremal Z_ψ-invariant as Z_ψ acts in an asymptotically Abelian manner, the set $\mathfrak{S}_{\mathrm{Z}_\psi}$ of all Z_ψ-invariant states is a simplex, and the decomposition of a Z_ψ-invariant state (as is the Gibbs state) in its extremal Z_ψ-invariant components is unique; the multiplicity of left-over extremal Z_ψ-invariant states which cannot be interpreted as pure thermodynamical phases is, however, appalling, so that even a third characterization of pure thermodynamical phases as extremal Z_ψ-invariant states would be far too permissive. Moreover, we showed[17,4] that this type of characterization is altogether inadequate when the initial group of invariance of the Hamiltonian is E^3, the Euclidean group in three dimensions.

The proposed dynamical characterization of thermodynamically pure phases involves the technical problem of the proper definition of the time evolution for an infinite system. We chose to define it for each representation separately and then particularized this definition successively to the representations associated with time-invariant states, with extremal KMS states, and with general KMS states. In the latter two cases, we showed that this definition could be made in an unambiguous manner, whereas this is not the case for more general representations. For extremal KMS states we proved that the time evolution is actually an automorphism of the C^*-algebra $\pi(\mathcal{A})$ whereas, already for general KMS states, it can only be shown to be an automorphism of the von Neumann algebra $\pi(\mathcal{A})''$.

Since the algebra $\mathcal{A}$ of the quasilocal observables on our physical system is simple, every representation of $\mathcal{A}$ is faithful, so that the time evolutions obtained for each extremal KMS state can be lifted up separately to continuous families of automorphisms of $\mathcal{A}$. These, however, do not coincide, and this feature is linked to the existence of several thermodynamical pure phases. The necessity for this generalized definition of the time evolution of an infinite system is similar to that encountered by Thirring and Wehrl[18] in their study of the BCS model (these authors, however, were not concerned with the KMS condition) and has also been met by the general frame proposed by Dubin and Sewell[19] (who assume the existence of the Gibbs state in the thermodynamical limit, define the time evolution for this state, and then prove that the KMS boundary condition is satisfied). We surmise from these results that, in the type of models considered here, the time evolution cannot be assumed to be given by a single family $\{\alpha^t\}$ of automorphisms of $\mathcal{A}$, but could at best be defined as a family of automorphisms of the enveloping von Neumann algebra[20] of $\mathcal{A}$, although this mathematical device does not seem to shed any light on the physical properties of the systems considered; we therefore preferred to attack the technical problem of the definition of the time evolution along the lines presented here.

Finally, we might remark that whereas our investigation was carried out explicitly for periodic Weiss–Ising interactions (where the comparison with traditional methods is particularly easy), the framework developed in Secs. 3 and 4 allows the treatment of general Kac potentials of the form

$$v_{ij} = \lim_{\gamma \to 0} \gamma f(\gamma |i-j|)$$

such as those studied by Lebowitz and Penrose[12] and Gates and Penrose[13]; the analysis of the latter authors is carried out for continuous gas models, whereas ours applies primarily to the lattice-gas translation of the magnetic models we considered; for the ordinary Weiss lattice gas the pure thermodynamical phases can, as usual, be interpreted as liquid and gas. The Gibbs states follow the actual isotherms, *including* the Maxwell plateau.

ACKNOWLEDGMENT

We want to thank Dr. E. J. Woods for suggesting some improvements in the presentation of our results.

APPENDIX A

Lemma: Let $W = (w_{ij})$ be a strictly positive, symmetric, real $n \times n$ matrix and let $x_1 \cdots x_n$ be n independent variables. Then

$$\exp\left(\sum_{i,j} x_i w_{ij} x_j\right) = \left[\prod_i \left(\frac{\lambda_i}{2\pi}\right)\right]^{\frac{1}{2}} \int_{-\infty}^{+\infty} d\zeta_1 \cdots \int_{-\infty}^{+\infty} d\zeta_n \times \exp\left(-\tfrac{1}{2}\sum_{i,j} \zeta_i w_{ij} \zeta_j + 2^{\frac{1}{2}} \sum_{i,j} x_i w_{ij} \zeta_j\right),$$

where λ_i are the eigenvalues of W.

Proof: We first notice with Kac[6] that, from

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} d\eta \exp\left(-\tfrac{1}{2}\eta^2\right) = 1,$$

it follows (upon substitution of η by $\zeta - 2^{\frac{1}{2}}a$) that for any real a

$$\exp a^2 = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} d\zeta \exp\left(-\tfrac{1}{2}\zeta^2 + 2^{\frac{1}{2}} a\zeta\right)$$

and then, for every real $a_1 \cdots a_n$ and $y_1 \cdots y_n$,

$$\exp\left(\sum_i a_i^2 y_i^2\right) = [(2\pi)^{-\frac{1}{2}}]^n \int_{-\infty}^{+\infty} d\zeta_1 \cdots \int_{-\infty}^{+\infty} d\zeta_n \times \exp\left(-\tfrac{1}{2}\sum_i \zeta_i^2 + 2^{\frac{1}{2}} \sum_i a_i y_i \zeta_i\right).$$

From our assumption on W, the quadratic form $\sum_i x_i w_{ij} x_j$ can be diagonalized to the form of the lhs of the above equality (with $a_i = \lambda_i^{\frac{1}{2}} > 0$). With the substitution $\zeta_i = a_i \chi_i$, we get

$$\exp\left(\sum_i x_i w_{ij} x_j\right) = \left[\prod_i \left(\frac{\lambda_i}{2\pi}\right)\right]^{\frac{1}{2}} \int_{-\infty}^{+\infty} d\chi_1 \cdots \int_{-\infty}^{+\infty} d\chi_n \times \exp\left(-\tfrac{1}{2}\sum_i a_i^2 \chi_i^2 + 2^{\frac{1}{2}} \sum_i a_i^2 y_i \chi_i\right),$$

the rhs of which we can change, by an orthogonal transformation, to

$$\left[\prod_i \left(\frac{\lambda_i}{2\pi}\right)\right]^{\frac{1}{2}} \int_{-\infty}^{+\infty} d\zeta_1 \cdots \int_{-\infty}^{+\infty} d\zeta_n \times \exp\left(-\tfrac{1}{2}\sum_{i,j} \zeta_i w_{ij} \zeta_j + 2^{\frac{1}{2}} \sum_{i,j} x_i w_{ij} \zeta_j\right),$$

thus proving our lemma.

APPENDIX B

Let $L_{l,\pi}(\Omega)$ and $L_{l,\pi}$ be the local Liouville operators, acting on $B(\mathcal{H}_\pi)$, respectively associated with the local Hamiltonians $H_{l,\pi}(\Omega) \equiv \pi(H_l(\Omega))$ and $H_{l,\pi}$. These two Liouville operators clearly generate two families of automorphisms $\alpha^t_{l,\pi}(\Omega)$ and $\alpha^t_{l,\pi}$. We first show that

for each $l \in \mathbf{Z}$

$$\underset{\substack{\Omega\to\infty \\ (\Omega\ni l)}}{\text{s-lim}}\ \alpha^t_{l,\pi}(\Omega)[\pi(A_l)] = \alpha^t_{l,\pi}[\pi(A_l)]. \tag{B1}$$

From

$$\underset{\substack{\Omega\to\infty \\ (\Omega\ni l)}}{\text{s-lim}}\ B_{l,\pi}(\Omega) = B_{l,\pi} \tag{B2}$$

and

$$\|B_{l,\pi}(\Omega)\| < c_l, \tag{B3}$$

we conclude that for each $n \in \mathbf{Z}$

$$\underset{\substack{\Omega\to\infty \\ (\Omega\ni l)}}{\text{s-lim}}\ [B + B_{l,\pi}(\Omega)]^n = (B + B_{l,\pi})^n. \tag{B4}$$

From

$$[B_{l,\pi}(\Omega),\ \pi(A_l)] = 0 \tag{B5}$$

and (B2), we have also

$$[B_{l,\pi},\ \pi(A_l)] = 0. \tag{B6}$$

From (B5) and (B6), respectively, we get

$$L^n_{l,\pi}(\Omega)[\pi(A_l)] = \delta^n_{l,\pi}[\pi(A_l)][B + B_{l,\pi}(\Omega)]^n, \tag{B7}$$

$$L^n_{l,\pi}[\pi(A_l)] = \delta^n_{l,\pi}[\pi(A_l)](B + B_{l,\pi})^n, \tag{B8}$$

with

$$\delta_{l,\pi}[\pi(A_l)] = -i[\pi(\sigma^z_l),\ \pi(A_l)]. \tag{B9}$$

We notice that

$$\|\delta_{l,\pi}[\pi(A_l)]\| < \|A_l\|. \tag{B10}$$

From (B4), (B7), (B8), and (B10), we conclude that

$$\underset{\substack{\Omega\to\infty \\ (\Omega\ni l)}}{\text{s-lim}}\ L^n_{l,\pi}(\Omega)[\pi(A_l)] = L^n_{l,\pi}[\pi(A_l)]. \tag{B11}$$

(B1) now follows from (B11) and from the fact that the exponential series

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} L^n[A] = \alpha^t[A]$$

converge in norm.

For every finite $\Omega_0 \subset \mathbf{Z}$, every finite sequence $\{l_1 \cdots l_m\} \subseteq \Omega_0$, and every $\Omega \supseteq \Omega_0$, we define

$$\alpha^t_\pi(\Omega)[\pi(A_{l_1} \otimes \cdots \otimes A_{l_m})] = \alpha^t_{l_1,\pi}[\pi(A_{l_1})] \cdots \alpha^t_{l_m,\pi}[\pi(A_{l_m})] \tag{B12}$$

which extends then by linearity to $\mathcal{A}(\Omega_0)$. This evidently coincides for all $\Omega \supseteq \Omega_0$ with the $\pi(\alpha^t(\Omega)[A])$ where $\alpha^t(\Omega)[A]$ is defined in Sec. 3. We then define in the same manner $\alpha^t_\pi[\pi(A)]$ for all A in $\mathcal{A}(\Omega_0)$. Since

$$\|\alpha^t_{l,\pi}(\Omega)[\pi(A_l)]\| = \|\pi(A_l)\| = \|\alpha^t_{l,\pi}[\pi(A_l)]\|, \tag{B13}$$

we can conclude from (B1), (B12), and (B13) that

$$\underset{\substack{\Omega\to\infty \\ (\Omega\supseteq\Omega_0)}}{\text{s-lim}}\ \pi(\alpha^t(\Omega)[A]) = \alpha^t_\pi[\pi(A)], \quad \text{for all } A \text{ in } \mathcal{A}(\Omega_0).$$

This is trivially extended to every A and $\mathcal{A}$.

Since $B_{l,\pi}$ belongs to $\pi(\mathcal{A})'$ and since every element $A(\Omega_0)$ of $\mathcal{A}(\Omega_0)$ can be obtained as a finite linear combination of finite products of the A_l with $l \in \Omega_0$, we see that for each $A \in \mathcal{A}(\Omega_0)$

$$\alpha^t_\pi(\pi(A)) = \sum_{n=0}^{\infty} \frac{t^n}{n!} L^n_{\Omega,\pi}[\pi(A)], \quad \text{for all } \Omega \supseteq \Omega_0,$$

where $L_{\Omega,\pi}$ is defined by

$$[L_{\Omega,\pi},\ A] = -i\Big[\sum_{l\in\Omega}(B + B_{l,\pi})\sigma^z_l,\ A\Big].$$

APPENDIX C

With the notation of Sec. 3 and ϕ time invariant, we now construct the operator U^t on the dense linear manifold

$$\{\pi(A)\Phi \mid A \in \mathcal{A}(\Omega);\ \Omega \subset \mathbf{Z}\}$$

of $\mathcal{H}_\pi$ as

$$U^t_\pi \pi(A)\Phi = U^t_{\Omega,\pi}\pi(A)U^{-t}_{\Omega,\pi}\Phi$$

for all A in $\mathcal{A}(\Omega)$, where Ω runs over all finite subsets of $\mathbf{Z}$. Due to the isotony property, U^t_π is clearly linear on its domain of definition. The time invariance of $\tilde{\phi}$ implies

$$\|U^t_\pi \pi(A)\Phi\| = \|\pi(A)\Phi\|,$$

so that U^t_π can be extended in a unique manner to a unitary operator on $\mathcal{H}_\pi$. We notice that this operator satisfies the following two properties:

$$U^t_\pi \Phi = \Phi,$$

$$U^t_\pi \pi(A) U^{-t}_\pi = \alpha^t_\pi[\pi(A)], \quad \text{for all } A \text{ in } \mathcal{A}.$$

Since the mapping $C \to U^t_\pi C U^{-t}_\pi$ is weakly continuous on $\mathcal{B}(\mathcal{H}_\pi)$, α^t_π extends to a spatial automorphism

$$\tilde{\alpha}^t_\pi[A] = U^t_\pi A U^{-t}_\pi, \quad \text{for all } A \text{ in } \pi(\mathcal{A})'',$$

of the von Neumann algebra $\pi(\mathcal{A})''$. This important result can also be obtained as follows: Since

$$(\pi(\mathcal{A})'')' = \pi(\mathcal{A})' = \bigcap_{\Omega\subset Z} \pi(\mathcal{A}(\Omega))',$$

we have, for all $\Omega \subset \mathbf{Z}$, all A in $\mathcal{A}(\Omega)$, and all X in $\pi(\mathcal{A})'$,

$$0 = [X, U^t_\pi(\Omega)\pi(A)U^{-t}_\pi(\Omega)] = [X, U^t_\pi \pi(A) U^{-t}_\pi] = U^t_\pi[U^{-t}_\pi X U^t_\pi,\ \pi(A)]U^{-t}_\pi,$$

which again implies $U^{-t}_\pi X U^t_\pi$ belongs to $\pi(\mathcal{A})'$. This implies that the natural extension $\tilde{\alpha}^t_\pi$ to $\pi(\mathcal{A})''$, defined by

$$\tilde{\alpha}^t_\pi[A] = U^t_\pi A U^{-t}_\pi, \quad \text{for all } A \text{ in } \pi(\mathcal{A})'',$$

maps $\pi(\mathcal{A})''$ onto itself and is then a spatial automorphism of this von Neumann algebra.

APPENDIX D

To show that the results of Sec. 3 and 4 contain those of Sec. 2 in the particular case considered there, we must check *first* that the ambiguity introduced when we extended ν_l and $\tilde{\nu}_l$ actually disappears in the final results and *second* that the time evolution α_π^t is indeed the limit of the time evolution for the finite system.

We notice that, when $\nu_{ij}(\Omega)$ is periodic in $|i-j|$ with period p, then the linear functional ν_l defined in Sec. 3 satisfies

$$\nu_l[f] = \nu_{l+p}[f], \quad \text{for all } f \text{ in } \mathcal{C}_\nu(\mathbf{Z}),$$

so that the natural consistency requirements impose

$$\tilde{\nu}_l[f] = \tilde{\nu}_{l+p}[f], \quad \text{for all } f \text{ in } \mathcal{C}(\mathbf{Z}).$$

Consequently, $B_{l,\psi} = B_{l+p,\psi}$, and hence $\langle\psi; \sigma_j^z\rangle$ is periodic of period p in j so that, with the assumptions of Sec. 2,

$$\begin{aligned} B_{i,\psi} &= 2\nu_l[\langle\psi; \sigma_j^z\rangle] = 2\sum_{j\in\mathbf{P}} f_{ij}\langle\psi; \sigma_j^z\rangle \\ &= 2\sum_{j\in\mathbf{P}} f_{ij} \tanh[\beta(B + B_{j,\psi})]. \end{aligned}$$

This establishes our first claim. To substantiate the second, it suffices (see Sec. 3) to show that $B_{l,\pi}(\Omega)$, which is uniformly bounded, converges strongly to $B_{l,\psi}$ on a dense domain of the representation space $\mathcal{H}_\pi$ of the GNS representation π associated with ψ extremal KMS. To prove this convergence, we consider, for any arbitrary but fixed finite subset Ω_0 of Z, any fixed A in $\mathcal{A}(\Omega_0)$, and all $\Omega \supset \Omega_0$,

$$\begin{aligned} &\left\| \left[\sum_{j\in\Omega} \nu_{ij}(\Omega)\pi(\sigma_j^z) - B_{i,\pi}\right]\pi(A)\Psi \right\| \\ &\quad \le \left\| \sum_{j\in\Omega_0} \nu_{ij}(\Omega)\sigma_j^z \right\| \|\pi(A)\Psi\| \\ &\quad\quad + \left\| \left[\sum_{j\in\Omega-\Omega_0} \nu_{ij}(\Omega)\pi(\sigma_j^z) - B_{i,\pi}\right]\pi(A)\psi \right\|. \end{aligned}$$

As $\Omega \to \infty$, the first term tends to zero. Upon using the facts that ψ is a product state, that $B_{i,\pi}$ is a c-number, and that $\langle\psi; \sigma_j^z\rangle$ is periodic in j, we see that the square of second term in the above relation tends, as $\Omega \to \infty$, to

$$\langle\psi; A^*A\rangle\left\{\sum_{j\in\Omega} \nu_{ij}(\Omega)\langle\psi; \sigma_j^z\rangle - B_{i,\psi}\right\}^2$$

which also tends to zero by construction of $B_{i,\psi}$.

APPENDIX E

We now analyze further the meaning of the self-consistency equations in the case $p = 2$ and external field $B = 0$. When we denote $f_{11} = f_{22} = a$ and $f_{12} = b$, we find for the free energy (per site) f of a state with mean free field B_1 at the even sites and B_2 at the odd sites:

$$\begin{aligned} f = &-a \tanh(\beta B_1)^2 - 2b \tanh(\beta B_1)\tanh(\beta B_2) \\ &- a \tanh(\beta B_2)^2 + B_1 \tanh(\beta B_1) + B_2 \tanh(\beta B_2) \\ &- kT \ln 2\cosh(\beta B_1) - kT \ln 2\cosh(\beta B_2). \end{aligned}$$

One finds then indeed that the extremal points of this function are determined by the self-consistency equations

$$\begin{aligned} 2a \tanh(\beta B_1) + 2b \tanh(\beta B_2) &= B_1, \\ 2a \tanh(\beta B_2) + 2b \tanh(\beta B_1) &= B_2. \end{aligned}$$

To be specific, let us consider the case $-a < b < 0$, and define the temperatures T_{I} and T_{II}, respectively, by $T_{\mathrm{I}} = 2(a-b)$ and $T_{\mathrm{II}} = 2(a+b)$; hence, $T_{\mathrm{II}} < T_{\mathrm{I}}$. The following solutions are obtained readily:

(i) For $T > T_{\mathrm{I}}$, only the solution $B_1 = B_2 = 0$ exists; this solution corresponds then to an (absolute) minimum.

(ii) For $T_{\mathrm{II}} < T < T_{\mathrm{I}}$, one has the three solutions

(a) $B_1 = B_2 = 0$ which is then a maximum,

(b) and (c) $B_1 = -B_2 = \pm B_{\mathrm{I}}$, in which B_{I} is solution of $2(a-b)\tanh(\beta B_{\mathrm{I}}) = B_{\mathrm{I}}$; these solutions are absolute minima.

(iii) For $T < T_{\mathrm{II}}$, one has always five solutions, namely,

(a) $B_1 = B_2 = 0$ (maximum),

(b) and (c) $B_1 = -B_2 = \pm B_{\mathrm{I}}$ (absolute minimum),

(d) and (e) $B_1 = B_2 = \pm B_{\mathrm{II}}$ in which B_{II} is solution of $2(a+b)\tanh(\beta B_{\mathrm{II}}) = B_{\mathrm{II}}$.

All these solutions have the property that, when $T \to T_{\mathrm{I}}$ (respectively T_{II}), one has $B_{\mathrm{I}} \to 0$ (respectively $B_{\mathrm{II}} \to 0$); therefore, they can be obtained by linearizing the self-consistency equations. It turns out, however, that there are still some peculiar solutions that cannot be obtained in this manner. Let T_{III} be the solution of $T\cosh^2[\beta B_{\mathrm{II}}(T)] = a - b$. Graphical methods (for instance) show that $T_{\mathrm{III}} < T_{\mathrm{II}}$ and that four new solutions to the self-consistency equations exist when $T < T_{\mathrm{III}}$, which are of the form $B_1 = \pm B_1^{\mathrm{III}}$, $B_2 = \pm B_2^{\mathrm{III}}$. When T tends to T_{III}, one has that B_1^{III} and B_2^{III} tend to $B_{\mathrm{II}} \neq 0$. Together with the extra solutions occurring at $T = T_{\mathrm{III}}$, one has that the status of the solution (iii.c) changes from saddle point to relative minimum.

APPENDIX F

We define, for every KMS state ϕ on $\mathcal{A}$, the GNS representation π_ϕ of $\mathcal{A}$, the time evolution $\tilde{\alpha}_\phi^t$ of

$\pi_\phi(\mathcal{A})''$ and the canonical extension $\tilde{\phi}$ of ϕ to $\pi_\phi(\mathcal{A})''$ which is KMS with respect to $\tilde{\alpha}^t_\phi$ (see Secs. 3 and 4). We know then[4] that $\tilde{\phi}$ can be uniquely decomposed into states $\tilde{\psi}$ on $\pi_\phi(\mathcal{A})''$ which are extremal KMS with respect to α^t_ϕ. We then construct the representation $\tilde{\pi}_\psi$ of $\pi_\phi(\mathcal{A})''$ the weak closure of which is then[9] a factor. This implies that $\tilde{\pi}_\psi(B_{l,\pi_\phi})$ is a c-number. From this we conclude, as in Sec. 4, that the restriction ψ of $\tilde{\psi}$ to $\mathcal{A}$ is a product state of factor representations. From this it follows[8] that the GNS representation π_ψ of $\mathcal{A}$ is also a factor. Since, on the other hand, ψ is KMS on $\mathcal{A}$ with respect to the time evolution defined from $\tilde{\pi}_\psi(B_{l,\pi_\phi})$, we conclude[9] that ψ is extremal KMS on $\mathcal{A}$ with respect to that time evolution. We now want to show that the latter coincide with the time evolution defined from ψ alone in Sec. 4. To this end, it is sufficient to show that

$$\tilde{\pi}_\psi(B_{l,\pi_\phi}) = B_{l,\psi}, \tag{F1}$$

which we shall now prove.

We recall[4] that μ is the central measure associated with $\tilde{\phi}$ and that to every μ-measurable set F corresponds an element D_F in $\pi_\phi(\mathcal{A})'' \cap \pi_\phi(\mathcal{A})'$ such that

$$\langle\tilde{\phi}; AD_F\rangle = \int_F \langle\tilde{\psi}; A\rangle\, d\mu(\psi), \quad \text{for all } A \text{ in } \pi_\phi(\mathcal{A})''. \tag{F2}$$

From the definition of B_{l,π_ϕ} we have, upon using (F2),

$$\begin{aligned}\int_F \langle\tilde{\psi}; B_{l,\pi_\phi}\rangle\, d\mu(\psi) &= \langle\tilde{\phi}; B_{l,\pi_\phi} D_F\rangle \\ &= 2\tilde{\nu}_l[\langle\tilde{\phi}; \sigma^z_j D_F\rangle] \\ &= 2\tilde{\nu}_l\Big[\int_F \langle\psi; \sigma^z_j\rangle\, d\mu(\psi)\Big].\end{aligned} \tag{F3}$$

As we already saw, the periodicity of $\tilde{\nu}_l$ implies that the extremal KMS states occurring in this decomposition are periodic with period p. Let S_p be the *closed* set of all states ψ on $\mathcal{A}$ such that $\langle\psi; A_{j+p}\rangle = \langle\psi; A_j\rangle$ for all j in Z and all A in $\mathcal{A}(j)$. Since μ is concentrated on the extremal KMS states, which are all in S_p, we have that the support of μ is contained in S_p. For these states,

$$\tilde{\nu}_l\langle\psi; \sigma^z_j\rangle = \sum_{k\in\mathrm{P}} f_{l,k}\langle\psi; \sigma^z_k\rangle, \tag{F4}$$

where $f_{l,k} = \nu_l(\chi_k)$ (the existence of the latter limit is part of the assumptions made at the end of Sec. 4). We can therefore interchange the mean $\tilde{\nu}_l$ and the integral in the rhs of (F3) to get

$$\begin{aligned}\int_F \langle\tilde{\psi}; B_{l,\pi_\phi}\rangle\, d\mu(\psi) &= \int_F 2\tilde{\nu}_l\langle\psi; \sigma^z_j\rangle\, d\mu(\psi) \\ &= \int_F B_{l,\psi}\, d\mu(\psi).\end{aligned} \tag{F5}$$

Since this holds for every μ-measurable F and since the integrands of both the lhs and the rhs of (F5) are w^*-continuous in $\tilde{\phi}$ [the rhs because of (F4)], we conclude that

$$\langle\tilde{\psi}; B_{l,\pi_\phi}\rangle = B_{l,\psi}, \tag{F6}$$

from which (F1) follows since $\tilde{\pi}_\psi(B_{l,\pi_\phi})$ is a c-number.

Finally, we notice that, since the self-consistency Eqs. (4.6) admit only a finite number of solutions, the decomposition of ϕ into its extremal KMS components is actually discrete; furthermore, the same reasoning as used in Appendix D leads to the strong convergence of $B_{l,\pi_\phi}(\Omega)$ whenever ϕ is KMS.

* Z.W.O. Postdoctoral fellow.

[1] R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957); P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[2] R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).

[3] For a partial list of references on this, see, for instance, Footnotes 10–19 in Ref. 4 below.

[4] G. G. Emch, H. J. F. Knops, and E. J. Verboven, J. Math. Phys. **11**, 1655 (1970).

[5] See, for instance, R. Brout, *Phase Transitions* (Benjamin, New York, 1965) or, for a nutshell statement of relevance, the last paragraph in the introduction to the paper of E. Lieb, J. Math. Phys. **7**, 1016 (1966).

[6] Section 3 in M. Kac, *Brandeis Summer Institute, 1966 Lecture Notes*, edited by M. Chretien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968), Vol. 1.

[7] M. Kac, in *Fundamental Problems in Statistical Mechanics II*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1968); A. J. F. Siegert and D. J. Vezetti, J. Math. Phys. **9**, 2173 (1968); C. J. Thompson, A. J. F. Siegert, and D. J. Vezzetti, J. Math. Phys **11**, 1018 (1970).

[8] M. A. Guichardet, Ann. Sci. École Norm. Super. **83**, 1 (1966).

[9] See, for instance, Lemma 1.3 in Ref. 4 or Theorem 10.3 in H. Araki, Commun. Math. Phys. **14**, 120 (1969).

[10] Theorem 3.6 in H. Araki and H. Miyata, Kyoto University, Preprint RIMS-37, 1968; see, also, discussion in Sec. 1 of Ref. 4.

[11] This type of potentials includes the purely periodic ones considered in Sec. 2 and, also, for instance, the "Kac potentials" considered by Penrose *et al.*[12,13] defined as $\nu_{ij}(\gamma) = \gamma\phi(\gamma\,|i-j|)$ in which the shape function ϕ is Riemann integrable; for the latter potentials $p = 1$ and $\nu_1(1) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \phi(x)\, dx$.

[12] J. L. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).

[13] D. J. Gates and O. Penrose, Commun. Math. Phys. **15**, 255 (1969).

[14] G. G. Emch, H. J. F. Knops, and E. J. Verboven, J. Phys. Soc. Japan, Suppl., **26**, 301 (1969).

[15] H. J. F. Knops, Ph.D. thesis, Nijmegen, 1969.

[16] See, for instance, D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).

[17] H. J. F. Knops, Commun. Math. Phys. **12**, 36 (1969).

[18] W. Thirring and A. Wehrl, Commun. Math. Phys. **4**, 303 (1967).

[19] D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970).

[20] Section XII in J. Dixmier, *Les C*-algebras et leurs représentations* (Gauthier-Villars, Paris, 1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Derivatives of Phase Shifts and Binding Energies by Use of Variational Principles*

Y. TIKOCHINSKY†
Department of Physics, Rutgers–The State University, New Brunswick, New Jersey 08903

(Received 1 December 1969)

The Kohn–Hulthén variational principle for the phase shifts, as well as the Rayleigh–Ritz principle for the binding energies, are used to determine the derivatives of $\delta_l = \delta(V, E, l, m, \hbar)$ and $E = E(V, l, m, \hbar)$ with respect to the listed parameters. A similar treatment utilizing Hamilton's variational principle leads to the corresponding classical results. The relation between the quantum mechanical and the classical expressions is examined. In particular, it is found that the quantum-mechanical binding energy corresponds to a certain path average of the classical energy. Some applications of resulting formulas are briefly reviewed. This work is an extension of ideas originated by Fock and Demkov.

I. INTRODUCTION

Variational principles are certainly among the most powerful tools at the disposal of the theoretical physicist. They are widely used in almost every branch of theoretical physics. It is therefore quite surprising that one of its most natural applications, namely, the calculation of derivatives of the stationary quantity with respect to the parameters of the system, seems to be overlooked. In fact, if one uses as trial functions the correct functions for a slightly different system (different mass, different force, etc.),one is bound to get the derivative of the stationary quantity with respect to the varied parameter. This process could be almost trivial but for the fact that the varied functions must usually satisfy some normalization or boundary conditions. The way to overcome this difficulty is to perform a change of scale—a technique devised by Fock and utilized by Demkov.[1]

We shall begin our discussion with scattered states. The derivatives of the phase shifts, as well as those of the complete scattering amplitude, will be obtained. Next, we shall consider the bound states, where the stationary quantity is the energy. Finally, we shall turn to classical mechanics and secure the corresponding derivatives for the action integral.

In examining the correspondence between the quantum-mechanical (q.m.) results and the classical (cl) results, we shall see that the q.m. binding energy E corresponds to $-\langle L\rangle_{\text{av}}$, where $\langle L\rangle_{\text{av}}$ is the time average of the Lagrangian. This is to be expected on the basis of the WKB approximation. However, there is a more striking correspondence due to a purely classical result, namely, $-\langle L\rangle_{\text{av}} = \langle E_{\text{cl}}\rangle$, where $\langle E_{\text{cl}}\rangle$ is a certain path average of the classical energy. This is, of course, reminiscent of Feynman's path integral method.[2] We shall also see, in passing, that even the simplest 1-particle classical problem has divergences built in which necessitate a careful limiting process.

Finally, a word concerning the novelty of the results. Many of the results derived here are well known. They have been derived, however, using diverse methods and thus are not as widely known as is warranted. This article should be regarded, therefore, at least in part, as a review.

II. QUANTUM MECHANICS

A. Scattered States

We shall start with the Kohn–Hulthén variational principle for the phase shifts δ_l,[3,4] where $\delta_l = \delta(V, E, l, m, \hbar)$. The principle reads

$$k\delta_l = \int_0^\infty U_t(r) L U_t(r)\, dr + k\delta_t + O((U_l - U_t)^2), \quad (1)$$

where $U_t(r)$ is a trial function satisfying the boundary conditions

$$U_t(0) = 0, \quad U_t(r) \underset{r\to\infty}{\sim} \sin(kr - \tfrac{1}{2}l\pi + \delta_t), \quad (2)$$

δ_t being the trial phase shift and $U_l(r)$ the solution of the radial equation

$$LU_l(r) \equiv \left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - \frac{2m}{\hbar^2}V(r)\right)U_l = 0 \quad (3)$$

with the boundary conditions

$$U_l(0) = 0, \quad U_l(r) \underset{r\to\infty}{\sim} \sin(kr - \tfrac{1}{2}l\pi + \delta_l). \quad (3')$$

1. Changing the Potential

Let $U_t = U_t(V + \Delta V, r)$ be a solution of (3) with a potential $V + \Delta V$. Then, to first order we have

$$k\delta_t(V) = k\delta_t(V + \Delta V) + \int_0^\infty U_t(V + \Delta V)\left(L - \frac{2m}{\hbar^2}\Delta V + \frac{2m}{\hbar^2}\Delta V\right) \times U_t(V + \Delta V)\, dr.$$

Hence, using Eq. (3),

$$\Delta\delta_l = -\frac{2m}{\hbar^2 k}\int_0^\infty U_l^2(V, r)\Delta V\, dr. \tag{4}$$

In particular, if V depends on a parameter α,

$$\frac{\partial\delta_l}{\partial\alpha} = -\frac{2m}{\hbar^2 k}\int_0^\infty U_l^2(\alpha, r)\frac{\partial V}{\partial\alpha}\, dr. \tag{4'}$$

2. Changing the Energy

Let $\bar{U}_t = U_l(k + \epsilon k, r)$ be a solution of (3) with a wavenumber $(1 + \epsilon)k$, where ϵ is an infinitesimal constant. (The corresponding change in E is $\Delta E = 2\epsilon E$.) Asymptotically

$$\bar{U}_t \sim \sin\left[(1 + \epsilon)kr - \tfrac{1}{2}l\pi + \delta_l(E + \Delta E)\right].$$

Since the last equation violates the boundary condition (2), $\bar{U}_t$ cannot serve as a trial function. However, if a change of scale is performed, namely,

$$U_t = \bar{U}_t(r/(1 + \epsilon)) = U_l((1 + \epsilon)k, r/(1 + \epsilon)),$$

the boundary condition (2) is satisfied. Inserting U_t into Eq. (1) and changing the variable of integration from r to $r' = r/(1 + \epsilon)$, we obtain

$$\begin{aligned} k\delta_l(E) = k\delta_l(E + \Delta E) &+ (1 + \epsilon)^{-1}\int_0^\infty dr' U_l((1 + \epsilon)k, r') \\ &\times \left(\frac{d^2}{dr'^2} + k^2(1 + \epsilon)^2 - \frac{l(l + 1)}{r'^2}\right. \\ &\left. - (1 + \epsilon)^2\frac{2m}{\hbar^2}V(r' + \epsilon r')\right)U_l((1 + \epsilon)k, r'). \end{aligned}$$

Finally, utilizing Eq. (3), we secure

$$\frac{\partial\delta_l}{\partial E} = \frac{k}{2E^2}\int_0^\infty U_l^2(k, r)\left(2V + r\frac{dV}{dr}\right)dr. \tag{5}$$

3. Changing the Angular Momentum

Let $U_t = U_l(r)$ be a solution of (3) with an angular momentum $\bar{l} = l + \Delta l$, where l is regarded as a continuous parameter. Asymptotically

$$U_t \sim \sin\left(kr - \tfrac{1}{2}l\pi + \delta_l - \tfrac{1}{2}\pi\Delta l\right);$$

hence, $\delta_t = \delta_{l+\Delta l} - \frac{1}{2}\pi\Delta l$. Inserting U_t into Eq. (1) we obtain

$$\frac{\partial\delta_l}{\partial l} = \tfrac{1}{2}\pi - (2l + 1)k^{-1}\int_0^\infty U_l^2 r^{-2}\, dr. \tag{6}$$

4. Changing the Mass

Changing the mass from m to $m + \Delta m$ and keeping the energy $E = \hbar^2k^2/(2m)$ fixed causes the wavenumber k to change by $\Delta k = k\Delta m/(2m)$. Choosing the trial wavefunction as $U_t = U_l(m + \Delta m, r/(1 + \epsilon))$, with $\epsilon = \Delta m/(2m)$, and proceeding exactly as in Sec. II.A.2, we obtain

$$\frac{\partial\delta_l}{\partial m} = (\hbar^2 k)^{-1}\int_0^\infty U_l^2 r\frac{dV}{dr}\, dr. \tag{7}$$

5. Changing Planck's Constant

Since m and $\hbar$ enter all expressions only through the combination $\xi = \hbar^2/(2m)$, we can immediately obtain (if only for mathematical curiosity)

$$\begin{aligned} \frac{\partial\delta_l}{\partial\hbar} &= \frac{\partial\delta_l}{\partial m}\frac{\partial\xi/\partial\hbar}{(\partial\xi/\partial m)} \\ &= -\frac{2m}{\hbar^3 k}\int_0^\infty U_l^2 r\frac{dV}{dr}\, dr. \end{aligned} \tag{8}$$

The derivatives (4)–(8) should vanish as the potential vanishes. This is obvious for Eqs. (4), (5), (7), and (8). It is also true for Eq. (6), since for $V \to 0$, $U_l \to krj_l(kr)$ and the following mathematical relations holds:

$$\int_0^\infty dx j_l^2(x) = \tfrac{1}{2}\pi\int_0^\infty dx\frac{J_{l+\frac{1}{2}}^2(x)}{x} = \frac{\frac{1}{2}\pi}{2l + 1}.$$

Before commenting on the significance of Eqs. (4)–(8), let us write down the corresponding relations for the complete scattering amplitude. These can be achieved by use of Kohn's variational principle[3]

$$T_{\mathbf{k}_f\mathbf{k}_i} = T^t_{\mathbf{k}_f\mathbf{k}_i} + \langle\phi^-_{\mathbf{k}_f t}|\, H - E\,|\psi^+_{\mathbf{k}_i t}\rangle + \text{second-order terms}, \tag{9}$$

and proceeding along the same lines as in Sec. II.A.1–5. In Eq. (9) $T^t_{\mathbf{k}'\mathbf{k}}$ is the trial scattering amplitude, $\psi^+_{\mathbf{k}t}$ is the trial wavefunction satisfying the boundary condition

$$\psi^+_{\mathbf{k}t} \sim e^{i\mathbf{k}\cdot\mathbf{r}} - [2m/(4\pi\hbar^2)]e^{ikr}/r\cdot T^t_{\mathbf{k}_r\mathbf{k}}, \tag{10}$$

where $\mathbf{k}_r = \mathbf{r}\,|\mathbf{k}|/|\mathbf{r}|$, and $\phi^{-*}_{\mathbf{k}t} = \phi^+_{-\mathbf{k}t}$ is an independent trial function satisfying Eq. (10) with $\mathbf{k}$ replaced by $-\mathbf{k}$. One obtains

$$\Delta T_{fi} = \langle\psi^-_f|\,\Delta V\,|\psi^+_i\rangle, \tag{11}$$

$$\frac{\partial T_{fi}}{\partial\alpha} = \langle\psi^-_f|\,\partial V/\partial\alpha\,|\psi^+_i\rangle, \tag{11'}$$

$$2E\frac{\partial T_{fi}}{\partial E} = -T_{fi} - \langle\psi^-_f|\,2V + \mathbf{r}\cdot\nabla V\,|\psi^+_i\rangle, \tag{12}$$

$$2m\frac{\partial T_{fi}}{\partial m} = -3T_{fi} - \langle\psi^-_f|\,\mathbf{r}\cdot\nabla V\,|\psi^+_i\rangle, \tag{13}$$

$$\hbar\frac{\partial T_{fi}}{\partial\hbar} = 3T_{fi} + \langle\psi^-_f|\,\mathbf{r}\cdot\nabla V\,|\psi^+_i\rangle, \tag{14}$$

where the notation has been slightly simplified. In these expressions the $\psi^{\pm}$ are exact solutions of the Schrödinger equation. Equations (4), (5), (7), and (8) follow from (11)–(14), if one assumes a spherically symmetric potential.

We shall now turn to discuss the results for the scattered states. Equation (11) is (under the most general conditions) an immediate consequence of the Gell-Mann–Goldberger transformation.[5] Let $T(V)$ and $\psi^{\pm}(V)$ denote the scattering amplitude and wavefunctions (with outgoing or ingoing spherical waves) for the potential V, and let χ denote a plane wave. Then by the Gell-Mann–Goldberger transformation

$$T_{fi}(V+\Delta V) = \langle\psi_f^-(V)|\, V\,|\chi_i\rangle + \langle\psi_f^-(V)|\,\Delta V\,|\psi_i^+(V+\Delta V)\rangle.$$

Since $T_{fi}(V) = \langle\psi_f^-(V)|\, V\,|\chi_i\rangle$, we have to first order

$$\Delta T_{fi}(V) = \langle\psi_f^-(V)|\,\Delta V\,|\psi_i^+(V)\rangle.$$

Curiously enough, it appears as if this result has never been stated in its present general form. See, however, Sugar and Blankenbecler,[6] Austern and Blair,[7] and Spruch.[8]

A special case of Eq. (11a) is of some interest. If we put $V(\mathbf{r}, \alpha) = \alpha V(\mathbf{r})$ and integrate Eq. (11a) between $\alpha = 0$ (no potential) and $\alpha = 1$ (full potential), we obtain [since $T_{fi}(\alpha = 0) = 0$]

$$T_{fi} \equiv T_{fi}(\alpha = 1) = \int_0^1 d\alpha\, \langle\psi_f^-(\alpha V)|\, V\,|\psi_i^+(\alpha V)\rangle. \quad (15)$$

This exact representation for the scattering amplitude has been used to discuss the validity of the high-energy approximation for medium energies.[9]

An interesting application of Eq. (4), namely, the derivation of the phase equation, has been discussed by Spruch.[8] Let

$$V_r(r') = V(r'), \quad r' < r,$$
$$= 0, \qquad r' > r,$$

and let $U_{lr}(r')$ and $\delta_l(r)$ denote the radial function and phase shift for this potential. Using Eq. (4) with $\Delta V(r') = V_{r+\Delta r}(r') - V_r(r')$, we have

$$\delta_l(r+\Delta r) - \delta_l(r) = -\frac{2m}{\hbar^2 k}\int_0^{\infty} U_{lr}^2(r')[V_{r+\Delta r}(r') - V_r(r')]\,dr' = -\frac{2m}{\hbar^2 k}\Delta r U_{lr}^2(r)V(r). \quad (16)$$

Now for $r' \geq r$, i.e., outside the range of the potential $V_r(r')$, the radial wavefunction is a linear combination of the spherical Bessel functions

$$U_{lr}(r') = \alpha_l(r)kr'j_l(kr') + \beta_l(r)kr'n_l(kr').$$

Comparison with the asymptotic form (3′) yields

$$\alpha_l(r) = \cos\delta_l(r), \quad \beta_l(r) = -\sin\delta_l(r).$$

Inserting these coefficients into Eq. (16), we obtain the phase equation[10]

$$\frac{d\delta_l(r)}{dr} = -\frac{2m}{\hbar^2 k}V(r)(kr)^2 \times [\cos\delta_l(r)\,j_l(kr) - \sin\delta_l(r)n_l(kr)]^2. \quad (17)$$

It has been suggested[8] that the same procedure may be used to obtain a similar differential equation for the complete scattering amplitude T_{fi}. Unfortunately, this is not possible, the reason being that the form (10) for the wavefunction $\psi_{\mathbf{k}}^+$ is correct only for $r \to \infty$ and *not* for any r outside the range of the potential.

Equation (4′) is also interesting from a calculational point of view. It can be used to facilitate the numerical search for the parameters of the potential needed to reproduce the given scattering data.

Equations (5) and (12) have been derived by Demkov.[1] They are called virial theorems after their classical counterpart. The physical significance of $\partial\delta_l/\partial E$ was discussed by Wigner[11] who showed that $\hbar\partial\delta_l/\partial E$ is the time delay of the lth-partial *scattered* wave in the potential well.

The first derivation of Eq. (6) is apparently due to Newton.[12] As an immediate consequence of (6) we have (for real V) the inequality

$$\partial\delta_l/\partial l \leq \tfrac{1}{2}\pi \quad (18)$$

and hence

$$\delta_{l+1} - \delta_l \leq \tfrac{1}{2}\pi, \quad (19)$$

a result first derived by Regge.[13] The inequality (18) has a simple semiclassical interpretation.[12] The WKB phase shift satisfies[14]

$$\theta = 2\frac{\partial\delta_l}{\partial l}, \quad (20)$$

where θ is the classical scattering angle corresponding to an energy E and an angular momentum $l_{\mathrm{cl}} = \hbar(l + \frac{1}{2})$. Hence, semiclassically, Eq. (18) states simply that $\theta \leq \pi$.

No immediate application of Eqs. (7) and (8) seems to suggest itself. However, Eq. (8) could be used, presumably, for further investigations of the classical limit of quantum mechanics.

B. Bound States

We shall begin our discussion by assuming a spherically symmetric potential. Later on this restriction will be removed. The Rayleigh–Ritz variational

principle for the binding energy $E = E(V, l, m, \hbar)$ is

$$E = \int_0^\infty U_t(r) H_r U_t(r)\, dr + O((U_t - U_l)^2), \quad (21)$$

where $U_t(r)$ is a normalized trial function, i.e.,

$$\int_0^\infty U_t^2(r)\, dr = 1 \quad (22)$$

and $U_l(r)$ is a normalized solution of the radial equation

$$H_r U_l(r) \equiv \left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + V(r)\right) U_l = E U_l, \quad (23)$$

$$\int_0^\infty U_l^2(r)\, dr = 1. \quad (24)$$

The derivation of the analogs of Eqs. (4)–(8) is essentially unchanged. There are, however, two new features, both of which will be illustrated in the derivation of the virial theorem. The first of these is connected with the normalization condition. In order to fulfill Eq. (22), any change of scale must be accompanied by a corresponding change of normalization. The second difference stems from the fact that the energy is no longer a parameter of the stationary quantity. Nevertheless, we shall see that a virial theorem does exist.

(a) *Changing the potential:* We obtain

$$\Delta E = \int_0^\infty U_l^2 \Delta V\, dr. \quad (25)$$

In particular, if V depends on a parameter α

$$\frac{\partial E}{\partial \alpha} = \int_0^\infty U_l^2 \frac{\partial V}{\partial \alpha}\, dr. \quad (25')$$

(b) *Changing the scale:* Let

$$U_t(r) = (1+\epsilon)^{\frac{1}{2}} U_l((1+\epsilon)r),$$

where $U_l(r)$ is a normalized solution of Eq. (23). The factor $(1+\epsilon)^{\frac{1}{2}}$ is introduced in order to preserve the normalization. Indeed,

$$\int_0^\infty U_t^2(r)\, dr = (1+\epsilon)\int_0^\infty U_l^2((1+\epsilon)r)\, dr = \int_0^\infty U_l^2(r')\, dr' = 1.$$

Inserting U_t into Eq. (21) and changing the variable of integration from r to $r' = (1+\epsilon)r$, we obtain

$$E = (1+\epsilon)^2 \int_0^\infty U_l(r')\left(-\frac{\hbar^2}{2m}\frac{d^2}{dr'^2} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r'^2} + (1+\epsilon)^{-2} V(r' - \epsilon r')\right) U_l(r')\, dr'.$$

Finally, utilizing Eq. (23) we secure

$$2E = \int_0^\infty U_l^2(r)\left(2V + r\frac{dV}{dr}\right) dr. \quad (26)$$

(c) *Changing the angular momentum:* We obtain

$$\frac{\partial E}{\partial l} = \frac{\hbar^2}{2m}(2l+1)\int_0^\infty \frac{U_l^2}{r^2}\, dr. \quad (27)$$

(d) *Changing the mass:* This leads to

$$\frac{\partial E}{\partial m} = -(2m)^{-1}\int_0^\infty U_l^2 r \frac{dV}{dr}\, dr. \quad (28)$$

(e) *Changing Planck's constant:* This results in

$$\frac{\partial E}{\partial \hbar} = \hbar^{-1}\int_0^\infty U_l^2 r \frac{dV}{dr}\, dr. \quad (29)$$

The limitation to spherically symmetric potentials is really not necessary. Utilizing the Rayleigh–Ritz variational principle

$$E = \langle \psi_t | H | \psi_t \rangle + \text{second-order terms} \quad (30)$$

(with normalized trial functions), we obtain

$$\Delta E = \langle \psi | \Delta V | \psi \rangle, \quad (31)$$

$$\frac{\partial E}{\partial \alpha} = \langle \psi | \frac{\partial V}{\partial \alpha} | \psi \rangle, \quad (31')$$

$$2E = \langle \psi | 2V + \mathbf{r}\cdot\nabla V | \psi \rangle, \quad (32)$$

$$\frac{\partial E}{\partial m} = -(2m)^{-1}\langle \psi | \mathbf{r}\cdot\nabla V | \psi \rangle, \quad (33)$$

$$\frac{\partial E}{\partial \hbar} = \hbar^{-1}\langle \psi | \mathbf{r}\cdot\nabla V | \psi \rangle, \quad (34)$$

where ψ is the exact normalized solution for the bound state.

Equations (25) and (31) are well known from first-order nondegenerate perturbation theory. For the degenerate case, however, Eq. (31) and perturbation theory give different results, the difference being of second order in the perturbing potential. [The use of Eq. (31) amounts to neglect of all off-diagonal elements of the perturbing potential.]

As in the scattering case, Eq. (25) can be used to facilitate the numerical search for the potential needed to bind a particle in a given level.

The virial theorems (26) and (32) are due to Fock.[15] (See also Ref. 1.)

Since by Eq. (27) $\partial E/\partial l > 0$, we have the well-known result that the energy of a state with a given number of nodes is a monotonic increasing function of the angular momentum. The restriction to a given number of nodes is essential since states with different number of nodes have separate $E(l)$ curves.

III. CLASSICAL MECHANICS

The variational principle underlying classical mechanics is Hamilton's principle for the action $I = I(q_2t_2, q_1t_1)$. It reads

$$I(q_2t_2, q_1t_1) = \int_{t_1}^{t_2} L(q_t, \dot{q}_t)\, dt + O([q(t) - q_t(t)]^2), \tag{35}$$

where $q(t)$ is the physical trajectory which passes through the points q_1 and q_2 at times t_1 and t_2, and $q_t(t)$ is an adjacent path going through the same end points at the same times. [For convenience of notation, we have lumped together all generalized coordinates into one vector $q(t)$.]

Without loss in generality, we can limit our discussion to motion in a central field. In fact, we shall treat only the radial equation, assuming that the angular momentum l is given. The Lagrangian for such a system is

$$L(r, \dot{r}) = \tfrac{1}{2}m\dot{r}^2 - V(r) - l^2/(2mr^2) \equiv T - U(r). \tag{36}$$

The action I depends, in turn, on the parameters V, l, and m.

It is convenient to reverse the order of presentation. We shall first discuss the bound states and then turn to examine the scattered states.

A. Bound States

(a) *Changing the potential:* Let $r_t = r(U + \Delta U, t)$ be the correct orbit for the potential $U + \Delta U$. (At the risk of repetition, we assume that,unless otherwise stated, all trial orbits go through the same end points at the same times.) Inserting $r(U + \Delta U, t)$ into Eq. (35), we have

$$\begin{aligned} I(U) &= \int_{t_1}^{t_2} [L(r(U + \Delta U, t), \dot{r}) - \Delta U + \Delta U]\, dt \\ &= I(U + \Delta U) + \int_{t_1}^{t_2} \Delta U\, dt. \end{aligned}$$

Denoting the time average of $F(t)$ by

$$\langle F\rangle_{\text{av}} \equiv (t_2 - t_1)^{-1}\int_{t_1}^{t_2} F\, dt, \quad (t_2 - t_1) \to \infty, \tag{37}$$

we obtain

$$-\langle \Delta L\rangle_{\text{av}} = \langle \Delta U\rangle_{\text{av}}. \tag{38}$$

In particular, if the potential U depends on a parameter α, we have

$$-\left\langle \frac{\partial L}{\partial \alpha}\right\rangle_{\text{av}} = \left\langle \frac{\partial U}{\partial \alpha}\right\rangle_{\text{av}}. \tag{38'}$$

(b) *Changing the scale:* Let $r_t(t) = (1 + \epsilon)r(t)$ where ϵ is an infinitesimal constant. Note that $r_t(t)$ does not go through the end points r_1, r_2.

$$\begin{aligned} \delta I &= \int_{t_1}^{t_2}\left(\frac{\partial L}{\partial r}\,\delta r + \frac{\partial L}{\partial \dot{r}}\,\delta \dot{r}\right) dt \\ &= \int_{t_1}^{t_2}\left(\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}}\right)\delta r\, dt + \frac{\partial L}{\partial \dot{r}}\,\delta r\Big|_{t_1}^{t_2} \\ &= \frac{\partial L}{\partial \dot{r}}\,\delta r\Big|_{t_1}^{t_2}. \end{aligned}$$

Since $\delta r = \epsilon r$, $\delta \dot{r} = \epsilon \dot{r}$, we obtain

$$\int_{t_1}^{t_2}\left(r\frac{\partial L}{\partial r} + \dot{r}\frac{\partial L}{\partial \dot{r}}\right) dt = r\frac{\partial L}{\partial \dot{r}}\Big|_{t_1}^{t_2}. \tag{39}$$

Dividing by $(t_2 - t_1)$ where $(t_2 - t_1) \to \infty$, we have (since the motion is bounded)

$$\left\langle r\frac{\partial L}{\partial r} + \dot{r}\frac{\partial L}{\partial \dot{r}}\right\rangle_{\text{av}} = \left\langle -r\frac{dU}{dr} + 2T\right\rangle_{\text{av}} = 0.$$

Finally, using the relations

$$2T = 2E - 2U \quad \text{and} \quad 2U + r\frac{dU}{dr} = 2V + r\frac{dV}{dr},$$

we obtain the virial theorem

$$2E = \left\langle 2V + r\frac{dV}{dr}\right\rangle_{\text{av}}. \tag{40}$$

(c) *Changing the angular momentum:* We have by Eq. (38')

$$\left\langle -\frac{\partial L}{\partial l}\right\rangle_{\text{av}} = \frac{l}{m}\langle r^{-2}\rangle_{\text{av}}. \tag{41}$$

(d) *Changing the mass:* Inserting $r_t(t) = r(m + \Delta m, t)$ into Eq. (35), we have

$$\begin{aligned} I(m) &= \int_{t_1}^{t_2}\left[\tfrac{1}{2}(m + \Delta m)\dot{r}_t^2 - V(r_t) - \frac{l^2}{2(m + \Delta m)r_t^2}\right. \\ &\quad \left. - \tfrac{1}{2}\Delta m\dot{r}^2 - \frac{\Delta m}{m}\frac{l^2}{2mr^2}\right] dt \\ &= I(m + \Delta m) - \frac{\Delta m}{m}\int_{t_1}^{t_2}(E - V)\, dt. \end{aligned} \tag{42}$$

Dividing by $(t_2 - t_1)$ and using the virial theorem (40), we secure

$$\left\langle -\frac{\partial L}{\partial m}\right\rangle_{\text{av}} = -(2m)^{-1}\left\langle r\frac{dV}{dr}\right\rangle_{\text{av}}. \tag{43}$$

The present derivation of the virial theorem [Eq. (40)] has been suggested by Fock and Krutkov.[16]

B. Scattered States

Since in a scattered state the particle eventually leaves the field of force, time averages like Eq. (38) are useless even if they exist. We shall, therefore, express our relations in terms of time integrals rather than time averages.

(a) *Changing the potential:* From (38) and (38′) we have

$$\int_{t_1}^{t_2} \Delta L\, dt = -\int_{t_1}^{t_2} \Delta U\, dt, \tag{44}$$

$$\int_{t_1}^{t_2} \frac{\partial L}{\partial \alpha}\, dt = -\int_{t_1}^{t_2} \frac{\partial U}{\partial \alpha}\, dt. \tag{44'}$$

(b) *Changing the scale:* As $(t_2 - t_1) \to \infty$ Eq. (39) diverges. However, by subtracting $2E = 2T + 2U = mv_0^2$ (v_0 being the magnitude of the initial velocity) from both sides of Eq. (39) we obtain a relation which remains finite for all times:

$$-\int_{t_1}^{t_2}\left(2U + r\frac{dU}{dr}\right) dt = -\int_{t_1}^{t_2}\left(2V + r\frac{dV}{dr}\right) dt = mvr - mv_0^2 t\Big|_{t_1}^{t_2}. \tag{45}$$

For $t \to \infty$, $v = v_0$ and $r = v_0 t + s$. By choosing the time scale such that the closest approach occurs at $t = 0$, we obtain a symmetric orbit satisfying $r(-t) = r(t)$. Letting $t_2 \to \infty$ and $t_1 \to -\infty$ in Eq. (45), we have

$$-2\int_0^{\infty}\left(2V + r\frac{dV}{dr}\right) dt = \lim_{t\to\infty}\{[mv_0(v_0 t + s) - mv_0^2 t] - [-mv_0(v_0 t + s) + mv_0^2 t]\} = 2mv_0 s.$$

Hence

$$-\frac{s}{v_0} = (2E)^{-1}\int_0^{\infty}\left(2V + r\frac{dV}{dr}\right) dt. \tag{46}$$

The quantity s has a simple physical interpretation.[1] It is the distance by which the scattered particle overtakes a free particle which starts at $t = 0$ from the origin and moves with a speed v_0.

(c) *Changing the angular momentum:* By Eq. (41) we have

$$\int_{t_1}^{t_2} \frac{\partial L}{\partial l}\, dt = -\frac{l}{m}\int_{t_1}^{t_2} r^{-2}\, dt. \tag{47}$$

(d) *Changing the mass:* Using Eq. (42) and the virial theorem (45), we obtain

$$\int_0^{\infty}\left(\frac{\partial L}{\partial m} - \frac{E}{m}\right) dt - \tfrac{1}{2}sv_0 = (2m)^{-1}\int_0^{\infty} r\frac{dV}{dr}\, dt. \tag{48}$$

IV. DISCUSSION

A glance at the results of Sec. IIB and Sec. IIIA reveals that the quantum-mechanical binding energy corresponds to minus the time-average of the classical Lagrangian. This is to be expected on the basis of the WKB approximation. Indeed, the WKB approximation for $\psi(\mathbf{r}, t) = \exp[(-i/\hbar)Et]\phi(\mathbf{r})$ is[17]

$$\psi^{\mathrm{WKB}}(\mathbf{r}, t) = A \exp[(i/\hbar)W(\mathbf{r}, t)],$$

where $W(\mathbf{r}, t)$ is a solution of the Hamilton–Jacobi equation

$$\frac{\partial W}{\partial t} + (2m)^{-1}(\nabla W)^2 + V(r) = 0. \tag{49}$$

A complete solution of Eq. (49) is given by[18]

$$W(\mathbf{r}, t) = \int_0^t L[\mathbf{r}(t), \dot{\mathbf{r}}(t)]\, dt \equiv \langle L\rangle_{\mathrm{av}} \cdot t,$$

where $\mathbf{r}(t)$ is the physical trajectory passing through $\mathbf{r}$ and $\mathbf{r}_0$ at times t and $t = 0$. Comparing the logarithmic derivative of the exact ψ with that of ψ^{WKB}, we have

$$\frac{\partial}{\partial t}(E \cdot t) \approx -\frac{\partial}{\partial t}(\langle L\rangle_{\mathrm{av}} \cdot t). \tag{50}$$

It is worthwhile to pursue Eq. (50) a little further. Using our previous notation, we have the well-known classical result[18]

$$\frac{\partial}{\partial t_2} I(\mathbf{r}_2 t_2, \mathbf{r}_1 t_1) \equiv \frac{\partial}{\partial t_2}\int_{t_1}^{t_2} L(\mathbf{r}, \dot{\mathbf{r}})\, dt = -E(\mathbf{r}_2 t_2, \mathbf{r}_1 t_1). \tag{51}$$

Integrating the last equation between $t_2' = t_0$ and $t_2' = t_2$, we have $(t_0 > t_1)$

$$I(\mathbf{r}_2 t_2, \mathbf{r}_1 t_1) - I(\mathbf{r}_2 t_0, \mathbf{r}_1 t_1) = -\int_{t_0}^{t_2} E(\mathbf{r}_2 t_2', \mathbf{r}_1 t_1)\, dt_2'. \tag{52}$$

In particular, let $t_0 = t_1 + \epsilon$, where $\epsilon > 0$ is an infinitesimal quantity. The quantity $I(\mathbf{r}_2 t_1 + \epsilon, \mathbf{r}_1 t_1)$ diverges as $\epsilon \to 0+$, since the speed of the particle must increase indefinitely in order to go a finite distance $|\mathbf{r}_2 - \mathbf{r}_1|$ in an infinitesimal amount of time. However, it diverges in a well-defined manner:

$$I(\mathbf{r}_2 t_1 + \epsilon, \mathbf{r}_1 t_1) = \int_{t_1}^{t_1+\epsilon} (T - V)\, dt = \epsilon\left[\frac{\tfrac{1}{2}m(\mathbf{r}_2 - \mathbf{r}_1)^2}{\epsilon^2} - V(\mathbf{r}_1)\right] = \frac{\tfrac{1}{2}m(\mathbf{r}_2 - \mathbf{r}_1)^2}{\epsilon}. \tag{53}$$

Substituting Eq. (53) into (52) and dividing both sides by $(t_2 - t_1)$, we obtain

$$\langle L\rangle_{\rm av} = -\lim_{\epsilon\to 0+}\lim_{(t_2-t_1)\to\infty}(t_2-t_1)^{-1}\int_{t_1+\epsilon}^{t_2} E(\mathbf{r}_2 t_2', \mathbf{r}_1 t_1)\,dt_2' \equiv -\langle E_{\rm cl}\rangle. \tag{54}$$

We are thus led to the definition of a new kind of average, namely, the *path average* of the classical energy. By definition, it is the average of the energy over all classical paths traversed by a particle as the end time t_2' ranges between t_1 and t_2. However, in order to secure a finite result, the limiting process described in Eq. (54) must be followed. It is this path average of the classical energy which corresponds to the quantum-mechanical binding energy:

$$E_{\rm q.m.} \leftrightarrow \langle E_{\rm cl}\rangle.$$

We turn now to examine the correspondence between the classical and quantum-mechanical results for the scattered states. This can be best understood using the framework of the WKB approximation. The WKB phase shift is[14]

$$\delta_l = \tfrac{1}{4}\pi + \tfrac{1}{2}l\pi - kr_0 + \int_{r_0}^{\infty}[k(r) - k]\,dr, \tag{55}$$

where

$$k = (2mE/\hbar^2)^{\frac{1}{2}},$$

$$k(r) = [2m/\hbar^2(E - V - \hbar^2(l+\tfrac{1}{2})^2/(2mr^2))]^{\frac{1}{2}},$$

and r_0 satisfies $k(r_0) = 0$. In Eq. (55), δ and r_0 depend parametrically on α, E, l, m, and $\hbar$. However, when taking the derivatives of Eq. (55) with respect to a parameter ξ, the quantity r_0 can be treated as a constant. This is true because

$$-k\frac{\partial r_0}{\partial \xi} - [k(r_0) - k]\frac{\partial r_0}{\partial \xi} = 0.$$

By differentiating Eq. (55), we obtain integrals of the form $\int_{r_0}^{\infty} F(r)k(r)^{-1}\,dr$. These quantities are closely related to the classical time integrals. In fact, if we associate with the quantum-mechanical angular momentum $\hbar l$ a classical angular momentum $l_{\rm cl}$ satisfying $l_{\rm cl} \equiv \hbar(l + \tfrac{1}{2})$, we have

$$\int_{r_0}^{\infty} F(r)k(r)^{-1}\,dr = \frac{\hbar}{m}\int_0^{\infty} F(r(t))\,dt,$$

where $r(t)$ is the classical orbit of a particle with energy E and angular momentum $l_{\rm cl}$. Performing the differentiations with respect to the listed parameters, we obtain

$$\frac{\partial\delta_l}{\partial\alpha} = -\frac{m}{\hbar^2}\int_{r_0}^{\infty}\frac{\partial V}{\partial\alpha}k(r)^{-1}\,dr = -\hbar^{-1}\int_0^{\infty}\frac{\partial V}{\partial\alpha}\,dt, \tag{56}$$

$$\frac{\partial\delta_l}{\partial k} = -r_0 + \int_{r_0}^{\infty}\left(\frac{k}{k(r)} - 1\right)dr = -r_0 - \int_0^{\infty}(v - v_0)\,dt, \tag{57}$$

$$\frac{\partial\delta_l}{\partial l} = \tfrac{1}{2}\pi - \int_{r_0}^{\infty}(l+\tfrac{1}{2})r^{-2}k(r)^{-1}\,dr = \tfrac{1}{2}\pi - \frac{l_{\rm cl}}{m}\int_0^{\infty} r^{-2}\,dt, \tag{58}$$

$$\frac{\partial\delta_l}{\partial m} = -\frac{r_0 k}{2m} + \int_{r_0}^{\infty}\left(\frac{k(r)}{2m} + [h^2k(r)]^{-1}\frac{l_{\rm cl}^2}{2mr^2} - \frac{k}{2m}\right)dr = \left[\delta_l - \left(\frac{l_{\rm cl}}{\hbar}\right)\frac{\partial\delta_l}{\partial l}\right]\Big/2m, \tag{59}$$

$$\frac{\partial\delta_l}{\partial\hbar} = -\left[\delta_l - \left(\frac{l_{\rm cl}}{\hbar}\right)\frac{\partial\delta_l}{\partial l}\right]\Big/\hbar. \tag{60}$$

The last two equations can be reduced to our normal form, namely, that of Eqs. (7) and (8). The first step is to rewrite Eq. (55) as

$$\delta_l = \tfrac{1}{4}\pi + \tfrac{1}{2}l\pi + \int_{r_0}^{r} k(r)\,dr - kr, \quad r\to\infty.$$

Converting the last integral into a time integral over the classical path and using $r = v_0 t + s$ as well as Eq. (45), we obtain

$$\hbar\delta_l = \tfrac{1}{2}l_{\rm cl}\pi + \int_0^{\infty} r\frac{dU}{dr}\,dt,$$

or, using Eq. (58),

$$\hbar\delta_l = l_{\rm cl}\frac{\partial\delta_l}{\partial l} + \int_0^{\infty} r\frac{dV}{dr}\,dt.$$

Finally, inserting the last equation into (59) and (60), we secure

$$\frac{\partial\delta_l}{\partial m} = (2m\hbar)^{-1}\int_0^{\infty} r\frac{dV}{dr}\,dt, \tag{59'}$$

$$\frac{\partial\delta}{\partial\hbar} = -\hbar^{-2}\int_0^{\infty} r\frac{dV}{dr}\,dt. \tag{60'}$$

Equations (56)–(60) are the semiclassical counterpart of the results derived in Sec. IIIB. It was mentioned earlier that the quantity $\hbar\partial\delta_l/\partial E = (\partial\delta_l/\partial k)/v_0$

[Eq. (5)] can be interpreted as the time delay of the lth scattered partial wave in the potential well. Comparing with the classical result Eq. (45), we should expect the correspondence $\partial\delta_l/\partial k \leftrightarrow -s$. This is indeed the case as can be seen from Eq. (57). For $t \to \infty$,

$$-r_0 - \int_0^t (v - v_0)\, dt = v_0 t - r(t) = -s.$$

Hence, in the semiclassical limit

$$\frac{\partial\delta_l}{\partial k} = -s.$$

The interpretation of Eq. (58) has already been discussed. [See Eq. (20).]

Finally, we shall use Eqs. (59) and (59′) to derive a semiclassical limit on the phase shifts. From Eq. (59′) we see that $\partial\delta_l/\partial m$ is positive for an attractive force and negative for a repulsive force. Also, δ_l is positive or negative according to the force being attractive or repulsive.[8] Using this fact in Eq. (59), we have in either case

$$\frac{dl}{(l + \frac{1}{2})} \geq \frac{d\delta_l}{\delta_l}.$$

Integrating the last inequality, we obtain

$$\frac{\delta_l}{\delta_0} \leq 2l + 1. \tag{61}$$

V. SUMMARY

Using a technique developed by Fock and Demkov, variational principles have been applied to calculate the derivatives of the stationary quantities with respect to the parameters of the system. The method is by no means limited to the variational principles employed in this article. Comparison of the quantum-mechanical results to the corresponding classical results leads to the introduction of a new kind of classical average, namely, the path-average of the classical energy. In view of the resemblance of this concept to Feynman's path integral method in quantum mechanics, it is felt that further investigations along these lines could be worthwhile.

ACKNOWLEDGMENT

I am indebted to Dr. D. Harrington for fruitful discussions.

* Work supported in part by the National Science Foundation.

† Present address: Department of Theoretical Physics, The Hebrew University of Jerusalem, Jerusalem, Israel.

[1] Yu. N. Demkov, *Variational Principles in the Theory of Collisions* (Macmillan, New York, 1963) and references therein.

[2] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[3] W. Kohn, Phys. Rev. **74**, 1763 (1948).

[4] L. Hulthén, Kgl. Fysiograf, Sallskapet. Lund Forh. **14**, No. 21 (1944).

[5] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

[6] R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B472 (1964).

[7] N. Austern and J. S. Blair, Ann. Phys. (N.Y.) **33**, 15 (1965) and references therein.

[8] L. Spruch, in *Lectures in Theoretical Physics*, W. E. Brittin, B. W. Downs, and J. Downs, Eds. (Interscience, New York, 1962), Vol. IV.

[9] Y. Tikochinsky, Phys. Letters, **29B**, 270 (1969).

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

[11] E. P. Wigner, Phys. Rev. **98**, 145 (1955).

[12] R. G. Newton, *The Complex j-Plane* (Benjamin, New York, 1964).

[13] T. Regge, Nuovo Cimento **14**, 951 (1959).

[14] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford U.P., New York, 1965).

[15] V. A. Fock, Z. Phys. **63**, 855 (1930).

[16] V. A. Fock and Yu. A. Krutkov, Phys. Z. Sowjetunion **1**, 756 (1932).

[17] See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955).

[18] See, for example, H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Labeling States and Constructing Matrix Representations of G_2

J. PATERA
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Canada

(Received 9 March 1970)

A method is developed for labeling G_2-internal states and for finding the matrices representing G_2-generators. The simple Lie algebra G_2 is embedded into A_6, whose representation spaces are labeled by Gel'fand patterns. For a given irreducible finite-dimensional representation $\Psi(G_2)$ of G_2, an optimal representation $\Phi(A_6) \supset \Psi(G_2)$ is chosen, and a lemma is formulated which enables us to select the subspace $R(\Psi)$ from the representation space $R(\Phi)$.

I. INTRODUCTION

One of the problems frequently encountered in applications of simple Lie groups and their finite-dimensional representations to nuclear and elementary particle physics is to distinguish (label) different vectors (internal states) of a basis in a representation space and to find matrices representing elements of the corresponding group. One natural labeling makes use of the weights. Unfortunately, their multiplicities are often higher than one so that they label subspaces of the representation space which are not, in general, 1 dimensional.

The state-labeling problem has been solved for important particular cases: the groups of unitary,[1] orthogonal,[2] and symplectic[3] matrices. In the first two cases the matrix representations were found explicitly. The solutions are based on the possibility of finding, in each case, a chain of subgroups $G_i \supset G_{i-1}$, $i = 1, 2, \cdots$, such that any irreducible representation $\Phi(G_i)$, when reduced to $\Phi(G_{i-1}) = \Psi_1(G_{i-1}) \oplus \Psi_2(G_{i-1}) \oplus \cdots$, contains each irreducible component $\Psi_j(G_{i-1})$ at most once. However, similar chains of subgroups do not exist for the five types of exceptional simple Lie groups. Therefore other methods are needed in these cases.

One such labeling method that has been used in special cases[4] was recently generalized[5] to apply, in principle, to any compact simple Lie group. It describes the basis vectors of the representation space as products of certain elementary factors. Practically, however, the use of the method is limited to rather low-dimensional representations of groups of low ranks. The difficulties here are in finding the elementary factors necessary to label different representations and in elimination of certain unwanted states.

The idea of our method is very simple. For complete labeling of representation spaces of a simple Lie group G, it is sufficient to embed G into a suitable classical group $\hat{G}$ which has all representation spaces labeled completely and to select out from the representation space of $\hat{G}$ its subspace in which an irreducible representation of subgroup G acts. Unlike the method of elementary factors, our approach enables us easily to find an explicit form of the matrices representing G by making use of the Gel'fand and Tsetlin results.[1,2]

To facilitate the use of our method it appears advantageous to consider each exceptional simple Lie group separately. Because there is a one-to-one correspondence between a simple Lie algebra and its connected Lie group and also because the representation spaces of the algebra and the group obviously coincide, we consider the Lie algebras and their representations only.

The purpose of this paper is to solve the state-labeling problem and to find explicitly matrix representations of the simple Lie algebra G_2. Our method requires only a standard use of Gel'fand patterns.

There are two possible ways to proceed in the case of G_2. Indeed, the lowest-dimensional representation of G_2 is orthogonal,[6] so that one can choose either A_6 or B_3[7] as the simple Lie algebra containing G_2. According to that choice, the G_2-states will be labeled either by Gel'fand patterns of A_6 or B_3. Our choice is A_6 because it has a particularly simple structure of the patterns needed in our problem [cf. (16)] and because other exceptional simple Lie algebras can be conveniently embedded into an algebra of type A_n.

Section II contains some mathematical preliminaries. In Sec. III an embedding of G_2 into A_6 is explicitly specified by finding a projection of A_6-root space onto the root space of G_2. For a given representation $\Psi(G_2)$ an optimal choice of representation $\Phi(A_6) \supset \Psi(G_2)$ is made. Section IV contains the state-labeling lemma and linear transformations determining the matrices of generators. In Sec. V two examples are considered. In the last section the relation between the completely labeled states of G_2 and the corresponding states of each simple subalgebra of G_2 is given explicitly.

II. MATHEMATICAL PRELIMINARIES

Let us first make several conventions and recall a few facts about simple Lie algebras. So that we can refer to this section in considering other cases, we do not restrict ourselves by the algebras of types A_6 and G_2 only.

Suppose that a system of simple roots $\{\alpha_i\}$, $i = 1, 2, \cdots, n$, has been chosen for any simple Lie algebra G and the roots are numbered as in Fig. 1 of Ref. 7. Let the system $\{\alpha_i\}$, which forms a basis in a real Euclidean space $R(G)$, be normalized by the condition $(\alpha_{\max}, \alpha_{\max}) = 2$, where $\alpha_{\max}$ is the longest simple root of G and $(\,,\,)$ denotes the scalar product in $R(G)$.

It is convenient to use as well another basis $\{\nu_i\}$, $i = 1, 2, \cdots, n$, in $R(G)$, related to $\{\alpha_i\}$ by

$$\alpha_i = 2\sum_{j=1}^{n} \frac{(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} \nu_j \quad \text{and by} \quad (\alpha_i, \nu_j) = \tfrac{1}{2}(\alpha_i, \alpha_i)\delta_{ij}. \tag{1}$$

The coordinates a_i and a^i of a vector

$$M = \sum_{i=1}^{n} a_i \nu_i = \sum_{i=1}^{n} a^i \alpha_i \in R(G)$$

are related as

$$a_i = \frac{2}{(\alpha_i, \alpha_i)}(M, \alpha_i) = \frac{4}{(\alpha_i, \alpha_i)} \sum_{j=1}^{n} \frac{(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} a^j. \tag{2}$$

In particular, the weights of a representation of G are vectors of $R(G)$ with integer coordinates a_i.

Let us now restrict ourselves to the simple Lie algebras of type A_n.

A Gel'fand pattern[1]

$$(g) = \begin{pmatrix} m_{1,n+1} & m_{2,n+1} & \cdots & & \cdots & m_{n+1,n+1} \\ & & & \cdot & & \\ & & & \cdot & & \\ & & & \cdot & & \\ & & m_{12} & & m_{22} & \\ & & & m_{11} & & \end{pmatrix}, \tag{3}$$

where m_{pq} are arbitrary integers such that

$$m_{p,q+1} \geq m_{pq} \geq m_{p+1,q+1}, \quad \text{and} \quad m_{n+1,n+1} = 0, \tag{4}$$

is a vector of the space $R(\Phi(A_n))$ in which an irreducible representation $\Phi(A_n)$ acts. It is also a weight vector of $\Phi(A_n)$ with weight M given by

$$\begin{aligned} M &= \sum_{i=1}^{n} \nu_i \left\{ -\sum_{k=1}^{i+1} m_{k,i+1} + 2\sum_{k=1}^{i} m_{ki} - \sum_{k=1}^{i-1} m_{k,i-1} \right\} \\ &= \sum_{i=1}^{n} \alpha_i \left\{ \sum_{k=1}^{i} m_{ki} - \frac{i}{n+1} \sum_{k=1}^{n+1} m_{k,n+1} \right\}, \end{aligned} \tag{5}$$

as can be verified, for instance, by comparing the set of patterns (3) corresponding to the same irreducible representation $\Phi(A_n)$ with the weight system of $\Phi(A_n)$.

If M is the highest weight of $\Phi(A_n)$, integers m_{pq} in (3) take their maximal values $m_{pq} = m_{p,n+1}$ for any $p \leq q \leq n+1$. Therefore, instead of (5), one has

$$\begin{aligned} M &= \sum_{i=1}^{n} \nu_i (m_{i,n+1} - m_{i+1,n+1}) \\ &= \sum_{i=1}^{n} \alpha_i \left(\frac{n+1-i}{n+1} \sum_{k=1}^{i} m_{k,n+1} - \frac{i}{n+1} \sum_{k=i+1}^{n+1} m_{k,n+1} \right). \end{aligned} \tag{5'}$$

III. EMBEDDING OF G_2 INTO A_6

In this section the embedding of G_2 into A_6 is specified by an explicit projection f^* of the 6-dimensional root space $R(A_6)$ onto the 2-dimensional $R(G_2)$. For any given irreducible representation $\Psi(G_2)$, we find the lowest-dimensional representation $\Phi(A_6)$ such that

$$\Phi(A_6) \supset \Phi(G_2) = \Psi(G_2) \oplus \Psi_1(G_2) \oplus \Psi_2(G_2) \oplus \cdots, \tag{6}$$

where $\Phi(G_2)$ is the reducible representation of G_2 induced by $\Phi(A_6)$ and the $\Psi_i(G_2)$, $i = 1, 2, \cdots$, are other irreducible components of it. Few details of the calculation are given because it is a particular case of a general method.[7]

Let us denote the simple roots of G_2 by β_1 and β_2, and the basis of $R(G_2)$ conjugate to $\{\beta_1, \beta_2\}$ by $\{\tau_1, \tau_2\}$. According to (1), we have

$$\tau_1 = 2\beta_1 + 3\beta_2 \quad \text{and} \quad \tau_2 = \beta_1 + 2\beta_2. \tag{7}$$

To find the projection $f^*(R(A_6)) = R(G_2)$ one has to insert the weights

$$\begin{aligned} &M_1 = \tau_2, \quad M_2 = \tau_1 - \tau_2, \quad M_3 = -\tau_1 + 2\tau_2, \\ &M_4 = 0, \quad M_5 = \tau_1 - 2\tau_2, \quad M_6 = -\tau_1 + \tau_2 \end{aligned} \tag{8}$$

of the lowest-dimensional representation of G_2 into the Table I of Ref. 7. From that table one has immediately

$$\begin{aligned} f^*(\nu_1) = \tfrac{1}{2} f^*(\nu_3) = \tfrac{1}{2} f^*(\nu_4) = f^*(\nu_6) = \tau_2, \\ f^*(\nu_2) = f^*(\nu_5) = \tau_1. \end{aligned} \tag{9}$$

Thus a vector $M \in R(A_6)$ is projected as follows:

$$\begin{aligned} f^*(M) &= \tau_1(a_2 + a_5) + \tau_2(a_1 + 2a_3 + 2a_4 + a_6) \\ &= \beta_1(a^2 + a^5) + \beta_2(a^1 + a^3 + a^4 + a^6), \end{aligned} \tag{10}$$

where the coordinates a_i and a^i are related by (2).

According to Theorem 0.11 of Ref. 8, we have

$$f^*\{\Delta[\Phi(A_6)]\} = \Delta[\Phi(G_2)], \tag{11}$$

where $\Delta[\Phi(G)]$ is the weight system of the representation Φ of G, and also

$$(g^M) = (g^{f^*(M)}), \tag{12}$$

where M and $f^*(M)$ indicate weights of (g). Consequently, the same pattern (g), which is a weight vector with an A_6-weight $M \in \Delta(\Phi(A_6))$ given by (5),

is also a G_2-weight vector corresponding to the weight $f^*(M) \in \Delta(\Phi(G_2))$.

The remaining problem to be solved in this section is to find for any given irreducible $\Psi(G_2)$ a representation $\Phi(A_6)$ such that (6) holds. Obviously, many representations of A_6 satisfy that requirement. Therefore, it is a matter of convenience to choose from them the one of lowest dimension and simplest structure of its Gel'fand patterns (3). Our choice is made by the following:

Lemma 1: Let $\Psi(G_2)$ and $\Phi(A_6)$ be irreducible representations of G_2 and A_6 with the highest weights

$$N_h = a_1\tau_1 + a_2\tau_2 \quad \text{and} \quad M_h = a_2\nu_1 + a_1\nu_2, \quad (13)$$

respectively. Then:

(a) $\Phi(A_6) \supset \Psi(G_2)$;

(b) if also $\Omega(A_6) \supset \Psi(G_2)$, then the dimension of $\Omega(A_6)$ is not lower than that of $\Phi(A_6)$;

(c) patterns (3) form a basis in $R(\Phi(A_6))$ provided

$$m_{ij} = 0 \quad \text{for any} \quad i \leq j \quad \text{and} \quad i = 3, 4, 5, 6, \text{ and } 7, \quad (14)$$

$$m_{17} = a_1 + a_2 \quad \text{and} \quad m_{27} = a_1. \quad (15)$$

Proof: Assertion (a) follows immediately from (9) and (11). Assertion (b) is verified by calculating the dimensions of all eight nonequivalent A_6-representation whose highest weights $M^{(i)}$, $i = 1, 2, \cdots, 8$, satisfy $f^*(M^{(i)}) = N_h$. Any other representation which has $M^{(i)}$ in its weight system but not as the highest weight is obviously of higher dimension than $\Phi(A_6)$. Finally, (c) follows from the form (13) of the highest weight M_h and from (4).

Because in the following we shall deal only with patterns satisfying (14) and (15), we can save considerable space by rearranging their nonzero elements and rewriting them as

$$\begin{pmatrix} a_1 + a_2 & m_{16} & m_{15} & m_{14} & m_{13} & m_{12} & m_{11} \\ a_1 & m_{26} & m_{25} & m_{24} & m_{23} & m_{22} & \end{pmatrix}. \quad (16)$$

We shall omit also the first column containing coordinates of the highest weight whenever this leads to no ambiguity. The A_6-weight M of (16) is given by (5). The G_2-weight N of (16) follows from (5) and (10):

$$\begin{aligned} N = f^*(M) = \tau_1\Big(-m_{11} + \sum_{j=1}^{2}(2m_{j2} - m_{j3} \\ - m_{j4} + 2m_{j5} - m_{j6})\Big) \\ + \tau_2\Big(2m_{11} + \sum_{j=1}^{2}(-3m_{j2} + 2m_{j3} + 2m_{j4} \\ - 3m_{j5} + 2m_{j6}) - 2a_1 - a_2\Big). \end{aligned} \quad (17)$$

IV. LABELING STATES AND CONSTRUCTING MATRIX REPRESENTATIONS

We start with the representation $\Phi(A_6)$ found in Sec. III for any given irreducible representation $\Psi(G_2)$, and with its representation space $R(\Phi)$ spanned by patterns (16). First, we express the G_2-generators as linear combinations of generators of A_6. Since all matrix representations of the latter algebra have been found explicitly,[1] we have thus obtained explicitly a (reducible) representation $\Phi(G_2)$. In order to find the subspace $R(\Psi)$ of $R(\Phi)$, which is irreducible with respect to the given representation $\Psi(G_2)$, we first find one vector $(g_h) \in R(\Psi)$. Applying linear transformations from $\Phi(G_2)$ to (g_h), one can generate the whole subspace $R(\Psi)$. In addition, this procedure automatically selects from the reducible representation $\Phi(G_2)$ its irreducible component $\Psi(G_2)$.

Following the notation of Gel'fand and Tsetlin,[1] the A_6-generators as represented in $\Phi(A_6)$ are matrices I_{ij} and $I_{ii} - I_{i+1,i+1}$, where $i \neq j$ and $i, j = 1, 2, \cdots, 7$, with the following commutation relations:

$$\begin{aligned} &[I_{ik}, I_{kp}] = I_{ip}, \quad i \neq p, \\ &[I_{ik}, I_{ki}] = I_{ii} - I_{kk}, \\ &[I_{ik}, I_{pj}] = 0, \quad \text{if} \quad k \neq p, \quad i \neq j. \end{aligned}$$

The coefficients of linear transformations $I_{ij}(g)$ are known explicitly for all patterns (g) and all values of the indices.[1] Consequently, the matrices I_{ij} are known.

It can be verified, e.g., using the commutation relations, that the representation $\Phi(G_2)$ of G_2-generators is

$$\begin{aligned} H_{\beta_1} &= I_{22} - I_{33} + I_{55} - I_{66}, \\ H_{\beta_2} &= \tfrac{1}{3}(I_{11} - I_{22} + 2I_{33} - 2I_{55} + I_{66} - I_{77}), \\ E_{-\beta_1} &= I_{32} + I_{65}, \\ E_{-\beta_2} &= \frac{1}{\sqrt{3}}(I_{21} + \sqrt{2}\, I_{43} + \sqrt{2}\, I_{54} + I_{76}), \\ E_{-\beta_1-\beta_2} &= \frac{1}{\sqrt{3}}(-I_{31} + \sqrt{2}\, I_{42} - \sqrt{2}\, I_{64} + I_{75}), \\ E_{-\beta_1-2\beta_2} &= \frac{1}{\sqrt{3}}(-\sqrt{2}\, I_{41} + I_{63} + I_{52} - \sqrt{2}\, I_{74}), \\ E_{-\beta_1-3\beta_2} &= -I_{51} + I_{73}, \\ E_{-2\beta_1-3\beta_2} &= I_{72} + I_{61}, \end{aligned} \quad (18)$$

and the generators $E_{\beta_1}, E_{\beta_2}, \cdots$, etc., corresponding to positive roots, are obtained from $E_{-\beta_1}, E_{-\beta_2}, \cdots$, etc., by permutation of indices of each I_{ij}.

Our solution of the G_2-state-labeling problem is contained in the following:

Lemma 2: Let $\Psi(G_2)$ be an irreducible representation of G_2 with the highest weight $N_h = a_1\tau_1 + a_2\tau_2$. Then a basis in the representation space $R(\Psi)$ consists of the vector

$$(g_h) = \begin{pmatrix} a_1 + a_2 & a_1 + a_2 & \cdots & a_1 + a_2 & a_1 + a_2 \\ a_1 & a_1 & & a_1 & \end{pmatrix} \tag{19}$$

and all linearly independent vectors

$$E_{-\beta_{i(k)}} \cdots E_{-\beta_{i(2)}} E_{-\beta_{i(1)}} (g_h), \tag{20}$$

where $i(k) = 1$ or 2 and $1 \leq k \leq 10a_1 + 6a_2$.

Proof: Let us first show that $(g_h) \in R(\Psi)$. According to our convention (16), (5'), and Lemma 1, the A_6-weight $M_h = a_2\nu_1 + a_1\nu_2$ of (g_h) is the highest weight of $\Phi(A_6)$. Its multiplicity is just one, and it is the only weight of $\Delta(\Phi(A_6))$ which is projected by f^* into N_h. Then, from (12), one concludes that (g_h) must belong to $R(\Psi)$. The linear transformations $E_{-\beta_1}$ and $E_{-\beta_2}$ are explicitly given by (18) and Ref. 1. Each operator $E_{-\beta_i}$ subtracts the simple root β_i, $i = 1, 2$, from the weight of the pattern. The assertion of the lemma then follows from Theorem 0.5 of Ref. 8. The maximal value of k is equal to the maximal number of simple roots β_i, $i = 1, 2$, which can be subtracted from the highest weight N_h. Its value may be found, for instance, in Ref. 6 or 8.

In fact, we have already solved also the second part of our problem, i.e., we have found the matrix elements of G_2-generators in the representation $\Psi(G_2)$. Indeed, the matrix elements are obtained by applying the linear transformation (18) to the vectors (19) and (20) of $R(\Psi)$.

Suppose (g) is one of the basis vectors (19) and (20). Then the generators in the representation $\Psi(G_2)$ are effectively given by the linear transformations

$$\begin{aligned}
H_{\beta_1}(g) &= \Big(-m_{11} + \sum_{j=1}^{2}(2m_{j2} - m_{j3} - m_{j4} \\
&\quad + 2m_{j5} - m_{j6})\Big)(g), \\
H_{\beta_2}(g) &= \tfrac{1}{3}\Big(2m_{11} + \sum_{j=1}^{2}(-3m_{j2} + 2m_{j3} + 2m_{j4} \\
&\quad - 3m_{j5} + 2m_{j6}) - 2a_1 - a_2\Big)(g), \\
E_{\beta_1}(g) &= \sum_{j=1}^{2}[a_{23}^{j}(g_{23}^{j}) + a_{56}^{j}(g_{56}^{j})], \\
E_{-\beta_1}(g) &= \sum_{j=1}^{2}[b_{32}^{j}(\bar{g}_{32}^{j}) + b_{65}^{j}(\bar{g}_{65}^{j})], \\
E_{\beta_2}(g) &= \frac{1}{\sqrt{3}}\Big(a_{12}^{1}(g_{12}^{1}) + \sum_{j=1}^{2}[\sqrt{2}\, a_{34}^{j}(g_{34}^{j}) \\
&\quad + \sqrt{2}\, a_{45}^{j}(g_{45}^{j}) + a_{67}^{j}(g_{67}^{j})]\Big), \\
E_{-\beta_2}(g) &= \frac{1}{\sqrt{3}}\Big(b_{21}^{1}(\bar{g}_{21}^{1}) + \sum_{j=1}^{2}[\sqrt{2}\, b_{43}^{j}(\bar{g}_{43}^{j}) \\
&\quad + \sqrt{2}\, b_{54}^{j}(\bar{g}_{54}^{j}) + b_{76}^{j}(\bar{g}_{76}^{j})]\Big).
\end{aligned} \tag{21}$$

Here $(g_{k-1,k}^{j})$ denotes a vector obtained from (g), where in each pattern $m_{j,k-1}$ has been replaced by $m_{j,k-1} + 1$. Similarly, $(\bar{g}_{k,k-1}^{j})$ is obtained from (g) by substituting $m_{j,k-1} - 1$ for $m_{j,k-1}$ in each pattern. If such a replacement should contradict the defining inequalities (4), the corresponding pattern would be zero. The coefficients a_{pq}^{j} and b_{qp}^{j} are given in Ref. 1. For the remaining generators one can either write the linear transformations using (18) as above or find the corresponding matrices by commutation of $E_{\pm\beta_1}$ and $E_{\pm\beta_2}$.

V. EXAMPLES

To illustrate our method, we consider two examples.

(1) Let $N_h = \tau_2$. Then, from (13), $M_h = \nu_1$ and both $\Psi(G_2)$ and $\Phi(A_6)$ are 7 dimensional.[8] Consequently, the state labeling is trivial: Both spaces $R(\Phi)$ and $R(\Psi)$ have common orthogonal basis consisting of patterns

$$(g) = \begin{pmatrix} m_{16} & m_{15} & \cdots & m_{11} \\ 0 & 0 & \cdots & \end{pmatrix}, \quad m_{1i} \leq m_{1,i+1} \leq 1. \tag{22}$$

The matrix representation of generators is obtained when (22) is used in (21).

(2) Let $N_h = \tau_1$. Then from (13), $M_h = \nu_2$ and the dimensions of $\Psi(G_2)$ and $\Phi(A_6)$ are 14 and 21, respectively.[8] The basis of $R(\Psi)$ is spanned by vectors

$$\begin{aligned}
(g_1) \equiv (g_h) &= \begin{pmatrix} 111111 \\ 11111 \end{pmatrix}, \\
(g_2) = E_{-\beta_1}(g_h) &\sim \begin{pmatrix} 111111 \\ 11110 \end{pmatrix}, \\
(g_3) = E_{-\beta_2}(g_2) &\sim \begin{pmatrix} 111110 \\ 11110 \end{pmatrix} + \sqrt{2}\begin{pmatrix} 111111 \\ 11100 \end{pmatrix}, \\
(g_4) = E_{-\beta_2}(g_3) &\sim \sqrt{2}\begin{pmatrix} 111110 \\ 11100 \end{pmatrix} + \begin{pmatrix} 111111 \\ 11000 \end{pmatrix}, \\
(g_5) = E_{-\beta_2}(g_4) &\sim \begin{pmatrix} 111110 \\ 11000 \end{pmatrix}, \\
(g_6) = E_{-\beta_1}(g_4) &\sim \sqrt{2}\begin{pmatrix} 111100 \\ 11100 \end{pmatrix} + \begin{pmatrix} 111111 \\ 10000 \end{pmatrix},
\end{aligned}$$

$$(g_7) = E_{-\beta_1}(g_5) \sim \begin{pmatrix}111100\\11000\end{pmatrix} + \begin{pmatrix}111110\\10000\end{pmatrix},$$

$$(g_8) = E_{-\beta_2}(g_6) \sim 2\begin{pmatrix}111100\\11000\end{pmatrix} + \begin{pmatrix}111110\\10000\end{pmatrix} + \begin{pmatrix}111111\\00000\end{pmatrix},$$

$$(g_9) = E_{-\beta_1}(g_7) \sim E_{-\beta_1}(g_8) \sim \begin{pmatrix}111100\\10000\end{pmatrix},$$

$$(g_{10}) = E_{-\beta_2}(g_7) \sim E_{-\beta_2}(g_8) \sim \sqrt{2}\begin{pmatrix}111000\\11000\end{pmatrix} + \begin{pmatrix}111110\\00000\end{pmatrix},$$

$$(g_{11}) = E_{-\beta_2}(g_9) \sim E_{-\beta}(g_{10}) \sim \sqrt{2}\begin{pmatrix}111000\\10000\end{pmatrix} + \begin{pmatrix}111100\\00000\end{pmatrix},$$

$$(g_{12}) = E_{-\beta_2}(g_{11}) \sim \begin{pmatrix}110000\\10000\end{pmatrix} + \sqrt{2}\begin{pmatrix}111000\\00000\end{pmatrix},$$

$$(g_{13}) = E_{-\beta_2}(g_{12}) \sim \begin{pmatrix}110000\\00000\end{pmatrix},$$

$$(g_{14}) = \begin{pmatrix}100000\\00000\end{pmatrix}. \tag{23}$$

Because the vectors (20) are not normalized, we have omitted common factors in (23); $\sim$ denotes proportionality. As in the first example, the matrix representation of generators is obtained when (23) is used in (21).

VI. SIMPLE SUBALGEBRAS OF G_2

It is in the nature of problems in physics that we are often interested not only in a particular group but also in some of its subgroups. Therefore it is useful to know not only a complete basis in $R(\Psi)$ but also its relation to representations of a subalgebra G contained in $\Psi(G_2)$. Once again one can repeat the embedding procedure of Sec. III for $G \subset G_2$, find the corresponding projection f^* of $R(G_2)$ onto $R(G)$, and refer to Dynkin's theorem in order to conclude that G_2-weight vectors are also weight vectors for G. Below, only results of such calculations are given (see the general method in Ref. 7), i.e., the G-weights of A_6-patterns (16) used for labeling of G_2-states.

There are five different simple subalgebras in G_2.[9] In order to specify completely each subalgebra, its type and index[9,10] j_G are indicated in each case.

(1) Subalgebra A_2; $j_G = 1$; γ_1 and γ_2 are the simple roots, and $\{\sigma_1, \sigma_2\}$ is the basis conjugate [cf. (1)] to $\{\gamma_1, \gamma_2\}$. Then

$$f^*(\tau_1) = \sigma_1 + \sigma_2 \quad \text{and} \quad f^*(\tau_2) = \sigma_2, \tag{24}$$

and consequently

$$\begin{aligned} f^*(N) = \sigma_1[&-m_{11} + 2(m_{12} + m_{22}) - (m_{13} + m_{23}) \\ &-(m_{14} + m_{24}) + 2(m_{15} + m_{25}) \\ &- (m_{16} + m_{26})] + \sigma_2[m_{11} - (m_{12} + m_{22}) \\ &+ m_{13} + m_{23} + m_{14} + m_{24} - (m_{15} + m_{25}) \\ &+ m_{16} + m_{26} - (2a_1 + a_2)], \end{aligned} \tag{25}$$

where N is given by (17).

(2) Subalgebra A_1; $j_G = 1$; γ is the simple root. Then

$$f^*(\tau_1) = \gamma \quad \text{and} \quad f^*(\tau_2) = \tfrac{1}{2}\gamma. \tag{26}$$

Using (17) and (26), one has

$$f^*(N) = \tfrac{1}{2}\gamma(m_{12} + m_{22} + m_{15} + m_{25} - 2a_1 - a_2). \tag{27}$$

(3) Subalgebra A_1; $j_G = 3$; γ is the simple root. Then

$$f^*(\tau_1) = \tfrac{3}{2}\gamma \quad \text{and} \quad f^*(\tau_2) = \gamma. \tag{28}$$

Using (17) and (28), one gets

$$\begin{aligned} f^*(N) = \tfrac{1}{2}\gamma(&m_{11} + m_{13} + m_{23} + m_{14} + m_{24} + m_{16} \\ &+ m_{26} - 4a_1 - 2a_2). \end{aligned} \tag{29}$$

(4) Subalgebra A_1; $j_G = 4$; γ is the simple root

$$f^*(\tau_1) = 2\gamma \quad \text{and} \quad f^*(\tau_2) = \gamma. \tag{30}$$

From (17) and (30), one has

$$f^*(N) = \gamma(m_{12} + m_{22} + m_{15} + m_{25} - 2a_1 - a_2). \tag{31}$$

(5) Subalgebra (principal) A_1; $j_G = 28$; γ is the simple root

$$f^*(\tau_1) = 5\gamma \quad \text{and} \quad f^*(\tau_2) = 3\gamma. \tag{32}$$

From (17) and (32) one gets

$$\begin{aligned} f^*(N) = \gamma(&m_{11} + m_{12} + m_{22} + m_{13} + m_{23} + m_{14} \\ &+ m_{24} + m_{15} + m_{25} + m_{16} + m_{26} - 6a_1 - 3a_2). \end{aligned} \tag{33}$$

ACKNOWLEDGMENT

I am grateful to Dr. A. K. Bose for several stimulating discussions.

[1] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR 71, 825 (1950); also Supplement II of 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). The results were rederived and signs of final formulas corrected in G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).

[2] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 1017 (1950); also Supplement I of Ref. 1.
[3] D. P. Zhelobenko, Usp. Mat. Nauk **17**, 27 (1962) [translated in Russian Math. Surv. **17**, 1 (1962), Chap. 13].
[4] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961); M. Moshinsky and V. S. Devi, J. Math. Phys. **10**, 455 (1969).
[5] R. T. Sharp and C. S. Lamm, J. Math. Phys. **10**, 2033 (1969).
[6] For example, see A. K. Bose and J. Patera, J. Math. Phys. **11**, 2231 (1970).
[7] A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).
[8] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 245 (1957).
[9] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 111 (1957), Tables 16 and 25.
[10] J. Patera, Nuovo Cimento **46**, 637 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Rough-Surface Scattering: Shadowing, Multiple Scatter, and Energy Conservation

P. J. LYNCH AND R. J. WAGNER
TRW Systems Group, Redondo Beach, California 90278

(Received 16 February 1970)

Multiple-scatter and shadowing effects are included in an extended theory of high-frequency scattering from a surface rough in one dimension. The single-scatter probability of slopes relation, corrected for shadowing, is an immediate consequence for any stationary random process. The double-scatter contribution (shadow corrected) is derived as well, and it provides a significant correction for surfaces with appreciable rms slope. The total power scattered by a perfectly reflecting rough surface is numerically evaluated as a test of energy conservation; the results show that the double-scatter formulation is substantially more accurate than the conventional single-scatter, unshadowed theory, particularly in the cases of large angles of incidence or very rough surfaces.

1. INTRODUCTION

The current interest in remote sensing[1,2] has focused attention on the interaction of radiation with surfaces. The classical theories[3,4] have dealt with surfaces of regular shape. The applications to the irregular, "rough" surfaces more appropriate to ocean and land, however, are usually based on assumptions tailored for computational ease. In particular, the Kirchhoff, or physical-optics, boundary conditions are used[5] to provide an expression for the angular scattered intensity which, upon subsequent ensemble averaging, reduces to a mathematically tractable form. The popular basis for the Kirchhoff approximation is that the surface curvature be negligible in a wavelength, but it is more instructive to view it as a "local" approximation. Thus, it is a valid representation when only the surface region in the immediate vicinity is responsible for the field at a surface point.[6] On the other hand, this approximation will give rise to serious errors when the nonlocal effects of shadowing and multiple scatter are prevalent. These latter effects are especially important near grazing incidence or for general angles of incidence on irregular surfaces with appreciable rms slopes.

The physical-optics theory provides some freedom in the choice of wavelength, but the additional restriction of geometrical optics yields an exceptionally simple angular distribution of scattered energy. Only those portions of the surface which can connect the incident and final directions by specular scatter take part in the scattering process. For a random rough surface, the average scattered intensity is proportional to the probability density of these specular slopes. This result has long been evident for normally distributed surfaces,[7] but Barrick[8] has extended the result to any stationary random process. Again, the theory is in error because of the neglect of multiple scatter and shadowing. Indeed, it is easy to show that energy is not conserved. The objective of this paper is a geometrical-optics theory of scatter from a 1-dimensional surface which contains all double-scatter corrections as well as a consistent formulation of shadowing. The vehicle for comparison with the single-scatter, unshadowed theory will be the energy-conservation integral. Though the choice of cylindrical symmetry is hardly physical, the results will certainly provide estimates of the severity of multiple scatter and/or shadowing as well as the critical range of angles for both effects.

The procedure is to make use of the incoherent nature of the scattering in the geometrical-optics realm and simply sum the scattered intensities from every (planar) surface element. We construct a ray trace, in principle, for each surface element and segregate the elements in classes according to the number of intersections the incident ray makes with

[2] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 1017 (1950); also Supplement I of Ref. 1.
[3] D. P. Zhelobenko, Usp. Mat. Nauk **17**, 27 (1962) [translated in Russian Math. Surv. **17**, 1 (1962), Chap. 13].
[4] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961); M. Moshinsky and V. S. Devi, J. Math. Phys. **10**, 455 (1969).
[5] R. T. Sharp and C. S. Lamm, J. Math. Phys. **10**, 2033 (1969).
[6] For example, see A. K. Bose and J. Patera, J. Math. Phys. **11**, 2231 (1970).
[7] A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).
[8] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 245 (1957).
[9] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 111 (1957), Tables 16 and 25.
[10] J. Patera, Nuovo Cimento **46**, 637 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Rough-Surface Scattering: Shadowing, Multiple Scatter, and Energy Conservation

P. J. LYNCH AND R. J. WAGNER
TRW Systems Group, Redondo Beach, California 90278

(Received 16 February 1970)

Multiple-scatter and shadowing effects are included in an extended theory of high-frequency scattering from a surface rough in one dimension. The single-scatter probability of slopes relation, corrected for shadowing, is an immediate consequence for any stationary random process. The double-scatter contribution (shadow corrected) is derived as well, and it provides a significant correction for surfaces with appreciable rms slope. The total power scattered by a perfectly reflecting rough surface is numerically evaluated as a test of energy conservation; the results show that the double-scatter formulation is substantially more accurate than the conventional single-scatter, unshadowed theory, particularly in the cases of large angles of incidence or very rough surfaces.

1. INTRODUCTION

The current interest in remote sensing[1,2] has focused attention on the interaction of radiation with surfaces. The classical theories[3,4] have dealt with surfaces of regular shape. The applications to the irregular, "rough" surfaces more appropriate to ocean and land, however, are usually based on assumptions tailored for computational ease. In particular, the Kirchhoff, or physical-optics, boundary conditions are used[5] to provide an expression for the angular scattered intensity which, upon subsequent ensemble averaging, reduces to a mathematically tractable form. The popular basis for the Kirchhoff approximation is that the surface curvature be negligible in a wavelength, but it is more instructive to view it as a "local" approximation. Thus, it is a valid representation when only the surface region in the immediate vicinity is responsible for the field at a surface point.[6] On the other hand, this approximation will give rise to serious errors when the nonlocal effects of shadowing and multiple scatter are prevalent. These latter effects are especially important near grazing incidence or for general angles of incidence on irregular surfaces with appreciable rms slopes.

The physical-optics theory provides some freedom in the choice of wavelength, but the additional restriction of geometrical optics yields an exceptionally simple angular distribution of scattered energy. Only those portions of the surface which can connect the incident and final directions by specular scatter take part in the scattering process. For a random rough surface, the average scattered intensity is proportional to the probability density of these specular slopes. This result has long been evident for normally distributed surfaces,[7] but Barrick[8] has extended the result to any stationary random process. Again, the theory is in error because of the neglect of multiple scatter and shadowing. Indeed, it is easy to show that energy is not conserved. The objective of this paper is a geometrical-optics theory of scatter from a 1-dimensional surface which contains all double-scatter corrections as well as a consistent formulation of shadowing. The vehicle for comparison with the single-scatter, unshadowed theory will be the energy-conservation integral. Though the choice of cylindrical symmetry is hardly physical, the results will certainly provide estimates of the severity of multiple scatter and/or shadowing as well as the critical range of angles for both effects.

The procedure is to make use of the incoherent nature of the scattering in the geometrical-optics realm and simply sum the scattered intensities from every (planar) surface element. We construct a ray trace, in principle, for each surface element and segregate the elements in classes according to the number of intersections the incident ray makes with

the surface. The average of the single-scatter intensity yields another general derivation of the probability of slopes theory, though it is automatically corrected for shadowing. The averaging of the double-scatter intensity is also carried through under the assumption of negligible correlation between the two surface scattering points. The result is complicated analytically, but it represents no difficulty to the computer. The integration over all physical angles of the scattered intensity yields the energy integral. A numerical evaluation for the double-scatter theory demonstrates energy conservation.

2. PARTIAL-SURFACE REPRESENTATION

We consider radiation of unit intensity and beam width S_0 incident on an irregular surface described by $z = \zeta(x, y) = \zeta(x)$. The incident and scattered rays are contained in the (x, z) plane, and the sense of the angles is clockwise (Figs. 1 and 2). We restrict the wavelength to the geometrical-optics limit; i.e., the wavelength is small compared to all surface parameters. The scattering process is incoherent in this limit as each surface element simply specularly reflects the incident radiation.[9] The radiation arriving at any surface element may come not only from the incident beam but also from prior scattering by another surface element. Also, a given surface element may be shielded from the incident beam if the illumination direction is nonnormal. These two effects of multiple scatter and shadowing must be accounted for in a proper summation of scattered intensities from each surface element.

We imagine a ray trace for each surface element, and we will include that element in one of a number of classes depending on the character of that trace. Initially, we divide the total surface into two parts, $\Sigma = \Omega + \Omega'$, where every surface element in Ω is visible to the incident beam with direction $\mathbf{k}_0$. The elements in Ω' are shielded from the incident beam, and their ray traces have zero weight. The illuminated subset is now split into two parts, $\Omega = \Omega_1 + \Omega_1'$, where every element in Ω_1 is characterized by a specularly scattered ray $\mathbf{k}_1$ which does not intersect the surface elsewhere (Fig. 1). For every element of the

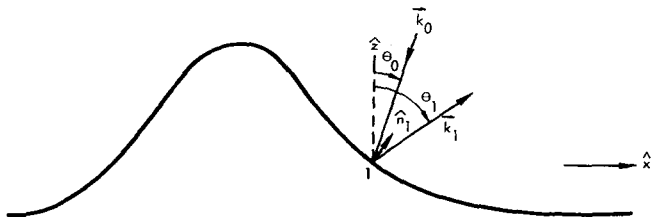

FIG. 1. The scattering configuration for a surface element in Ω_1.

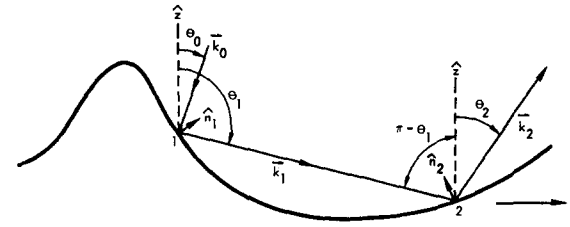

FIG. 2. The scattering configuration for a surface element dS_1 in Ω_2.

subset $\Omega_1' = \Omega_2 + \Omega_2'$, the ray $\mathbf{k}_1$ does intersect the surface again, but now Ω_2 is defined as a smaller subset with the property that the follow-up ray $\mathbf{k}_2$ does not intersect the surface elsewhere (Fig. 2). We continue this bookkeeping until we achieve an empty subset, say Ω_{n+1} (a maximum of n surface scatterings):

$$\Omega = \sum_{i=1}^{n} \Omega_i. \tag{2.1}$$

The ray trace associated with any surface element in Ω begins with the incident beam and follows all the subsequent surface interactions. The initial polarization is taken as either vertical or horizontal, and the scattered radiation will retain this polarization because of the cylindrical symmetry. The incident power intercepted by a surface element dS_1 in Ω is unit intensity times the elemental area projected onto the incident wavefront, i.e., $(-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)\, dS_1$. If the reflectivity is $r(\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)$ (for either vertical or horizontal polarization), then the power leaving dS_1 in direction $\hat{\mathbf{k}}_1$, due to the excitation by the incident beam, is

$$dp(x_1) \equiv dp_1 = r(\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)(-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)\, dS_1. \tag{2.2}$$

If dS_1 is an element of Ω_1, then dp_1 leaves the surface for good. The probability that dp_1 is visible to an observer oriented at angle θ is $P_{\Omega_1}(\theta) = \delta[\theta - \theta_1(x_1)]$, where $\theta_1 = -\theta_0 - 2\tan^{-1} s_1$ by Eq. (A3). Here s_1 is the surface slope at point x_1. Thus, the power reflected into θ from a $dS_1 \in \Omega_1$ is $P_{\Omega_1}(\theta)\, dp_1$, and the total power reflected into θ from Ω_1 is

$$S_0\gamma_1(\theta, \theta_0) \equiv \int_{\Omega_1} dS_1(-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1) r(\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1) \times \delta[\theta - \theta_1(\theta_0, s_1)]. \tag{2.3}$$

If the element dS_1 does not belong to Ω_1, it belongs to Ω_1', and the infinitesimal column bearing dp_1 intersects the surface about some point x_2, cutting out an element of area dS_2'. The prime is to emphasize that this area is dependent on the size and orientation of dS_1. The power leaving dS_2' in direction $\hat{\mathbf{k}}_2$, due to the excitation of dS_1 by the incident beam, is

$$dp_2' = r(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{n}}_2)\, dp_1. \tag{2.4}$$

If $dS_1 \in \Omega_2$, dp_2' leaves the surface. The probability of reflection into a given observation angle θ is $P_{\Omega_2}(\theta) = \delta[\theta - \theta_2(x_2)]$, where $\theta_2 = (\pi + \theta_0) + 2(\tan^{-1} s_1 - \tan^{-1} s_2)$ by Eqs. (A3) and (A7). The total power reflected into θ from Ω_2 is then

$$S_0\gamma_2(\theta, \theta_0) \equiv \int_{\Omega_2} dS_1(-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)r(\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1)r(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{n}}_2) \times \delta[\theta - \theta_2(\theta_0, s_1, s_2)]. \quad (2.5)$$

We can generalize the above for n scatterings. The scattering effect of the rough surface on the incident radiation can then be written as

$$\begin{aligned} S_0\gamma(\theta, \theta_0) &\equiv S_0 \sum_{i=1}^{n} \gamma_i(\theta, \theta_0) \\ &= \int_{\Omega_1} dS_1 \cos(\alpha_1) r_1 \delta(\theta - \theta_1) \\ &\quad + \int_{\Omega_2} dS_1 \cos(\alpha_1) r_1 r_2 \delta(\theta - \theta_2) + \cdots \\ &\quad + \int_{\Omega_n} dS_1 \cos(\alpha_1) r_1 r_2 \cdots r_n \delta(\theta - \theta_n), \end{aligned} \quad (2.6)$$

where $(-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1) \equiv \cos \alpha_1$, and an obvious abbreviation for the reflectivities has been introduced. Now, $\cos \alpha_1$ must be positive, for all the Ω_i lie in the illuminated subset. Also, the reflectivities are positive, so that each term in Eq. (2.6) is positive. Thus, by taking just the first m terms as an approximation, we have a lower bound to the scattered intensity.

The total scattered power is found by integrating Eq. (2.6) over all observable angles $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$. Since the θ_i in the δ functions are restricted to the observable range by the partial-surface construction, we have

$$\begin{aligned} S_0 \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta\, \gamma(\theta, \theta_0) &= \int_{\Omega_1} dS_1 \cos(\alpha_1) r_1 + \int_{\Omega_2} dS_1 \cos(\alpha_1) r_1 r_2 \\ &\quad + \cdots + \int_{\Omega_n} dS_1 \cos(\alpha_1) r_1 r_2 \cdots r_n. \end{aligned} \quad (2.7)$$

Again, each term is positive, so that the retention of the first m terms provides a lower bound to the total scattered power. There is no absorption for a perfectly conducting surface, and so the right-hand side of Eq. (2.7) must equal the incident power when $r = 1$. From Eqs. (2.1) and (2.7), we have

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta\, \gamma[\theta, \theta_0; r = 1] = S_0^{-1} \int_{\Omega} dS_1 \cos \alpha_1 = 1, \quad (2.8)$$

for the projected area of the illuminated subset must make up the beam front. This statement of energy conservation provides a useful standard for testing the validity of a theory of rough-surface scattering; in particular, it provides a basis for comparison of Eq. (2.6) with the uncorrected single-scatter theory.

3. ENSEMBLE AVERAGE OF SCATTERED INTENSITY

As it stands, Eq. (2.6) appears to be no more than a formal device for keeping track of the final destinations and magnitudes of the incident rays. However, the application of this equation to the problem of scatter from random surfaces permits the replacement of deterministic surface classes by known probability functions. We will view $\zeta(x)$ as one of an ensemble of possible surfaces generated by a stationary random process and calculate the ensemble average of $\gamma(\theta, \theta_0)$.

We can convert the integrations over the Ω_i into integrations over the entire surface Σ by defining two functions which take on only the values of zero and unity. Thus, we introduce an illumination function $\epsilon(x; \mathbf{k}_0)$, which has value unity if the surface point at x is illuminated from $\hat{\mathbf{k}}_0$ and which has value zero if that point is shadowed. In addition, we define $\nu(x, \mathbf{k}_i)$ to have value unity if the specularly scattered ray which leaves point x in direction $\hat{\mathbf{k}}_i$ intersects the surface at some other point, whereas it has value zero if $\hat{\mathbf{k}}_i$ does not intersect the surface again. With these definitions, we can now write Eqs. (2.3) and (2.5) as

$$S_0\gamma_1(\theta, \theta_0) = \int_{\Sigma} dS_1 \cos(\alpha_1) r(\cos \alpha_1) \epsilon(x_1; \mathbf{k}_0) \times [1 - \nu(x_1; \mathbf{k}_1)]\delta[\theta - \theta_1(s_1)], \quad (3.1)$$

$$S_0\gamma_2(\theta, \theta_0) = \int_{\Sigma} dS_1 \cos(\alpha_1) r(\cos \alpha_1) r(\cos \alpha_2) \times \epsilon(x_1; \mathbf{k}_0)\nu(x_1; \mathbf{k}_1) \times [1 - \nu(x_2; \mathbf{k}_2)]\delta[\theta - \theta_2(s_1, s_2)]. \quad (3.2)$$

Here, x_2 is the point of intersection of the ray $\hat{\mathbf{k}}_1$, and, as such, it depends on the initial illumination point x_1. The integrands are independent of the y coordinate because of the cylindrical symmetry.

It is a trivial matter to average Eq. (3.1). The surface element is $dS_1 = (1 + s_1^2)^{\frac{1}{2}}\, dx_1\, dy_1$ and $\cos \alpha_1$ is dependent on x_1 only through the slope s_1 [Eq. (A4)]. Thus, Eq. (3.1) depends on x_1 only through $s(x_1)$, $\epsilon(x_1)$, and $\nu(x_1)$. For any stationary random process there are no preferred points, so that the probability distribution $P(s_1, \epsilon_1, \nu_1; x_1)$ $[= P(s_1, \epsilon_1, \nu_1)]$ is independent of x_1. Therefore, the averaging process

leaves a trivial coordinate integration:

$$S_0\langle\gamma_1(\theta,\theta_0)\rangle = \int_{-L'}^{L'} dy_1 \int_{-L}^{L} dx_1 \int_{-\infty}^{\infty} ds_1 \times \sum_{\epsilon_1=0}^{1}\sum_{\nu_1=0}^{1} P[s_1,\epsilon_1(\mathbf{k}_0),\nu_1(\mathbf{k}_1)]f_1(s_1) \times \epsilon_1(1-\nu_1)\delta[\theta-\theta_1(s_1)]$$
$$= (2L)(2L')\int_{-\infty}^{\infty} ds_1 P[s_1,\epsilon_1(\mathbf{k}_0)=1,\nu_1(\mathbf{k}_1)=0] \times f_1(s_1)\delta[\theta-\theta_1(s_1)], \quad (3.3)$$

where

$$f_1(s_1) = (1+s_1^2)^{\frac{1}{2}}[\cos\alpha_1(s_1)]r_1(s_1). \quad (3.4)$$

The segments along the x and y axes cut out by the incident beam are taken here as $2L$ and $2L'$, respectively, so that, in these terms, $S_0 = (2L')(2L)\cos\theta_0$. The ray $\mathbf{k}_1$ must intersect the surface if $|\theta_1| > \frac{1}{2}\pi$ (Fig. 2). Thus, $P[\nu_1(\mathbf{k}_1)=1]=1$ for $|\theta_1|>\frac{1}{2}\pi$, and $P(s_1,\ \epsilon_1=1,\ \nu_1=0)$ must vanish for this range. For $|\theta_1|<\frac{1}{2}\pi$, the probability that the ray $\mathbf{k}_1$ does not intersect the surface elsewhere is equivalent to the probability that the point x_1 is illuminated by the ray $-\mathbf{k}_1$. By use of Eq. (A3), $s_1(\theta_1) = -\tan[\frac{1}{2}(\theta_0+\theta_1)]$, we can change the integration variable in Eq. (3.3) to θ_1:

$$\langle\gamma_1(\theta,\theta_0)\rangle = \sec\theta_0\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta_1 \left|\frac{ds_1(\theta_1)}{d\theta_1}\right| P[s_1(\theta_1),\epsilon_1(\mathbf{k}_0,-\mathbf{k}_1)=1] \times f_1(s_1)\delta(\theta-\theta_1)$$
$$= \sec\theta_0\left|\frac{ds_1(\theta)}{d\theta}\right| P\left[s_1(\theta) = -\tan\left(\frac{\theta_0+\theta}{2}\right)\right] \times P[\epsilon_1(\mathbf{k}_0,-\mathbf{k})=1 \mid s_1(\theta)]f_1[s_1(\theta)]. \quad (3.5)$$

The conditional probability in Eq. (3.5) is the probability that a surface point is illuminated from both the incident and final directions, given that the slope at the point is the specular slope. Equation (3.5) can be checked against the familiar result for normal statistics by the assumption of full illumination and the use of Eqs. (3.4) and (A4) and the identity

$$\tan\left(\frac{\theta_0+\theta}{2}\right) = \left(\frac{\sin\theta+\sin\theta_0}{\cos\theta+\cos\theta_0}\right). \quad (3.6)$$

The conditional probability of Eq. (3.5) has been the subject of several theoretical investigations, each with somewhat different results. The formulation by Sancer,[10] which combines techniques of both Smith[11] and Wagner,[12] seems to be the strongest theoretically, and it yields the best numerical results by far for the energy-conservation calculation in Sec. 6. We list his results for a normal random process:

$$P[\epsilon(\theta_0,\theta)=1 \mid s_1(\theta)] = [1+2(B+B_0)]^{-1} \equiv S(\theta,\theta_0), \quad -\tfrac{1}{2}\pi<\theta<0,$$
$$= (1+2B_0)^{-1} \equiv S(\theta_0), \quad 0<\theta<\theta_0,$$
$$= (1+2B)^{-1} \equiv S(\theta), \quad \theta_0<\theta<\tfrac{1}{2}\pi, \quad (3.7)$$

where

$$B = [4(\pi)^{\frac{1}{2}}V]^{-1}[\exp(-V^2) - (\pi)^{\frac{1}{2}}V(1-\operatorname{erf} V)], \quad (3.8)$$

$$V = [\delta|\tan\theta|]^{-1}, \quad (3.9)$$

$$\delta = 2^{\frac{1}{2}}s_0. \quad (3.10)$$

The quantity B_0 follows from B by replacement of θ by θ_0. The rms slope of the surface is s_0.

An approximation must be made in carrying out the averaging process for Eq. (3.2). The integrand depends on x_1 and $x_2(x_1)$ through $s(x_1)$, $s(x_2)$, $\epsilon(x_1)$, $\nu(x_1)$, and $\nu(x_2)$. Therefore, for a stationary random process, the appropriate probability density depends on $|x_2-x_1|$. We will neglect this dependence and assume no correlation between the random variables at points 1 and 2. Though this might seem to be a severe restriction at first, it actually makes considerable physical sense. As the angle of observation must lie in the physical range $|\theta|<\frac{1}{2}\pi$, the double-scatter process, θ_0 to θ_1 and θ_1 to θ, will almost always consist of a pair of slopes s_1 and s_2 which are appreciably different in value. But the slope of a random surface changes appreciably in a correlation distance T. Thus, we expect $|x_2-x_1|\geq T$ for the typical case, and the correlation effects will then be perturbations. With this assumption, the coordinate integration is again trivial and, from Eq. (3.2), the average intensity takes the form

$$S_0\langle\gamma_2(\theta,\theta_0)\rangle \simeq \int_{-L'}^{L'} dy_1 \int_{-L}^{L} dx_1 \int_{-\infty}^{\infty} ds_2 \int_{-\infty}^{\infty} ds_1 \times \sum_{\epsilon_1=0}^{1}\sum_{\nu_1=0}^{1}\sum_{\nu_2=0}^{1} P[s_1,s_2,\epsilon_1(\mathbf{k}_0),\nu_1(\mathbf{k}_1),\nu_2(\mathbf{k}_2)] \times f_2(s_1,s_2)\epsilon_1\nu_1(1-\nu_2)\delta[\theta-\theta_2(s_1,s_2)]$$
$$= (2L)(2L')\int_{-\infty}^{\infty} ds_2 \int_{-\infty}^{\infty} ds_1 \times P[s_1,s_2,\epsilon_1(\mathbf{k}_0)=1,\nu_1(\mathbf{k}_1)=1,\nu_2(\mathbf{k}_2)=0] \times f_2(s_1,s_2)\delta[\theta-\theta_2(s_1,s_2)], \quad (3.11)$$

where

$$f_2(s_1,s_2) = (1+s_1^2)^{\frac{1}{2}}[\cos\alpha_1(s_1)]r_1(s_1)r_2(s_1,s_2). \quad (3.12)$$

Once again the probability density must vanish if $|\theta_2|>\frac{1}{2}\pi$, for the probability that $\mathbf{k}_2$ intersects the

surface at some further point x_3 is then unity. For $|\theta_2| < \frac{1}{2}\pi$, the probability that $\mathbf{k}_2$ does not intersect the surface elsewhere is equivalent to the probability that point 2 is illuminated by $-\mathbf{k}_2$. It is clearly advisable to change variables in Eq. (3.11) from (s_2, s_1) to (θ_2, s_1'), where the equations of transformation are

$$s_1 = s_1', \quad s_2 = \frac{s_1' \tan [\frac{1}{2}(\theta_0 - \theta_2)] - 1}{s_1' + \tan [\frac{1}{2}(\theta_0 - \theta_2)]}. \quad (3.13)$$

The last equation is derived in the Appendix [Eq. (A8)]. The new integration ranges are $-\pi < \theta_2 < \pi$, $-\infty < s_1' < \infty$, for every s_1' in the infinite range can be coupled with some s_2 to reach a fixed θ_2 by multiple scatter. On the basis of the preceding remarks on the probability density, Eq. (3.11) now takes the form

$$\begin{aligned}\langle \gamma_2(\theta, \theta_0)\rangle &= \sec\theta_0 \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta_2 \int_{-\infty}^{\infty} ds_1 \left|\frac{\partial s_2(s_1, \theta_2)}{\partial \theta_2}\right| \\ &\quad \times f_2[s_1, s_2(s_1, \theta_2)]\delta(\theta - \theta_2) \\ &\quad \times P[s_1, s_2(s_1, \theta_2), \epsilon_1(\mathbf{k}_0) = 1, \\ &\qquad \epsilon_2(-\mathbf{k}_2) = 1, \nu_1(\mathbf{k}_1) = 1] \\ &= \sec\theta_0 \int_{-\infty}^{\infty} ds_1 \left|\frac{\partial s_2^0}{\partial \theta}\right| f_2(s_1, s_2^0) \\ &\quad \times P[s_1, s_2^0, \epsilon_1(\mathbf{k}_0) = 1, \epsilon_2(-\mathbf{k}) = 1, \nu_1(\mathbf{k}_1) = 1], \end{aligned} \quad (3.14)$$

where

$$s_2^0 = \frac{s_1 \tan [\frac{1}{2}(\theta_0 - \theta)] \mp 1}{s_1 + \tan [\frac{1}{2}(\theta_0 - \theta)]}$$

and

$$\left|\frac{\partial s_2^0}{\partial \theta}\right| = \frac{1}{2} \frac{(1 + s_1^2) \sec^2 [\frac{1}{2}(\theta_0 - \theta)]}{\{s_1 + \tan [\frac{1}{2}(\theta_0 - \theta)]\}^2}. \quad (3.15)$$

By comparison with Eq. (3.5), we see that both the simple linear combinations of incident and final angle, $(\theta_0 + \theta)$ and $(\theta_0 - \theta)$, must enter for an accurate description of the scattering. It remains to express the fifth-order probability density in terms of known functions. We now specialize to normal statistics.

4. DOUBLE-SCATTER PROBABILITY DENSITY

We will simplify the notation slightly and replace the statements $\epsilon_i(\mathbf{k}_j) = 1$ and $\nu_1(\mathbf{k}_1) = 1$ by the symbols $I_i(\mathbf{k}_j)$ and $I_{12}(\mathbf{k}_1)$, respectively. The probability density of Eq. (3.14) can be broken up and written as the product of five probability functions, four of them conditional:

$$P[s_1, s_2^0, I_1(\mathbf{k}_0), I_{12}(\mathbf{k}_1), I_2(-\mathbf{k})] = P(s_1)P_1P_2P_3P_4, \quad (4.1)$$

where

$$P_1 = P[I_1(\mathbf{k}_0) \mid s_1], \quad (4.2)$$

$$P_2 = P[I_{12}(\mathbf{k}_1) \mid s_1, I_1(\mathbf{k}_0)], \quad (4.3)$$

$$P_3 = P[s_2^0 \mid s_1, I_1(\mathbf{k}_0), I_{12}(\mathbf{k}_1)], \quad (4.4)$$

$$P_4 = P[I_2(-\mathbf{k}) \mid s_1, s_2^0, I_1(\mathbf{k}_0), I_{12}(\mathbf{k}_1)]. \quad (4.5)$$

The quantity P_1 is simply the probability that a surface point with given slope s_1 is illuminated by the incident beam. For normal statistics, it has the form[10]

$$P_1 = S(\theta_0)u(\cot\theta_0 - s_1), \quad (4.6)$$

where $S(\theta_0)$ was defined in Eq. (3.7) and u is a step function with value unity for positive argument and value zero for negative argument.

The function P_2 is the probability that the ray $\mathbf{k}_1$ intersects the surface at point 2, given the slope and the fact of illumination at point 1. If $|\theta_1| > \frac{1}{2}\pi$, this probability is exactly unity. For $|\theta_1| < \frac{1}{2}\pi$, P_2 can be recast exactly in terms of monostatic and bistatic illumination probabilities. By the argument used previously in Sec. 3, we have

$$\begin{aligned} P_2 &= P[\nu_1(\mathbf{k}_1) = 1 \mid s_1, I_1(\mathbf{k}_0)] \\ &= 1 - P[\nu_1(\mathbf{k}_1) = 0 \mid s_1, I_1(\mathbf{k}_0)] \\ &= 1 - P[I_1(-\mathbf{k}_1) \mid s_1, I_1(\mathbf{k}_0)], \quad |\theta_1| < \tfrac{1}{2}\pi. \end{aligned} \quad (4.7)$$

But, $P[I_1(\mathbf{k}_0, -\mathbf{k}_1) \mid s_1] = P[I_1(\mathbf{k}_0) \mid s_1]P[I_1(-\mathbf{k}_1) \mid s_1, I_1(\mathbf{k}_0)]$, so that, for the entire angular range of θ_1,

$$\begin{aligned} P_2 &= 1 - P[I_1(\mathbf{k}_0, -\mathbf{k}_1) \mid s_1]/P_1, \quad |\theta_1| < \tfrac{1}{2}\pi, \\ P_2 &= 1, \quad |\theta_1| > \tfrac{1}{2}\pi. \end{aligned} \quad (4.8)$$

The shadow functions are expressed mathematically in Eqs. (3.7) and (4.6).

Approximations must be made for the probabilities P_3 and P_4. We note, first, that the explicit condition s_1 can be suppressed in both since the two conditions $I_1(\mathbf{k}_0)$ and $I_{12}(\mathbf{k}_1)$ are sufficient to determine the slope at point 1 from Eq. (A3). In words, $P_3 = P[s_2^0 \mid I_1(\mathbf{k}_0), I_{12}(\mathbf{k}_1)]$ is the probability density for the slope at the intersection point 2, given that point 1 is illuminated by the source and given that the specular ray $\mathbf{k}_1$ does intersect the surface at point 2. The condition $I_1(\mathbf{k}_0)$ is important for θ_1 equal to and somewhat larger than the incident angle (P_2 vanishes for $0 < \theta_1 < \theta_0$), for, here, the requirement that $\mathbf{k}_0$ not intersect the surface while $\mathbf{k}_1$ does will restrict s_2^0 to slopes near that of $\mathbf{k}_0$. For other θ_1, with the continuing assumption that points 1 and 2 are well separated (s_1 and s_2^0 are then uncorrelated), slope statistics at point 2 are primarily affected by the condition of "illumination" at point 2. In the absence of an extended shadowing theory, we make what seems to be a reasonable and accurate

approximation and neglect the condition $I_1(\mathbf{k}_0)$:

$$P_3 \simeq P[s_2^0 \mid I_{12}(\mathbf{k}_1)]. \tag{4.9}$$

A mathematical statement for the right-hand side of Eq. (4.9) has been derived by Wagner[12] for normal statistics, and the result is consistent with the work of Smith.[11] The distribution of slopes which can be illuminated by $\mathbf{k}_1$ is again normal. As the range of illuminated slopes varies with θ_1, the normalization factor for the conditional statistics is dependent on θ_1:

$$\begin{aligned} P_3 &\simeq \left(\frac{2}{1-\operatorname{erf} V_1}\right) u(s_2^0 - \cot\theta_1) P(s_2^0), \quad 0 < \theta_1 < \tfrac{1}{2}\pi, \\ &\simeq \left(\frac{2}{1-\operatorname{erf} V_1}\right) u(\cot\theta_1 - s_2^0) P(s_2^0), \\ &\qquad\qquad -\tfrac{1}{2}\pi < \theta_1 < 0, \\ &\simeq \left(\frac{2}{1+\operatorname{erf} V_1}\right) P(s_2^0), \quad |\theta_1| > \tfrac{1}{2}\pi, \end{aligned} \tag{4.10}$$

where

$$V_1 = [\delta\,|\tan\theta_1|]^{-1},$$

$$P(s) = (\pi\delta^2)^{-\frac{1}{2}} \exp(-s^2/\delta^2), \quad \delta = 2^{\frac{1}{2}} s_0. \tag{4.11}$$

Once again, the u's are step functions, with value unity for positive argument and value zero for negative arguments.

The complicated P_4 is exactly unity in the double range $-\pi < \theta_1 < 0, 0 < \theta < \theta_0$. For this combination of (θ, θ_1), point 2 is to the left of point 1 (Fig. 3). The surface to the right of point 1 cannot intersect $\mathbf{k}_0$ because of the condition $I_1(\mathbf{k}_0)$. The ray $\mathbf{k}_1$ connects points 1 and 2 because of the condition $I_{12}(\mathbf{k}_1)$. The ray $\mathbf{k}$ will always be to the left of $\mathbf{k}_0$ if $0 < \theta < \theta_0$ and so cannot intersect the surface for this range of observation angles.

There are no certainties for other combinations of (θ, θ_1). As an example, consider the case $0 < \theta_1 < \pi$, where the point 2 is to the right of point 1 (Fig. 2). Even when $0 < \theta < \theta_0$, the distance between the two points may be such that the surface can intersect the ray $\mathbf{k}$ to the right of point 2 while the ray $-\mathbf{k}_0$ achieves sufficient altitude to clear. From this last, it is clear that $I_1(\mathbf{k}_0)$ is still an important condition for angles outside of the special double range, but there is no theory available for P_4 as it stands. In order to proceed, we drop the conditions on point 1 for (θ, θ_1) outside the double range. This leaves

$$\begin{aligned} P_4 &\sim P[I_2(\mathbf{k}) \mid s_2^0, I_{12}(\mathbf{k}_1)], \\ &-\tfrac{1}{2}\pi < \theta < 0 \quad \text{or} \quad \theta > \theta_0, \end{aligned}$$

and

$$0 < \theta < \theta_0, \quad 0 < \theta_1 < \pi. \tag{4.12}$$

Equation (4.12) has the superficial appearance of a bistatic illumination probability, but the ray $\mathbf{k}_1$ originates from a surface point rather than from a source at infinity. This distinction effectively divides the θ_1 angular range into subranges $|\theta_1| < \frac{1}{2}\pi$ and $|\theta_1| > \frac{1}{2}\pi$. First we note that s_2^0 is the slope connecting $\mathbf{k}_1$ and $\mathbf{k}$. But, from Eq. (4.9), P_3 is zero unless $s_2^0(s_1)$ can be illuminated by $\mathbf{k}_1$ [the step functions in Eq. (4.10) permit only those values of s_1 for which this is true]. Thus the only s_2^0 permitted in Eq. (4.12) are those which are illuminated by $\mathbf{k}_1$, and the only distinguishing feature, with respect to illumination, between the conditions s_2^0 and $I_{12}(\mathbf{k}_1)$ is the origin of $\mathbf{k}_1$ at a surface point. For $|\theta_1| < \frac{1}{2}\pi$, the two conditions are very nearly redundant (see next paragraph), and $I_{12}(\mathbf{k}_1)$ can be dropped with negligible error. Equation (4.12) then has the form of Eq. (4.2) with $\theta_0 \to \theta$ and $s_1 \to s_2^0$. If s_2^0 is illuminated by $\mathbf{k}_1$, it is illuminated by $-\mathbf{k}$, so that the mathematical statement of Eq. (4.12) is

$$\begin{aligned} P[I_2(\mathbf{k}) \mid s_2^0, I_{12}(\mathbf{k}_1)] &\simeq P[I_2(\mathbf{k}) \mid s_2^0] = S(\theta), \\ |\theta| \le \tfrac{1}{2}\pi, &\quad 0 < \theta_1 < \tfrac{1}{2}\pi, \end{aligned}$$

and

$$-\tfrac{1}{2}\pi < \theta < 0 \quad \text{or} \quad \theta > \theta_0, \quad -\tfrac{1}{2}\pi < \theta_1 < 0. \tag{4.13}$$

When $|\theta_1| > \frac{1}{2}\pi$, the fact that $\mathbf{k}_1$ originates from a surface point can affect the probability that $\mathbf{k}$ reaches the observer. Thus, there are ranges of θ, $(\pi + \theta_1) < \theta < \frac{1}{2}\pi$, for θ_1 negative (Fig. 3) and $-\frac{1}{2}\pi < \theta < -(\pi - \theta_1)$ for θ_1 positive (Fig. 2), for which $\mathbf{k}$ lies below $-\mathbf{k}_1$ and so must strike the surface with certainty yielding zero for Eq. (4.12). For $\mathbf{k}$ just above $-\mathbf{k}_1$, the effect of $I_{12}(\mathbf{k}_1)$ on the shadowing of $\mathbf{k}$ should decrease fairly rapidly with increasing angular difference. In order to proceed mathematically, we will assume a stepfunction behavior at $\mathbf{k} = -\mathbf{k}_1$ with $P_4 = 0$ when $\mathbf{k}$ lies below $-\mathbf{k}_1$ and $P_4 = S(\theta)$ for other $\mathbf{k}$. This concludes the breakdown of P_4 into illumination probabilities. We can now group the preceding

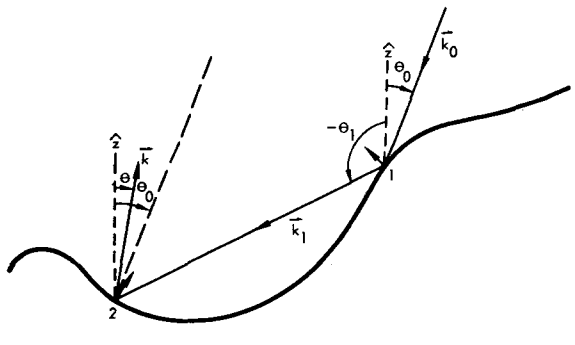

FIG. 3. P_4 is exactly unity for this configuration of angles.

results together:

$$P_4 \simeq \begin{cases} S(\theta), & 0 < \theta_1 < \tfrac{1}{2}\pi, \\ \begin{cases} 1, & 0 < \theta < \theta_0, \\ S(\theta), & \text{other } \theta, \end{cases} & -\tfrac{1}{2}\pi < \theta_1 < 0, \\ \begin{cases} 1, & 0 < \theta < \theta_0, \\ u[(\pi + \theta_1) - \theta]S(\theta), & \text{other } \theta, \end{cases} & \theta_1 < -\tfrac{1}{2}\pi, \\ u[\theta + (\pi - \theta_1)]S(\theta), & \theta_1 > \tfrac{1}{2}\pi. \end{cases} \tag{4.14}$$

The quantity $S(\theta)$ is defined in Eq. (3.7).

5. $\langle \gamma_2(\theta,\theta_0) \rangle$

We wish to combine the results of Sec. 4 with Eq. (3.14). As a preliminary, we streamline Eq. (4.14) by the introduction of yet another function $\omega(\theta, \theta_1)$ which takes on only values of zero and unity:

$$\begin{aligned} \omega(\theta, \theta_1) &= 0, \quad 0 < \theta < \theta_0 \quad \text{and} \quad -\pi < \theta_1 < 0, \\ &= 1, \quad \text{all other} \quad (\theta, \theta_1). \end{aligned} \tag{5.1}$$

We can now construct the function

$$S[\theta, \omega(\theta, \theta_1)] \equiv 1 + \omega(\theta, \theta_1)[S(\theta) - 1], \tag{5.2}$$

which assumes the forms $S(\theta)$ or unity in accordance with Eq. (5.1). By Eqs. (5.1) and (5.2), Eq. (4.14) becomes

$$P_4 \simeq S[\theta, \omega(\theta, \theta_1)] \begin{cases} 1, & 0 < \theta_1 < \tfrac{1}{2}\pi, \\ 1, & -\tfrac{1}{2}\pi < \theta_1 < 0, \\ u[(\pi + \theta_1) - \theta], & \theta_1 < -\tfrac{1}{2}\pi, \\ u[\theta + (\pi - \theta_1)], & \theta_1 > \tfrac{1}{2}\pi, \end{cases} \tag{5.3}$$

where the fact that $u[(\pi + \theta_1) - \theta] = 1$ for $0 < \theta < \theta_0$, $\theta_1 < -\frac{1}{2}\pi$ has been used.

The substitution of Eqs. (4.1), (4.6), (4.8), (4.10), and (5.3) into Eq. (3.14) yields the following expression for the double-scatter coefficient:

$$\langle \gamma_2(\theta, \theta_0) \rangle = \sum_{j=1}^{4} \int_{a_j}^{a_{j+1}} ds_1 T(s_1, \omega) A_j(s_1), \tag{5.4}$$

where

$$T(s_1, \omega) = P(s_1)S(\theta_0)\left(\frac{2}{1 - \operatorname{erf} V_1}\right)P(s_2^0)S[\theta, \omega(\theta, \theta_1)] \times \sec\theta_0 \left|\frac{\partial s_2^0}{\partial \theta}\right| f_2(s_1, s_2^0) \tag{5.5}$$

and

$$(A_j) = \begin{pmatrix} \left(\dfrac{1 - \operatorname{erf} V_1}{1 + \operatorname{erf} V_1}\right)u[(\theta + \pi) - \theta_1], \\ \left(1 - \dfrac{S(\theta_1)}{S(\theta_0)}\right)u(-s_1 - \tan\theta_0) \times u(s_2^0 - \cot\theta_1), \\ \left(1 - \dfrac{S(\theta_0, \theta_1)}{S(\theta_0)}\right)u(\cot\theta_1 - s_2^0), \\ \left(\dfrac{1 - \operatorname{erf} V_1}{1 + \operatorname{erf} V_1}\right)u[(\pi - \theta) + \theta_1] \end{pmatrix}. \tag{5.6}$$

The integration limits in Eq. (5.4) are

$$a_1 = -\infty,\ a_2 = -\tan[\tfrac{1}{2}(\theta_0 + \tfrac{1}{2}\pi)],\ a_3 = -\tan\tfrac{1}{2}\theta_0,$$
$$a_4 = \tan[\tfrac{1}{2}(\tfrac{1}{2}\pi - \theta_0)], \text{ and } a_5 = \cot\theta_0.$$

6. ENERGY CONSERVATION

An important application of the scattering formalism is the calculation of the total power scattered into the physical range of angles. It is especially instructive to consider a conducting surface, for we know what the result should be from Eq. (2.8):

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta \langle \gamma(\theta, \theta_0; r = 1) \rangle = 1. \tag{6.1}$$

It has been shown[13] for a normally distributed surface that the uncorrected probability of slopes theory does not, in general, satisfy Eq. (6.1), with the deviation from unity dependent on incident angle and roughness. When shadowing alone was accounted for, the conservation condition was fulfilled at grazing incidence. In this section we present and discuss a computer evaluation of Eq. (6.1) for the single- and double-scatter coefficients as applied to the case of normal statistics. A number of roughnesses and several incident angles are considered. The results are then compared with the uncorrected theory and the conservation condition.

The equation of interest is

$$\begin{aligned} N(\theta_0) &\equiv N_1(\theta_0) + N_2(\theta_0) \\ &= \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta \, [\langle \gamma_1(\theta, \theta_0; r = 1) \rangle + \langle \gamma_2(\theta, \theta_0; r = 1) \rangle]. \end{aligned} \tag{6.2}$$

We use Eqs. (3.5) and (3.6) for the single-scatter theory $N_1(\theta_0)$ and Eqs. (5.4)–(5.6) for the double-scatter correction $N_2(\theta_0)$. In addition, we evaluate the totally uncorrected theory $M(\theta_0)$, which follows from Eq. (3.5) by setting the illumination probability equal to unity. For a perfect reflector,

$$M(\theta_0) = \int_{b_1}^{b_2} ds_1 P_{\text{normal}}(s_1)(1 + s_1 \tan\theta_0), \tag{6.3}$$

TABLE I. Energy-conservation coefficients.

θ_0	$Q(\theta_0)$	$s_0 = \tan 10°$	$\tan 15°$	$\tan 30°$	$\tan 45°$
1°	M	1.0000	0.9998	0.9167	0.6828
	N_1	1.0000	0.9984	0.8308	0.5262
	N_2	0.0000	0.0015	0.1450	0.3752
	N	1.0000	0.9999	0.9758	0.9014
15°	M	1.0000	0.9984	0.9168	0.7162
	N_1	0.9999	0.9946	0.8359	0.5522
	N_2	0.0001	0.0048	0.1397	0.3595
	N	1.0000	0.9994	0.9757	0.9116
30°	M	0.9997	0.9905	0.9192	0.8201
	N_1	0.9988	0.9791	0.8514	0.6338
	N_2	0.0011	0.0180	0.1251	0.3071
	N	0.9999	0.9971	0.9765	0.9410
45°	M	0.9950	0.9713	0.9415	0.9973
	N_1	0.9889	0.9482	0.8788	0.7645
	N_2	0.0096	0.0436	0.1021	0.2063
	N	0.9985	0.9918	0.9809	0.9708
60°	M	0.9741	0.9536	1.0369	1.2715
	N_1	0.9532	0.9150	0.9123	0.9055
	N_2	0.0394	0.0712	0.0753	0.0846
	N	0.9926	0.9862	0.9877	0.9901
75°	M	0.9710	1.0420	1.4277	2.0284
	N_1	0.9123	0.9172	0.9497	0.9711
	N_2	0.0742	0.0714	0.0449	0.0262
	N	0.9865	0.9886	0.9945	0.9973
89°	M	4.5449	6.6339	13.7000	23.3580
	N_1	0.9862	0.9909	0.9960	0.9980
	N_2	0.0126	0.0084	0.0037	0.0019
	N	0.9989	0.9993	0.9997	0.9999

where

$$b_1 = -\tan\left[\tfrac{1}{2}(\tfrac{1}{2}\pi - \theta_0)\right]$$

and

$$b_2 = \tan\left[\tfrac{1}{2}(\theta_0 + \tfrac{1}{2}\pi)\right]$$

from Eq. (A3). The numerical results are presented in Table I. The quantities $Q = M$, N_1, N_2, and N are evaluated for the seven angles at the left and the four rms slopes between tan 10° and tan 45°.

It will be noted that $M(\theta_0)$ is too small for angles near normal incidence and much too large for angles near grazing. The former "energy loss" is a consequence of the single-scatter theory in that all surface slopes are available to scatter the incident radiation. Therefore, surface elements with slopes outside the range $b_1 < s_1 < b_2$ scatter the incident rays into final directions outside of the physical range $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$. The double-scatter theory N_2 returns much of the lost radiation to the observation range. The infinite catastrophe for near-grazing incidence is due to the neglect of shadowing corrections. The incident beam intercepts a surface length of sec θ_0 times the beam width. The same amount of energy is incident regardless of θ_0, so that the large intercepted surface for $\theta_0 \to \frac{1}{2}\pi$ is greatly shadowed. An appropriate shadowing theory will eliminate the sec θ_0 factor. The quantity $1 - M(\theta_0)$ is displayed in Fig. 4 for several rms slopes.

The shadow-corrected, single-scatter theory has the property $N_1(\theta_0) < M(\theta_0)$ over the entire range of θ_0. This makes sense, for not only is there the energy loss associated with M, but in addition the shadowing correction in N_1 reduces the output into any nonzero, observable direction. Of course, any scattered ray which is shadowed is a multiply reflected ray, and N_2 must account for part of this shadowed radiation. This is clearly the situation as, from Table I, we note that $N(\theta_0)$ is much closer to unity than $M(\theta_0)$ for all s_0, save for the odd angular range where M passes unity on its way to infinity. In this latter respect, we note that $N(\theta_0) \leq 1$ for all θ_0. This is most gratifying, and it is a result dependent on the choice of shadow correction, Eq. (3.7). The shadowing theory of Wagner, e.g., leads to numerical results for N which approach unity at grazing incidence from above. Thus, the energy-conservation integral provides a standard against which the various shadowing theories may be judged. We display $1 - N(\theta_0')$ in Fig. 5 for comparison with Fig. 4.

The numerical results differentiate between the relatively smooth 10°, 15° surfaces and the rough 30°, 45° surfaces. Thus, for the surfaces with modest s_0, the quantity $N_2(\theta_0)$ starts off at or near zero for normal incidence, then rises to some maximum as θ_0 increases, and finally falls back slowly to zero as θ_0 approaches grazing. Now, the most likely slopes for scatter are $-\tan\alpha < s < \tan\alpha$, where $\tan\alpha = \delta = 2^{\frac{1}{2}}s_0$ and α is defined relative to grazing. For a reasonably smooth surface and normal incidence, the scattered rays are not likely to strike the surface elsewhere. As θ_0 increases from zero, the normals to the important range of "far" slopes (the positive slopes of Fig. 1) make larger angles with the incident rays. Therefore, the scattered rays leave these slopes at larger angles from the z direction. The probability of multiple scatter thus increases as θ_0 increases. However, the far slopes are also being shadowed, with only the unimportant steep slopes initially but with the shadowing condition eventually extending to positive slopes $\leq \tan\alpha$ as θ_0 increases. The maximum multiple scatter occurs near $\frac{1}{2}\pi - \theta_0 = \alpha$; for increasing θ_0, the slopes responsible for multiple scatter are increasingly shadowed.

For the rough surfaces, on the other hand, there is multiple scatter for all angles of incidence. The quantity N_2 is a maximum at normal incidence and decreases monotonically with θ_0. This decrease for increasing θ_0 arises because the normals to the

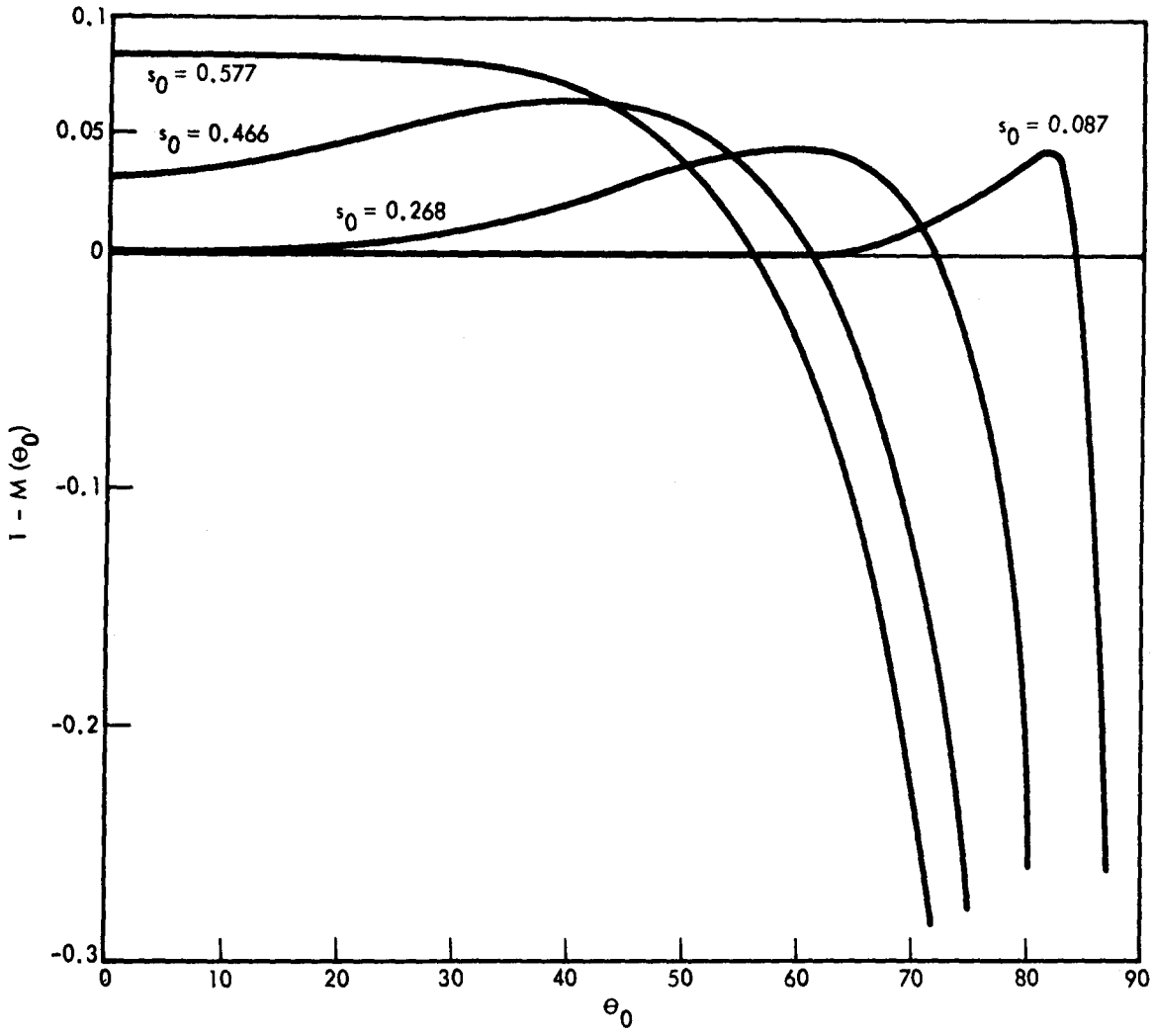

FIG. 4. Energy nonconservation for the uncorrected scattering theory, with rms slope as parameter.

important "near" slopes (negative slopes of Fig. 1) make smaller angles with the incident rays. The corresponding scattered rays leave the surface at smaller angles relative to the z direction and, thus, are less likely to intersect the surface again. At $\frac{1}{2}\pi - \theta_0 \leq \alpha$, this mechanism is joined by the substantial shadowing of far-side slopes responsible for multiple scatter. The table indicates that the double-scatter formalism is not sufficient for $s_0 \geq 1$ and near-normal incidence; i.e., the higher-order scattering coefficients in Eq. (2.6) are not negligible. The monotonic decrease of multiple-scatter effects with θ_0, however, means that an accurate description is still possible for some range of angles prior to grazing.

7. CONCLUSIONS

A ray-optics approach has been applied to the theory of scattering from 1-dimensional, random rough surfaces. In principle, the theory can account for shadowing and all orders of multiple scatter. As a practical matter, shadow-consistent descriptions of both single and double scattering are presented in terms of known functions. The former is just the familiar proportionality of scattered intensity with the slope probability density, but now the appropriate illumination probability is included as well. The double-scatter correction provides computational accuracy; it has appreciable value for large angles of incidence on relatively smooth surfaces and for all angles of incidence, save grazing, on the rough surfaces ($s_0 \geq \tan 20°$). In order to test the ability of the theory to conserve energy, a numerical evaluation of the energy integral is carried out for a perfect reflector. The double-scatter formulation successfully accounts for the incident energy when $s_0 \leq \tan 30°$, and it is far superior to the conventional single-scatter, unshadowed theory for any rms slope.

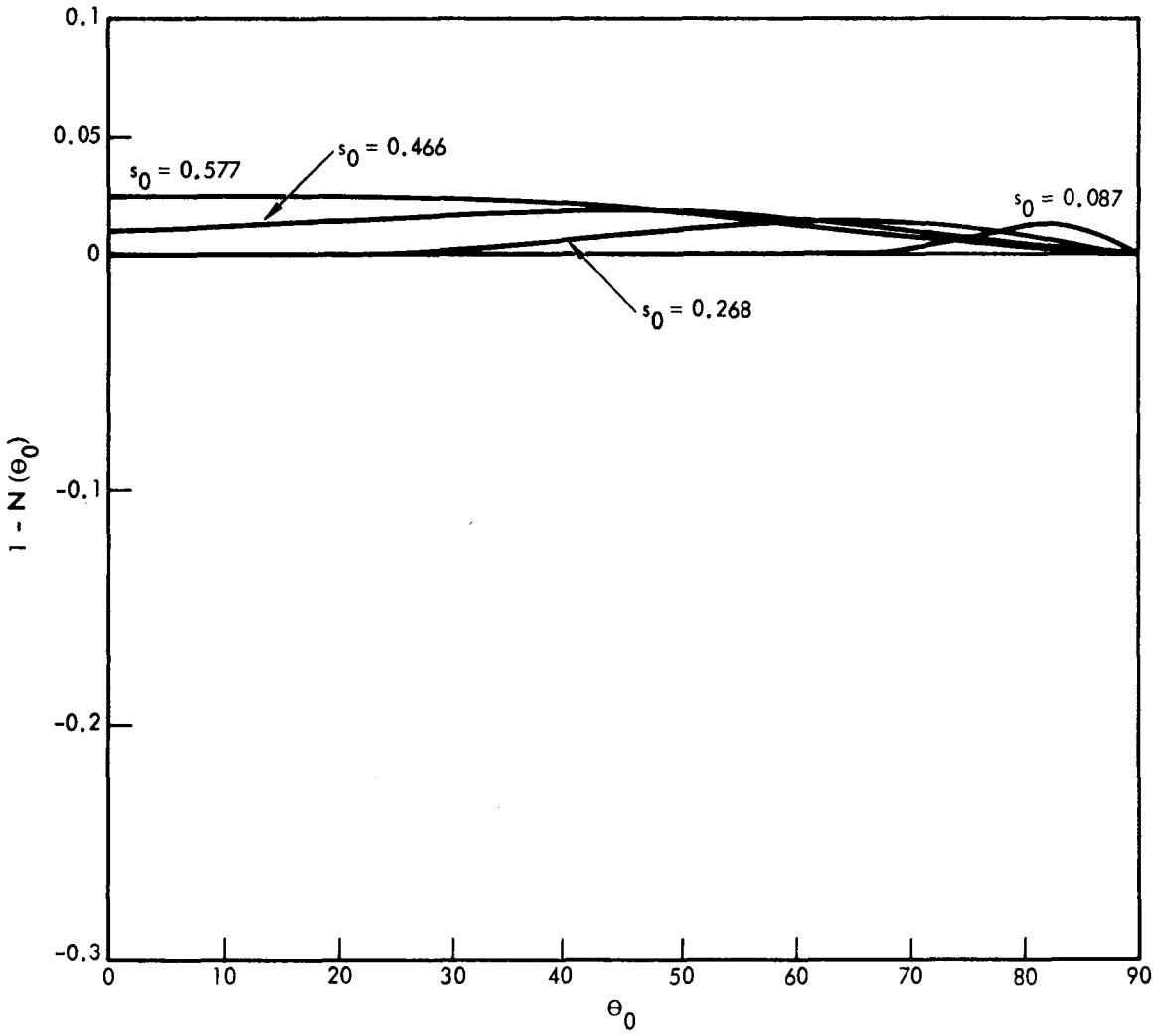

FIG. 5. Energy-conservation discrepancies for the corrected scattering theory and the s_0 of the previous figure.

APPENDIX

We derive here the basic geometrical quantities appearing in Eqs. (3.5) and (3.14). We first consider the parameters associated with point 1 and define

$$-\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{n}}_1 = \cos \alpha_1, \quad \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{z}} = \cos \beta_1, \tag{A1}$$

where β_1 is measured clockwise from the z axis. It follows immediately that the slope at point 1 is $s_1 = -\tan \beta_1$. From Fig. 2,

$$\theta_1 = \theta_0 + 2\alpha_1, \quad \beta_1 = \theta_0 + \alpha_1, \tag{A2}$$

so that

$$\beta_1 = \tfrac{1}{2}(\theta_0 + \theta_1)$$

and

$$s_1 = -\tan \beta_1 = -\tan [\tfrac{1}{2}(\theta_0 + \theta_1)]. \tag{A3}$$

We can also write $\cos \alpha_1$ in terms of θ_0 and s_1:

$$\begin{aligned} \cos \alpha_1 &= \cos (\beta_1 - \theta_0) \\ &= (1 + s_1^2)^{-\frac{1}{2}}(\cos \theta_0 - s_1 \sin \theta_0). \end{aligned} \tag{A4}$$

The procedure for point 2 is the same. We define

$$-\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{n}}_2 = \cos \alpha_2, \quad \hat{\mathbf{n}}_2 \cdot \hat{\mathbf{z}} = \cos \beta_2, \tag{A5}$$

and β_2 is measured clockwise from the z direction. The slope at point 2 is $s_2 = -\tan \beta_2$. From Fig. 2, the angle between $-\mathbf{k}_1$ and the z axis is $\pi - \theta_1$; thus, we have

$$\alpha_2 - \beta_2 = \pi - \theta_1, \quad 2\alpha_2 = (\pi - \theta_1) + \theta_2. \tag{A6}$$

The solution of the preceding pair of equations yields $\beta_2 = -\frac{1}{2}(\pi - \theta_1 - \theta_2)$, and

$$s_2 = -\tan \beta_2 = \tan [\tfrac{1}{2}(\pi - \theta_1 - \theta_2)]. \tag{A7}$$

By use of the relation $\theta_1 = -\theta_0 - 2 \tan^{-1} s_1$ from Eq. (A3), we obtain s_2 as a function of s_1 and θ_2:

$$s_2 = \frac{s_1 \tan [\frac{1}{2}(\theta_0 - \theta_2)] - 1}{s_1 + \tan [\frac{1}{2}(\theta_0 - \theta_2)]}. \tag{A8}$$

The remaining quantity needed is $\cos \alpha_2$ as a function of s_1 and θ_2, for this is the argument of one of the reflectivities. Thus, from Eq. (A6),

$$\begin{aligned}\cos \alpha_2 &= \cos [\tfrac{1}{2}\pi + \tfrac{1}{2}(\theta_2 - \theta_1)] \\ &= \sin [\tfrac{1}{2}(\theta_1 - \theta_2)] \\ &= -\sin [\tfrac{1}{2}(\theta_0 + \theta_2) - \beta_1] \\ &= -(1 + s_1^2)^{-\frac{1}{2}}\{\sin [\tfrac{1}{2}(\theta_0 + \theta_2)] \\ &\quad + s_1 \cos [\tfrac{1}{2}(\theta_0 + \theta_2)]\}. \qquad \text{(A9)}\end{aligned}$$

The reflectivities will be written here for convenience. If we distinguish them as $r_V = |R_V|^2$ and $r_H = |R_H|^2$ to denote vertical and horizontal polarization, respectively, then we have

$$R_V(\cos \alpha_i) = \frac{K \cos \alpha_i - (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}}{K \cos \alpha_i + (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}},$$

$$R_H(\cos \alpha_i) = \frac{\cos \alpha_i - (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}}{\cos \alpha_i + (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}},$$

$$K = \frac{\epsilon}{\mu}.$$

Here, ϵ is the relative complex permittivity, and μ is the relative permeability.

[1] W. H. Peake, IRE Trans. Antennas Propagation **7**, S324 (1959).
[2] A. Stogryn, IEEE Trans. Antennas Propagation **15**, 278 (1967).
[3] H. Van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957).
[4] V. Fock, *Electromagnetic Diffraction and Propagation Problems* (Pergamon, London, 1965).
[5] P. Beckmann and A. Spizzichino, *Scattering of Electromagnetic Waves from Rough Surfaces* (Macmillan, New York, 1963), pp. 17–98.
[6] P. Lynch, J. Acoust. Soc. Am. **47**, 804 (1970).
[7] T. Hagfors, J. Geophys. Res. **71**, 379 (1966).
[8] D. Barrick, IEEE Trans. Antennas Propagation **16**, 449 (1968).
[9] R. Kodis, IEEE Trans. Antennas Propagation **14**, 77 (1966).
[10] M. Sancer, IEEE Trans. Antennas Propagation **17**, 577 (1969).
[11] B. Smith, IEEE Trans. Antennas Propagation **15**, 668 (1967).
[12] R. Wagner, J. Acoust. Soc. Am. **41**, 138 (1967).
[13] P. Lynch and R. Wagner, J. Acoust. Soc. Am. **47**, 816 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Boundary-Value Problems of Linear-Transport Theory—Green's Function Approach

MADHOO KANAL
Space-Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48105

(Received 27 January 1970)

Case's technique utilizing Green's functions for dealing with boundary-value problems of the neutron linear-transport theory is exploited. We show that the Fourier coefficients of the Green's function over the Case spectrum are precisely the normal modes. In particular, if we assume that the scattering kernel is rotationally invariant (which indeed we do assume) and approximate it by a degenerate kernel consisting of spherical harmonics, the set of modes is deficient for problems lacking azimuthal symmetry. We also show that the expansion of the scattering kernel, in terms of spherical harmonics (or any set of orthogonal functions for that matter), permits the linear factorization of the Fourier coefficients of the Green's function in terms of the lowest element, with the proportionality functions consisting of complete orthogonal polynomials. As a consequence of this attribute of Fourier coefficients, the eigenfunctions (continuum and discrete) also factorize, which then permits decoupling of the appropriate singular integral equations. To illustrate our idea, we solve half-space and slab problems. However, the basic procedure is kept sufficiently general so that the extension to problems involving other geometrics remains straightforward.

1. INTRODUCTION

The normal-mode (eigenfunction) expansion technique of Case,[1] in dealing with boundary-value problems, has achieved considerable success in the types of problems for which the normal modes (continuum plus discrete) form a complete orthogonal set. However, there are several problems of interest, for instance, in the theory of neutron diffusion and kinetic theory of gases,[2] where the sets of modes are either deficient or the appropriate integral equations are regular. In particular, in a recent paper by Case *et al.*,[3] it has been shown for spherical geometry that one cannot directly adapt the above-mentioned technique. In this paper, we consider the Green's function approach also due to Case.[4] We show that the Fourier coefficients of the Green's function for the appropriate neutron 1-speed transport equation over the Case spectrum are precisely the normal modes. In particular, if we assume that the scattering kernel is rotationally invariant (which indeed we do assume) and approximate it by a

The remaining quantity needed is $\cos \alpha_2$ as a function of s_1 and θ_2, for this is the argument of one of the reflectivities. Thus, from Eq. (A6),

$$\begin{aligned}\cos \alpha_2 &= \cos [\tfrac{1}{2}\pi + \tfrac{1}{2}(\theta_2 - \theta_1)] \\ &= \sin [\tfrac{1}{2}(\theta_1 - \theta_2)] \\ &= -\sin [\tfrac{1}{2}(\theta_0 + \theta_2) - \beta_1] \\ &= -(1 + s_1^2)^{-\frac{1}{2}}\{\sin [\tfrac{1}{2}(\theta_0 + \theta_2)] \\ &\quad + s_1 \cos [\tfrac{1}{2}(\theta_0 + \theta_2)]\}. \qquad \text{(A9)}\end{aligned}$$

The reflectivities will be written here for convenience. If we distinguish them as $r_V = |R_V|^2$ and $r_H = |R_H|^2$ to denote vertical and horizontal polarization, respectively, then we have

$$R_V(\cos \alpha_i) = \frac{K \cos \alpha_i - (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}}{K \cos \alpha_i + (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}},$$

$$R_H(\cos \alpha_i) = \frac{\cos \alpha_i - (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}}{\cos \alpha_i + (K - 1 + \cos^2 \alpha_i)^{\frac{1}{2}}},$$

$$K = \frac{\epsilon}{\mu}.$$

Here, ϵ is the relative complex permittivity, and μ is the relative permeability.

[1] W. H. Peake, IRE Trans. Antennas Propagation **7**, S324 (1959).
[2] A. Stogryn, IEEE Trans. Antennas Propagation **15**, 278 (1967).
[3] H. Van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957).
[4] V. Fock, *Electromagnetic Diffraction and Propagation Problems* (Pergamon, London, 1965).
[5] P. Beckmann and A. Spizzichino, *Scattering of Electromagnetic Waves from Rough Surfaces* (Macmillan, New York, 1963), pp. 17–98.
[6] P. Lynch, J. Acoust. Soc. Am. **47**, 804 (1970).
[7] T. Hagfors, J. Geophys. Res. **71**, 379 (1966).
[8] D. Barrick, IEEE Trans. Antennas Propagation **16**, 449 (1968).
[9] R. Kodis, IEEE Trans. Antennas Propagation **14**, 77 (1966).
[10] M. Sancer, IEEE Trans. Antennas Propagation **17**, 577 (1969).
[11] B. Smith, IEEE Trans. Antennas Propagation **15**, 668 (1967).
[12] R. Wagner, J. Acoust. Soc. Am. **41**, 138 (1967).
[13] P. Lynch and R. Wagner, J. Acoust. Soc. Am. **47**, 816 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Boundary-Value Problems of Linear-Transport Theory—Green's Function Approach

MADHOO KANAL
Space-Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48105

(Received 27 January 1970)

Case's technique utilizing Green's functions for dealing with boundary-value problems of the neutron linear-transport theory is exploited. We show that the Fourier coefficients of the Green's function over the Case spectrum are precisely the normal modes. In particular, if we assume that the scattering kernel is rotationally invariant (which indeed we do assume) and approximate it by a degenerate kernel consisting of spherical harmonics, the set of modes is deficient for problems lacking azimuthal symmetry. We also show that the expansion of the scattering kernel, in terms of spherical harmonics (or any set of orthogonal functions for that matter), permits the linear factorization of the Fourier coefficients of the Green's function in terms of the lowest element, with the proportionality functions consisting of complete orthogonal polynomials. As a consequence of this attribute of Fourier coefficients, the eigenfunctions (continuum and discrete) also factorize, which then permits decoupling of the appropriate singular integral equations. To illustrate our idea, we solve half-space and slab problems. However, the basic procedure is kept sufficiently general so that the extension to problems involving other geometrics remains straightforward.

1. INTRODUCTION

The normal-mode (eigenfunction) expansion technique of Case,[1] in dealing with boundary-value problems, has achieved considerable success in the types of problems for which the normal modes (continuum plus discrete) form a complete orthogonal set. However, there are several problems of interest, for instance, in the theory of neutron diffusion and kinetic theory of gases,[2] where the sets of modes are either deficient or the appropriate integral equations are regular. In particular, in a recent paper by Case *et al.*,[3] it has been shown for spherical geometry that one cannot directly adapt the above-mentioned technique. In this paper, we consider the Green's function approach also due to Case.[4] We show that the Fourier coefficients of the Green's function for the appropriate neutron 1-speed transport equation over the Case spectrum are precisely the normal modes. In particular, if we assume that the scattering kernel is rotationally invariant (which indeed we do assume) and approximate it by a

degenerate kernel consisting of the spherical harmonics, then the set of modes is deficient for problems lacking azimuthal symmetry. However, if the index of degeneracy is allowed to approach infinity, then the deficiency of that set vanishes. Furthermore, we also show that the expansion of the scattering kernel, in terms of spherical harmonics (or any set of orthogonal functions for that matter), permits the linear factorization of the Fourier coefficients of the Green's function in terms of the lowest element with the proportionality functions which consist of complete orthogonal polynomials. This attribute of Fourier coefficients then leads to the factorization of eigenfunctions (continuum and discrete) and the eventual decoupling of the singular integral equations. The main advantage of Green's function technique over the normal-mode expansion technique is that the normal modes appear "naturally" in the Green's function, with the additional terms (if any) which make the set complete also appearing as an integral part of it.

To illustrate our idea, we solve half-space and slab problems. The latter type of problems are treated in somewhat greater detail than the former. In particular, two limiting cases of thick and thin slabs are considered. We begin by first presenting the basic formulas[4] and relevant mathematical tools.

2. BASIC FORMULAS

In the 1-speed approximation,[4] the neutron-transport equation we consider is

$$(1 + \mathbf{\Omega}\cdot\nabla)\Psi(\mathbf{r},\mathbf{\Omega}) = \int d\Omega' f(\mathbf{\Omega}\cdot\mathbf{\Omega}')\Psi(\mathbf{r},\mathbf{\Omega}') + Q(\mathbf{r},\mathbf{\Omega}), \quad (1)$$

where $\mathbf{\Omega}$ is the unit velocity vector, Ψ is the angular density, Q is some given source function, and $f(\mathbf{\Omega}\cdot\mathbf{\Omega}')$ is a rotationally invariant scattering kernel. The appropriate Green's function satisfies

$$(1 + \mathbf{\Omega}\cdot\nabla)G(\mathbf{r},\mathbf{\Omega};\mathbf{r}_0,\mathbf{\Omega}_0) = \int d\Omega' f(\mathbf{\Omega}\cdot\mathbf{\Omega}')G(\mathbf{r},\mathbf{\Omega}';\mathbf{r}_0,\mathbf{\Omega}_0) + \delta(\mathbf{r}-\mathbf{r}_0)\delta(\mathbf{\Omega}\cdot\mathbf{\Omega}_0). \quad (2)$$

The quadrature for the angular density is

$$\Psi(\mathbf{r},\mathbf{\Omega}) = \int_V d\Omega' d^3r' G(\mathbf{r},\mathbf{\Omega};\mathbf{r}',\mathbf{\Omega}')Q(\mathbf{r}',\mathbf{\Omega}') + \int_S d\Omega' dS' G(\mathbf{r},\mathbf{\Omega};\mathbf{r}_s',\mathbf{\Omega}')\hat{n}_i(\mathbf{r}_s')\cdot\mathbf{\Omega}'\Psi(\mathbf{r}_s',\mathbf{\Omega}'), \quad (3)$$

where V is the volume in which the angular density is to be determined, S is the boundary of V, $\mathbf{r}_s'$ is a point on S, and $\hat{n}_i$ is a unit normal pointing into V. The integral equation for the surface distribution $\Psi(\mathbf{r}_s,\mathbf{\Omega})$ is

$$\Psi(\mathbf{r}_s,\mathbf{\Omega}) = \int_V d\Omega' d^3r' G(\mathbf{r}_s,\mathbf{\Omega};\mathbf{r}',\mathbf{\Omega}')Q(\mathbf{r}',\mathbf{\Omega}') + \int_{S'} d\Omega' dS' G_{\pm}(\mathbf{r}_s,\mathbf{\Omega};\mathbf{r}_s',\mathbf{\Omega}')\hat{n}_i(\mathbf{r}_s')\cdot\mathbf{\Omega}'\Psi(\mathbf{r}_s',\mathbf{\Omega}') = 0. \quad (4)$$

The object is to construct the Green's function from Eq. (2) and solve the integral Eq. (4) for the surface distribution[5] $\Psi(\mathbf{r}_s,\mathbf{\Omega})$. Having obtained $\Psi(\mathbf{r}_s,\mathbf{\Omega})$, we then determine the angular density $\Psi(\mathbf{r},\mathbf{\Omega})$ by Eq. (3). The basic mathematical tools relevant to such a treatment are the elementary use of Fourier transforms and the theory of singular integral equations of the type

$$\tfrac{1}{2}B(\mu)\Gamma(\mu) + \frac{1}{2\pi i}\mathcal{S}\int_L \frac{d\nu}{\nu-\mu}A(\mu,\nu)\Gamma(\nu) = f(\mu). \quad (5)$$

Reduction of Eq. (4) to the integral equation (5) should become obvious soon.

3. GREEN'S FUNCTION FOR THE 1-SPEED TRANSPORT EQUATION AND EIGENFUNCTIONS

In this section, we take a cursory look at the relationship between the eigenfunctions of the 1-speed transport equation and the Fourier components of the corresponding Green's function. We express the scattering kernel $f(\mathbf{\Omega}\cdot\mathbf{\Omega}')$ in Eq. (1) in the degenerate form

$$f(\mathbf{\Omega}\cdot\mathbf{\Omega}') = \sum_{l=0}^{N}\frac{2l+1}{4\pi}b_l P_l(\mathbf{\Omega}\cdot\mathbf{\Omega}'), \quad (6)$$

where N is arbitrary. Using the addition theorem for spherical harmonics, i.e.,

$$P_l(\mathbf{\Omega}\cdot\mathbf{\Omega}') = \sum_{m=-l}^{l}\frac{4\pi}{2l+1}Y_{lm}^*(\mathbf{\Omega})Y_{lm}(\mathbf{\Omega}') \quad (7)$$

in Eq. (6), the 1-speed transport equation then may be written as

$$(1 + \mathbf{\Omega}\cdot\nabla)\Psi(\mathbf{r},\mathbf{\Omega}) = \sum_{l=0}^{N}\sum_{m=-l}^{l} b_l Y_{lm}^*(\mathbf{\Omega})\langle\Psi Y_{lm}\rangle, \quad (8)$$

where the inner product is defined by

$$\langle fg\rangle = \int d\Omega f(\mathbf{\Omega})g(\mathbf{\Omega}). \quad (9)$$

Let us consider the Fourier transform of Eq. (8), i.e., set

$$\Psi(\mathbf{r},\mathbf{\Omega}) = \frac{1}{(2\pi)^3}\int d^3k e^{i\mathbf{k}\cdot\mathbf{r}}\psi^k(\mathbf{\Omega}). \quad (10)$$

Then, Eq. (8) becomes

$$(1 + i\mathbf{k}\cdot\boldsymbol{\Omega})\psi^k(\boldsymbol{\Omega}) = \sum_{l=0}^{N}\sum_{m=-l}^{l} b_l Y_{lm}^*(\boldsymbol{\Omega})\langle\psi^k Y_{lm}\rangle. \quad (11)$$

The appropriate Green's function satisfies

$$(1 + \boldsymbol{\Omega}\cdot\nabla)G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \sum_{l=0}^{N}\sum_{m=-l}^{l} b_l Y_{lm}^*(\boldsymbol{\Omega})\langle G Y_{lm}\rangle + \delta(\mathbf{r} - \mathbf{r}_0)\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0). \quad (12)$$

To construct the Green's function, let us take the Fourier transform of Eq. (12); i.e., set

$$G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \frac{1}{(2\pi)^3}\int d^3k e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)} g(k, \boldsymbol{\Omega}; \boldsymbol{\Omega}_0). \quad (13)$$

The result is

$$g(k, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0) = \sum_{l,m} b_l \frac{Y_{lm}^*(\boldsymbol{\Omega})}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}\langle g Y_{lm}\rangle + \frac{\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}. \quad (14)$$

Now, every solution of Eq. (14) must be of the form

$$g = \sum_{l,m} b_l \frac{Y_{lm}^*(\boldsymbol{\Omega})}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}\xi_{lm}(k, \boldsymbol{\Omega}_0) + \frac{\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}, \quad (15)$$

where $\xi_{lm} \equiv \langle g Y_{lm}\rangle$ are to be determined. If we multiply both sides of Eq. (15) with $Y_{l'm'}(\boldsymbol{\Omega})$ and integrate over $\boldsymbol{\Omega}$, we get a system of linear inhomogeneous equations for ξ_{lm}. They are

$$\sum_{l,m}\xi_{lm}(k, \boldsymbol{\Omega}_0)\left(\delta_{ll'}\delta_{mm'} - b_l\left\langle\frac{Y_{lm}^* Y_{l'm'}}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}\right\rangle\right) = \frac{Y_{l'm'}(\boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}_0}. \quad (16)$$

Simple calculations will show that

$$\left\langle\frac{Y_{lm}^* Y_{l'm'}}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}\right\rangle = 2\pi\delta_{mm'}\int_{-1}^{1}\frac{d\mu'}{1 + ik\mu} Y_{lm}(\mu, 0)Y_{l'm'}(\mu, 0).$$

By using this simplification in Eq. (16), we get

$$\sum_{l=|m|}^{N}\xi_{lm}\mathcal{A}_{ll'}^m(k) = \frac{Y_{l'm}(\boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}_0}, \quad (17)$$

where

$$\mathcal{A}_{ll'}^m(k) = \delta_{ll'} - 2\pi b_l\int_{-1}^{1}\frac{d\mu}{1 + ik\mu} Y_{lm}(\mu, 0)Y_{l'm}(\mu, 0). \quad (18)$$

When the determinant (the dispersion function)

$$\Lambda_m(k) = \det|\mathcal{A}_{ll'}^m| \quad (19)$$

of the system (17) is nonzero, for any fixed m, we have the unique solution

$$\xi_{lm} = \sum_{l'=|m|}^{N}\frac{d_m(l')}{\Lambda_m}\frac{Y_{l'm}(\boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}_0}, \quad m \le l, \quad (20)$$

where $d_m(l')$ denotes the signed minor of the matrix $(\mathcal{A}_{ll'}^m)$ associated with the lth row and the l'th column. In particular, the homogeneous equations

$$g - \sum_{l,m} b_l\frac{Y_{lm}^*(\boldsymbol{\Omega})}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}\xi_{lm}(k, \boldsymbol{\Omega}_0) = 0 \quad (21)$$

and

$$\sum_{l=|m|}^{N}\xi_{lm}\mathcal{A}_{ll'}^m = 0 \quad (22)$$

then have the unique solutions $g = 0$ and $\xi_{lm} = 0$. On the other hand, when $\Lambda_m = 0$, Eq. (22) and, consequently, Eq. (16) have nonzero solutions, and the number of linearly independent solutions is equal to the nullity of the matrix $(\mathcal{A}_{ll'}^m)$ (i.e., the difference between its order and its rank). In any event, the most general Fourier representation of G is of the form

$$G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \frac{1}{(2\pi)^3}\sum_{l,m} b_l Y_{lm}^*(\boldsymbol{\Omega}) \times\int d^3k e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)}\frac{\xi_{lm}(k, \boldsymbol{\Omega}_0)}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}} + \delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)\frac{1}{(2\pi)^3}\int d^3k\frac{e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)}}{1 + i\mathbf{k}\cdot\boldsymbol{\Omega}}. \quad (23)$$

We note that the Fourier components ξ_{lm} of G, given by Eq. (20), are sectionally holomorphic functions in the complex k-vector space, with a branch cut for $k = -i\infty$ to $-i$ and i to $i\infty$, and they have poles at the zeros of the dispersion function Λ_m. In what follows, we look at ξ_{lm} in terms of their relation to the eigenfunctions of Eq. (11) over this spectrum (the Case spectrum),[6] and also examine a certain recurrence relation leading to the factorization of ξ_{lm} in terms of the lowest element ξ_{mm}.

Our first immediate observation is that, for a fixed direction of $\mathbf{k}$, the difference of boundary values of ξ_{lm} about its branch cut are precisely the continuum eigenfunctions[2] of Eq. (11); i.e., if we denote such functionals by $E_{lm}(k, \boldsymbol{\Omega})$, then

$$E_{lm}(k, \Omega_k, \phi) = \xi_{lm}^-(k, \Omega_k, \phi) - \xi_{lm}^+(k, \Omega_k, \phi) \quad (24)$$

or, explicitly,

$$E_{lm} = \sum_{l'=|m|}^{N} Y_{l'm}(\Omega_k, \phi) \times\left(\frac{d_m^-(l')}{\Lambda_m^-(1 + ik\Omega_k)_-} - \frac{d_m^+(l')}{\Lambda_m^+(1 + ik\Omega_k)_+}\right) \quad (25)$$

satisfy Eq. (11). Here, $\Omega_k = \hat{\mathbf{k}} \cdot \boldsymbol{\Omega}$, and + (−) denotes the boundary value as k approaches the branch cut from the left (right) side. On the other hand, if k_j is a simple zero[7] of $\Lambda_m(k)$, then the discrete eigenfunction [of Eq. (11)] is given by

$$F_{lm}(k_j, \Omega_k, \phi) = \lim_{k \to k_j} (k - k_j)\xi_{lm}(k, \Omega_k, \phi), \quad (26)$$

i.e.,

$$F_{lm}(k_j, \Omega_k, \phi) = \frac{1}{\Lambda'_m(k_j)} \sum_{l'=|m|}^{N} d_m\begin{pmatrix} l & \\ l' & k_j \end{pmatrix} \frac{Y_{l'm}(\Omega_k, \phi)}{1 + ik_j\Omega_k}, \quad (27)$$

where $\Lambda'_m(k_j)$ is the derivative of $\Lambda_m(k)$ evaluated at $k = k_j$.

It may seem peculiar at first sight that, for a fixed point in the Case spectrum, there are N number of eigenfunctions for l ranges from $|m|$ to N. However, we shall see presently that all such eigenfunctions are not distinct. In fact, they differ from the lowest eigenfunction ($l = |m|$) by a multiplicative factor which is a polynomial in (i/k). To see that, consider Eq. (16) rewritten in the form

$$\sum_{l,m} \xi_{lm}\left(\delta_{ll'}\delta_{mm'}(1 - b_l) + b_l\left\langle Y^*_{lm} \frac{i\mathbf{k} \cdot \boldsymbol{\Omega}}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}} Y_{l'm'}\right\rangle\right) = \frac{Y_{l'm'}(\boldsymbol{\Omega}_0)}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}_0}. \quad (28)$$

Using the recurrence relation for spherical harmonics,

$$\Omega_k Y_{lm}(\boldsymbol{\Omega}) = A_{lm}Y_{l+1\,m}(\boldsymbol{\Omega}) + A_{l-1\,m}Y_{l-1\,m}(\boldsymbol{\Omega}), \quad (29)$$

where

$$A_{lm} = \left(\frac{(l + 1 - m)(l + 1 + m)}{(2l + 1)(2l + 3)}\right)^{\frac{1}{2}}, \quad (30)$$

we obtain

$$z(b_l - 1)\xi_{lm} + A_{lm}\xi_{l+1\,m} + A_{l-1\,m}\xi_{l-1\,m} = -zY_{lm}(\boldsymbol{\Omega}_0), \quad (31)$$

where, for convenience, we have put $k = i/z$. From this equation, we conclude that

$$\xi_{lm} = h_{lm}(z)\xi_{mm} + W_{lm}(z, \boldsymbol{\Omega}_0), \quad (32)$$

where $h_{lm}(z)$ are complete orthogonal polynomials (in the Stieltjes sense) satisfying the following 3-term recurrence relation:

$$A_{lm}h_{l+1\,m}(z) + z(b_l - 1)h_{lm}(z) + A_{l-1\,m}h_{l-1\,m}(z) = 0, \quad (33)$$

and $W_{lm}(z, \boldsymbol{\Omega})$ are also polynomial in z. Equation (32) gives us the desired factorization of ξ_{lm} (mentioned above) in terms of the lowest element ξ_{mm}. Two immediate consequences of this equation are (1) the factorization of eigenfunctions and (2) a convenient representation of the dispersion function Λ_m. In other words, we have

$$E_{lm}(\nu, \boldsymbol{\Omega}) = h_{lm}(\nu)E_{mm}(\nu, \boldsymbol{\Omega}), \quad (34)$$

$$F_{lm}(\nu_i, \boldsymbol{\Omega}) = h_{lm}(\nu_i)F_{mm}(\nu_i, \boldsymbol{\Omega}), \quad (35)$$

and

$$\Lambda_m(z) = 1 - z \sum_{l=|m|} b_l h_{lm}(z)\left\langle \frac{Y^*_{lm}Y_{mm}}{z - \Omega_k}\right\rangle. \quad (36)$$

Equations (34) and (35), of course, follow by definitions (24) and (26), while Eq. (36) is obtained merely by substituting ξ_{lm} in Eq. (28) by means of Eq. (32). In particular, for the lowest element ξ_{mm}, we have

$$\Lambda_m(z)\xi_{mm}(z, \boldsymbol{\Omega}) = Y_{mm}(z)\frac{z}{z - \Omega_k} - \sum_{l=|m|}^{N} W_{lm}(z, \boldsymbol{\Omega}) \times \left(\delta_{lm} - b_l\left\langle Y^*_{lm} \frac{z}{z - \Omega_k} Y_{mm}\right\rangle\right), \quad (37)$$

from which we may readily construe the explicit forms of the lowest eigenfunctions.

The results of this section may be summarized as follows:

(i) The Green's function for a degenerate kernel of the form given by Eq. (1) was Fourier transformed. For the Fourier components (ξ_{lm}) of G, we obtained a set of inhomogeneous linear algebraic relations.

(ii) It was then shown that the difference of boundary values of ξ_{lm} about the Case-spectral line gave rise to the continuum eigenfunctions of Eq. (11), while the discrete ones consisted of the

$$\lim_{z \to \nu_j} (z - \nu_j)\xi_{lm};$$

ν_j is a simple zero of Λ_m.

(iii) Using the recurrence relation for spherical harmonics, we obtained a 3-term inhomogeneous recurrence relation for ξ_{lm} which permitted us to express all ξ_{lm} linearly in terms of the lowest coefficient ξ_{mm}. As a consequence of this factorization, all the eigenfunctions for fixed m and ν (or ν_j) become proportional to the corresponding lowest eigenfunction, with the factors being orthogonal polynomials in ν (or ν_j).

We may remark here that result (ii) is valid independently of the geometry, the type of functions used to express the scattering kernel, and the rank N of degeneracy. Result (iii), on the other hand, though valid for any geometry, is crucially dependent on the fact that we expanded the scattering kernel in terms of orthogonal functions. In other words, the Fourier

coefficients of G satisfy a 3-*term* inhomogeneous recurrence relation of the type given by Eq. (31) if and only if the scattering kernel is expanded in terms of a set of orthogonal functions. The coefficients then factorize in the way given by Eq. (32) and the corresponding eigenfunctions as given by Eqs. (34) and (35). As a final remark, we wish to state that the above factorization of ξ_{lm}, in terms of a *single* lowest element, is not possible if the scattering kernel is a function of all velocity components, such as in the energy-dependent case.[2]

In what follows we shall restrict our treatment to 1-dimensional problems. In particular, for the purpose of illustrating the general formulation discussed above, we shall consider half-space and slab problems. For the latter, the angular density in two asymptotic limits of thick and thin slabs will be given.

4. ONE-DIMENSIONAL PROBLEMS (GENERAL FORMULATION)

The 1-dimensional version of the Fourier representation of the Green's function [Eq. (23)] is

$$G(x, \mathbf{\Omega}; x_0, \mathbf{\Omega}_0) = \frac{1}{2\pi} \sum_{l,m} b_l Y_{lm}^*(\mathbf{\Omega}) \times \int_{-\infty}^{\infty} dk \frac{e^{ik(x-x_0)}}{1 + ik\mu} \xi_{lm}(k, \mathbf{\Omega}_0) + \delta(\mathbf{\Omega} \cdot \mathbf{\Omega}_0) \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \frac{e^{ik(x-x_0)}}{1 + ik\mu}, \tag{38}$$

where $\mu = \hat{x} \cdot \mathbf{\Omega}$. Let $G_>$ denote G for $x > x_0$, and $G_<$ for $x < x_0$; the point source is presumed to be at x_0. First, consider $x > x_0$. In order to express G in terms of eigenfunctions of Eq. (11), as discussed previously, consider the integral in Eq. (38) over the contour C shown in Fig. 1. Assuming that Λ_m has no zeros on the real k axis, the sum of the integrals from $-\infty$ to ∞ and that around the branch cut equals the residue arising from the zeros of the dispersion function $\Lambda_m(k)$ in the upper-half k plane[8]; since $x > x_0$, the integral along the semicircle at infinity gives zero contribution. Hence

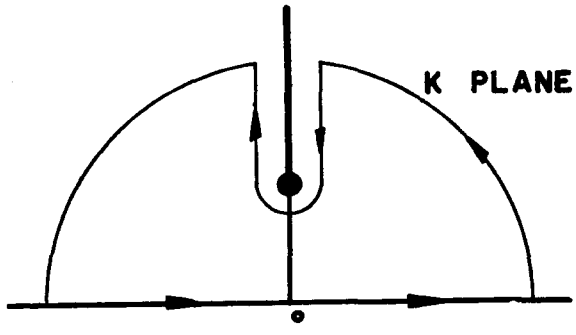

FIG. 1. Contour $x > x_0$.

$$G_> = \frac{1}{2\pi} \sum_{l,m} b_l Y_{lm}^*(\mathbf{\Omega}) \times \int_i^{i\infty} dk e^{ik(x-x_0)} \left(\frac{\xi_{lm}^-}{(1 + ik\mu)_-} - \frac{\xi_{lm}^+}{(1 + ik\mu)_+} \right) + i \sum_{l,m} b_l Y_{lm}^*(\mathbf{\Omega}) \sum_{j=1}^{M} e^{-(x-x_0)/\nu_{m_j}} F_{lm}(\nu_{m_j}, \mathbf{\Omega}_0) + \delta(\mathbf{\Omega} \cdot \mathbf{\Omega}_0) e^{-(x-x_0)/\mu} \frac{\Theta(\mu)}{\mu}, \tag{39}$$

where M is the total number of zeros ν_{m_j} of Λ_m in the upper-half k plane, $\Theta(\mu)$ is the Heaviside step function, and $F_{lm}(\nu_{m_j}, \mathbf{\Omega}_0)$ are the discrete eigenfunctions of Eq. (11). The explicit form of F_{lm} is as given by Eq. (27), with k_j replaced by i/ν_{m_j}.

Putting $k = i/\nu$ in Eq. (30) and using the Plemelj formula

$$1/(\nu - \mu)_\pm = \mathcal{P}[1/(\nu - \mu)] \mp i\pi\delta(\nu - \mu), \tag{40}$$

we re-express $G_>$ in the form

$$G_> = \frac{1}{2\pi i} \sum_{m=-N}^{N} \sum_{l=|m|}^{N} b_l Y_{lm}^*(\mathbf{\Omega}) \times \left(\mathcal{P} \int_0^1 \frac{d\nu}{(\nu - \mu)\nu} e^{-(x-x_0)/\nu} E_{lm}(\nu, \mathbf{\Omega}_0) + \pi i [\xi_{lm}^-(\mu, \mathbf{\Omega}_0) + \xi_{lm}^+(\mu, \mathbf{\Omega}_0)] e^{-(x-x_0)/\mu} \frac{\Theta(\mu)}{\mu} + 2\pi i \sum_{j=1}^{M} e^{-(x-x_0)/\nu_{m_j}} F_{lm}(\nu_{m_j}, \mathbf{\Omega}_0) \right) + \delta(\mathbf{\Omega} \cdot \mathbf{\Omega}_0) e^{-(x-x_0)/\mu} \frac{\Theta(\mu)}{\mu}, \tag{41}$$

where we have now identified $\xi_{lm}^-(\nu, \mathbf{\Omega}_0) - \xi_{lm}^+(\nu, \mathbf{\Omega}_0)$ with the continuum eigenfunctions $E_{lm}(\nu, \mathbf{\Omega}_0)$ [Eqs. (24) and (25)] and have used the identity

$$\sum_{l=0}^{N} \sum_{m=-l}^{l} \mathcal{A}_{lm} = \sum_{m=-N}^{N} \sum_{l=|m|}^{N} \mathcal{A}_{lm}. \tag{42}$$

We note that the singular part of $G_>$ is appropriately expressed in terms of the continuum eigenfunctions and has a Cauchy-type kernel, but that the second term on the right-hand side contains the sum of the boundary values of ξ_{lm}, which are not eigenfunctions.

However, if we write Eq. (17) in the form

$$\sum_{l=|m|}^{N} \xi_{lm}(z, \Omega_0)\mathcal{A}_{ll'}^{m}(z) = \frac{z}{z-\mu_0} Y_{l'm}(\Omega_0) \quad (43)$$

and consider the difference of its boundary values as z approaches the cut ($-1 \leq \nu \leq 1$) from the top and the bottom, then we obtain

$$\sum_{l=|m|}^{N} [\mathcal{A}_{ll'}^{m+}(\mu) - \mathcal{A}_{ll'}^{m-}(\mu)](\xi_{lm}^{+} + \xi_{lm}^{-}) = \sum_{l=|m|}^{N} (\mathcal{A}_{ll'}^{m+} + \mathcal{A}_{ll'}^{m-})(\xi_{lm}^{+} - \xi_{lm}^{-}) - 4\pi i\mu\delta(\mu-\mu_0)Y_{l'm}(\Omega_0), \quad (44)$$

which relates $\xi_{lm}^{+} + \xi_{lm}^{-}$ to the eigenfunctions

$$\xi_{lm}^{+} - \xi_{lm}^{-} (\equiv E_{lm}).$$

By means of this equation, we may now replace the second term in Eq. (41) by the right-hand side of Eq. (44), if we note that [see Eq. (18)]

$$\mathcal{A}_{ll'}^{m+}(\mu) - \mathcal{A}_{ll'}^{m-}(\mu) = 4\pi^2 i b_l \mu Y_{lm}^{*}(\Omega) Y_{l'm}(\Omega). \quad (45)$$

Thus, using Eqs. (44), (45), and the factorizations given by Eqs. (34) and (35) in Eq. (41), we get

$$\begin{aligned} G_{>} = &\sum_{m=-N}^{N} \frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)} \\ &\times \left(\frac{1}{2\pi i} \mathcal{P}\int_0^1 \frac{d\nu}{\nu-\mu} e^{-(x-x_0)/\nu} A_m(\mu, \nu) \frac{E_{mm}(\nu, \Omega_0)}{4\pi^2 i\nu} \right. \\ &+ \tfrac{1}{2} B_m(\mu) \frac{E_{mm}(\mu, \Omega_0)}{4\pi^2 i\mu} e^{-(x-x_0)/\mu}\Theta(\mu) \\ &\left. + \sum_{j=1}^{M} e^{-(x-x_0)/\nu_{mj}} A_m(\mu, \nu_{mj}) \frac{F_{mm}(\nu_{mj}, \Omega_0)}{4\pi^2} \right) \\ &+ \frac{1}{\mu}\delta(\mu-\mu_0)e^{-(x-x_0)/\mu}\Theta(\mu) \\ &\times \left(\delta(\phi-\phi_0) - \frac{1}{2\pi} \sum_{m=-N}^{N} e^{im(\phi_0-\phi)} \right), \end{aligned} \quad (46)$$

where

$$A_m(\mu, \nu) = \sum_{l=|m|}^{N} [\mathcal{A}_{lm}^{m+}(\mu) - \mathcal{A}_{lm}^{m-}(\mu)] h_{lm}(\nu), \quad (47)$$

$$B_m(\mu) = \sum_{l=|m|}^{N} [\mathcal{A}_{lm}^{m+}(\mu) + \mathcal{A}_{lm}^{m-}(\mu)] h_{lm}(\mu), \quad (48)$$

$$\mathcal{A}_{lm}^{m\pm}(\mu) = \delta_{lm} - 2\pi b_l \times \int_{-1}^{1} d\mu' \frac{\mu}{(\mu-\mu')_{\pm}} Y_{lm}(\mu', 0) Y_{mm}(\mu', 0), \quad (49)$$

and $h_{lm}(\nu)$ are polynomials given by the recurrence relation (33).

Similarly, for $x < x_0$, we have

$$\begin{aligned} G_{<} = &-\sum_{m=-N}^{N} \frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)} \\ &\times \left(\frac{1}{2\pi i} \mathcal{P}\int_{-1}^{0} \frac{d\nu}{\nu-\mu} e^{-(x-x_0)/\nu} A_m(\mu, \nu) \frac{E_{mm}(\nu, \Omega_0)}{4\pi^2 i\nu} \right. \\ &+ \tfrac{1}{2} B_m(\mu) \frac{E_{mm}(\mu, \Omega_0)}{4\pi^2 i\mu} e^{-(x-x_0)/\mu}\Theta(-\mu) \\ &\left. + \sum_{j=1}^{M} e^{-(x-x_0)/\nu_{mj}} A_m(\mu, -\nu_{mj}) \frac{F_{mm}(-\nu_{mj}, \Omega_0)}{4\pi^2} \right) \\ &- \mu\delta(\mu-\mu_0)e^{-(x-x_0)/\mu}\Theta(-\mu) \\ &\times \left(\delta(\phi-\phi_0) - \sum_{m=-N}^{N} e^{im(\phi_0-\phi)} \right). \end{aligned} \quad (50)$$

A few remarks are due here. In the expression (46), the last two terms cannot cancel so long as N, the rank of degeneracy of the scattering kernel, is finite. In other words, for the problems lacking azimuthal symmetry, the set of eigenfunctions (E_{mm}, F_{mm}) do not possess half-range completeness for degenerate kernels. This was to be expected, because any arbitrary function of ϕ cannot be expanded in terms of a finite set of $e^{im\phi}$. Consequently, the last two terms are there to substantiate the deficiency of the set (E_{mm}, F_{mm}), as may be seen by letting N approach infinity; the terms cancel, and hence the deficiency becomes zero. On the other hand, for azimuthally symmetric problems, the above set is complete over the half-range of ν; this is readily seen by integrating Eq. (46) with respect to ϕ from 0 to 2π. The same remarks apply to $G_{>}$.

5. APPLICATIONS

A. Half-Space Problems

As an application of the above formulation, let us first consider the half-space problems. Shifting the point source to the origin ($x_0 = 0$), we may write the integral for $\Psi(x, \Omega)$ [see Eq. (3)] in the form

$$\begin{aligned} \Psi(x, \Omega) = f(x, \Omega) + &\sum_{m=-N}^{N} \frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)} \\ &\times \left(\frac{1}{2\pi i} \mathcal{P}\int_0^1 \frac{d\nu}{\nu-\mu} e^{-x/\nu} A_m(\mu, \nu)\Gamma_m(\nu) \right. \\ &+ \tfrac{1}{2} B_m(\mu) e^{-x/\mu}\Theta(\mu)\Gamma_m(\mu) \\ &\left. + \sum_{j=1}^{M} e^{-x/\nu_{mj}} A_m(\mu, \nu_{mj})\Gamma_m^0(\nu_{mj}) \right) \\ &+ e^{-x/\mu}\Theta(\mu)\left(\Psi(0, \Omega) - \frac{1}{2\pi} \sum_{m=-N}^{N} e^{-im\phi} \right. \\ &\left. \times \int_0^{2\pi} d\phi' e^{im\phi'} \Psi(0, \mu, \phi') \right), \end{aligned} \quad (51)$$

where $f(x, \Omega)$ is the angular density due to source

and where

$$\Gamma_m(\nu) = \int d\Omega' \mu' \Psi(0, \Omega') \frac{E_{mm}(\nu, \Omega')}{4\pi^2 i \nu} \tag{52}$$

and

$$\Gamma^0_m(\nu_{m_j}) = \int d\Omega' \mu' \Psi(0, \Omega') \frac{F_{mm}(\nu_{m_j}, \Omega')}{4\pi^2} \tag{53}$$

are the coefficients to be determined from the given boundary condition. An equation that determines them is

$$\begin{aligned}\Psi(0, \Omega) = f(0, \Omega) + \sum_{m=-N}^{N} \frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)} \\ \times \left(\frac{1}{2\pi i} \mathcal{J} \int_0^1 \frac{d\nu}{\nu - \mu} A_m(\mu, \nu) \Gamma_m(\nu) \right. \\ + \tfrac{1}{2} B_m(\mu) \Gamma_m(\mu) \Theta(\mu) \\ \left. + \sum_{j=1}^{M} A_m(\mu, \nu_{m_j}) \Gamma^0_m(\nu_{m_j}) \right) \\ + \Theta(\mu) \left(\Psi(0, \Omega) - \frac{1}{2\pi} \sum_{m=-N}^{N} e^{-im\phi} \right. \\ \left. \times \int_0^{2\pi} d\phi' e^{im\phi'} \Psi(0, \mu, \phi') \right).\end{aligned} \tag{54}$$

In solving this integral equation for any specific problem, we assume that $\Psi(0, \Omega)$ for $\mu > 0$ is known, so that

$$\Psi_m(0, \mu) = \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{im\phi} \Psi(0, \Omega), \quad \mu > 0, \tag{55}$$

is also known. This entails a considerable amount of simplification in the solution of the integral Eq. (54). If we multiply it by $e^{im'\phi}$ and integrate over ϕ from 0 to 2π, we obtain a set of $2N + 1$ decoupled integral equations of the form

$$\begin{aligned}\frac{1}{2\pi i} \mathcal{J} \int_0^1 \frac{d\nu}{\nu - \mu} A_m(\mu, \nu) \Gamma_m(\nu) \\ + \tfrac{1}{2} B_m(\mu) \Gamma_m(\mu) = \Phi_m(\mu),\end{aligned} \tag{56}$$

where

$$\begin{aligned}\Phi_m(\mu) = \mu Y_{mm}(\mu, 0)[\Psi_m(0, \mu) - f_m(\mu)] \\ - \sum_{j=1}^{M} A_m(\mu, \nu_{m_j}) \Gamma^0_m(\nu_{m_j})\end{aligned} \tag{57}$$

and

$$f_m(\mu) = \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{im\phi} f(0, \Omega). \tag{58}$$

The set of Eqs. (56) are singular integral equations, which may be solved by the standard procedure due to Muskhelishvili.[9] In fact, an elaborate solution for $m = 0$, but arbitrary N, has been given by Mika.[10] Since the procedure for $m \neq 0$ is the same as for $m = 0$, we merely state the pertinent results.

Let us assume that the zeros of $\Lambda_m(z)$ [see Eq. (36)] are nondegenerate and the polynomials $h_{lm}(z)$ [see Eq. (33)] are simple, i.e., of degree precisely $l - m$. Splitting the kernel in Eq. (56) into the singular and the regular parts, we rewrite it in the form

$$\begin{aligned}\frac{\Lambda^+_m(\mu) - \Lambda^-_m(\mu)}{2\pi i} \mathcal{J} \int_0^1 \frac{d\nu}{\nu - \mu} \Gamma_m(\nu) \\ + \tfrac{1}{2}[\Lambda^+_m(\mu) + \Lambda^-_m(\mu)] \Gamma_m(\mu) = \tilde{\Phi}_m(\mu),\end{aligned} \tag{59}$$

where

$$\begin{aligned}\tilde{\Phi}_m(\mu) = \Phi_m(\mu) \\ - \frac{1}{2\pi i} \int_0^1 d\nu \Gamma_m(\nu) \frac{A_m(\mu, \nu) - A_m(\mu, \mu)}{\nu - \mu}\end{aligned} \tag{60}$$

and where we have used the fact that

$$A_m(\mu, \mu) = \Lambda^+_m(\mu) - \Lambda^-_m(\mu) \tag{61}$$

and

$$B_m(\mu) = \Lambda^+_m(\mu) + \Lambda^-_m(\mu) \tag{62}$$

[compare Eqs. (15), (16), and (17) with Eq. (36)]. In Eq. (60), the integral may be written as a sum over the moments of $\Gamma_m(\nu)$ as follows:

$$\begin{aligned}\frac{A_m(\mu, \nu) - A_m(\mu, \mu)}{\nu - \mu} \\ = 4\pi^2 i \mu Y_{mm}(\mu, 0) \sum_{l=|m|}^{N} b_l \frac{Y_{lm}(\mu, 0)}{\nu - \mu} [h_{lm}(\nu) - h_{lm}(\mu)].\end{aligned} \tag{63}$$

If we write $h_{lm}(\nu)$ as

$$h_{lm}(\nu) = \sum_{k=0}^{l-|m|} C_k(l, m) \nu^k, \tag{64}$$

then

$$\frac{h_{lm}(\nu) - h_{lm}(\mu)}{\nu - \mu} = \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m) \mu^{k-j-1} \nu^j. \tag{65}$$

Substituting the appropriate ratio in Eq. (63) by means of Eq. (65), we obtain

$$\begin{aligned}\frac{A_m(\mu, \nu) - A_m(\mu, \mu)}{\nu - \mu} \\ = 4\pi^2 i \mu Y_{mm}(\mu, 0) \sum_{l=|m|+1}^{N} b_l Y_{lm}(\mu, 0) \\ \times \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m) \mu^{k-j-1} \nu^j.\end{aligned} \tag{66}$$

Denoting the moments of $\Gamma_m(\nu)$ by g_{jm}, i.e.,

$$g_{jm} = \int_0^1 d\nu \nu^j \Gamma_m(\nu), \tag{67}$$

we re-express $\tilde{\Phi}_m$ [Eq. (60)] in the form

$$\begin{aligned}\tilde{\Phi}_m(\mu) = \Phi_m(\mu) - 2\pi i \mu Y_{mm}(\mu, 0) \sum_{l=|m|+1}^{N} b_l Y_{lm}(\mu, 0) \\ \times \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m) \mu^{k-j-1} g_{jm}.\end{aligned} \tag{68}$$

The solution of Eq. (59) is

$$\Gamma_m(\nu) = \frac{\Lambda_m^+(\nu) + \Lambda_m^-(\nu)}{2\Lambda_m^+(\nu)x_m^-(\nu)}\tilde{\Phi}_m(\nu) - \frac{\Lambda_m^+(\nu) - \Lambda_m^-(\nu)}{\Lambda_m^+(\nu)\Lambda_m^-(\nu)}\mathcal{P}\frac{1}{2\pi i}\int_0^1 d\mu\,\frac{\tilde{\Phi}_m(\mu)}{\nu - \mu}\frac{x_m^-(\mu)}{\Lambda_m^-(\mu)}, \tag{69}$$

while the conditions that determine $\Gamma_m^0(\nu_{m_j})$ are

$$\int_0^1 d\mu\mu^j \frac{x_m^-(\mu)}{\Lambda_m^-(\mu)}\tilde{\Phi}_m(\mu) = 0, \quad j = 0, 1, \cdots, M-1, \tag{70}$$

where

$$x_m(z) = \frac{1}{(1-z)^M}\exp\left(\frac{1}{\pi}\int_0^1 \frac{d\mu}{\mu - z}\arg\Lambda_m^+(\mu)\right). \tag{71}$$

Equations (70) give just the sufficient number of conditions to determine the unknown discrete coefficients $\Gamma_m^0(\nu_{m_j})$. The moments g_{jm} may be evaluated by using Eq. (67).

B. Slab Problems

Let us take the volume V under consideration to be the slab between $x = -\frac{1}{2}L$ and $x = \frac{1}{2}L$. Assuming that there are no sources ($Q = 0$), we see that the integral representation of $\Psi(x, \mathbf{\Omega})$, by virtue of Eqs. (3), (46), and (50), is then

$$\begin{aligned}\Psi(x, \mathbf{\Omega}) = &\sum_{m=-N}^{N}\frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)}\\ &\times\left(\mathcal{P}\frac{1}{2\pi i}\int_0^1\frac{d\nu}{\nu-\mu}e^{-(x+\frac{1}{2}L)/\nu}A_m(\mu,\nu)\Gamma_m^{(1)}(\nu)\right.\\ &+\tfrac{1}{2}B_m(\mu)e^{-(x+\frac{1}{2}L)/\mu}\Theta(\mu)\Gamma_m^{(1)}(\mu)\\ &\left.+\sum_{j=1}^{M}e^{-(x+\frac{1}{2}L)/\nu_{m_j}}A_m(\mu,\nu_{m_j})D_m^{(1)}(\nu_{m_j})\right)\\ &+e^{-(x+\frac{1}{2}L)/\mu}\Theta(\mu)\left(\Psi(-\tfrac{1}{2}L,\mathbf{\Omega})-\frac{1}{2\pi}\right.\\ &\left.\times\sum_{m=-N}^{N}e^{-im\phi}\int_0^{2\pi}d\phi' e^{im\phi'}\Psi(-\tfrac{1}{2}L,\mu,\phi')\right)\\ &-\sum_{m=-N}^{N}\frac{e^{-im\phi}}{\mu Y_{mm}(\mu,0)}\\ &\times\left(\mathcal{P}\frac{1}{2\pi i}\int_{-1}^{0}\frac{d\nu}{\nu-\mu}e^{-(x-\frac{1}{2}L)/\nu}\Gamma_m^{(2)}(\nu)A_m(\mu,\nu)\right.\\ &+\tfrac{1}{2}B_m(\mu)e^{-(x-\frac{1}{2}L)/\mu}\Theta(-\mu)\Gamma_m^{(2)}(\mu)\\ &\left.+\sum_{j=1}^{M}e^{-(x-\frac{1}{2}L)/\nu_{m_j}}A_m(\mu,-\nu_{m_j})D_m^{(2)}(\nu_{m_j})\right)\\ &-e^{-(x-\frac{1}{2}L)/\mu}\Theta(-\mu)\left(\Psi(\tfrac{1}{2}L,\mathbf{\Omega})-\frac{1}{2\pi}\right.\\ &\left.\times\sum_{m=-N}^{N}e^{-im\phi}\int_0^{2\pi}d\phi' e^{im\phi'}\Psi(\tfrac{1}{2}L,\mu,\phi')\right).\end{aligned} \tag{72}$$

The coefficients $\Gamma_m^{(1),(2)}(\nu)$ and $D_m^{(1),(2)}(\nu_{m_j})$ in Eq. (72), which are to be determined, are defined as

$$\Gamma_m^{(1),(2)}(\nu) = \int d\Omega'\mu'\Psi(\mp\tfrac{1}{2}L,\mathbf{\Omega}')\frac{E_{mm}(\nu,\mathbf{\Omega}')}{4\pi^2 i\nu} \tag{73}$$

and

$$D_m^{(1),(2)}(\nu_{m_j}) = \int d\Omega'\mu'\Psi(\mp\tfrac{1}{2}L,\mathbf{\Omega}')\frac{F_{mm}(\pm\nu_{m_j},\mathbf{\Omega}')}{4\pi^2}, \tag{74}$$

where $\Psi(\mp\frac{1}{2}L, \mathbf{\Omega})$ is the surface distribution at $x = \mp\frac{1}{2}L$. The rest of the symbols have the same meaning as previously.

In dealing with any particular problem, we assume that $\Psi(\mp\frac{1}{2}L, \mathbf{\Omega})$ for $(\substack{\mu>0\\ \mu<0})$ are known. In that case, we may reduce Eq. (73) to two sets of decoupled singular integral equations by letting x approach $\mp\frac{1}{2}L$, multiplying both sides by $e^{im'\phi}$, and integrating over ϕ' from 0 to 2π. The result is

$$\begin{aligned}&\tfrac{1}{2}B_m(\mu)\Gamma_m^{(1)}(\mu) + \mathcal{P}\frac{1}{2\pi i}\int_0^1\frac{d\nu}{\nu-\mu}A_m(\mu,\nu)\Gamma_m^{(1)}(\nu)\\ &\quad+\sum_{j=1}^{M}A_m(\mu,\nu_{m_j})D_m^{(1)}(\nu_{m_j})\\ &\quad-\sum_{j=1}^{M}e^{-L/\nu_{m_j}}A_m(\mu,-\nu_{m_j})D_m^{(2)}(\nu_{m_j})\\ &\quad-\frac{1}{2\pi i}\int_{-1}^{0}\frac{d\nu}{\nu-\mu}e^{L/\nu}A_m(\mu,\nu)\Gamma_m^{(2)}(\nu)\\ &=\mu Y_{mm}(\mu,0)\Psi_m(-\tfrac{1}{2}L,\mu),\quad \mu>0,\end{aligned} \tag{75}$$

$$\begin{aligned}&\tfrac{1}{2}B_m(\mu)\Gamma_m^{(2)}(\mu) + \mathcal{P}\frac{1}{2\pi i}\int_{-1}^{0}\frac{d\nu}{\nu-\mu}A_m(\mu,\nu)\Gamma_m^{(2)}(\nu)\\ &\quad+\sum_{j=1}^{M}A_m(\mu,-\nu_{m_j})D_m^{(2)}(\nu_{m_j})\\ &\quad-\sum_{j=1}^{M}e^{-L/\nu_{m_j}}A_m(\mu,\nu_{m_j})D_m^{(1)}(\nu_{m_j})\\ &\quad-\frac{1}{2\pi i}\int_0^{1}\frac{d\nu}{\nu-\mu}e^{-L/\nu}A_m(\mu,\nu)\Gamma_m^{(1)}(\nu)\\ &=-\mu Y_{mm}(\mu,0)\Psi_m(\tfrac{1}{2}L,\mu),\quad \mu<0.\end{aligned} \tag{76}$$

Clearly, exact solutions of these integral equations are not feasible. However, they are well suited for approximations in the asymptotic limits.

1. Thick Slabs ($L \gg 1$)

For this limiting case we can solve Eqs. (75) and (76) for the coefficients by the iterative procedure discussed in Ref. 4. Thus, in the zeroth approximation,

we ignore the terms involving the exponentials $e^{-L/\nu}$. Equations (75) and (76) then reduce to Eq. (56) for the half-space problems. Let us, therefore, assume for a moment that $\Gamma_m^{(2)}(\nu)$, $D_m^{(2)}(\nu_{m_j})$, and $\Gamma_m^{(1)}(\nu)$, $D_m^{(1)}(\nu_{m_j})$ are known in Eqs. (75) and (76), respectively. Then, formally, solutions of Eqs. (75) and (76) are

$$\Gamma_m^{(1)}(\nu) = \frac{\Lambda_m^+(\nu) + \Lambda_m^-(\nu)}{2\Lambda_m^+(\nu)x_m^-(\nu)}\Phi_m^{(1)}(\nu) - \frac{\Lambda_m^+(\nu) - \Lambda_m^-(\nu)}{\Lambda_m^+(\nu)\Lambda_m^-(\nu)} I_m^{(1)}(\nu), \quad (77)$$

where

$$\begin{aligned}\Phi_m^{(1)}(\nu) = {} & \nu Y_{mm}(\nu, 0)\Psi_m(-\tfrac{1}{2}L, \nu) \\ & - \sum_{j=1}^{M} A_m(\nu, \nu_{m_j}) D_m^{(1)}(\nu_{m_j}) \\ & + \sum_{j=1}^{M} e^{-L/\nu_{m_j}} A_m(\nu, -\nu_{m_j}) D_m^{(2)}(\nu_{m_j}) \\ & + \frac{1}{2\pi i}\int_{-1}^{0} \frac{d\nu'}{\nu' - \nu} e^{L/\nu'} A_m(\nu, \nu')\Gamma_m^{(2)}(\nu') \\ & - 2\pi i \nu Y_{mm}(\nu, 0) \sum_{l=|m|+1}^{N} b_l Y_{lm}(\nu, 0) \\ & \times \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m)\nu^{k-j-1} g_{jm}^{(1)}, \end{aligned} \quad (78)$$

$$g_{jm}^{(1)} = \int_0^1 d\nu \nu^j \Gamma_m^{(1)}(\nu), \quad (79)$$

and

$$\begin{aligned} I_m^{(1)}(\nu) = {} & \mathcal{S}\frac{1}{2\pi i}\int_0^1 \frac{d\mu'}{\nu - \mu'} \frac{x_m^-(\mu')}{\Lambda_m^-(\mu')} \\ & \times \mu' Y_{mm}(\mu', 0)\Psi_m(-\tfrac{1}{2}L, \mu') \\ & - \sum_{j=1}^{M} [D_m^{(1)}(\nu_{m_j}) R_m^{(1)}(\nu, \nu_{m_j}) \\ & - e^{-L/\nu_{m_j}} D_m^{(2)}(\nu_{m_j}) R_m^{(1)}(\nu, -\nu_{m_j})] \\ & + \frac{1}{2\pi i}\int_{-1}^{0} \frac{d\nu'}{\nu - \nu'} e^{L/\nu'}\Gamma_m^{(2)}(\nu') \\ & \times [R_m^{(1)}(\nu', \nu') - R_m^{(1)}(\nu, \nu')], \end{aligned} \quad (80)$$

with

$$R_m^{(1)}(\nu, \nu') = \mathcal{S}\frac{1}{2\pi i}\int_0^1 \frac{d\mu'}{\nu - \mu'} \frac{x_m^-(\mu')}{\Lambda_m^-(\mu')} A_m(\mu', \nu'). \quad (81)$$

The additional conditions that determine the discrete coefficients are

$$\int_0^1 d\mu \mu^j \frac{x_m^-(\mu)}{\Lambda_m^-(\mu)} \Phi_m^{(1)}(\mu) = 0, \quad j = 0, 1, \cdots, M - 1. \quad (82)$$

Similarily, for Eq. (76), we have

$$\Gamma_m^{(2)}(\nu) = \frac{\Lambda_m^+(\nu) + \Lambda_m^-(\nu)}{2\Lambda_m^+(\nu)x_m^-(\nu)}\Phi_m^{(2)}(\nu) - \frac{\Lambda_m^+(\nu) - \Lambda_m^-(\nu)}{\Lambda_m^+(\nu)\Lambda_m^-(\nu)} I_m^{(2)}(\nu), \quad (83)$$

where

$$\begin{aligned}\Phi_m^{(2)}(\mu) = {} & -\Big(\mu Y_{mm}(\mu, 0)\Psi_m(\tfrac{1}{2}L, \mu) \\ & - \sum_{j=1}^{M} A_m(\mu, -\nu_{m_j}) D_m^{(2)}(\nu_{m_j}) \\ & + \sum_{j=1}^{M} e^{-L/\nu_{m_j}} A_m(\mu, \nu_{m_j}) D_m^{(1)}(\nu_{m_j}) \\ & + \int_0^1 \frac{d\nu}{\nu - \mu} e^{-L/\nu} A_m(\mu, \nu)\Gamma_m^{(1)}(\nu) \\ & - 2\pi i \mu Y_{mm}(\mu, 0) \sum_{l=|m|+1}^{N} b_l Y_{lm}(\mu, 0) \\ & \times \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m)\mu^{k-j-1} g_{jm}^{(2)}\Big), \end{aligned} \quad (84)$$

$$g_{jm}^{(2)} = \int_{-1}^{0} d\nu \nu^j \Gamma_m^{(2)}(\nu), \quad (85)$$

$$\begin{aligned} I_m^{(2)}(\nu) = {} & -\mathcal{S}\frac{1}{2\pi i}\int_{-1}^{0} \frac{d\mu'}{\nu - \mu'} \frac{x_m^-(\mu')}{\Lambda_m^-(\mu')} \\ & \times \mu' Y_{mm}(\mu', 0)\Psi_m(\tfrac{1}{2}L, \mu') \\ & + \sum_{j=1}^{M} [D_m^{(2)}(\nu_{m_j}) R_m^{(2)}(\nu, -\nu_{m_j}) \\ & - e^{-L/\nu_{m_j}} D_m^{(1)}(\nu_{m_j}) R_m^{(2)}(\nu, \nu_{m_j})] \\ & - \frac{1}{2\pi i}\int_0^1 \frac{d\nu'}{\nu - \nu'} e^{-L/\nu'}\Gamma_m^{(1)}(\nu') \\ & \times [R_m^{(2)}(\nu', \nu') - R_m^{(2)}(\nu, \nu')], \end{aligned} \quad (86)$$

$$R_m^{(2)}(\nu, \nu') = \mathcal{S}\frac{1}{2\pi i}\int_{-1}^{0} \frac{d\mu'}{\nu - \mu'} \frac{x_m^-(\mu')}{\Lambda_m^-(\mu')} A_m(\mu', \nu'). \quad (87)$$

The additional conditions are

$$\int_{-1}^{0} d\mu \mu^j \frac{x_m^-(\mu)}{\Lambda_m^-(\mu)} \Phi_m^{(2)}(\mu) = 0, \quad j = 0, 1, \cdots, M.$$

Consider Eq. (77) first. In the zeroth approximation, ignore all the terms involving the exponentials. The coefficients $\Gamma_m^{(1)}(\nu)$ {denoting the degree of approximation as $[\Gamma_m^{(1)}(\nu)]_n$} are then given by

$$[\Gamma_m^{(1)}(\nu)]_0 = \frac{\Lambda_m^+(\nu) + \Lambda_m^-(\nu)}{2\Lambda_m^+(\nu)x_m^-(\nu)}[\Phi_m^{(1)}(\nu)]_0 - \frac{\Lambda_m^+(\nu) - \Lambda_m^-(\nu)}{\Lambda_m^+(\nu)\Lambda_m^-(\nu)}[I_m^{(1)}(\nu)]_0, \quad (88)$$

where

$$[\Phi_m^{(1)}(\mu)]_0 = \mu Y_{mm}(\mu, 0)\Psi_m(-\tfrac{1}{2}L, \mu) - \sum_{j=1}^{M} A_m(\mu, \nu_{m_j}) D_m^{(1)}(\nu_{m_j}) - 2\pi i \mu Y_{mm}(\mu, 0) \sum_{l=|m|+1}^{N} b_l Y_{lm}(\mu, 0) \times \sum_{k=1}^{l-|m|} \sum_{j=0}^{k-1} C_k(l, m)\mu^{k-j-1} g_{jm}^{(2)} \quad (89)$$

and

$$[I_m^{(1)}(\nu)]_0 = \mathcal{S} \frac{1}{2\pi i} \int_0^1 \frac{d\mu'}{\nu - \mu'} \frac{x_m^-(\mu')}{\Lambda_m^-(\mu')} \times \mu' Y_{mm}(\mu', 0)\Psi_m(-\tfrac{1}{2}L, \mu') - \sum_{j=1}^{M} D_m^{(1)}(\nu_{m_j}) R_m^{(1)}(\nu, \nu_{m_j}). \quad (90)$$

Similar quantities for $\Gamma_m^{(2)}$ should be obvious. In the first approximation, the correction to Eq. (89) for $\Gamma_m^{(1)}(\nu)$ is obtained simply by retaining the exponential terms in Eqs. (78) and (80), with $\Gamma_m^{(2)}(\nu)$ and $D_m^{(2)}(\nu_{m_j})$ replaced by $[\Gamma_m^{(2)}(\nu)]_0$ and $[D_m^{(2)}(\nu_{m_j})]_0$, respectively. Thus,

$$[\Gamma_m^{(1)}(\nu)]_1 = \frac{\Lambda_m^+(\nu) + \Lambda_m^-(\nu)}{2\Lambda_m^+(\nu)x_m^-(\nu)} [\Phi_m^{(1)}(\nu)]_1 - \frac{\Lambda_m^+(\nu) - \Lambda_m^-(\nu)}{\Lambda_m^+(\nu)\Lambda_m^-(\nu)} [I_m^{(1)}(\nu)]_1, \quad (91)$$

where

$$[\Phi_m^{(1)}(\nu)]_1 = [\Phi_m^{(1)}(\nu)]_0 + \sum_{j=1}^{M} e^{-L/\nu_{m_j}} A_m(\nu, -\nu_{m_j})[D_m^{(2)}(\nu_{m_j})]_0 + \frac{1}{2\pi i} \int_{-1}^{0} \frac{d\nu'}{\nu' - \nu} e^{L/\nu'} A_m(\nu, \nu')[\Gamma_m^{(2)}(\nu')]_0 \quad (92)$$

and

$$[I_m^{(1)}(\nu)]_1 = [I_m^{(1)}(\nu)]_0 + \sum_{j=1}^{M} e^{-L/\nu_{m_j}} [D_m^{(2)}(\nu_{m_j})]_0 R_m^{(1)}(\nu, -\nu_{m_j}) + \frac{1}{2\pi i} \int_{-1}^{0} \frac{d\nu'}{\nu - \nu'} e^{L/\nu'} [\Gamma_m^{(2)}(\nu')]_0 \times [R_m^{(1)}(\nu', \nu') - R_m^{(1)}(\nu, \nu')]. \quad (93)$$

The same iterative procedure may be followed to approximate the discrete coefficients $D_m^{(1)}(\nu_{m_j})$ which are determined by Eq. (82). The procedure for obtaining $\Gamma_m^{(2)}$ and $D_m^{(2)}$ is exactly the same. Here we omit the details.

2. *Thin Slabs*

Because this situation is physically much simpler than the limiting case ($L \gg 1$) considered previously, one can obtain the integral representation for $\Psi(x, \Omega)$ by dealing directly with Eq. (3). The approximation procedure for various other situations is discussed in Refs. 2 and 4. To avoid repetition, we merely state the pertinent results here. Thus, if we write

$$\Psi(-\tfrac{1}{2}L, \Omega) = \Psi(\tfrac{1}{2}L, \Omega) + \tilde{\Psi}(-\tfrac{1}{2}L, \Omega), \quad \mu < 0, \quad (94)$$

$$\Psi(\tfrac{1}{2}L, \Omega) = \Psi(-\tfrac{1}{2}L, \Omega) + \tilde{\Psi}(\tfrac{1}{2}L, \Omega), \quad \mu > 0, \quad (95)$$

where $\tilde{\Psi}(\mp\frac{1}{2}L, \Omega)$ are to be of order L, then one can show that[11]

$$\tilde{\Psi}(\tfrac{1}{2}L, \Omega) = \int d\Omega' \mu' [\Psi(-\tfrac{1}{2}L, \Omega')\Theta(\mu') + \Psi(\tfrac{1}{2}L, \Omega')\Theta(-\mu')] \cdot \{G_>(\tfrac{1}{2}L, \Omega; -\tfrac{1}{2}L, \Omega') - G_+(-\tfrac{1}{2}L, \Omega; -\tfrac{1}{2}L, \Omega') + G_<(-\tfrac{1}{2}L, \Omega; \tfrac{1}{2}L, \Omega') - G_-(-\tfrac{1}{2}L, \Omega; -\tfrac{1}{2}L, \Omega')\}, \quad (96)$$

where

$$G_\pm(-\tfrac{1}{2}L, \Omega; -\tfrac{1}{2}L, \Omega') = \lim_{x \to -\frac{1}{2}L \{\text{From within } V \atop \text{From without } V\}} G(x, \Omega; -\tfrac{1}{2}L, \Omega'). \quad (97)$$

For a homogeneous medium we have

$$\tilde{\Psi}(-\tfrac{1}{2}L, \Omega) = -\tilde{\Psi}(\tfrac{1}{2}L, \Omega). \quad (98)$$

The combination of Green's functions, occurring in the right-hand side of Eq. (96), may be calculated explicitly by means of Eqs. (46) and (50). It is given by

$$G_> - G_+ + G_< - G_- = \sum_{m=-N}^{N} \frac{e^{-im\phi}}{\mu Y_{mm}(\mu, 0)} \times \left(\int_0^1 d\nu (e^{-L/\nu} - 1) E_{mm}(\nu, \Omega') H_m(\nu, \mu) + \frac{1}{4\pi^2} \sum_{j=1}^{M} (e^{-L/\nu_{m_j}} - 1)(A_m(\mu, \nu_{m_j}) F_{mm}(\nu_{m_j}, \Omega') - A_m(\mu, -\nu_{m_j}) F_{mm}(-\nu_{m_j}, \Omega')) \right) + \frac{1}{\mu} \delta(\mu - \mu') \left(\delta(\phi - \phi') - \frac{1}{2\pi} \sum_{m=-N}^{N} e^{im(\phi' - \phi)} \right) \times [(e^{-L/\mu} - 1)\Theta(\mu) - (e^{L/\mu} - 1)\Theta(-\mu)], \quad (99)$$

where for convenience we have defined

$$H_m(\nu,\mu)=\frac{1}{2\pi i}\,\mathcal{P}\,\frac{1}{\nu-\mu}\,\frac{A(\mu,\nu)}{4\pi^2 i\nu}+\frac{B_m(\mu)}{8\pi^2 i\mu}\,\delta(\nu-\mu). \tag{100}$$

In general, the contribution from the terms involving discrete eigenfunctions [in Eq. (99)] is small compared to the terms involving the continuum eigenfunctions. Let us therefore ignore that term and further approximate the terms involving exponentials as follows:

$$\begin{aligned}\int_0^1 d\nu(e^{-L/\nu}-1)F(\nu)&=\int_0^1 d\nu[F(\nu)-F(0)](e^{-L/\nu}-1)\\&\quad+F(0)\int_0^1 d\nu(e^{-L/\nu}-1)\\&=\sum_{n=1}^{\infty}\frac{(-1)^n L^n}{n!}\int_0^1 d\nu\nu^{-n}[F(\nu)-F(0)]\\&\quad+F(0)\int_0^1 d\nu(e^{-L/\nu}-1).\end{aligned}$$

Since

$$\int_0^1 d\nu e^{-L/\nu}=\sum_{\substack{n=0\\ n\neq 1}}^{\infty}\frac{(-L)^n}{n!\,(1-n)}+L(\log L-1+\gamma),$$

where $\gamma=0.577216$ is the Euler's constant, we get

$$\begin{aligned}&\int_0^1 d\nu(e^{-L/\nu}-1)F(\nu)\\&\quad=\sum_{n=1}^{\infty}\frac{(-L)^n}{n!}\int_0^1 d\nu\nu^{-n}(F(\nu)-F(0))\\&\qquad+F(0)\left(\sum_{n=2}^{\infty}\frac{(-L)^n}{n!\,(1-n)}+L(\log L-1+\gamma)\right).\end{aligned} \tag{101}$$

Retaining terms only up to quadratic in L, we see that Eq. (99), by means of Eq. (101), becomes

$$\begin{aligned}&G_>-G_++G_<-G_-\\&=\sum_{m=-N}^{N}\frac{e^{-im\phi}}{\mu Y_{mm}(\mu,0)}\Big[L\log LE_{mm}(0,\mathbf{\Omega}')H_m(0,\mu)\\&\quad+L\Big((\gamma-1)E_{mm}(0,\mathbf{\Omega}')H_m(0,\mu)\\&\quad-\int_0^1 d\nu\nu^{-1}(E_{mm}(\nu,\mathbf{\Omega}')\cdot H_m(\nu,\mu)\\&\quad-E_{mm}(0,\mathbf{\Omega}')H_m(0,\mu))\Big)\\&\quad+\tfrac{1}{2}L^2\Big(\int_0^1 d\nu\nu^{-2}(E_{mm}(\nu,\mathbf{\Omega}')H_m(\nu,\mu)\\&\quad-E_{mm}(0,\mathbf{\Omega}')H_m(0,\mu))-E_{mm}(0,\mathbf{\Omega}')H_m(0,\mu)\Big)\Big].\end{aligned} \tag{102}$$

The angular density Ψ may now be calculated simply by inserting the expression (102) for the given combination of Green's function in Eq. (97).

ACKNOWLEDGMENTS

The author wishes to express his deep gratitude to Professor K. M. Case for his infinite patience and help while this work was being pursued. Equal thanks are also extended to G. R. Carignan (Director of the Space Physics Research Laboratory) for his encouragement. This work was supported in part by NASA Grant NGR-23-005-15.

[1] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).

[2] M. Kanal, thesis, University of Michigan, Ann Arbor, Mich., 1969.

[3] K. M. Case, R. Zelazny, and M. Kanal, J. Math. Phys. **11**, 223 (1970).

[4] K. M. Case, "On the Boundary Value Problems of Linear Transport Theory," Department of Physics, University of Michigan, Ann Arbor, Mich., Report, 1967.

[5] The two integral equations (4) for $\Psi(\mathbf{r}_s,\mathbf{\Omega})$ are actually identical by virtue of the jump condition.

$$\hat{n}_i(r_s')\cdot\mathbf{\Omega}'[G_+(\mathbf{r}_s,\mathbf{\Omega};r_s',\mathbf{\Omega}')-G_-(\mathbf{r}_s,\mathbf{\Omega};\mathbf{r}_s',\mathbf{\Omega}')]=\delta^{(3)}(\mathbf{r}_s-\mathbf{r}_s')\delta(\mathbf{\Omega}\cdot\mathbf{\Omega}'),$$

where $G_\pm$ is the boundary value of G as $\mathbf{r}$ approaches the surface from (within/without) V.

[6] We shall also consider the Case spectrum in the complex z plane with $z=i/k$. In this plane, all the appropriate singular parts of the Green's function have branch cuts on the real axis for $-1\leq\nu\leq 1$, while the zeros of Λ_m are also transformed accordingly.

[7] For real zeros of Λ_m one needs to modify the procedure as given in Refs. 2 and 3.

[8] We assume that $\Lambda_m^+(k)\Lambda_m^-(k)\neq 0$ for any value of m; i.e., no zeros of $\Lambda_m(k)$ are imbedded in its branch line $k=\pm i\infty$ to $\pm i$.

[9] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1946).

[10] J. R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).

[11] In arriving at Eq. (97), we assumed that $L/\mu\ll 1$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

General Expansion of the Determinant of the Maxwell Dyadic

HARRY GELMAN
The MITRE Corporation, Bedford, Massachusetts 01730

(Received 23 March 1970)

Electromagnetic propagation, influenced by arbitrary tensor constitutive functions in an unbounded medium, is considered. The general expansion of the determinantal eigenvalue equation for the dispersion relations is obtained, exhibiting, for the first time, the functional dependence of the eigenvalue equation on the constitutive tensors.

I. INTRODUCTION

The dispersion relations for the propagation of characteristic electromagnetic waves, governed by tensor constitutive functions in an unbounded medium, come from the determinantal eigenvalue equation

$$\det \mathbf{M} = 0, \tag{1}$$

where the Maxwell dyadic $\mathbf{M}$ is given by[1]

$$\mathbf{M} \equiv n^2(\hat{\mathbf{k}}\hat{\mathbf{k}} - \mathbf{I}) + \boldsymbol{\Gamma}. \tag{2}$$

In Eq. (2), $\hat{\mathbf{k}}$ is a unit vector in the direction of propagation, $\mathbf{I}$ is the unit dyadic, n is the index of refraction defined as $n \equiv ck/\omega$, where k is the wavenumber and ω the angular frequency, and $\boldsymbol{\Gamma}$ is a dyadic defined from the permittivity dyadic $\boldsymbol{\epsilon}$ and the conductivity dyadic $\boldsymbol{\sigma}$ by the equation[2]

$$\boldsymbol{\Gamma} \equiv \epsilon_0^{-1}[\boldsymbol{\epsilon} + (i/\omega)\boldsymbol{\sigma}]. \tag{3}$$

Equation (1) is, of course, the condition that the homogeneous wave equation $\mathbf{M} \cdot \mathbf{E} = 0$ have nontrivial solutions (the characteristic modes).

When the $\boldsymbol{\Gamma}$ for a particular situation has a sufficiently simple structure, it is common practice to arrive at the eigenvalue equation by directly expanding $\det \mathbf{M}$ in a suitable coordinate system. There are situations, however, in which the structure of $\boldsymbol{\Gamma}$ is complicated to the point that one would prefer an alternative to directly expanding $\det \mathbf{M}$ in order to find the eigenvalue equation. Such situations occur, for example, in the consideration of propagation in certain anisotropic plasmas.

The purpose of this paper is to derive the general expansion of $\det \mathbf{M}$ for arbitrary $\boldsymbol{\Gamma}$, an expansion which has not been obtained before. In situations with a complicated $\boldsymbol{\Gamma}$, this expansion leads to the eigenvalue equation with less algebraic manipulation than would be required if $\det \mathbf{M}$ were evaluated directly. In addition, the expansion shows, for the first time, the structure of the general eigenvalue equation, exhibiting those functionals of $\boldsymbol{\Gamma}$ which actually enter into the calculation of dispersion relations. The expansion is presented in Sec. II.

The considerations of this paper exclude material media with either a permanent or induced magnetization. The constitutive equations considered are the linear relations

$$\mathbf{J}(\mathbf{k}, \omega) = \boldsymbol{\sigma}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{k}, \omega), \tag{4a}$$

$$\mathbf{D}(\mathbf{k}, \omega) = \boldsymbol{\epsilon}(\mathbf{k}, \omega) \cdot \mathbf{E}(\mathbf{k}, \omega), \tag{4b}$$

in wave-vector ($\mathbf{k}$)–angular-frequency (ω) space, where $\mathbf{J}$, $\mathbf{D}$, and $\mathbf{E}$ are the current density, displacement, and electric field strength, respectively. Equations (4) are, of course, not the most general linear constitutive relations, but they are the ones most often encountered in practice.

II. EXPANSION OF THE DETERMINANT OF M

We evaluate $\det \mathbf{M}$ in a coordinate-independent manner to emphasize the generality of the result. The definition of $\det \mathbf{M}$ is[3]

$$\det \mathbf{M} \equiv (1/3!)\epsilon_{ijp}\epsilon_{qrs}M_{iq}M_{jr}M_{ps}, \tag{5}$$

where the ϵ's in Eq. (5) are the completely antisymmetric Levi-Civita symbols. We may write $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{T} + n^2\hat{\mathbf{k}}\hat{\mathbf{k}}, \tag{6}$$

where the dyadic $\mathbf{T}$ is defined as

$$\mathbf{T} \equiv \boldsymbol{\Gamma} - n^2\mathbf{I}. \tag{7}$$

When Eq. (6) is inserted into Eq. (5) and the result multiplied out, the only terms that survive are the term containing no components of $\hat{\mathbf{k}}$, which by Eq. (5) is $\det \mathbf{T}$, and the three terms containing a product of two components of $\hat{\mathbf{k}}$. These three terms are all equal since they differ among themselves by cyclic permutations of their indices. The remaining terms all contain components of $\hat{\mathbf{k}} \times \hat{\mathbf{k}}$ and, hence, vanish. We have, therefore,

$$\det \mathbf{M} = \det \mathbf{T} + \tfrac{1}{2}n^2\epsilon_{ijp}\epsilon_{qrs}T_{iq}T_{jr}\hat{k}_p\hat{k}_s. \tag{8}$$

The second term in Eq. (8) can be evaluated by expressing the product of Levi–Civita symbols as a sum of products of Kronecker deltas. When this is done, we arrive at

$$\det \mathbf{M} = \det \mathbf{T} + \tfrac{1}{2}n^2[(\mathrm{Tr}\,\mathbf{T})^2 - \mathrm{Tr}\,(\mathbf{T}^2) + 2\hat{\mathbf{k}}\cdot\mathbf{T}^2\cdot\hat{\mathbf{k}} - 2(\mathrm{Tr}\,\mathbf{T})\hat{\mathbf{k}}\cdot\mathbf{T}\cdot\hat{\mathbf{k}}]. \quad (9)$$

The various functionals of $\mathbf{T}$ called for in Eq. (9) are readily evaluated in terms of $\boldsymbol{\Gamma}$, from Eq. (7). The most complicated of these functionals to evaluate is det $\mathbf{T}$. The evaluation can be carried out either by using the definition (5) or by directly expanding det $\mathbf{T}$ in some coordinate system, since $\mathbf{T}$ has the same matrix form in all coordinate systems. The result for det $\mathbf{T}$ is

$$\det \mathbf{T} = \det \boldsymbol{\Gamma} - \tfrac{1}{2}n^2[(\mathrm{Tr}\,\boldsymbol{\Gamma})^2 - \mathrm{Tr}\,(\boldsymbol{\Gamma}^2)] + n^4\,\mathrm{Tr}\,\boldsymbol{\Gamma} - n^6. \quad (10)$$

The other functionals required by Eq. (9) are

$$\mathrm{Tr}\,\mathbf{T} = \mathrm{Tr}\,\boldsymbol{\Gamma} - 3n^2, \quad (11a)$$

$$\mathrm{Tr}\,(\mathbf{T}^2) = \mathrm{Tr}\,(\boldsymbol{\Gamma}^2) - 2n^2\,\mathrm{Tr}\,\boldsymbol{\Gamma} + 3n^4, \quad (11b)$$

$$\hat{\mathbf{k}}\cdot\mathbf{T}\cdot\hat{\mathbf{k}} = \hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}} - n^2, \quad (11c)$$

$$\hat{\mathbf{k}}\cdot\mathbf{T}^2\cdot\hat{\mathbf{k}} = \hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}^2\cdot\hat{\mathbf{k}} - 2n^2\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}} + n^4. \quad (11d)$$

When Eqs. (10) and (11) are inserted in Eq. (9), we have

$$\det \mathbf{M} = (\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}})n^4 + [\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}^2\cdot\hat{\mathbf{k}} - (\mathrm{Tr}\,\boldsymbol{\Gamma}) \times \hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}}]n^2 + \det\boldsymbol{\Gamma}. \quad (12)$$

Equation (12) is the general expansion of the determinant of the Maxwell operator, for arbitrary $\boldsymbol{\Gamma}$. It is seen to depend on the four functionals Tr $\boldsymbol{\Gamma}$, det $\boldsymbol{\Gamma}$, $\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}}$, and $\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}^2\cdot\hat{\mathbf{k}}$—the latter of which, incidentally, does not require one to square $\boldsymbol{\Gamma}$ since it can be evaluated as

$$\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}^2\cdot\hat{\mathbf{k}} = (\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma})\cdot(\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}}). \quad (13)$$

The eigenvalue equation for n, which ensues by setting the right-hand side of Eq. (12) equal to zero, is not biquadratic in n, as it superficially appears to be, since, in general, $\boldsymbol{\Gamma}$ depends on $\mathbf{k}$.

We have derived Eq. (12) without reference to any particular coordinate system. Considering the simplicity of the result, we see that the derivation is perhaps longer than it should be. However, the method has the advantage that it emphasizes the generality of the result and leads to it unambiguously. We may verify the expansion in a way which also serves as a simpler derivation. We choose a coordinate system in which $\hat{\mathbf{k}}$ lies along one of the coordinate axes, say, the x axis. Then det $\mathbf{M}$ may be written, in explicit determinant form, as

$$\det \mathbf{M} = \begin{vmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & \Gamma_{22} - n^2 & \Gamma_{23} \\ \Gamma_{31} & \Gamma_{32} & \Gamma_{33} - n^2 \end{vmatrix}. \quad (14)$$

The polynomial form of this determinant is

$$\det \mathbf{M} = An^4 + Bn^2 + C, \quad (15)$$

where A, B, and C are given by

$$A = \Gamma_{11}, \quad (16a)$$

$$B = \Gamma_{13}\Gamma_{31} - \Gamma_{11}\Gamma_{33} + \Gamma_{12}\Gamma_{21} - \Gamma_{11}\Gamma_{22}, \quad (16b)$$

$$C = \det\boldsymbol{\Gamma}. \quad (16c)$$

In the coordinate system we are using the element Γ_{11} may be written $\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}}$. The structure of the right-hand side of Eq. (16b) suggests that we add and subtract the term Γ_{11}^2. The form of B is then

$$B = \Gamma_{1j}\Gamma_{j1} - \Gamma_{11}(\mathrm{Tr}\,\boldsymbol{\Gamma}) = (\boldsymbol{\Gamma}^2)_{11} - \Gamma_{11}(\mathrm{Tr}\,\boldsymbol{\Gamma}). \quad (17)$$

For $\hat{\mathbf{k}}$ along the x axis, B may be written

$$B = \hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}^2\cdot\hat{\mathbf{k}} - (\mathrm{Tr}\,\boldsymbol{\Gamma})\hat{\mathbf{k}}\cdot\boldsymbol{\Gamma}\cdot\hat{\mathbf{k}},$$

and Eq. (15) is identical to Eq. (12) in this coordinate system. Since det $\mathbf{M}$ and the form of the coefficients A, B, and C are invariant under orthogonal transformation, the identity holds in all coordinate systems.

ACKNOWLEDGMENT

I would like to thank Dr. P. Schram for suggesting the alternative derivation of the main result presented in Sec. II.

[1] D. B. Melrose, Astrophys. Space Sci. **2**, 171 (1968).

[2] Equation (3) is written in rationalized MKS units. The rest of the paper is written without reference to particular systems of units.

[3] G. Goertzel and N. Tralli, *Some Mathematical Methods of Physics* (McGraw-Hill, New York, 1960), Appendix 1A.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Polynomial Algebras

I. V. V. RAGHAVACHARYULU AND NALINI B. MENON
Matscience, The Institute of Mathematical Sciences, Madras-20, India

(Received 10 September 1969; Revised Manuscript Received 3 February 1970)

The present work is concerned with what are called polynomial algebras as an extension of the work of Ramakrishnan and his colleagues on the algebras of matrices satisfying conditions like $L^m = I$ and $L^m = L^k$. Assuming L_m to be an m-dimensional linear space, we generate a class of associative algebras called polynomial algebras by requiring that *every element* L of L_m satisfy a polynomial equation $L^n + P_1L^{n-1} + \cdots + P_n = 0$. We show that some very important algebras which physicists have found useful can be obtained by various restrictions on the polynomial. A few general properties of these algebras are established.

1. INTRODUCTION

In a series of contributions, Ramakrishnan and his colleagues have initiated and studied the matrix algebras obtained by imposing restrictive polynomial conditions[1] like

$$L^m = I \quad \text{and} \quad L^m = L^k. \tag{1.1}$$

The work presented here extends these studies by imposing more general polynomial conditions, leading to what we shall call polynomial algebras.

Let L_m be an m-dimensional linear space over a field F. We generate a class of associative algebras called polynomial algebras $A[\alpha^1, \alpha^2, \cdots, \alpha^m]$ with $\{\alpha^i \mid i = 1, 2, \cdots, m\}$ as generating elements by requiring that *every element*[1]

$$L(x) = x_1\alpha^1 + x_2\alpha^2 + \cdots + x_m\alpha^m \tag{1.2}$$

belonging to L_m satisfy a polynomial equation

$$P[x; L] \equiv L^n + P_1L^{n-1} + \cdots + P_n = 0, \tag{1.3}$$

where n is independent of m. We show that some very important algebras in physics such as Clifford and Grassman algebras (ordinary and generalized) and spin and parafield algebras are indeed polynomial algebras.

In the second and third sections, we define polynomial algebras and study operations on them. In the fourth section, we recover a number of different algebras of physical importance as polynomial algebras. In the fifth and last section, we study the transformations of the algebras and obtain their automorphism groups.

2. CHARACTERISTIC EQUATIONS OF POLYNOMIAL ALGEBRAS

Let $A[\alpha^1, \alpha^2, \cdots, \alpha^m]$ be an F-algebra defined by a set of generating elements $\{\alpha^i \mid i = 1, \cdots, m\}$ over a field F. We write $A[\alpha^1, \cdots, \hat{\alpha}^i, \cdots, \alpha^m]$ for the algebra obtained by restricting the generating relations of A to those containing $\alpha^1, \cdots, \alpha^{i-1}, \alpha^{i+1}, \cdots, \alpha^m$ only. Further, we denote by $A[\alpha^1, \cdots, \alpha^i, \alpha^i, \alpha^{i+2}, \cdots, \alpha^m]$ the algebra, if it exists, obtained by *substituting* α^i for α^{i+1} in the elements of A. Obviously, the restriction and substitution operations can be extended to a finite number of generators of A. We write $A[\alpha^1, \cdots, \alpha^m] \cdot\equiv\cdot B[\beta^1, \cdots, \beta^n]$ for two isomorphic algebras A and B under $\alpha^i \leftrightarrow \beta^i$ and $m = n$.

Let I denote the set of positive integers. F-algebras $A[\alpha]: \{A[\alpha^1, \cdots, a^m], m \in I\}$ are called simplicial algebras (S A) if they satisfy the following conditions:

(a) For every $A[\alpha^1, \cdots, \alpha^m] \in A[\alpha]$ algebras

$$A[\alpha^1, \cdots, \hat{\alpha}^{i_1}, \cdots, \hat{\alpha}^{i_k}, \cdots, \alpha^m]$$

and

$$A[\alpha^1, \cdots, \alpha^{j_1}, \alpha^{j_1}, \cdots, \alpha^{j_l}, \alpha^{j_l}, \cdots, \alpha^m]$$

exist for all $0 < k, l \leq m$;

(b) $\delta_i: A[\alpha^1, \cdots, \alpha^m]$

$$\cdot\equiv\cdot A[\alpha^1, \cdots, \alpha^{i-1}, \hat{\alpha}^i, \alpha^{i+1}, \cdots, \alpha^{m+1}];$$

(c) $\sigma_i: A[\alpha^1, \cdots, \alpha^m]$

$$\cdot\equiv\cdot A[\alpha^1, \cdots, \alpha^i, \alpha^i, \alpha^{i+1}, \cdots, \alpha^m],$$

for all i.

δ and σ are called the face (restriction) and degeneracy (substitution) operations, respectively.[2] Simplicial algebras can be directly defined without referring to generating elements. Then under δ and σ the generating elements should be stable.

Note that the simplicial conditions can be extended to the index set $I \otimes I \cdots \otimes I$ when considering algebras $A[\{\alpha\}, \{\beta\}, \cdots, \{\gamma\}]$, where $[\{\alpha\}, \{\beta\}, \cdots, \{\gamma\}]$ are sets of generating elements. This is accomplished by extending the above definitions under the mapping $i \leftrightarrow (i, \cdots, i)$ and considering the set $\{\alpha^i, \beta^i, \cdots, \gamma^i\}$ of generating elements simultaneously under δ_i and σ_i. Hereafter $A[\alpha]$ denotes the set of algebras $A[\alpha]$ and those obtained from $A[\alpha]$ by substitution and restriction operations. Now let an algebra $\Delta_n \in A[\alpha]$ have n generators and let $\mu: \Delta_n \cdot\equiv\cdot \Delta_m$ be an isomorphism. Suppose that $\alpha^{i_1}, \cdots, \alpha^{i_s}$

written in the reverse order are the generators of Δ_m not contained in $\mu[\Delta_n]$ and $\alpha^{j_1}, \cdots, \alpha^{j_t}$ are the generators of Δ_n such that $\mu[\alpha^j] = \mu[\alpha^{j+1}]$; then

$$\mu = \delta_{i_1} \circ \delta_{i_2} \circ \cdots \circ \delta_{i_s} \circ \sigma_{j_1} \cdots \sigma_{j_t}, \qquad (2.1)$$

where $1 \leq i_s < \cdots < i_1 \leq m$, $1 \leq j_1 < \cdots < j_t \leq n$ and $n + s = m + t$. Further, the factorization is unique. Now, for simplicial algebras $A[\alpha]$, we have the following:

Theorem: The set of necessary and sufficient conditions for

$$A[\alpha^{p_1}, \cdots, \alpha^{p_n}] \cdot\equiv\cdot A[\alpha^{q_1}, \cdots, \alpha^{q_n}], \qquad (2.2)$$

where $p_i, q_i \in I$ and $1 \leq p_1 < p_2 \cdots < p_m$ and $1 \leq q_1 < q_2 \cdots < q_m$, is that the $A[\alpha]$ satisfy the simplicial conditions.

Proof: Obviously, that (2.2) implies the simplicial conditions and the simplicial conditions imply (2.2) is seen from (2.1).

Suppose $A[\alpha^1, \cdots, \alpha^m]$ is an algebra not necessarily associative with a finite basis over an infinite field F. Let[3]

$$L(x) = x_1\alpha^1 + x_2\alpha^2 + \cdots + x_m\alpha^m \qquad (2.3)$$

be an element of the linear space L_m over F with $\{\alpha^i \mid i = 1, \cdots, m\}$ as basis elements. $L(x)$ satisfies a minimal equation

$$P[x; L] \equiv L^n + P_1L^{n-1} + \cdots + P_n = 0, \qquad (2.4)$$

where P_r, $r = 1, \cdots, m$, is a homogeneous polynomial of rth degree in $x_1, x_2, \cdots, x_m$ and $x = (x_1, \cdots, x_m)$. *It is important to note that (2.4) holds for every general element* $L(x) \in L_m$.

Let $A[\alpha]$ be simplicial algebras *not necessarily associative* with finite basis over an infinite field F. Since

$$A[\alpha^1, \cdots, \alpha^n] \cdot\equiv\cdot A[\alpha^1, \cdots, \hat{\alpha}^{i_1}, \cdots, \hat{\alpha}^{i_{m-n}}, \cdots, \alpha^m] \quad \text{for} \quad m > n$$

when $m - n$ elements are deleted, it easily follows by considering the cases with $n = 1, 2, \cdots, m - 1$ that the coefficients P_r of the minimal equation $P[x; L] = 0$ of $L(x) \in L_m$ form the rth-degree symmetric homogeneous polynomials (SHP's) in x. Hence, the P_r, $r = 1, \cdots, n$, are given by

$$\begin{aligned} P_1 &= a_1^1 \sum x_i, \\ P_2 &= a_{[1^2]}^2 \sum_{i_1 < i_2} x_{i_1}x_{i_2} + a_{[2]}^2 \sum x_i^2, \\ &\cdots \\ P_r &= \sum_{[a_1 \cdots a_r]} a_{[a_1, a_2, \cdots, a_r]}^r \sum_{i_1 < i_2 \cdots < i_r} x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_r}^{a_r}, \end{aligned} \qquad (2.5)$$

where $i_j \in [1, 2, \cdots, m]$ and $[a_1, a_2, \cdots, a_r]$ is a composition of r, i.e., integers $a_i > 0$ are such that $\sum a_i = r$.

Example 1: Consider the algebra $A[e^1, \cdots, e^m]$ with generating relations $e^ie^j = \delta_{ij}e^i$ and $1_{(1,\cdots,m)} = e^1 + e^2 + \cdots + e^m$. The characteristic equation of $L(x) \equiv \sum x_ie^i \in L_m$ is $P[x; L] \equiv \Pi(L - x_i) = 0$.

Obviously, by considering index sets $[i_1, \cdots, i_m] \subset I$ in the above example with $e^{i_1} + e^{i_2} + \cdots + e^{i_m} = 1_{(i_1,\cdots,i_m)}$, we obtain simplicial algebras $A[e]$.

Example 2: Let $A[\alpha^1, \cdots, \alpha^m]$ be either a generalized Clifford or Grassman algebra (see Sec. 4). It is important to note that the degree of the minimal equation of $L(x) = \sum x_i\alpha^i \in L_m$ is independent of m. This is not true in the case of simplicial algebras in Example 1 which do not intuitively correspond to polynomial algebras.

Polynomial condition: Simplicial algebras

$$A[\alpha]: \{A[\alpha^1, \cdots, \alpha^m]; m \in I\}$$

are said to satisfy the polynomial condition if all the polynomials $P[A]: \{P_m[x; L]; L(x) \in L_m\}$ associated with $A[\alpha]$ have the same degree k (say). For $m > k$ the minimal equation $P_{m+1}[x_1, \cdots, x_{m+1}; L] = 0$ of $x_1\alpha^1 + \cdots + x_{m+1}\alpha^{m+1} \in L_{m+1}$ is obtained from that of $x_1\alpha^1 + \cdots + x_m\alpha^m \in L_m$ by adding terms containing x_{m+1} to the SHP's that are the coefficients of $P_m[x; L]$, without altering their order and weight. In what follows, all the SHP's so obtained from an SHP $\mathfrak{L}$ are represented by $\mathfrak{L}$ itself without specifying its variables.

Isotropy condition: Simplicial algebras $A[\alpha]$ are said to satisfy the isotropy condition if there exist functions $\{F_m(x_1, \cdots, x_m); m \in I\}$ such that $\lambda F_m^{-1}L(x)$ satisfy the minimal polynomial equation

$$P[\lambda, 0, \cdots, 0; L] = 0.$$

The isotropy condition is very restrictive and implies the polynomial condition. $F_m(x_1, \cdots, x_m)$ is a homogeneous function of degree one. Obviously, F^s should be an SHP for some $s \leq k$. Let r be the minimal integer with this property and $F^r = \mathfrak{L}$ an SHP. Then it follows that, for simplicial algebras satisfying the isotropy condition,

$$P[x; L] \equiv L^p \prod_{i=1}^{q} (L^r + b_i\mathfrak{L}) = 0, \qquad (2.6)$$

where $b_i \in K$, a suitable extension of the field F.

Simplicial algebras $A[\alpha]$ are said to satisfy the factorizability condition if there exists an SHP $\mathfrak{L}$ such that every $P[x; L]$ can be factorized as in (2.6) over a field $K \supset F$ with fixed integers p and q. Obviously, simplicial algebras satisfying the isotropy condition are factorizable. *Simplicial algebras which satisfy the factorizability condition are called polynomial algebras.*

Hence, from definition, with every set of polynomial algebras $A[\alpha]$ an SHP $\mathfrak{L}$ is uniquely associated, but for multiplication by a constant. Polynomial algebras $A^{\mathfrak{L}}[\alpha]$ associated with the same $\mathfrak{L}$ are called associated polynomial algebras with polynomials $P^{\mathfrak{L}}$. In particular, if $P^{\mathfrak{L}}$ is of degree r, which is also the degree of $\mathfrak{L}$, then $P^{\mathfrak{L}}$ can be taken as $P^{\mathfrak{L}}: L^r = \mathfrak{L}$.[1] Algebras with the polynomial condition $L^r = \mathfrak{L}$ are called basic polynomial algebras. For $r = 1$, $A^{\mathfrak{L}}[\alpha]$ reduces to F.

3. OPERATIONS ON POLYNOMIAL ALGEBRAS

If the polynomial algebras $A[\alpha]$ with moduli over an infinite field F are a direct sum $+$ of the polynomial algebras $A_i[\alpha_i]$, $i = 1, \cdots, t$, and if $P[x; L] = 0$ and $P_i[x; L] = 0$ are the polynomial equations of $A[\alpha]$ and $A_i[\alpha_i]$, $i = 1, \cdots, t$, respectively, then

$$P[x; L] = \prod_{i=1}^{t} P_i[x; L]. \tag{3.1}$$

From this it follows that, if the $A[\alpha]$ are polynomial algebras with moduli, then $A_i[\alpha_i]$ for each i is a set of polynomial algebras and conversely. If $A[\alpha]$ is associated with the SHP $\mathfrak{L}$, then each one of the PA's $A_i[\alpha_i]$ is associated with $\mathfrak{L}$ and conversely.

Consider two sets of simplicial algebras $A[\alpha]$ and $B[\beta]$. We define their sum $A[\alpha] \oplus B[\beta]$ to be the algebras

$$C[\alpha \oplus \beta]: \{C_m[\alpha^1 \otimes I + I \otimes \beta^1; \alpha^2 \otimes I + I \otimes \beta^2; \cdots; \alpha^m \otimes I + I \otimes \beta^m]; m \in I\} \tag{3.2}$$

and their product $A[\alpha] \otimes B[\beta]$ to be the algebras

$$D[\alpha \otimes \beta]: \{D_m[\alpha^1 \otimes \beta^1; \alpha^2 \otimes \beta^2; \cdots; \alpha^m \otimes \beta^m]\}. \tag{3.3}$$

The extension of simplicial operations to $A \oplus B$ and $A \otimes B$ is direct. The minimal polynomials associated with $C[\alpha \oplus \beta]$ and $D[\alpha \otimes \beta]$ are symbolically given by $P_A[x; L] \oplus P_B[x; L]$ and $P_A[x; L] \otimes P_B[x; L]$ which have as their roots the sums (products) of the roots of the polynomials $P_A[x; L]$ and $P_B[x; L]$, with minimum multiplicity such that they are the minimal polynomials associated with the algebras $A \oplus B$ and $A \otimes B$, respectively. Hence, if A and B are polynomial algebras associated with the SHP's $\mathfrak{L}$ and $\mathfrak{M}$, then $A \oplus B$ and $A \otimes B$ are associated with the SHP's $\mathfrak{L} + \mathfrak{M}$ and $\mathfrak{L}\mathfrak{M}$, respectively.

4. POLYNOMIAL ALGEBRAS IN PHYSICS

In this section, we consider a few polynomial algebras which are of interest to physicists. First we consider the polynomial algebras

$$A[\alpha]: \{A[\alpha^1, \cdots, \alpha^m]; m \in I\}$$

satisfying a second-degree polynomial equation. We assume that F is the real field and $L(x) = x_1\alpha^1 + \cdots + x_m\alpha^m \in L_m$ satisfies, without loss of generality, the minimal equation

$$L^2 + P_2 = 0, \tag{4.1}$$

where $P_2 = a_{21} \sum' x_i x_j + a_{22} \sum x_i^2$. Substituting for L and equating the coefficients of $x_i x_j$ to zero, we obtain the generating relations

$$\begin{aligned} \alpha^i\alpha^j + \alpha^j\alpha^i &= -2a_{22}I, \quad \text{if} \quad i = j, \\ &= -a_{21}I, \quad \text{if} \quad i \neq j, \end{aligned} \tag{4.2}$$

of $A[\alpha]$. Obviously, the $A[\alpha]$ satisfy the simplicial conditions. In fact, if we waive these conditions, the above set of generating relations can be generalized to

$$\alpha^i\alpha^j + \alpha^j\alpha^i = -a_{ij}I, \quad i, j = 1, 2, \cdots, m, \tag{4.2'}$$

where $[a_{ij}]$ is a symmetric matrix. The representation theory of this algebra was considered by Landsberg.[4]

Now let us recover some familiar algebras.

Grassman algebras: Algebras isomorphic to Grassman algebras[5] are obtained by taking $a_{21} = a_{22} = 0$ in (4.2) when the generating relations become $\alpha^i\alpha^j + \alpha^j\alpha^i = 0$.

Clifford algebras: We choose $a_{22} = -1$ and $a_{21} = 0$ in (4.2); then we obtain Clifford algebras, with the generating relations

$$\alpha^i\alpha^j + \alpha^j\alpha^i = 2\delta^{ij}. \tag{4.3}$$

Obviously, this is an isotropic polynomial algebra with

$$\mathfrak{L} = x^2 = x_1^2 + x_2^2 + \cdots. \tag{4.4}$$

Algebras of annihilation and creation operators: Now let us consider a mixed algebra obtained from (4.2'). For that, let $m = 2\nu$ and

$$\begin{aligned} a_{ij} &= 2, \quad \text{if} \quad |i - j| = \nu, \\ &= 0, \quad \text{otherwise.} \end{aligned} \tag{4.5}$$

This is the νth-order algebra of annihilation and creation operators. To put this in the familiar form, we introduce

$$a^i = \alpha^i \quad \text{and} \quad a^{i\dagger} = \alpha^{\nu+i}, \quad i = 1, 2, \cdots, \nu.$$

Then the commutation relations (4.2′) under (4.5) become

$$a^i a^j + a^j a^i = a^{i\dagger} a^{j\dagger} + a^{j\dagger} a^{i\dagger} = 0,$$
$$a^i a^{j\dagger} + a^{j\dagger} a^i = 2\delta^{ij} I. \tag{4.6}$$

Now consider the index set $I \otimes I$ and the algebras $A[a, a^\dagger]$ satisfying the polynomial condition

$$P[z, \bar{z}; L] \equiv L^2 = 2\{z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots + z_\nu\bar{z}_\nu\}, \tag{4.7}$$

where

$$L = \sum (z_i a^i + \bar{z}_i a^{i\dagger}). \tag{4.8}$$

By equating coefficients of products of z_i and $\bar{z}_i$, we recover the commutation relations (4.6). Algebras of annihilation and creation operators are isotropic algebras with

$$\mathfrak{L} = |z|^2 = (z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots). \tag{4.9}$$

As is well known, there is an intimate connection between Clifford algebras of order 2ν and the algebras of annihilation and creation operators of order ν. To establish it, we put

$$\alpha^{2i-1} = (a^i + a^{i\dagger})/2^{\frac{1}{2}}, \quad \alpha^{2i} = i(a^i - a^{i\dagger})/2^{\frac{1}{2}}. \tag{4.10}$$

Then $L = \sum x_i\alpha^i$ becomes $L = \sum (z_i a^i + \bar{z}_i a^{i\dagger})$ and $x^2 = \sum x_i^2$ becomes $|z|^2 = \sum z_i\bar{z}_i$, where

$$z_i = 2^{-\frac{1}{2}}(x_{2i-1} + ix_{2i})$$

and $\bar{z}_i$ is the complex conjugate of z_i when we obtain Eq. (4.7) from (4.1) with $a_{22} = -1$ and $a_{21} = 0$.

Generalized Clifford and Grassman algebras: Now we consider a few polynomial algebras $A[\alpha]$ satisfying polynomial equations of degree greater than two. The generating relations of these algebras cannot be obtained from the polynomial condition alone. The conditions on α's, obtained by equating the coefficients of products x_i to zero, generate algebras of infinite order.

Now consider the algebras[6] $A[\alpha]:\{A[\alpha^1, \cdots, \alpha^m]; m \in I\}$ satisfying the generating relations $\alpha^i\alpha^j = \omega\alpha^j\alpha^i$, $i < j$, and $(\alpha^i)^p = \underline{a}$ over an infinite field F containing $\underline{a}$ and a pth primitive root ω of unity. These algebras obviously satisfy the simplicial conditions. If $\underline{a} = 1(0)$, $A[\alpha]$ is called the simplicial set of generalized Clifford (Grassman) algebras $A_{GC}[p](A_{GG}[p])$ and, when $n = 2$, the ordinary simplicial set of Clifford (Grassman) algebras are recovered. Note that, for $p > 2$, these algebras are order dependent on the generating elements. From the generating relations it can be verified directly that

$$\sum_{\{j\}} [\alpha^{j(1)}]^{p_1}[\alpha^{j(2)}]^{p_2} \cdots [\alpha^{j(m)}]^{p_m} = 0 \tag{4.11}$$

if at least two $p_i \neq 0$, where $[p_1, p_2, \cdots, p_m]$ is a composition of p and $\{j\} \cdot\equiv\cdot S_m$ the set of all permutations of $[1, 2, \cdots, m]$. Hence, from (4.11), we obtain

$$L^p \equiv (x_1\alpha^1 + x_2\alpha^2 + \cdots + x_m\alpha^m)^p$$
$$= a(x_1^p + x_2^p + \cdots + x_m^p). \tag{4.12}$$

Since the roots of the Eq. (4.12) in L are distinct, it is the minimal equation of $L(x) \in L_m$. Hence the $A[\alpha]$ define polynomial algebras since the degree of the minimal equation of $A[\alpha^1, \cdots, \alpha^m]$ is independent of m. In fact, the algebras $A[\alpha]$ are the basic polynomial algebras associated with $\mathfrak{L} = x_1^p + x_2^p + \cdots$.

Given two simplicial sets of generalized Clifford algebras $A_{GC}[p]$ and $A_{GC}[q]$, where p and q are relatively prime, then $A_{GC}[p] \otimes A_{GC}[q] = A_{GC}[pq]$. Hence, if $p = p_1^{m_1}p_2^{m_2} \cdots p_r^{m_r}$ is a prime power decomposition of p, then

$$A_{GC}[p] = \prod_{\otimes}^{r} A_{GC}[p_i^{m_i}]. \tag{4.13}$$

Generalized Spin and Parafield Algebras: Given the simplicial (polynomial) algebras $A[\alpha]$, let us consider the algebras $\sum_{\oplus}^{s} A[\alpha] \equiv A[\alpha] \oplus \cdots \oplus A[\alpha]$, s-times introduced by the usual inductive process. These are also simplicial (polynomial) algebras. The roots of the minimal polynomial equations of $\sum^s A[\alpha]$ are sums of the roots of the minimal polynomial equations of $\sum^{s-1} A[\alpha]$ and $A[\alpha]$. Now, to obtain the commutation relations of $\sum^s A$ that are independent of s, we consider $\mathbf{I}^{ij} = \boldsymbol{\alpha}^i\boldsymbol{\alpha}^j - \boldsymbol{\alpha}^j\boldsymbol{\alpha}^i$, where

$$\boldsymbol{\alpha}^i = \alpha^i \otimes I \otimes \cdots \otimes I + I \otimes \alpha^i \otimes \cdots \otimes I$$
$$+ \cdots + I \otimes I \otimes \cdots \otimes \alpha^i$$

are the sth-order helicity operators associated with α^i, and find the commutation relations $[\boldsymbol{\alpha}^k, \mathbf{I}^{ij}]$. These commutation relations are easily seen to be independent of s. Hence, to obtain their explicit form, we have to only consider the case when $s = 1$ and evaluate $[\alpha^k, \mathbf{I}^{ij}]$ from the defining relations of $A[\alpha]$.

When we take for $A[\alpha]$ the Clifford algebras A_C (annihilation and creation operator algebras A_{ac}), we obtain the spin (parafield) algebras. The commutation relations of spin algebras are easily obtained from Eqs. (4.3):

$$[\boldsymbol{\alpha}^k, \mathbf{I}^{ij}] = \delta^{ik}\boldsymbol{\alpha}^i - \delta^{ik}\boldsymbol{\alpha}^j.$$

Since the index set for A_{ac} is $I \otimes I$, we have to consider $\mathbf{M}^{ij} = \mathbf{a}^i\mathbf{a}^j - \mathbf{a}^j\mathbf{a}^i$ and $\mathbf{N}^{ij} = \mathbf{a}^{i\dagger}\mathbf{a}^j - \mathbf{a}^j\mathbf{a}^{i\dagger}$ and

their complex conjugates. The following commutation relations are easily obtained from (4.6):

$$[\mathbf{a}^k, \mathbf{N}^{ij}] = \delta^{ki}\mathbf{a}^j \quad \text{and} \quad [\mathbf{a}^k, \mathbf{M}^{ij}] = 0.$$

The minimal polynomial equations satisfied by spin $\sum^s A_C$ and parafield $\sum^s A_{ac}$ algebras are easily obtained as

$$(L^2 - \tfrac{1}{4}s^2\mathfrak{L})(L^2 - \tfrac{1}{4}(s-1)^2\mathfrak{L})\cdots(L^2 - \tfrac{1}{4}\mathfrak{L}) = 0, \quad \text{if } s \text{ is half-integral},$$

$$(L^2 - s^2\mathfrak{L})(L^2 - (s-1)^2\mathfrak{L})\cdots(L^2 - \mathfrak{L})L = 0, \quad \text{if } s \text{ is integral}, \tag{4.14}$$

where L and $\mathfrak{L}$ are given by Eqs. (4.1), (4.4) and (4.8), (4.9) for spin and parafield algebras, respectively. It is interesting to note that, by considering the coefficient of $z_i^{2\nu}$ from (4.14), we obtain the Ryan–Sudarshan relation[7] $(a^i)^{2\nu+1} = 0$, $(a^i)^p \neq 0$ for $p \leq 2\nu$.

Note that the $\sum^s A_{GC}[p]$ are a generalization of spin algebras which reduce to ordinary spin algebras when $n = 2$. These generalized spin algebras and their parafield algebras will be considered elsewhere.[8]

5. AUTOMORPHISMS OF $A[\alpha]$

Let $A[\alpha]$ be a set of simplicial algebras satisfying the polynomial condition. In this section, we shall consider the groups $G^1[A]:\{G^1[\alpha^1, \cdots, \alpha^m]; m \in I\}$ of all nonsingular transformations

$$\tau: {}'\alpha^i = \sum \tau_j^i \alpha^j, \tag{5.1}$$

when $A['\alpha]$ give rise to the same type of generating relations as that of $A[\alpha]$ and the groups

$$G^2[A]:\{G^2[x_1, \cdots, x_m]; m \in I\},$$

of all nonsingular transformations

$$t: {}'x^i = \sum t_i^j x_j, \tag{5.2}$$

leaving invariant the coefficients $P_i[x_1, \cdots, x_m]$ of the minimal polynomial expression $P[x; L]$. Because every transformation of $G_m^1[A] \equiv G^1[\alpha^1, \cdots, \alpha^m]$ defines a transformation of $G_m^2[A] \equiv G^2[x_1, \cdots, x_m]$ in an obvious way, we have

$$G_m^1[A] \subset G_m^2[A] \subset \text{aut}\,\{L_m\}, \tag{5.3}$$

where aut $\{L_m\}$ is the set of all automorphisms of the vector space L_m.

It is instructive to find[9] $G_m^i[A_{GC}[p]]$, $i = 1, 2$, for the simplicial set of generalized Clifford algebras $A_{GC}[p]$. Obviously $G_m^1[p] = \prod_{\otimes}^m C_p$, where C_p is the cyclic group of order p. To find $G_m^2[p]$, the transformation t should leave invariant $\sum x_i^p$, i.e.,

$$\sum{}' x_i^p = \sum x_i^p. \tag{5.4}$$

Substituting (5.2) in (5.4) and comparing the coefficients of $x_j^p x_j^{p-2} x_k^2$ and $x_j^{p-2} x_k x_j$, $j \neq k \neq l$, we have

$$\sum_i (t_i^j)^p = 1, \tag{5.5}$$

$$\sum_i (t_i^j)^{p-2}(t_i^k)^2 = 0, \tag{5.6}$$

and

$$\sum_i (t_i^j)^{p-2} t_i^k t_i^l = a \quad \text{(say)}, \tag{5.7}$$

respectively. If we write (5.5) and (5.6) together, we have

$$\sum_i (t_i^j)^{p-2}(t_i^k)^2 = \delta^{jk}. \tag{5.8}$$

Hence, it follows that

$$\det([t_i^j]^{p-2}) \neq 0. \tag{5.9}$$

The algebras turn out to be quite distinct depending upon whether $a = 0$ or $a \neq 0$. From (5.7) it follows that

$$\begin{pmatrix} t_1^k t_1^l \\ t_2^k t_2^l \\ \cdot \\ \cdot \\ \cdot \\ t_m^k t_m^l \end{pmatrix} = a((t_i^j)^{p-2})^{-1} \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}. \tag{5.10}$$

If $a = 0$, it follows from (5.9) that there exists no index i such that $t_i^j = 0$ for $j = 1, \cdots, m$ and from (5.10) that there exists only one $j(i)$ for each i such that

$$t_i^{j(i)} \neq 0. \tag{5.11}$$

From (5.9), it is obvious that $j(i)$, $i = 1, \cdots, m$, is a permutation of $(1, 2, \cdots, m)$. From (5.5) and (5.11) we have

$$(t_i^{j(i)})^p = 1.$$

Hence $t_i^{j(i)} = \omega^{p_i}$ where ω is a primitive pth root of unity in F. Now, it is easily seen that $p^m m!$ linear transformations exist leaving $\sum x_i^p$ invariant when $p > 2$.

Theorem: The groups $G_m^i[p] \equiv G^i[A_{GC}[p]]$, $i = 1, 2$, of the simplicial set of generalized Clifford algebras $A_{GC}[p]$ are given by

$$G_m^1[2] \equiv G_m^2[2] \cdot\equiv\cdot O_m,$$

the mth-order orthogonal group over F,

$$G_m^1\,[p \neq 2] \cdot\equiv\cdot \prod_{\otimes}^m C_p \quad \text{is given by} \quad {}'\alpha^i = \omega^{p_i}\alpha^i,$$

where p_i is any integer mod p, and

$$G_m^2\,[p \neq 2] \cdot\equiv\cdot \prod_{\otimes}^m C_p \otimes S_m$$

$$\text{is given by} \quad {}'x_i = \omega^{p_i} x_j(i), \tag{5.12}$$

where $j(i)$, $i = 1, 2, \cdots, m$, is any permutation of $(1, 2, \cdots, m)$.

Let us now consider general polynomial algebras (PA's) $A[\alpha]$ and introduce the groups $H^i[A] \subset G^i[A]$, $i = 1, 2$, defined by the transformations $'\alpha^i = \underline{\tau}\alpha^i$ and $x_i = \underline{t}x_i$, where $\underline{\tau}, \underline{t} \in \mathbb{C}$, the field of complex numbers. Obviously, $H^i[A]$ are cyclic groups and are of the form $\prod_{\otimes}^{m} C_r$ and $\prod_{\otimes}^{m} C_s$ for suitable integers r and s where $r \mid s$. This means that the minimal equations satisfied by an L are of the form

$$P[x; L] \equiv L^p(L^{tr} + Q_1 L^{(t-1)r} + \cdots + Q_t) = 0$$

and the degree of each variable x_i in the SHP $Q_i[x_1, \cdots, x_m]$ is divisible by r. For example, if q is divisible by r and $r \geq 3$, then the terms $(x_i)^{q-2}(x_j)^2$ and $(x_i)^{q-2}x_jx_k$ will not occur in the coefficients $x_1, \cdots, x_m$ of the polynomial expression $P[x; L]$. Hence, for the PA with polynomial equations in which $a^s_{[s]} \neq 0$ in (2.5),

$$G^2_m[PA] \cdot\equiv\cdot \prod_{\otimes}^{m} C_s \otimes S_m$$

and is given by

$$'x_i = \omega^{s_i} x_j(i),$$

where $j(i)$ is any permutation of $(1, 2, \cdots, m)$ and ω is a primitive sth root of unity.

Now since $G^1_m[A] \subset G^2_m[A]$, for simplicial algebras A there are two exclusive possibilities. The generating relations of A are either order dependent or order independent. In the first case, the symmetry groups are $G^1_m[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_r$ and in the second

$$G^1_m[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_r \otimes S_m.$$

If $q = 2$, there are once again two extreme possibilities: either $G^i_m[A] \cdot\equiv\cdot O_m$, the mth-order orthogonal group over F, or

$$G^i_m[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_2 \otimes S_m.$$

In the first case $A[\alpha]$ is necessarily a set of polynomial algebras associated with the symmetric function $\mathfrak{L} = x_1^2 + x_2^2 + \cdots$. In the second case the function $\mathfrak{L}$ of even degree $2s$ (say) is such that

$$\mathfrak{L} \neq a(x_1^2 + \cdots + x_m^2)^s.$$

We summarize the above discussion in the following:

Theorem: Let $A[\alpha]$ be simplicial algebras satisfying the polynomial condition. If the group $H^2[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_s$, then all the coefficients P_r of $P[x; L]$ vanish unless $s \mid r$ and the terms that are present in P_r are symmetric polynomials wherein each term contains x_j with a degree divisible by s. Further, if one $a^s_{[s]} \neq 0$ in (2.5) for s/r, $r \leq m$, and $s > 2$, then the group $G^2_m[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_2 \otimes S_m$ with the transformations (5.12). In the case when $s = 2$ and $r = 2j$, if, for one r, $P_r \neq a_r\{x_1^2 + \cdots + x_m^2\}^j$, then

$$G^2_m[A] \cdot\equiv\cdot \prod_{\otimes}^{m} C_2 \otimes S_m;$$

and, if every $P_r = a_r\{x_1^2 + \cdots + x_m^2\}$, then

$$G^2_m[A] \cdot\equiv\cdot O_m,$$

the group of all orthogonal transformations over F, and $A[\alpha]$ belongs to $\mathfrak{L}$ where $\mathfrak{L} = (x_1^2 + x_2^2 + \cdots)$.

ACKNOWLEDGMENTS

The authors would like to thank Professor A. Ramakrishnan for giving them an opportunity to work at Matscience. One of the authors (N. B. M.) is grateful to the Council of Scientific and Industrial Research for a Junior Research Fellowship.

[1] A. Ramakrishnan, R. Vasudevan, N. R. Ranganathan, and P. S. Chandrasekaran, J. Math. Anal. Appl. **23**, 10 (1968); A. Ramakrishnan and R. Vasudevan, Matscience Preprint 26, 1969, to be published in J. Math. Anal. Appl. (1970).

[2] S. Eilenberg and J. A. Zilber, Ann. Math. **51**, 499 (1950).

[3] A. Ramakrishnan, J. Math. Anal. Appl. **20**, 9 (1967).

[4] P. T. Landsberg, Proc. Math. Soc. (London) **49**, 458 (1947).

[5] M. Schonberg, Anales Acad. Cienc. (Brazil) **28**, 11 (1957); H. Flanders, *Differential Forms* (Academic, New York, 1963).

[6] A. O. Morris, Quart. J. Math. (Oxford) **18**, 7 (1967); A. Ramakrishnan, R. Vasudevan, N. R. Ranganathan, and P. S. Chandrasekaran, J. Math. Anal. Appl. **23**, 10 (1968).

[7] C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963); S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1962); A. B. Govorkov, Zh. Eksp. Teor. Fiz. **54**, 1785 (1968) [Sov. Phys. JETP **27**, 960 (1968)].

[8] I. V. V. Raghavacharyulu, Mastech. Symp. **1**, 201 (1969).

[9] For the case when $n = 2$, see H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1963), p. 246, and for $n > 2$ see K. Morinaga and T. Nono, J. Sci. Hiroshima Univ. **16**, 13 (1952).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Symmetries of the Racah Coefficients

B. M. MINTON
Department of Physics, The University of Calgary, Calgary, Alberta, Canada

(Received 7 April 1970)

A new symmetry of the Racah coefficients is derived using a property of a generalized hypergeometric function of unit argument. The symmetry is similar in appearance, though not derivation, to that given by Regge.

One representation of Racah's coefficient $W(abcd; ef)$ is given by the series[1]

$$W(abcd; ef) = \Delta(abe)\Delta(cde)\Delta(acf)\Delta(bdf) \times \sum_p \left\{(-1)^p \Gamma\begin{bmatrix} a+b+c+d+2-p \\ a+b+1-e-p, c+d+1-e-p, a+c+1-f-p, b+d+1-f-p \end{bmatrix} \times \Gamma\begin{bmatrix} 1 \\ e+f+1-a-d+p, e+f+1-b-c+p, 1+p \end{bmatrix}\right\}, \quad (1)$$

where

$$\Delta(xyz) = \left\{\Gamma\begin{bmatrix} x+y+1-z, x+z+1-y, y+z+1-x \\ x+y+z+2 \end{bmatrix}\right\}^{\frac{1}{2}} \quad (2)$$

and where we have used the notation

$$\frac{\Gamma(a)\Gamma(b)\cdots}{\Gamma(p)\Gamma(q)\cdots} = \Gamma\begin{bmatrix} a, b, \cdots \\ p, q, \cdots \end{bmatrix}.$$

The coefficient is defined with the restriction that the triads (abe), (cde), (acf), and (bdf) have integer sum and a, b, c, d, e, and f are integral or half-integral. The series in Eq. (1) terminates when one of the Γ functions in the denominator has a pole, i.e., after $a+b-e$, $c+d-e$, $a+c-f$, or $b+d-f$ terms, whichever is the smaller. We shall subsequently assume, for convenience, that the series terminates after either $a+c-f$ or $b+d-f$ terms.

We may rewrite Eq. (1) in terms of the ${}_4F_3$ generalized hypergeometric function of unit argument. Then

$$W(abcd; ef) = \Delta(abe)\Delta(cde)\Delta(acf)\Delta(bdf) \times \Gamma\begin{bmatrix} a+b+c+d+2 \\ a+b+1-e, c+d+1-e, a+c+1-f, b+d+1-f, e+f+1-a-d, e+f+1-b-c \end{bmatrix} \times {}_4F_3[W; 1], \quad (3a)$$

with

$${}_4F_3[W; 1] \equiv {}_4F_3[e-a-b, e-c-d, f-c-a, f-b-d; -a-b-c-d-1, e+f+1-a-d, e+f+1-b-c; 1]. \quad (3b)$$

The ${}_4F_3$ function is defined in the usual manner[2] and terminates after the same number of terms as the series of Eq. (1).

It is well known that $W(abcd; ef)$ is unchanged by certain permutations of its parameters. For example,

$$W(abcd; ef) = W(badc; ef). \quad (4)$$

In addition, using a different series representation, Regge[3] has given further symmetries of the form

$$W(abcd; ef) = W(a, \tfrac{1}{2}[b+c+e-f], \tfrac{1}{2}[b+c+f-e], d; \tfrac{1}{2}[b+e+f-c], \tfrac{1}{2}[c+e+f-b]). \quad (5)$$

Equations (4) and (5) may be verified by direct substitution into Eq. (1) or Eq. (3) since they leave either the series or hypergeometric function unchanged.

We shall give a further set of symmetries for $W(abcd; ef)$ which are similar in form to those of Eq. (5). They differ, however, in that the definitions of $W(abcd; ef)$ afforded by Eq. (1) or Eq. (3) are not obviously invariant under these new symmetries, and we shall need to use a transformation property of the hypergeometric function in order to verify them.

We note first that the hypergeometric function ${}_4F_3[W; 1]$ is of the terminating Saalschutzian variety, i.e., the sum of the lower parameters exceeds the sum of the upper parameters by one. For this type of function, we know that[4]

$$
\begin{aligned}
{}_4F_3[A, B, C, D; E, F, G; 1] = \Gamma\begin{bmatrix} E+F-A-B-D, E+F-A-B-C, F-C-D, F \\ E+F-A-B, E+F-A-B-C-D, F-C, F-D \end{bmatrix} \\
\times {}_4F_3[E-A, E-B, C, D; E, E+F-A-B, E+G-A-B; 1], \quad (6)
\end{aligned}
$$

where $A+B+C+D+1 = E+F+G$ and either C or D is a negative integer, so that both functions terminate. We see then that, from Eqs. (3b) and (6), we may write

$$
\begin{aligned}
{}_4F_3[W; 1] = \Gamma\begin{bmatrix} b+f+d+2, a+c+f+2, a+d+e-f+1, e+f-b-c+1 \\ 2f+2, a+b+c+d+2, a+e-b+1, e+d-c+1 \end{bmatrix} \\
\times {}_4F_3[b+f-d+1, c+f-a+1, f-a-c, f-b-d; e+f-a-d+1, 2f+2, f-e-a-d; 1]. \quad (7)
\end{aligned}
$$

We have assumed that the series in Eqs. (1) or (3) terminates after $a+c-f$ or $b+d-f$ terms in making this transformation. We may clearly write a similar transformation if there are $a+b-e$ or $c+d-e$ terms.

We may now return to Eq. (3). We then have, for example, that

$$
\begin{aligned}
&W(\tfrac{1}{2}[a+c+d-b], \tfrac{1}{2}[e-f-a+d-1], \tfrac{1}{2}[e-f+a-d-1], \tfrac{1}{2}[a+b+d-c]; \\
&\qquad \tfrac{1}{2}[e+f+b+c+1], \tfrac{1}{2}[e+f-b-c-1]) \\
&= \left\{\Gamma\begin{bmatrix} d-f-b, a+c+f+2, b+e-a+1, a-c-f, 1+c+e-d, b+d+f+2 \\ c+d+e+2, a+b+e+2 \end{bmatrix}\right. \\
&\quad \left.\times \Gamma\begin{bmatrix} 1+a+c-f, 1+d+f-b, e-c-d, b+d-f+1, e-a-b, a+f-c+1 \\ a+e-b+1, e+d-c+1 \end{bmatrix}\right\}^{\frac{1}{2}} \\
&\quad \times \Gamma\begin{bmatrix} a+d+e-f+1 \\ d-f-b, a-c-f, a+c-f+1, b+d+1-f, e+f-a-d+1, 2f+2 \end{bmatrix} \\
&\quad \times {}_4F_3[b+f-d+1, c+f-a+1, f-a-c, f-b-d; f-a-d-e, 1+e+f-a-d, \\
&\qquad 2f+2; 1] \quad (8) \\
&= \{\ \}^{\frac{1}{2}}\Gamma\begin{bmatrix} a+b+c+d+2, e+d-c+1, e+a-b+1 \\ d-f-b, a-c-f, 1+a+c-f, 1+b+d-f, 1+e+f-a-d, b+d+f+2 \end{bmatrix} \\
&\quad \times \Gamma\begin{bmatrix} 1 \\ a+c+f+2, e+f-b-c+1 \end{bmatrix} {}_4F_3[W; 1], \quad (9)
\end{aligned}
$$

where $\{\ \}^{\frac{1}{2}}$ represents the similarly bracketed term in Eq. (8) and we have substituted from Eq. (7) into Eq. (8) to get Eq. (9). After some simplification, we finally obtain, using Eq. (3), that

$$
\begin{aligned}
&W(\tfrac{1}{2}[a+c+d-b], \tfrac{1}{2}[e-f-a+d-1], \tfrac{1}{2}[e-f+a-d-1], \tfrac{1}{2}[a+b+d-c]; \\
&\qquad \tfrac{1}{2}[e+f+b+c+1], \tfrac{1}{2}[e+f-b-c-1]) = W(abcd; ef). \quad (10)
\end{aligned}
$$

This symmetry is similar in appearance to the Regge symmetry [Eq. (5)] apart from the ± 1 terms, although it is essentially different in origin. Other symmetries may be obtained by combining Eq. (10) with Eqs. (4) and (5).

[1] See, for example, L. C. Biedenharn, J. Math. & Phys. **31**, 287 (1953).
[2] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966), Chap. 2.
[3] T. Regge, Nuovo Cimento **11**, 116 (1959).
[4] Reference 2, p. 64, Eq. (2.4.1.7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

New Solutions of the Kinematic Dynamo Problem

STEPHEN CHILDRESS
New York University, Courant Institute of Mathematical Sciences, New York, New York 10012

(Received 28 January 1970)

The steady-state kinematic dynamo problem in a homogeneous 3-dimensional core is studied. The existence of a class of smooth solenoidal dynamos, satisfying a no-slip condition on the core boundary, is proved using perturbation theory. The dynamos are of the form $\mathbf{q} = \mathbf{q}^{(1)} + \mathbf{q}^{(2)} + \mathbf{q}^{(3)}$, where $\mathbf{q}^{(1)}$ is spatially periodic on a sufficiently small scale of length, $\mathbf{q}^{(2)}$ is zero except near the core boundary, and $\mathbf{q}^{(3)}$ is an arbitrary sufficiently small motion. The term $\mathbf{q}^{(1)}$ is also a spatially periodic dynamo in an appropriate sense for an infinite core. The last property allows a simple characterization of the bounded dynamos in terms of the admissible $\mathbf{q}^{(1)}$.

1. INTRODUCTION

In this paper, we study the steady-state induction of an electromagnetic field by a motion within a spherical fluid conductor of unit radius. The principal result of the investigation is an explicit construction of a class of bounded kinematic dynamos, i.e., suitably regular motions of the fluid capable of maintaining indefinitely a magnetic field of bounded nonzero energy.[1] These solutions of the kinematic dynamo problem have a number of features which are consistent with the physical aspects of the theory of the earth's magnetic field; they are related, in particular, to a certain class of laminar and turbulent motions of a perfect fluid which might be expected to appear in a rotating mass of fluid.

These dynamos have the structure

$$\mathbf{q}(\mathbf{r};\epsilon) = -\epsilon\nabla \times [\omega(r;\epsilon)\mathbf{v}(\mathbf{r}/\epsilon)] + \epsilon^{\beta+\frac{1}{2}}\mathbf{w}(\mathbf{r}), \qquad 0 \le r \le 1, \quad (1.1)$$

where $\mathbf{q}$ = velocity, $\mathbf{r} = (x, y, z) = (x_1, x_2, x_3)$, $r = |\mathbf{r}|$, and β is any positive number. The function $\mathbf{w}$ may be any continuously differentiable solenoidal field satisfying the no-slip condition $\mathbf{w} = 0$ on the core boundary $r = 1$. The function $\mathbf{v}(\mathbf{r})$ is of the form

$$\mathbf{v}(\mathbf{r}) = \sum_K \boldsymbol{\mu}(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{r}). \quad (1.2)$$

In (1.2), K is the set of vectors $\mathbf{k} = m_1\mathbf{k}_1 + m_2\mathbf{k}_2 + m_3\mathbf{k}_3$, where $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ is an orthogonal basis in 3-space and m_1, m_2, m_3 are nonzero integers. Further conditions to be imposed on $\mathbf{v}$ are described in Sec. 4. The scalar function ω in (1.1) is an infinitely differentiable "cutoff" which vanishes on $r = 1$ and is unity for $0 \le r \le 1 - \epsilon$. A suitable choice is

$$\omega(r;\epsilon) = \pi^{-\frac{1}{2}} \int_{-\infty}^{p(r)} e^{-s^2}\, ds, \quad (1.3a)$$

where

$$p(r) = (1 - \tfrac{1}{2}\epsilon - r)/(r - 1 + \epsilon)(1 - r). \quad (1.3b)$$

Finally, ϵ is an arbitrary positive parameter. We prove that, with suitable additional conditions on $\mathbf{v}$, motions of the form (1.1)–(1.3) are steady-state kinematic dynamos in the unit sphere provided that ϵ is sufficiently small.

The field $\mathbf{v}$ is studied in a separate analysis, by considering a related kinematic dynamo problem in an infinite conductor. This yields, in a natural manner, additional necessary conditions on $\mathbf{v}$, which fully characterize this class of bounded dynamos. An example of an admissible $\mathbf{v}$ is, in component form,

$$\mathbf{v}(\mathbf{r}) = (\sin y + \cos z, \sin z + \cos x, \sin x + \cos y), \quad (1.4)$$

as is shown in Sec. 4.

The dynamo theory required to study (1.1) is not, however, easily related to that of (1.4), since there is no natural algebra comparable to that of the periodic functions upon which to base the analysis. Instead, we use perturbation theory in an operator formalism, the unperturbed operator corresponding to a certain *comparison* dynamo problem which, for the case of the spherical conductor, is easily solved. In either case, the operator equation has the well-known eigenvalue character.[1] The spectrum of the (self-adjoint) comparison operator is discrete, real, and symmetric about the origin. In the examples to be treated, the exact eigenvalues are obtained by perturbing a *positive* comparison eigenvalue. A second family of dynamos and electromagnetic eigensolutions, associated with the negative comparison eigenvalues, may be obtained by reflection since, if $\mathbf{q}(\mathbf{r};\epsilon)$ is a steady dynamo in a sphere, then the same is true of $-\mathbf{q}(-\mathbf{r};\epsilon)$ [cf. (2.10)–(2.4) below]. For simplicity and without loss of generality, we may, therefore, restrict attention to dynamos of like parity, as determined by the positivity of the unperturbed eigenvalue. As in other formulations of the kinematic dynamo problem, the primary goal is the existence of a *real* perturbed eigenvalue.

The method we use may be described schematically in the following way: The effect of a motion $\mathbf{q}(\mathbf{r};\epsilon)$ on a magnetic field $\mathbf{h}$ defined in the core is to induce a second field $L\mathbf{h}$ defined in all space (see Sec. 5). This definition determines the operator L in the form

$$L = HQ, \quad Q\mathbf{h} \equiv \mathbf{q} \times \mathbf{h}, \tag{1.5}$$

where H is the integral operator whose kernel is the Green's tensor for Maxwell's equations without displacement current. The exact problem is therefore obtained on a suitable Banach space of functions $\mathbf{h}$ defined in the core, and the form is

$$L\mathbf{h} = \lambda\mathbf{h}. \tag{1.6}$$

For the dynamos studied here, the comparison problem has the elementary form

$$H\mathbf{h} = \mu\mathbf{h}, \quad \mu > 0. \tag{1.7}$$

The connection between (1.6) and (1.7) can be clarified by exhibiting the eigenvalue problem

$$H\mathbf{h} = \nu\mathbf{h} - V\mathbf{h}, \quad V = L^2 - H, \tag{1.8}$$

in which V occurs as the perturbing operator and $L^2 \sim H$ and $\nu \sim \mu$ in the proposed scheme.

Thus, the method may be characterized by the fact that it is the second iterate of L, rather than L itself, which is compared with a self-adjoint operator. Smallness of V is achieved by an appropriate choice of linear space. The perturbation is ultimately carried out in Banach space, using a maximum norm containing a fractional power of ϵ as a factor of the estimate on first derivatives of $\mathbf{h}$ (see Sec. 5). We emphasize this point, not only because the choice of appropriate space proved to be a central issue in the analysis, but also because some of the original proposals dealing with the kinematic dynamo theory, in effect, utilized almost exclusively the Hilbert-space norm associated with magnetic fields of finite total energy.[2,3] For motions of the form (1.1) (and these are believed to be among the simplest fluid motions for which an analytic dynamo theory can be developed) such spaces appear to be "wide." A magnetic field which is bounded in the inner-product norm, uniformly as $\epsilon \to 0$, may interact with a motion of this type to induce a field which becomes unbounded with respect to the same norm as $\epsilon \to 0$. This leads to very essential difficulties in formulating a rational theory of the perturbed spectrum. However, on Banach spaces of the kind mentioned above, it becomes possible to render L uniformly bounded as $\epsilon \to 0$. Thus, in this respect, boundedness of the exact operator is here regarded as determining on the class of norms that are useful.

The above reasoning and, indeed, all essential parts of the subsequent analysis can be illustrated quite simply with the aid of a complex-valued 1-dimensional analog of Maxwell's equations for a moving medium. Therefore, we present this parallel analysis in one space dimension in Sec. 3, even though the fields are there complex and cannot be related, by Cowling's theorem,[1] to a real version of the dynamo problem.

The physical basis for the regenerative cycle which emerges from the dynamos studied in this paper is very similar to that underlying the model proposed by Parker.[4] Although Parker considers a time-dependent dynamo cycle, the idea of using a (time) rapid small-scale motion capable of inducing a predominantly large-scale current in the conductor is clearly contained in his model. As is clear from (1.1), the present examples are predominantly small scale with respect to spatial variation in the velocity, and, since H is independent of ϵ in (1.8), the induced magnetic field is predominantly large scale (i.e., on the scale of the core). The parameter ϵ may therefore be regarded as a ratio of length scales determined from the dominant components of the fields. The notion of a "scale separation" of this kind is, in fact, rather common in dynamo theory. In the model of Backus, it occurs as a separation of the cycle into well-defined periods of motion and (sufficiently long) periods of free decay.[3] In Herzenberg's 2-sphere model, the small parameter is the ratio of sphere radius to conductor radius.[5,6] Compared with these examples, our results may be viewed as extensions of known existence theory to a class of essentially fluid motions which cannot be built up from a finite collection of rigid rotators or which do not require an implausible distribution of body forces for their operation.

The results given in the present paper were originally announced in April, 1967.[7,8] In the intervening time, the analysis has been further clarified with the aid of 1-dimensional analogs, and a few errors have been corrected, although the proof given here remains essentially unchanged. Our original discussion of the self-adjointness of H contained an error in the treatment of boundary terms. Examination of this point led to a simplification in geometry and a new definition of the Hilbert space appropriate to H. The convexity condition on periodic dynamos and the argument employing modified operators in the existence proof for the bounded conductor (see Sec. 7) also required revision. In order to make the present paper self-contained, portions of the construction of spatially periodic dynamos, described elsewhere, are repeated in Sec. 4.[8]

2. FORMULATION

Let D denote the interior of the unit sphere centered at the origin, S the boundary $r = 1$, and V the exterior region $r > 1$. We also set $\bar{D} = D + S$ and $\mathcal{E} = \bar{D} + V$. We refer to $\bar{D}$ as the conductor, or *core*, and endow it with a constant scalar electrical conductivity σ and magnetic permeability μ. The magnetic field $\mathbf{h}$ and the electric field $\mathbf{e}$ generated by the motion $\mathbf{q}$ of D are assumed to satisfy Maxwell's equations for a moving medium in the magnetohydromagnetic limit. In dimensionless notation, these are

$$\mathbf{J} \equiv \nabla \times \mathbf{h} = \mathbf{e} + R\mathbf{q} \times \mathbf{h}, \tag{2.1a}$$

$$\nabla \cdot \mathbf{h} = 0, \tag{2.1b}$$

$$\nabla \times \mathbf{e} = -R \frac{\partial \mathbf{h}}{\partial t}, \tag{2.1c}$$

when $\mathbf{r} \in D$; since V is regarded as empty space,

$$\nabla \times \mathbf{h} = \nabla \cdot \mathbf{h} = \nabla \cdot \mathbf{e} = 0, \tag{2.2a}$$

$$\nabla \times \mathbf{e} = -R \frac{\partial \mathbf{h}}{\partial t}, \tag{2.2b}$$

when $\mathbf{r} \in V$. Here R is a real dimensionless parameter usually called the magnetic Reynolds number.[9] In the kinematic dynamo theory, $\mathbf{q}$ is a prescribed function of $\mathbf{r}$ and t. Given $\mathbf{q}$, we then seek solutions of (2.1) and (2.2) which satisfy

$$\mathbf{r} \times \mathbf{e}, \quad \mathbf{h} \text{ continuous on } S, \tag{2.3}$$

$$r^3 h, \quad r^2 e \text{ bounded in } V. \tag{2.4}$$

A subsidiary condition,

$$\mathbf{J} \cdot \mathbf{r} \to 0 \quad \text{as} \quad r \to 1 \quad \text{in } D, \tag{2.5}$$

which follows from (2.2) and (2.3), is also used. If $\mathbf{q}$, $\mathbf{h}$, and $\mathbf{e}$ are independent of time, we refer to the problem and its solutions as *steady*. The steady kinematic dynamo problem, which is the principal problem studied in this paper, can then be defined as follows: To determine a class Q of reasonable motions $\mathbf{q}(\mathbf{r})$ such that $\mathbf{q} \in Q$ implies the existence of a real value of R for which (2.1)–(2.4) has a nontrivial steady solution. Here, "reasonable" is understood to imply that $\mathbf{q}$ is solenoidal in D and zero on S, and it is at least continuously differentiable in $\bar{D}$. These conditions imply that such $\mathbf{q}$ represents a possible smooth source-free flow of an incompressible fluid which adheres to the boundary.

It is readily observed from (2.1) that the mathematical problem posed here reduces to an eigenvalue problem for the linear elliptic system

$$\nabla^2 \mathbf{h} + R\nabla \times (\mathbf{q} \times \mathbf{h}) = 0, \quad \nabla \cdot \mathbf{h} = 0, \tag{2.6}$$

with variable coefficients determined by $\mathbf{q}$ and the eigenparameter R. Stated in another way, the dimensional *speed* of the dynamo is the eigenvalue. The basic linearity of the problem as indicated by (2.6) is, however, misleading since, as formulated above, the kinematic problem requires an analysis of the (nonlinear) functional connection between the real point spectrum of the differential operator in (2.6) and the motion (coefficient) $\mathbf{q}$ by which this operator is determined. The class Q studied here is of the form (1.1).

Our analysis of the steady problem is based on a reformulation of (2.1)–(2.4) as an integral equation. Let $\mathbf{K}(\mathbf{r}, \mathbf{r}')$ and $\mathbf{E}(\mathbf{r}, \mathbf{r}')$ be the magnetic and electric fields obtained by solving (2.1)–(2.4) when the term $R(\mathbf{q} \times \mathbf{h})$ in (2.1a) is replaced by $\mathbf{I}\delta(\mathbf{r} - \mathbf{r}')$, where $\mathbf{I}$ = idemfactor and $\mathbf{r} \in \mathcal{E}$, $\mathbf{r}' \in D$. Applying the superposition principle, we obtain the integral eigenvalue problem for $\mathbf{h}$ in the form (1.6):

$$L\mathbf{h} \equiv \epsilon^{-\frac{1}{2}} \int_D \mathbf{K}(\mathbf{r}, \mathbf{r}') \cdot [\mathbf{q}(\mathbf{r}') \times \mathbf{h}(\mathbf{r}')]\, d\mathbf{r}' = \lambda \mathbf{h}(\mathbf{r}), \quad \mathbf{r} \in \bar{D}, \quad \lambda = R_0^{-1}, \tag{2.7}$$

where we have set $R(\epsilon) = R_0(\epsilon)\epsilon^{-\frac{1}{2}}$ in (2.1a). With $\mathbf{E} = \nabla\Phi$, the corresponding electric potential is given by

$$\phi(\mathbf{r}) = \lambda^{-1}\epsilon^{-\frac{1}{2}} \int_D \Phi(\mathbf{r}, \mathbf{r}') \cdot [\mathbf{q}(\mathbf{r}') \times \mathbf{h}(\mathbf{r}')]\, d\mathbf{r}' \equiv \lambda^{-1} M\mathbf{h}. \tag{2.8}$$

For the spherical core, an explicit construction of $\mathbf{K}$ and $\mathbf{E}$ can be given (see the Appendix).

3. AN ANALOGOUS PROBLEM IN ONE DIMENSION

A. Equations

In this section, the steady eigenvalue problem (2.7), with $\mathbf{q}$ given by (1.1), is studied in a complex-valued 1-dimensional analog defined on one space dimension. Here extension to the complex plane is necessary in order to avoid the implications of Cowling's theorem.[1] Notation, terminology, and the steps in the derivation of the existence theorem (Theorem 2 below) parallel the study of (2.7) in Secs. 5–7.

In this analog, $h(x, t)$ (magnetic field), $e(x, t)$ (electric field), and $q(x, t)$ (velocity field) are complex functions of the real variables x and t. The infinite conductor becomes the real x line, while the analogous core is taken to be the segment $|x| < 1$. The analog of

the Maxwell equations is taken to be

$$i\frac{\partial h}{\partial x} = e + Rq^*h^* \tag{3.1a}$$

and

$$i\frac{\partial e}{\partial x} = -R\frac{\partial h}{\partial t} \tag{3.1b}$$

in the conductor. In (3.1a), the star denotes complex conjugate.

If all variables are independent of time and $q = \exp(ix/\epsilon)$ in (3.1), then there exists a family of x-periodic solutions defined when $|\epsilon R| < \frac{1}{2}$. These are found by noting that, in this case, e = constant and that, from (3.1),

$$\frac{d}{dx}q\frac{dh}{dx} = R^2qh + Rq^*e - \epsilon^{-1}e. \tag{3.2}$$

The solution of (3.2) is straightforward, and we obtain

$$h = C[\exp(i\lambda_2 x) - R\lambda_1^{-1}\exp(i\lambda_1 x)] - R^{-1}q^*e^* + \epsilon^{-1}R^{-2}e,$$

$$\lambda_{1,2} = -(2\epsilon)^{-1}[1 \pm (1 - 4\epsilon^2R^2)^{\frac{1}{2}}], \tag{3.3}$$

where C is an arbitrary real constant. For the same q, (3.2) becomes an eigenvalue problem for $R(\epsilon)$ provided that we add a periodicity condition. We treat only the case $e = 0$ and require that

$$h(-1) = h(+1). \tag{3.4}$$

With (3.4), the equation for R is

$$R = \lambda_1(\sin\lambda_2/\sin\lambda_1), \tag{3.5}$$

which, for sufficiently small ϵ, has solutions $R = R_0^{(n)}(\epsilon)\epsilon^{-\frac{1}{2}}$, where

$$[R_0^{(n)}]^2 = n\pi - \epsilon^{\frac{1}{2}}\sin 1/\epsilon\cos n\pi R_0^{(n)} + O(\epsilon) \tag{3.6}$$

as $\epsilon \to 0$ with $n = 0, 1, 2, \cdots$. If q^* replaces q in the above, similar equations are obtained with $-R$ replacing $+R$ in (3.5). This property of the model problem is analogous to the invariance of the steady dynamo effect under reflection, as noted in Sec. 1.

B. Integral Formulation

The model dynamo problem for a 1-dimensional core may now be defined as before, for q in the form

$$q = -i\epsilon\frac{d}{dx}[\omega(|x|;\epsilon)v(x;\epsilon)] + \epsilon^{\beta+\frac{1}{2}}w(x), \tag{3.7}$$

where ω is given by (1.3) and w is continuously differentiable for $|x| \leq 1$. We add to (3.1a) and (3.1b) the boundary condition

$$h(-1) = h(+1) = 0 \tag{3.1c}$$

and introduce the model version of (2.7):

$$Lh \equiv \epsilon^{-\frac{1}{2}}\int_{-1}^{+1} K(x, x')q^*(x')h^*(x')\,dx' = \lambda h(x), \quad |x| < 1, \quad \lambda = R_0^{-1}, \tag{3.7'}$$

$$K(x, x') = -\tfrac{1}{2}i\,\mathrm{sgn}(x - x') + \tfrac{1}{2}ix. \tag{3.7''}$$

Now, adopting the choice $v(x, \epsilon) = \exp(ix/\epsilon)$ in (3.7), we define the comparison eigenvalue problem

$$Hh \equiv \int_{-1}^{+1} K(x, x')h(x')\,dx' = \mu h(x) \tag{3.8}$$

and the quantities V and ν as in (1.8).

C. Estimates

The operators H, L, and V are characterized by their action on two principal linear spaces. Let C_1 denote the complex space of functions $h(x)$ which are continuously differentiable on $|x| \leq 1$ and satisfy (3.1c). A Hilbert space $\mathcal{H}$ is then obtained by completing C_1 in the norm

$$\|h\| = (h, h)^{\frac{1}{2}}, \quad (f, g) = \int_{-1}^{+1}\left(\frac{df}{dx}\frac{dg^*}{dx}\right)dx. \tag{3.9}$$

In addition, admitting now a parametric dependence on ϵ, we introduce the 2-parameter family of Banach spaces $C_1^\gamma(\epsilon)$, defined for each positive ϵ and γ by adding to C_1 the norm

$$\|h\|_\infty^\gamma = \|h\|_\infty + \epsilon^\gamma\|h'\|_\infty, \quad \|h\|_\infty = \sup_{|x|\leq 1}|h|, \quad h' = \frac{dh}{dx}. \tag{3.10}$$

If, for fixed γ, $\|h\|_\infty^\gamma = O(1)$ or $o(1)$ as $\epsilon \to 0$, we say that $h(x;\epsilon)$ is uniformly bounded or uniformly small, respectively, in $C_1^\gamma(\epsilon)$. Similar definitions will be implicit in our discussion of other properties of functions or operators when uniformity with respect to ϵ is required. In particular, an operator will be uniformly compact on C_1^γ if it maps a uniformly bounded sequence in C_1^γ into a uniformly equicontinuous sequence in C_1^γ.

The following lemmas are then elementary:

Lemma 1: H is a self-adjoint on $\mathcal{H}$ and is uniformly compact on $C_1^\gamma(\epsilon)$;

Lemma 2: L is uniformly bounded on $C_1^\gamma(\epsilon)$ if $\gamma = \frac{1}{2}$.

Lemma 3:

$$\|Vh\|_\infty = o(\|h\|_\infty^\gamma) \quad \text{as} \quad \epsilon \to 0 \quad \text{if} \quad \gamma < \min[1, \tfrac{1}{2} + \beta].$$

Proof of Lemma 1: This is straightforward; we note that the uniform compactness follows from the inequality

$$|Hh(x_2) - Hh(x_1)| + \epsilon^\gamma |[Hh(x_2) - Hh(x_1)]'| \leq |x_2 - x_1| \, \|h\|_\infty^\gamma$$

for $|x_1|, |x_2| \leq 1$, which is easily proved using the mean value theorem. The eigenvalues of H acting on $\mathcal{H}$ are simple with $\mu = \mu_n = 1/n\pi$, $n = \pm 1, \pm 2, \cdots$, and the associated eigenfunctions are complete in $\mathcal{H}$.

Proof of Lemma 2: We use (3.7), (3.9), and an integration by parts to obtain

$$\|Lh\|_\infty \leq (\epsilon^{\frac{1}{2}} + \epsilon^\beta \|w\|_\infty) \|h\|_\infty + \epsilon^{\frac{1}{2}} \|h'\|_\infty, \quad (3.10'a)$$

$$\|(Lh)'\|_\infty \leq 2\epsilon^{-\frac{1}{2}} \|q\|_\infty \|h\|_\infty. \quad (3.10'b)$$

By our construction, $\|q\|_\infty = O(1)$ as $\epsilon \to 0$, so that the result follows if $\beta \geq 0$.

Proof of Lemma 3: We can derive the similar estimates

$$\|Vh\|_\infty \leq (6\epsilon + 3\epsilon^{\frac{1}{2}+\beta} \|w\|_\infty) \|h\|_\infty + \epsilon \|h'\|_\infty + (\epsilon^{\frac{1}{2}} + \epsilon^\beta \|w\|_\infty) \|Lh\|_\infty, \quad (3.11a)$$

$$\|(Vh)'\|_\infty \leq 2(1 + \|\epsilon\omega'\|_\infty + \epsilon^{\beta+\frac{1}{2}} \|w\|_\infty) \times (\|h\|_\infty + \epsilon^{-\frac{1}{2}} \|Lh\|_\infty) + \epsilon^{\frac{1}{2}} \|(Lh)'\|_\infty. \quad (3.11b)$$

Combining (3.10) and (3.11) and using the fact that $\|\epsilon\omega'\|_\infty$ is $O(1)$ as $\epsilon \to 0$, we then obtain

$$\|Vh\|_\infty = O(\epsilon \|h\|_\infty) + O(\epsilon^{\frac{1}{2}+\beta} \|h\|_\infty) + O(\epsilon^{2\beta} \|h\|_\infty) + O(\epsilon \|h'\|) + O(\epsilon^{\beta+\frac{1}{2}} \|h'\|), \quad (3.12a)$$

$$\|(Vh)'\|_\infty = O(\|h\|_\infty) + O(\epsilon^{\beta-\frac{1}{2}} \|h\|_\infty) + O(\|h'\|_\infty). \quad (3.12b)$$

Now (3.12a) implies Lemma 3, while (3.12b) proves, in addition, that V is uniformly bounded on $C_1^\gamma(\epsilon)$ if $\max [0, \frac{1}{2} - \beta] \leq \gamma \leq \min [1, \frac{1}{2} + \beta]$. However, V is not uniformly small for any $\gamma > 0$, since (3.12b) contains an estimate of order $\|h'\|_\infty$. A device which enables us to overcome this difficulty is described in the next section.

D. The Modified Operators

The family of modified operators $\tilde{V}(c)$ is now defined by

$$\tilde{V}h = \int_{-1}^{+1} K(x, x')(\omega v^* q - 1)h \, dx' - \left(\int_{-1}^{+1} K(x, x') v^* \omega h \, dx'\right)\left(\int_{-1}^{+1} qh \, dx'\right) + \epsilon^{\frac{1}{2}} v^* \omega c^{-1} (L^3 h)^*, \quad (3.13)$$

where c is a new real nonzero parameter. Inspection of (3.13) shows that h has been replaced by $c^{-1}L^2 h$ in the term of V which contributed the estimate of order $\|h'\|_\infty$ in (3.12b). Thus, if h satisfies $L^2 h = \nu h$ with $\nu = c$, then $\tilde{V}h = Vh$. Conversely, it is not difficult to see from (3.13) that, if

$$[H + \tilde{V}(c)]h(c) = \nu(c)h(c) \quad (3.14)$$

for some $\nu(c)$ and $h(c)$ uniformly on a positive closed interval in which $\nu(c) = c$ has a solution $\nu_0 = \lambda^2(\epsilon)$, then $h(\lambda^2)$ and $\nu(\lambda^2)$ provide an eigenfunction of L^2 with positive eigenvalue. In this case, at least one of the two numbers $\pm\lambda$ is a real eigenvalue of L. Therefore, it is sufficient, under these conditions, to study the modified problem (3.14).

The importance of the above modification is reflected in an improved estimate on $C_1^\gamma(\epsilon)$.

Theorem 1: For each fixed positive c, $\tilde{V}(c)$ is uniformly small on $C_1^\gamma(\epsilon)$ provided that

$$\max [0, \tfrac{1}{2} - \beta] < \gamma < \min [1, \tfrac{1}{2} + \beta].$$

Proof: This result is an immediate consequence of (3.10), (3.11), and (3.13), which together imply

$$\|\tilde{V}\|_\infty^\gamma = O(\epsilon^{2\beta}) + O(\epsilon^\gamma) + O(\epsilon^{1-\gamma}) + O(\epsilon^{\beta+\gamma-\frac{1}{2}}) + O(\epsilon^{\beta+\frac{1}{2}-\gamma}). \quad (3.15)$$

E. Perturbation of H

Let f be any solution of (3.8) with positive eigenvalue μ_k. We write (3.14) in the form

$$h = -(H - \nu)^{-1}\tilde{V}h \equiv Th = (T - PT)h + PTh \equiv \tilde{T}h + PTh, \quad (3.16)$$

where P is the projection in $\mathcal{H}$ onto scalar multiples of f. If $(I - \tilde{T})^{-1}f \in \mathcal{H}$, then a solution of (3.16) is obtained in the form $h = (I - \tilde{T})^{-1}f$ provided that

$$f = PT(I - \tilde{T})^{-1}f. \quad (3.17)$$

(This argument is given in detail in Sec. 4B.) Now, (3.17) is equivalent to

$$\nu = \mu_k + (f, \tilde{V}(I - \tilde{T})^{-1}f)(f, f)^{-1} \equiv \mu_k + R(\nu, c, \epsilon), \quad (3.18)$$

the inner product being that appropriate to $\mathcal{H}$. From the manner in which c and ν occur in $\tilde{V}$ and $\tilde{T}$, respectively [cf. (3.13) and (3.16)], it is seen (by partial differential with respect to either parameter) that R is continuous in both parameters provided that V and $(I - \tilde{T})^{-1}$ exist as operators on C_1^γ for all nonzero c. Now, suppose that this is the case whenever ν lies in some closed subset Δ of the real line and that $|R| = o(1)$ as $\epsilon \to 0$ uniformly for ν and c in Δ. It then follows from the implicit function theorem that (3.18) has a real solution $\nu(c, \epsilon)$ for ϵ sufficiently small.

Moreover, by a similar argument, $\nu(c, \epsilon) = c$ has a solution $c(\epsilon)$ provided that ϵ is sufficiently small. Then $\nu(\epsilon) = \nu(c(\epsilon), \epsilon)$ is the desired positive eigenvalue of L^2, and existence is proved.

Therefore, it follows, from (3.15) and (3.18), that it is sufficient to show that Δ can be chosen so that $\|\tilde{T}\|_\infty^\gamma = o(1)$ uniformly for nonzero ν and c in Δ. We first observe that, if Δ is chosen to be positive, of finite length, and to contain among the eigenvalues of H only μ_k, then $(H - \nu)^{-1}(I - P)$ is a bounded operator on C_1^γ. (Here P is regarded as an operator on C_1^γ.) Indeed, if this were not the case, then, by Lemma 1, H would have an eigenvalue in Δ with eigenfunction of the form $(I - P)h$, h being some element of C_1^γ. But then the eigenfunction would belong to $\mathcal{H}$, and this is impossible by the above choice of Δ. Combining this result with (3.15), we see that

$$\|\tilde{T}\|_\infty^\gamma = o(1),$$

if $\beta > 0$ and $\frac{1}{2} < \gamma < \frac{1}{2} + \beta$. Therefore, we have proved

Theorem 2: If $\beta > 0$, the functions q of the form (3.7) with $v = \exp(ix/\epsilon)$ are "dynamos" if ϵ is sufficiently small. This is the existence theorem in the 1-dimensional analog.

F. Remarks

The difficulty mentioned in Sec. 1, concerning the exclusive use of Hilbert-space norms, can be clarified in the present model. The relation between Lemma 2 and Theorem 1 suggests that boundedness of L should be obtainable in a family of norms which provide useful estimates on the operator V. The obvious family of Hilbert spaces is determined by the norms

$$\|h\|_2^\gamma = (h, h)^{\frac{1}{2}} + \epsilon^\gamma (h', h')^{\frac{1}{2}}. \tag{3.19}$$

We now show that, with $q = \exp(ix/\epsilon)$ replacing (3.7), $\|L\|_2^\gamma$ cannot be $O(1)$ as $\epsilon \to 0$ for any choice of $\gamma \geq 0$. Consider the function

$$\begin{aligned} g &= 0, && |x| > \epsilon^\alpha, \\ &= q^* \epsilon^{-\frac{1}{2}\alpha}, && |x| \leq \epsilon^\alpha, \end{aligned}$$

with $\alpha > 0$. Then $\|g\|_2^\gamma = 2^{\frac{1}{2}}(1 + \epsilon^{\gamma-1})$ and $\|Lg\|_2^\gamma \geq \epsilon^{\gamma - \frac{1}{2} - \frac{1}{2}\alpha}$. Thus, $\|Lg\|_2^\gamma / \|g\|_2^\gamma \to \infty$ for $\alpha > 2\gamma - 1$ if $\gamma \geq 1$ and for $\alpha > 1$ if $\gamma < 1$, as $\epsilon \to 0$.

Theorem 2 admits a number of generalizations, the simplest and most useful of which allows $v(x; \epsilon)$ to be an arbitrary trigonometric function of x,

$$v(x) = \sum_{k=-\infty}^{+\infty} a_k \exp(ikx/\epsilon),$$

where only a finite number of the complex amplitudes a_k are nonzero. In this case, the only essential change in the previous results is that H must be redefined and is given by

$$Hh \equiv \sum_{k=-\infty}^{+\infty} \frac{|a_k|^2}{k} \int_{-1}^{+1} K(x, x') h(x')\, dx'. \tag{3.20}$$

It is not difficult to see how (3.21) arises from the manner in which Lemma 3 and Theorem 1 were obtained. The same conclusions hold for arbitrary periodic functions v with period $2\pi\epsilon$ provided that the infinite sum in (3.20) converges and is nonzero.

4. SPATIALLY PERIODIC DYNAMOS IN THREE DIMENSIONS

A. Admissible v

We now consider, in detail, the periodic functions $\mathbf{v}$ [cf. (1.2)] which determine the set Q of 3-dimensional dynamos $\mathbf{q}$ having the form (1.1). For simplicity, we impose the further condition that $\mathbf{v}$ be trigonometric, i.e., (as in the extension of the analog mentioned in Sec. 3E) only a finite number of complex amplitudes attached to K will be nonzero. It is also understood that $\mathbf{v}$ is real, so that necessarily $\boldsymbol{\mu}(-\mathbf{k}) = \boldsymbol{\mu}^*(\mathbf{k})$ in (1.2). For a given field of this form, we may attempt to solve (2.1) with $\mathbf{q} = \mathbf{v}$ in (2.1a) for the steady-state fields in an infinite conductor. The analogous problem in Sec. 3 yielded the periodic solutions (3.3). In the actual 3-dimensional problem, explicit solutions for arbitrary ϵ are not available, and, instead, we seek a convergent expansion for small ϵ.

The expansion is carried in the complex linear space $S(\mathbf{n})$ spanned by the basis $\mathfrak{B} = \{\boldsymbol{\beta}\}$, where $\boldsymbol{\beta} = \mathbf{i}_\alpha \times \exp[i\mathbf{n} \cdot \mathbf{r} + \epsilon^{-1}\mathbf{m} \cdot \mathbf{r}] \equiv \mathbf{i}_\alpha \beta(\mathbf{m})$, $\alpha = 1, 2, 3$, $\mathbf{m} \in K_0$; $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$ is orthonormal and K_0 is K plus the zero vector; $\mathbf{n}$ is some real nonzero 3-vector. The norm on $S(\mathbf{n})$ is defined by

$$\|\mathbf{h}\|_s = \sum_{K_0} |\boldsymbol{\Gamma}(\mathbf{m})| \quad \text{if} \quad \mathbf{h} = \sum_{K_0} \boldsymbol{\Gamma}(\mathbf{m}) \beta(\mathbf{m}). \tag{4.1}$$

The dynamo equations (2.1) may be written, in $S(\mathbf{n})$, in the form

$$\mathbf{h} = T_\infty \mathbf{h} \equiv -R\nabla^{-2}\nabla \times (\mathbf{v} \times \mathbf{h}), \tag{4.2}$$

where $\nabla^{-2}\beta(\mathbf{m}) = -|\mathbf{m}\epsilon^{-1} + \mathbf{n}|^{-2}\beta(\mathbf{m})$, and the subscript ∞ refers to the fact that the conductor is now infinite. We propose to solve (4.2) with a series in ϵ of the form

$$\mathbf{h} = \sum_{j=0}^{\infty} \epsilon^{j/J} \mathbf{h}_j(\mathbf{r}; \epsilon), \tag{4.3}$$

where J is a positive integer, and

$$\begin{aligned} \mathbf{h}_0 &= \boldsymbol{\Gamma}(\mathbf{n}; \epsilon) \exp(i\mathbf{n} \cdot \mathbf{r}), \quad \mathbf{h} \in S(\mathbf{n}), \\ \|\mathbf{h}_j\|_s &= O(1), \quad \text{as} \quad \epsilon \to 0, \end{aligned} \tag{4.4}$$

for $j = 1, 2, \cdots$. We also require

$$\lim_{\epsilon \to 0} \|\mathbf{h} - \mathbf{h}_0\|_s = 0, \quad R(\epsilon) = R_0 \epsilon^{1/J-1}, \tag{4.5}$$

for R_0 some positive constant, for some $J \geq 2$.

B. Analysis

We first seek to estimate the effect of T_∞ when the mean or average effect taken over the small $[O(\epsilon)]$ length scale is removed. Let the projection P be defined for $\boldsymbol{\beta} \in \mathfrak{B}$ by

$$P\boldsymbol{\beta} = 0, \quad \mathbf{m} \neq 0,$$
$$= \boldsymbol{\beta}, \quad \mathbf{m} = 0.$$

If $\tilde{T}_\infty = T_\infty - PT_\infty$, then it follows from (4.1), (4.2), and (4.5) that

$$\|\tilde{T}_\infty\|_s \leq R_0 \|\mathbf{v}\|_s \epsilon^{1/J}(1 - \epsilon n)^{-1} \tag{4.6}$$

and hence that $(I - \tilde{T}_\infty)^{-1}$ exists on PS if $\epsilon < \epsilon_0$, where

$$\epsilon_0 = \min \{(2n)^{-1}, (2 \|\mathbf{v}\|_s R_0)^{-J}\}. \tag{4.7}$$

The proof of (4.6) is straightforward and is omitted.

Suppose now that $\epsilon < \epsilon_0$ where ϵ_0 is defined by (4.7). Let $\mathbf{f}$ be some element of PS, and set $\mathbf{g} = (I - \tilde{T}_\infty)^{-1}\mathbf{f}$. Then

$$\mathbf{g} - T_\infty \mathbf{g} = \mathbf{f} - PT_\infty(I - \tilde{T}_\infty)^{-1}\mathbf{f},$$

so that $\mathbf{g}$ will solve (4.2)–(4.5) if and only if $\mathbf{f}$ solves

$$\mathbf{f} = PT_\infty(I - \tilde{T}_\infty)^{-1}\mathbf{f}, \tag{4.8}$$

in which case $\mathbf{f} = P\mathbf{g}$. In this way, solutions of the periodic dynamo problem may be generated (for ϵ sufficiently small) by solutions of the *compatibility* Eq. (4.8). A solution of (4.8) corresponding to a given $\mathbf{n}$ represents, moreover, a Fourier mode having $\mathbf{n}$ as its wavenumber vector. For small ϵ, it follows from (4.6) that this wave is the dominant part of $\mathbf{g}$. This separation of scales, wherein the dominant small-scale velocity field gives rise to a magnetic field with a dominant part varying on a scale of order unity, also occurred in the analog of Sec. 3, as is clear from (3.3) and (3.16).

To solve the compatibility equation, we write it in wavenumber vector form, assuming that $\mathbf{f} = \boldsymbol{\Gamma} \exp(i\mathbf{n} \cdot \mathbf{r})$,

$$i\mathbf{n} \times \boldsymbol{\Gamma} = i\mathbf{n}\Phi + \mathbf{A} \cdot \boldsymbol{\Gamma}, \quad i\mathbf{n} \cdot \boldsymbol{\Gamma} = 0, \tag{4.9}$$

where $\Phi = in^{-2}\mathbf{n} \cdot \mathbf{A} \cdot \Gamma$ and $\mathbf{A}$ is the 3×3 complex matrix defined by

$$\mathbf{A} \cdot \boldsymbol{\Gamma} = \sum_{j=2}^{\infty} R_0^j (i)^{j-1} \epsilon^{j/J-1} \sum_{(j)} \left[\frac{\boldsymbol{\mu}_j \times \mathbf{m}_{j-1} \times \cdots \times \boldsymbol{\Gamma}}{\mathbf{m}_{j-1}^2 \cdots \mathbf{m}_1^2} \right]$$
$$\equiv \sum_{j=2}^{\infty} \epsilon^{j/J-1} R_0^j \mathbf{A}_0^{(j)} \cdot \boldsymbol{\Gamma}. \tag{4.10}$$

In (4.10), $\mathbf{m}_j = \mathbf{k}_1 + \mathbf{k}_2 + \cdots + \mathbf{k}_j + \epsilon\mathbf{n}$, where $\mathbf{k}_j$ is an arbitrary vector from the jth copy of K and "j-summation" is defined to be summation over all elements $\mathbf{m}_j$ satisfying $\mathbf{k}_1 + \cdots + \mathbf{k}_i \neq 0$, $i = 1, 2, \cdots, j-1$, $\mathbf{k}_1 + \cdots + \mathbf{k}_j = 0$. We have also written $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$ for $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$. Note that in $\mathbf{A}^{(j)}(\epsilon)$, as defined by (4.10), ϵ and $\mathbf{n}$ occur only in the combination $\epsilon\mathbf{n}$. Now, if $\mathbf{A}$ is to have a bounded nonzero limit as $\epsilon \to 0$, then it is necessary and sufficient that J equal the index of the first nonzero matrix in the series $(\mathbf{A}^{(2)}(O), \mathbf{A}^{(3)}(O), \cdots)$. At this point, we adopt this particular value of J and assume that it is finite.

Elsewhere, we have shown that $\mathbf{A}^{(j)}$ is Hermitian (anti-Hermitian) if j is even (odd) and that $\mathbf{A}^{(j)}(O)$ is real.[8] Moreover, the dispersion relation for $\mathbf{n}$ following from (4.9) is given by

$$n^4 - \sum_{\alpha,\beta} a'_{\alpha\beta} n_\alpha n_\beta - in^2 \sum_{\alpha,\beta,\gamma=1}^{3} \epsilon_{\alpha\beta\gamma} a_{\alpha\beta} n_\gamma, \tag{4.11}$$

where $\mathbf{A} = \{a_{\alpha\beta}\}$ and $\mathbf{A}' = \{a'_{\alpha\beta}\}$ is the matrix of cofactors of $\mathbf{A}$. The reality of $\mathbf{n}$ when ϵ is small will therefore depend on J and the choice of the $\boldsymbol{\mu}(\mathbf{k})$ in (1.2).

C. Construction of Periodic Dynamos

Let $\mathbf{A}_0$ and $\mathbf{A}'_0$ denote the values of $\mathbf{A}$ and $\mathbf{A}'$, respectively, when $\epsilon = 0$. The field $\mathbf{v}$ will be said to be a *spatially periodic dynamo* if $\mathbf{A}$ is Hermitian and $\mathbf{A}'_0$ positive definite, the former uniformly for ϵ in an interval of the form $0 \leq \epsilon \leq \epsilon_1 \leq \epsilon_0$, $\epsilon_1 > 0$. Clearly, then, $\mathbf{v}$ is a spatially periodic dynamo if J is even, $\mathbf{A}'_0$ positive definite, and $\mathbf{A}^{(j)} = 0$ for all odd $j \geq 3$.

If $\mathbf{v}$ is a dynamo in this sense, it also has properties of physical interest. From (4.6) and (4.14) it can be shown that there will exist, for ϵ sufficiently small, a closed surface B swept out by real wavenumber vectors $\mathbf{n}$ satisfying (4.14). This surface has the property that

$$B \to B_0 \quad \text{as} \quad \epsilon \to 0, \tag{4.12}$$

where B_0 is defined by nonzero solutions of

$$n^4 - \mathbf{n} \cdot \mathbf{A}'_0 \cdot \mathbf{n} = 0. \tag{4.13}$$

Superposition of these plane-wave solutions gives the magnetic field

$$\mathbf{h} = \int_B (I - \tilde{T}_\infty)^{-1} \boldsymbol{\Gamma}(\mathbf{n}; \epsilon) \exp(i\mathbf{n} \cdot \mathbf{r})\, ds, \tag{4.14}$$

which solves (2.1) and tends to zero as $r \to \infty$.[10] Such fields are the natural eigensolutions for steady dynamos in an infinite conductor, in spite of the fact that the total magnetic energy of such a solution will not, in general, be finite (see Sec. 8). The above

conditions, isolating the spatially periodic dynamos from other admissible fields v, have been interpreted physically elsewhere.[6,8]

D. The Motions Q

The class Q of velocity fields covered by the existence theory of Secs. 5–7 can now be characterized by the admissible v in the following way: In (1.1), v is to be admissible in the sense of Sec. 4a; it is also to be a periodic dynamo in the sense of Sec. 4C, with $J = 2$ and $\mathbf{A}_0 = \mathbf{I}$. Finally, v is required to have the Beltrami property $\mathbf{v} \times (\nabla \times \mathbf{v}) = 0$. Without loss of generality, we may impose this last condition in the form[11]

$$\nabla \times \mathbf{v} = -\mathbf{v}. \tag{4.15}$$

This family of v can be studied explicitly by noting from (4.10) that, if (4.15) is satisfied,

$$\mathbf{A}_0^{(2)} = i \sum_K k^{-2}[(\boldsymbol{\mu}^* \times \boldsymbol{\mu}) \circ \mathbf{k}]. \tag{4.16}$$

However, in view of (4.15), we see that $|\boldsymbol{\mu}(\mathbf{k})| \neq 0$ implies $k = 1$ and that

$$\boldsymbol{\mu}(\mathbf{k}) = a[i\mathbf{k} \times \mathbf{c} - \mathbf{c}], \tag{4.17}$$

where a is a complex number and c is a real unit vector normal to k. Substituting (4.17) into (4.16), we obtain

$$\mathbf{A}_0^{(2)} = 2 \sum_K |a(\mathbf{k})|^2 \, \mathbf{k} \circ \mathbf{k}. \tag{4.18}$$

The remaining condition on v is, thus, that the mapping onto $A_0^{(2)}$, as given by (4.18), yields the identity matrix.

5. PERTURBATION OF H

The comparison eigenvalue problem (1.7) is defined for (2.1)–(2.4) as in (3.8), by setting

$$H\mathbf{h} \equiv \int_D \mathbf{K}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{h}(\mathbf{r}') \, d\mathbf{r}'. \tag{5.1}$$

A suitable linear space $\mathcal{H}$ for the operator H may be defined as follows: Let h be irrotational and $r^3 h$ bounded in V, solenoidal in $\mathcal{E}$, continuously differentiable in D and V separately, and continuous in $\mathcal{E}$. These functions constitute a space C_1, on which the inner product

$$(\mathbf{f}, \mathbf{g}) = \int_D (\nabla \times \mathbf{f}) \cdot (\nabla \times \mathbf{g}^*) \, d\mathbf{r} \tag{5.2}$$

may be defined. Completion of C_1 in the norm

$$\|\mathbf{h}\|_1 = (\mathbf{h}, \mathbf{h})^{\frac{1}{2}} = \|\nabla \times \mathbf{h}\| \tag{5.3}$$

provides the Hilbert space $\mathcal{H}$.[12] It can be shown[6] that an orthonormal basis in $\mathcal{H}$ is given by the set $\{\mathbf{h}_{mn}^{(k)}, k = \pm 1, \pm 2, \cdots; n = 1, 2, \cdots; m = -n, -n+1, \cdots, n-1, n\}$, where

$$\mathbf{h}_{mn}^{(k)} = \nabla \times \nabla \times \mathbf{r} P_{mn}^{(k)} + \nabla \times \mathbf{r} T_{mn}^{(k)}, \tag{5.4a}$$

$$T_{mn}^{(k)} = C_n^{(k)} j_n(\sigma_n^{(k)} r) Y_{mn}, \quad 0 \leq r \leq 1,$$
$$= 0, \quad r > 1, \tag{5.4b}$$

$$P_{mn}^{(k)} = (\sigma_n^{(k)})^{-1}(T_{mn}^{(k)} - \psi_{mn}^{(k)}), \quad 0 \leq r < 1, \tag{5.4c}$$
$$= -C_n^{(k)}(n+1)^{-1} j_n'(\sigma_n^{(k)}) r^{-(n+1)} Y_{mn}, \quad r > 1,$$

$$\psi_{mn}^{(k)} = C_n^{(k)}(n+1)^{-1} \sigma_n^{(k)} j_n'(\sigma_n^{(k)}) r^n Y_{mn}. \tag{5.4d}$$

Also,

$$(C_n^{(k)})^{-1} = [n(n+1)]^{\frac{1}{2}} \sigma_n^{(k)} j_{n+1}(\sigma_n^{(k)}), \tag{5.4e}$$

where $\sigma_n^{(k)}$ is a zero of the Bessel function $j_n(r)$ defined by

$$j_n(r) = \left(\frac{\pi}{2r}\right)^{\frac{1}{2}} J_{n+\frac{1}{2}}(r) = (-1)^n r^n \left(\frac{1}{r}\frac{d}{dr}\right)^n \frac{\sin r}{r}, \tag{5.4f}$$

and Y_{mn} is the surface harmonic

$$Y_{mn} = (-1)^m \left(\frac{(2n+1)}{4\pi}\right)^{\frac{1}{2}} \left(\frac{(n-m)!}{(n+m)!}\right)^{\frac{1}{2}} \times P_n^m(\cos\theta) \exp(im\phi). \tag{5.4g}$$

We also want to regard H as an operator on the Banach spaces $C_1^\gamma(\epsilon)$, $\epsilon, \gamma > 0$, obtained by equipping C_1 with a norm analogous to (3.10): If $\mathbf{T} = \{T_{ij\cdots k}\}$ is any tensor, we define

$$\|\mathbf{T}\|_\infty^\gamma = \|\mathbf{T}\|_\infty + \epsilon^\gamma \|\nabla \mathbf{T}\|_\infty, \tag{5.5a}$$

$$\|\mathbf{T}\|_\infty = \max_{i,j,\cdots,k} \left\{ \sup_D |T_{ij\cdots k}| \right\}. \tag{5.5b}$$

We now prove the next lemma.

Lemma 1:* H is bounded and self-adjoint on $\mathcal{H}$ and is uniformly compact on C_1^γ.

Proof.: To prove self-adjointness, we note that, if $\mathbf{f}, \mathbf{g} \in C_1$, there will exist scalar functions ϕ and ψ so that [cf. (5.2)]

$$(\mathbf{f}, H\mathbf{g}) - (H\mathbf{f}, \mathbf{g})$$
$$= \int_D (\nabla \times \mathbf{f} \cdot \nabla\phi - \nabla \times \mathbf{g} \cdot \nabla\psi + \nabla \cdot \mathbf{f} \times \mathbf{g}) \, d\mathbf{r}. \tag{5.6}$$

Since (2.5) is satisfied for all functions in C_1, the right-hand side of (5.6) reduces to a surface integral of the normal component of $\mathbf{f} \times \mathbf{g}$. But f and g are continuous on S and are irrotational in V, so that

$$(\mathbf{f}, H\mathbf{g}) - (H\mathbf{f}, \mathbf{g}) = \int_V \nabla \cdot (\mathbf{g} \times \mathbf{f}) = 0.$$

Boundedness follows from the inequality

$$\|H\mathbf{f}\|_1^2 = \int_D (\nabla\phi + \mathbf{f}) \cdot (\nabla \times H\mathbf{f})\, dr$$
$$\leq \|\mathbf{f}\|\, \|H\mathbf{f}\|_1$$
$$\leq \pi^{-1} \|\mathbf{f}\|_1 \|H\mathbf{f}\|_1.$$

In particular, $\|H\|_1 \leq \pi^{-1}$. The uniform compactness follows from the representation of the kernel $\mathbf{K}$ given in the Appendix [see (A10), (A13), and (A19)]. Thus, we have

$$\mathbf{K}(\mathbf{r}, \mathbf{r}') = \nabla \times \mathbf{A}(\mathbf{r}, \mathbf{r}'), \tag{5.7a}$$

where $\mathbf{A}$ is a tensor which is regular for $\mathbf{r}, \mathbf{r}' \in D$ except at $\mathbf{r} = \mathbf{r}'$, and

$$|A_{ij}| \leq a_{ij} |\mathbf{r} - \mathbf{r}'|^{-1}, \tag{5.7b}$$

$$\left|\frac{\partial A_{ij}}{\partial x_k}\right| \leq b_{ijk} |\mathbf{r} - \mathbf{r}'|^{-2}, \tag{5.7c}$$

for some constants a_{ij}, $b_{ij}k$. Using (5.7) and Gauss' theorem in (5.1), we may represent $H\mathbf{h}$, $\mathbf{h} \in C_1$, as a surface integral similar to a single layer potential, the distribution being continuously differentiable, together with a volume integral similar to a continuous volume distribution in potential theory. Known results in potential theory carry over, and it is found that a uniformly bounded sequence in C_1^γ is mapped under H into a sequence which is uniformly equicontinuous with respect to the same norm. Thus, H is uniformly compact on C_1^γ.

We can also establish the following versions of Lemmas 2 and 3 and Theorem 1 of Sec. 3.

Lemma 2:* The operator L, defined by (2.7), is uniformly bounded on $C_1^{\frac{1}{2}}(\epsilon)$.

Lemma 3:* If $\mathbf{h} \in C_1$ and if α is any number between $\frac{3}{4}$ and 1, then

$$\|V\mathbf{h}\|_\infty = O(\epsilon^{1-\alpha} \|\mathbf{h}\|_\infty) + O(\epsilon^{2\alpha-1} \|\nabla\mathbf{h}\|_\infty) + O(\epsilon^{\beta+\frac{1}{2}} \|\mathbf{w}\|_\infty \|\nabla\mathbf{h}\|_\infty), \tag{5.8}$$

as $\epsilon \to 0$, where $V = L^2 - H$ and $\mathbf{q}$ is of the form (1.1) with $\beta > 0$. As in the model problem, V fails to be uniformly small on C_1^γ but, again using modified operators, we can obtain:

Theorem 1:* There is a family of operators $\tilde{H}(c)$ on C_1 with the following properties: (i) If $c \neq 0$, then $\tilde{H}(c) - H(c)$ is uniformly small on C_1^γ for some γ between $\frac{1}{2}$ and 1; (ii) if $\tilde{H}(\nu)\mathbf{h} = \nu\mathbf{h}$ for some $\mathbf{h}(\mathbf{r}; \epsilon) \in C_1$ and $\nu(\epsilon) \geq \text{const} > 0$, uniformly for $0 < \epsilon \leq \epsilon_1$, $\epsilon_1 > 0$, then $L^2\mathbf{h} = \nu\mathbf{h}$ for ϵ sufficiently small.

We give the proofs of these results in the next two sections.

Using Theorem 1*, we now solve the eigenvalue problem (1.6). Let $\tilde{V}(c) = \tilde{H} - H$ and formally set

$$T = (H - \nu)^{-1}\tilde{V}, \tag{5.9}$$

so that the modified eigenvalue problem becomes $\mathbf{h} = T\mathbf{h}$. Let μ_k be any positive eigenvalue of H with an associated finite set $\{\mathbf{f}_i, i = 1, 2, \ldots, m\}$ of eigenfunctions belonging to C_1 and orthonormal in $\mathcal{K}$. We define the projections P_i and P by

$$P_i\mathbf{h} = \mathbf{f}_i \int_D (\nabla \times \mathbf{f}_i) \cdot \mathbf{h}\, d\mathbf{r}, \tag{5.10a}$$

$$P = \sum_{i=1}^m P_i, \tag{5.10b}$$

and set $\tilde{T} = T - PT$. Suppose that $\mathbf{f}$ is of the form

$$\mathbf{f} = \sum_{i=1}^m \alpha_i(\epsilon, c)\mathbf{f}_i; \tag{5.11}$$

then $\mathbf{g}(\epsilon, c) = (I - \tilde{T})^{-1}\mathbf{f}$ solves the modified problem provided that

$$\alpha_i\mathbf{f}_i = P_iT(I - \tilde{T})^{-1}\mathbf{f}, \quad i = 1, 2, \cdots, m, \tag{5.12}$$

is satisfied by $\mathbf{f}$. This system is the compatibility equation for the bounded conductor [cf. (4.8)]. We rewrite (5.12) in the equivalent form

$$(\mu_k - \nu)\alpha_i = \sum_{i=1}^m R_{ij}\alpha_j, \quad i = 1, 2, \cdots, m, \tag{5.13a}$$

$$R_{ij} = \int_D \mathbf{f}_i \cdot [\tilde{V}(I - \tilde{T})^{-1}\mathbf{f}_j]\, d\mathbf{r}. \tag{5.13b}$$

Then we see that (5.12) has a nontrivial solution if and only if

$$\det\{(\nu - \mu_k)\delta_{ij} + R_{ij}\} \equiv (\nu - \mu_k)^m + \mathcal{R}(\nu, \epsilon, c) = 0. \tag{5.14}$$

If $\mathcal{R} = o(1)$ as $\epsilon \to 0$ and is continuous with respect to ϵ and c in some interval of the form

$$-\delta < \nu - \mu_k < +\delta, \quad \delta > 0, \quad 0 < \epsilon \leq \epsilon_1, \quad \epsilon_1 \leq 0,$$
$$c_1 < c < c_2, \quad c_1 < \mu_k, \quad c_2 > \mu_k$$

and if m is *odd*, then (5.14), with $\mathcal{R}(\nu, \epsilon, \nu)$ replacing $\mathcal{R}(\nu, \epsilon, c)$, has at least one solution $\nu(\epsilon)$ for ϵ sufficiently small.

In order to solve the equation $\mathbf{h} = T\mathbf{h}$ for $c = \nu$, it is therefore sufficient to show that $(I - \tilde{T})^{-1}$ exists and that the required conditions on $\mathcal{R}$ are satisfied. We first choose δ, in the above interval, to be positive and less than μ_k, so that, among the eigenvalues of H, $\Delta = [\mu_k - \delta, \mu_k + \delta]$ contains only μ_k. Thus, if $\nu \in \Delta$, then $(H - \nu)^{-1}$ is bounded on the subspace $(I - P)C_1$. Indeed, if this were not the

case, then, from Lemma 1*, this subspace would contain an eigenfunction of H belonging to C_1 and hence to $\mathcal{K}$, which is impossible because of the choice of Δ and the definition[13] of P. Thus, for $\nu \neq \mu_k$, we have an estimate of the form

$$\|\tilde{T}\|_\infty^\gamma \leq C \|\tilde{V}\|_\infty^\gamma, \tag{5.15}$$

where C is a constant independent of ν and ϵ. Letting $\nu \to \mu_k$ in (5.15), we obtain a uniform estimate on T for all $\nu \in \Delta$. By Theorem 1*, it then follows that $\tilde{T}$ is uniformly small on C_1^γ. Finally, it is apparent from the definition of $\tilde{T}$ that $\mathcal{R}$ is actually a differentiable function of ν and c on an interval of the above form. By Theorem 1* again, it then follows that $\mathbf{g}(\epsilon, \lambda^2)$ satisfies $L^2\mathbf{h} = \lambda^2\mathbf{h}$ for ϵ sufficiently small, so that one of the two functions $(L \pm \lambda)\mathbf{g}(\epsilon, \lambda^2)$ is an eigenfunction of L corresponding to the real eigenvalue $\pm\lambda$. It remains to be checked that H has an eigenvalue μ_k with the requisite properties. However, from (5.4), we see that the eigenvalues $\{(\sigma_n^{(k)})^{-1}\}$ have multiplicity $2n + 1$, with associated eigenfunctions $\{h_{mn}^{(k)}, m = -n, \cdots, +n\}$ belonging to C_1. Thus, we have proved

Theorem 2:* If $\beta > 0$ and ϵ is sufficiently small, then the motions (1.1) with $\mathbf{q} \in Q$ are steady-state kinematic dynamos. This is the desired existence theorem and the main result of this paper.

6. PRELIMINARY ESTIMATES

In this section, we shall prove Lemmas 2* and 3*. The first of these is a consequence of estimates on L obtained by dividing the domain of integration into two parts, so that global and local estimates (the latter associated with the singularity for $\mathbf{r} = \mathbf{r}'$) may be treated separately. Let D_α be the intersection of the interior of a sphere of radius ϵ^α and center at $\mathbf{r}$ with D. Here $\mathbf{r}$ is any point in $\bar{D}$, and $0 < \alpha \leq 1$. Also, let $D_\alpha' = D - \bar{D}_\alpha$ and $\partial D_\alpha = \bar{D}_\alpha - D_\alpha$. Then we write

$$\begin{aligned} L\mathbf{h} &= \epsilon^{-\frac{1}{2}} \int_{D_1} \mathbf{K} \cdot (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' + \epsilon^{-\frac{1}{2}} \int_{D_1'} \mathbf{K} \cdot (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' \\ &\equiv I_1 + I_2, \end{aligned} \tag{6.1}$$

where here and below the I_i are defined, in order, by the terms on the right. In treating I_1, we use the decomposition (A10) derived in the Appendix and note from (A13) and (A19) that

$$|K_{ij}^*| \leq \alpha_{ij}(1 - rr')^{-2}, \tag{6.2a}$$

$$\left|\frac{\partial K_{ij}^*}{\partial x_k}\right| \leq \beta_{ijk}(1 - rr')^{-3} \tag{6.2b}$$

for positive constants α_{ij} and β_{ijk}. With

$$\begin{aligned} I_1 &= \frac{\epsilon^{-\frac{1}{2}}}{4\pi} \int_{D_1} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' \\ &\quad + \epsilon^{-\frac{1}{2}} \int_{D_1} \mathbf{K}^* \cdot (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' \\ &\equiv I_{11} + I_{12}, \end{aligned} \tag{6.3}$$

we then obtain, using (6.2) and (1.1),

$$\begin{aligned} \|I_{11}\|_\infty = O(\epsilon^{\frac{1}{2}} \|\mathbf{h}\|_\infty), \quad \|I_{12}\|_\infty = O(\epsilon^{\frac{1}{2}} \|\mathbf{h}\|_\infty), \\ \|\nabla I_{12}\|_\infty = O(\epsilon^{-\frac{1}{2}} \|\mathbf{h}\|_\infty), \end{aligned} \tag{6.4}$$

as $\epsilon \to 0$.[14] To obtain an estimate on ∇I_{11}, we consider the identity

$$\begin{aligned} &\int_{D_1} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' \\ &= \int_{D_1} |\mathbf{r} - \mathbf{r}'|^{-1} \nabla' \times (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' + \int_{\partial D_1} (\mathbf{q} \times \mathbf{h}) \times d\mathbf{s}'. \end{aligned} \tag{6.5}$$

Using (6.5) in (6.3), we obtain

$$\|\nabla I_{11}\|_\infty = O(\epsilon^{-\frac{1}{2}} \|\mathbf{h}\|_\infty) = O(\epsilon^{\frac{1}{2}} \|\nabla\mathbf{h}\|_\infty). \tag{6.6}$$

To treat I_2 in (6.1), we use (1.1) and integrate by parts to obtain

$$\begin{aligned} &\epsilon^{-\frac{1}{2}} \int_{D_1'} \mathbf{K} \cdot (\mathbf{q} \times \mathbf{h})\, d\mathbf{r}' \\ &= \epsilon^{\frac{1}{2}} \int_{\partial D_1} \omega[\mathbf{h} \cdot \mathbf{v}\mathbf{K} - \mathbf{K} \cdot \mathbf{v}\mathbf{h} - \mathbf{K} \cdot \mathbf{h}\mathbf{v}] \cdot d\mathbf{s}' \\ &\quad + \epsilon^{\frac{1}{2}} \int_{D_1'} \omega[(\mathbf{v} \cdot \nabla'\mathbf{K}) \cdot \mathbf{h} \\ &\quad + (\mathbf{h} \cdot \nabla'\mathbf{K}) \cdot \mathbf{v} - \mathbf{h} \cdot \mathbf{v}\nabla' \cdot \mathbf{K}]\, d\mathbf{r}' \\ &\quad - \epsilon^{\frac{1}{2}} \int_{D_1'} \mathbf{K} \cdot \omega\mathbf{v} \times (\nabla' \times \mathbf{h})\, d\mathbf{r}' \\ &\quad + \epsilon^{\frac{1}{2}} \int_{D_1'} \mathbf{K} \cdot \mathbf{h}(\mathbf{v} \cdot \nabla'\omega)\, d\mathbf{r}' + \epsilon^{\beta} \int_{D_1'} \mathbf{K} \cdot (\mathbf{w} \times \mathbf{h})\, d\mathbf{r}' \\ &\equiv \sum_{j=1}^{5} I_{2j}. \end{aligned} \tag{6.7}$$

If (5.7) is used to estimate $\mathbf{K}$ and $\nabla\mathbf{K}$, we have

$$\begin{aligned} \|I_{21}\|_\infty &= O(\epsilon^{\frac{1}{2}} \|\mathbf{h}\|_\infty), \\ \|\nabla I_{21}\|_\infty &= O(\epsilon^{\frac{1}{2}} \log \epsilon^{\frac{1}{2}} \|\mathbf{h}\|_\infty), \\ \|I_{22}\|_\infty &= O(\epsilon^{\frac{1}{2}} \log \epsilon \|\mathbf{h}\|_\infty), \\ \|\nabla I_{22}\|_\infty &= O(\epsilon^{-\frac{1}{2}} \|\mathbf{h}\|_\infty), \\ \|I_{23}\|_\infty &= O(\epsilon^{\frac{1}{2}} \|\nabla\mathbf{h}\|_\infty), \\ \|\nabla I_{23}\|_\infty &= O(\epsilon^{\frac{1}{2}} \log \epsilon \|\nabla\mathbf{h}\|_\infty), \\ \|I_{24}\|_\infty &= O(\epsilon^{\frac{1}{2}} \|\mathbf{h}\|_\infty), \\ \|\nabla I_{24}\|_\infty &= O(\epsilon^{\frac{1}{2}} \log \epsilon \|\mathbf{h}\|_\infty), \\ \|I_{25}\|_\infty &= O(\epsilon^{\beta} \|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty), \\ \|\nabla I_{25}\|_\infty &= O(\epsilon^{\beta} \log \epsilon \|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty). \end{aligned} \tag{6.8}$$

Combining (6.4), (6.6), and (6.8), we then obtain

$$\|L\mathbf{h}\|_\infty = O(\epsilon^{\frac{1}{2}} \log \epsilon \, \|\mathbf{h}\|) + O(\epsilon^{\frac{1}{2}} \|\nabla \mathbf{h}\|) + O(\epsilon^{\beta} \|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty), \tag{6.9}$$

$$\|\nabla L\mathbf{h}\| = O(\epsilon^{-\frac{1}{2}} \|\mathbf{h}\|) + O(\epsilon^{\frac{1}{2}} \log \epsilon \, \|\nabla \mathbf{h}\|), \tag{6.10}$$

which imply, in view of (5.5), Lemma 2*.

To obtain Lemma 3*, we replace D_1 by D_α in (6.1) and retain α as a parameter. This change is indicated by adding the superscript (α) to the I_i. It is not difficult to see that this does not change the estimates (6.8) on I_{2i}, $i = 1, \cdots, 5$, except for the estimate on ∇I_{22}. This must be replaced by

$$\|\nabla I_{22}^{(\alpha)}\|_\infty = O(\epsilon^{\frac{1}{2}-\alpha} \|\mathbf{h}\|_\infty). \tag{6.11}$$

Using (6.8), (6.9), and (6.11), we then have

$$\|LI_2\|_\infty = O(\epsilon^{1-\alpha} \|\mathbf{h}\|_\infty) + O(\epsilon \log \epsilon \, \|\nabla \mathbf{h}\|_\infty) + O(\epsilon^{\beta+\frac{1}{2}} \|\mathbf{w}\|_\infty \|\nabla \mathbf{h}\|_\infty), \tag{6.12}$$

when $\beta > 0$ and $\frac{1}{2} \leq \alpha \leq 1$. For the integrals over D, we have, clearly,

$$\|I_{11}^{(\alpha)}\|_\infty = O(\epsilon^{\alpha-\frac{1}{2}} \|\mathbf{h}\|_\infty), \tag{6.13}$$

$$\|I_{12}^{(\alpha)}\|_\infty = O(\epsilon^{\alpha-\frac{1}{2}} \|\mathbf{h}\|_\infty), \tag{6.14a}$$

$$\|\nabla I_{12}^{(\alpha)}\|_\infty = O(\epsilon^{-\frac{1}{2}} \|\mathbf{h}\|_\infty). \tag{6.14b}$$

However, (6.14b) must be strengthened in order to obtain a suitable estimate of $LI_{12}^{(\alpha)}$. Specifically, if $\partial D_\alpha \cap S = 0$, then $\mathbf{K}$ is regular over the domain of integration in I_{12}, and the estimates (6.14) come from a spherical shell of thickness $2\epsilon^\alpha$ adjacent to S. Moreover, if r is constrained to lie within this shell in obtaining the sup with respect to $\mathbf{r}$, these estimates drop to $O(\epsilon^{3\alpha-\frac{1}{2}} \|\mathbf{h}\|_\infty)$. We can therefore refine (6.14) to obtain, when used in conjunction with (6.9),

$$\|LI_{12}^{(\alpha)}\|_\infty = O(\epsilon^{\alpha} \log \epsilon \, \|\mathbf{h}\|_\infty) + O(\epsilon^{\beta+2\alpha-\frac{1}{2}} \|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty). \tag{6.15}$$

In order to study $LI_{11}^{(\alpha)}$, we substitute for $\mathbf{h}(\mathbf{r}')$ the expression

$$\mathbf{h}(\mathbf{r}') = \mathbf{h}(\mathbf{r}) + (\mathbf{r}' - \mathbf{r}) \cdot \nabla \mathbf{h}(\mathbf{r}_0), \tag{6.16}$$

where $\mathbf{r}_0$ is some point on the line segment between $\mathbf{r}'$ and $\mathbf{r}$ determined (since $\mathbf{r}$ is fixed) by $\mathbf{r}'$. The effect of the two terms in (6.16), taken in order, is to decompose $I_{11}^{(\alpha)}$ into two parts,

$$I_{11}^{(\alpha)} = I_{111}^{(\alpha)} + I_{112}^{(\alpha)},$$

where, clearly,

$$\|I_{112}^{(\alpha)}\|_\infty = O(\epsilon^{2\alpha-1} \|\mathbf{h}\|_\infty). \tag{6.17}$$

Now, consider $I_{111}^{(\alpha)}$, with $\mathbf{q}$ given by (1.1). In the computation which follows, we take $\mathbf{v}(\mathbf{r})$ to have the simple form

$$\mathbf{v}(\mathbf{r}) = (\sin y + \cos z, \sin z + \cos x, \sin x + \cos y). \tag{6.18}$$

However, the final estimate [Eq. (6.27) below] is valid for an arbitrary element of Q. We write

$$I_{111}^{(\alpha)} = \sum_{i=1}^{5} \Omega_i,$$

and study $L\Omega_i$, $i = 1, \cdots, 5$, in turn. Here

$$\Omega_1 \equiv \frac{-\epsilon^{\frac{1}{2}}}{4\pi} \int_{D_\alpha} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times \left\{ \left[\nabla' \omega(r') \times \mathbf{v}\left(\frac{\mathbf{r}'}{\epsilon}\right) \right] \times \mathbf{h}(\mathbf{r}) \right\} d\mathbf{r}'$$

vanishes identically if the distance between $\mathbf{r}$ and S exceeds $\epsilon + \epsilon^\alpha$. Otherwise, its components are each $O(\epsilon^{\alpha-\frac{1}{2}} \|\mathbf{h}\|_\infty)$. Thus,

$$\|L\Omega_1\|_\infty = O(\epsilon^{2\alpha-1} \|\mathbf{h}\|_\infty). \tag{6.19}$$

Next, we set

$$\Omega_2 \equiv \frac{\epsilon^{-\frac{1}{2}}}{\pi} \int_{D_\alpha} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times \left[[\omega(r') - 1] \mathbf{v}\left(\frac{\mathbf{r}'}{\epsilon}\right) \times \mathbf{h}(\mathbf{r}) \right] d\mathbf{r}'$$

and observe that

$$\|L\Omega_2\|_\infty = O(\epsilon^{\alpha} \|\mathbf{h}\|_\infty), \tag{6.20}$$

by a similar argument. Setting

$$\frac{\epsilon^{-\frac{1}{2}}}{4\pi} \int_{D_\alpha} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times [\mathbf{v}(\mathbf{r}'/\epsilon) \times \mathbf{h}(\mathbf{r})] \, d\mathbf{r}' \equiv \Omega_3 + \Omega_4,$$

where

$$\Omega_3 = \frac{\epsilon^{-\frac{1}{2}}}{4\pi} \int_{\rho \leq \epsilon^\alpha} \rho^{-3} \boldsymbol{\rho} \times \left[\mathbf{v}\left(\frac{\mathbf{r}}{\epsilon} + \frac{\boldsymbol{\rho}}{\epsilon}\right) \times \mathbf{h}(\mathbf{r}) \right] d\boldsymbol{\rho}. \tag{6.21}$$

We may estimate $L\Omega_4$ in the same way, giving

$$\|L\Omega_4\|_\infty = O(\epsilon^{2\alpha-1} \|h\|_\infty). \tag{6.22}$$

Lastly, we have

$$\Omega_5 = \frac{\epsilon^{\beta}}{4\pi} \int_{D_\alpha} \nabla |\mathbf{r} - \mathbf{r}'|^{-1} \times [\mathbf{w}(\mathbf{r}') \times \mathbf{h}(\mathbf{r}')] \, d\mathbf{r}'$$

and the obvious estimate

$$\|L\Omega_5\| = O(\epsilon^{\beta+\alpha-\frac{1}{2}} \|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty). \tag{6.23}$$

Then, considering (6.21) with $\mathbf{v}$ given by (6.18), we obtain by an explicit calculation

$$\epsilon^{-\frac{1}{2}} \mathbf{v}\left(\frac{\mathbf{r}}{\epsilon}\right) \times \Omega_3 = \mathbf{h} - \eta(\epsilon)\mathbf{h} + \mathcal{F}, \tag{6.24}$$

where

$$\eta(\epsilon) = \epsilon^{1-\alpha} \sin \epsilon^{\alpha-1}$$

and $\mathcal{F}$ is a finite linear combination of terms of the form

$$\mathbf{i}_\alpha h_\beta(\mathbf{r})[\sin(\mathbf{k}\cdot\mathbf{r}/\epsilon)],\ \cos(\mathbf{k}\cdot\mathbf{r}/\epsilon)],\ k_\alpha \text{ integral},\ \mathbf{k}\neq 0.$$

Thus, the method used to derive (6.9) may be repeated to obtain

$$\|H\omega\mathcal{F}\|_\infty = O(\epsilon \log \epsilon \|\mathbf{h}\|_\infty) + O(\epsilon \|\nabla\mathbf{h}\|_\infty). \quad (6.25)$$

Combining (6.24) and (6.25), we see that

$$\begin{aligned}\|L\Omega_3 - Hh\|_\infty &\le \|H(\omega-1)\mathbf{h}\|_\infty \|H\omega\eta\mathbf{h}\|_\infty \\ &\quad + \|H\omega\mathcal{F}\|_\infty + O(\epsilon^{\beta-\frac{1}{2}+\alpha}\|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty) \\ &= O(\epsilon^{1-\alpha}\|\mathbf{h}\|_\infty) + O(\epsilon^{\beta-\frac{1}{2}+\alpha}\|\mathbf{w}\|_\infty \|\mathbf{h}\|_\infty) \\ &\quad + O(\epsilon \|\nabla\mathbf{h}\|_\infty). \quad (6.26)\end{aligned}$$

Now, using (6.19), (6.20), (6.22), and (6.23) together with (6.26), we obtain

$$\|LI_{111}^{(\alpha)} - H\mathbf{h}\|_\infty = O(\epsilon^{1-\alpha}\|\mathbf{h}\|_\infty) + O(\epsilon\|\nabla\mathbf{h}\|_\infty), \quad (6.27)$$

provided that $\frac{3}{4} < \alpha \le 1$. Combining (6.12), (6.15), (6.17), and (6.27), we obtain Lemma 3*.

7. ESTIMATES IN THE MODIFIED EIGENVALUE PROBLEM

In this section, we establish Theorem 1* and thus complete the proof of Theorem 2*, as outlined in Sec. 5. The problem we treat here also arose in the 1-dimensional model and was discussed in Secs. 2C and 2D. We first seek an estimate for $\|V\|_\infty^\gamma$ with $\gamma > 0$, using those derived in the preceding section. From (6.3) and (6.10), we can obtain

$$\begin{aligned}\|\nabla LI_1^{(\alpha)}\|_\infty &= O(\epsilon^{\alpha-1}\log\epsilon\|\mathbf{h}\|_\infty) + O(\epsilon^\alpha \log\epsilon\|\nabla\mathbf{h}\|_\infty) \\ &\quad + O(\epsilon^{\alpha+\beta-\frac{1}{2}}\log\epsilon\|\mathbf{w}\|_\infty\|\mathbf{h}\|_\infty), \quad (7.1)\end{aligned}$$

and, from (6.8) and (6.10), we have

$$\begin{aligned}\|\nabla L(I_2^{(\alpha)} - I_{23}^{(\alpha)})\|_\infty \\ = O(\log\epsilon\|\mathbf{h}\|_\infty) + O(\epsilon^{\beta-\frac{1}{2}}\|\mathbf{w}\|_\infty\|\mathbf{h}\|_\infty). \quad (7.2)\end{aligned}$$

In conjunction with Lemma 3*, (7.1) and (7.2) imply

$$\begin{aligned}\|L^2 - H - LI_{23}^{(\alpha)}\|_\infty^\gamma \\ = O(\epsilon^{1-\alpha}) + O(\epsilon^{2\alpha-\gamma-1}) + O(\|\mathbf{w}\|_\infty \epsilon^{\beta+\frac{1}{2}-\gamma}) \quad (7.3)\end{aligned}$$

provided that $\frac{3}{4} < \alpha \le 1$ and $\frac{1}{2} < \gamma \le 1$. Thus, if $\frac{3}{4} < \alpha < 1$ and $\frac{1}{2} < \gamma < 2\alpha - 1$ ($\|\mathbf{w}\|_\infty = 0$) or if $\frac{1}{2} < \gamma < \min[2\alpha - 1, \beta + \frac{1}{2}]$ ($\|\mathbf{w}\|_\infty \neq 0$), the right-hand side of (7.3) is $o(1)$ as $\epsilon \to 0$. However, (6.8) and (6.10) can only provide the estimate

$$\|LI_{23}^{(\alpha)}\|_\infty^\gamma = O(1),$$

so that Lemma 3* cannot be extended to obtain uniform smallness of V on C_1^γ.

We are therefore led to define the family of modified operators $\tilde{H}(c)$ for $c \neq 0$ by

$$\begin{aligned}\tilde{H}(c)\mathbf{h} &\equiv L^2\mathbf{h} + c^{-1}LI_{23}^{(\alpha)}(L^2\mathbf{h} - c\mathbf{h}) \\ &= (L^2 - LI_{23}^{(\alpha)})\mathbf{h} + c^{-1}LI_{23}^{(\alpha)}(L^2\mathbf{h}), \quad (7.4)\end{aligned}$$

with

$$I_{23}^{(\alpha)}(\mathbf{h}) = -\epsilon^{\frac{1}{2}}\int_{D_{\alpha'}} \mathbf{K}\cdot\omega\mathbf{v}\times(\nabla'\times\mathbf{h})\,d\mathbf{r}'$$

following from (6.7). Using (6.9) and (6.10) repeatedly, we then obtain

$$\begin{aligned}\|LI_{23}^{(\alpha)}(L^2\mathbf{h})\|_\infty^\gamma &= O(\epsilon^{1-\gamma}\log\epsilon\|\mathbf{h}\|_\infty^\gamma) \\ &\quad + O(\epsilon^{\beta+\frac{1}{2}-\gamma}\|\mathbf{w}\|_\infty\|\mathbf{h}\|_\infty^\gamma). \quad (7.5)\end{aligned}$$

Combining (7.3)–(7.5) with $c \neq 0$, we obtain part (i) of Theorem 1*.

Now, suppose that $\tilde{H}(\nu)\mathbf{h} = \nu\mathbf{h}$ with $\mathbf{h} \in C_1$ and ν positive, uniformly as $\epsilon \to 0$. From (7.4), we then see that $\mathbf{h}' \equiv (L^2 - \nu)\mathbf{h}$ satisfies

$$LI_{23}^{(\alpha)}(\mathbf{h}') = -\nu\mathbf{h}'. \quad (7.6)$$

Now, if part (i) of Theorem 1* is used, we see that this $\mathbf{h}$ has the property that $\|\mathbf{h} - \mathbf{f}_0\|_\infty^\gamma \to 0$ as $\epsilon \to 0$ for γ in some positive open interval chosen as described above, where $\mathbf{f}_0$ is a solution of (1.7) with $\mu = \mu_k$. In particular, $\|\mathbf{h}\|_\infty^\gamma = O(1)$ for γ in this interval. We also find, using (6.9) and (6.10), that $\|L^2\|_\infty^\gamma = O(1)$ for these γ. Thus, we may assume that $\|\mathbf{h}'\|_\infty^\gamma = 1$ for all γ in this interval which exceed some value γ'. However, applying (6.8) and (6.9) to (7.6), it follows that

$$\nu\|\mathbf{h}'\|_\infty = O(\epsilon^{1-\gamma}\log\epsilon\|\mathbf{h}'\|_\infty^\gamma),$$

for some γ less than 1. Thus, we have a contradiction unless $\mathbf{h}' = 0$. This proves part (ii) of Theorem 1*.

8. REMARKS

For periodic motions of the kind used to construct Q (cf. Sec. 4D), the "unperturbed" ($\epsilon \equiv 0$) magnetic fields can be obtained by solving (4.9) with $\mathbf{A} = \mathbf{A}_0^{(2)}$. An example is

$$\mathbf{h}_0 = n\nabla\times\mathbf{a}r^{-1}\sin nr + \nabla\times\nabla\times\mathbf{a}r^{-1}\sin nr, \quad (8.1)$$

where $\mathbf{a}$ is an arbitrary constant vector. We see from (8.1) that, in general, the induced fields (4.12), while vanishing at infinity, do not have a bounded total magnetic energy. Nevertheless, (8.1) is similar to the eigenfunctions of the comparison problem (1.7). The largest positive eigenvalue μ_1 is $(20.19)^{-\frac{1}{2}}$,

approximately, and it is associated with three linearly independent dipolelike eigensolutions:

$$\begin{aligned} \mathbf{h}_0 &= (k^2\psi - 2)\mathbf{a} + \nabla\mathbf{a}\cdot\nabla\psi - k\mathbf{a}\times\nabla\psi, \\ \phi_0 &= 2k\mathbf{a}\cdot\mathbf{r}, \quad \mathbf{r}\in D, \\ \psi &= \frac{3}{k^2\sin k}\frac{\sin kr}{r}, \quad k = (\mu_1)^{-1}. \end{aligned} \tag{8.2}$$

It is, therefore, of interest to attempt to apply the present model to the geomagnetic dynamo problem.

In several respects, the motions (1.1) are physically compatible with a rotating fluid system. Condition (4.15) not only makes $\mathbf{v}$ into a force-free perfect fluid motion, but also introduces a preferred parity, which could be induced, in practice, by the direction of rotation. Moreover, the second term in (1.1) accounts for possible weak large-scale (geostrophic) motion, which can be expected to appear in most contained rotating fluid systems. The model can, however, be criticized in several respects. In the first place, there is no *a priori* reason to suppose that small-scale components of the core motion should dominate over large-scale ones, which is the case in (1.1) when $\epsilon \ll 1$. Indeed, in most physical theories of the geomagnetic dynamo, a typical dynamo speed is usually associated (in a spherical core) with an axisymmetric-toroidal shearing motion.[4] Such motions actually cooperate in the achieving the dynamo effect in models proposed by Parker[4] and Braginskii,[15] while they are entirely subsidiary to it in (1.1). In the second place, Parker has also suggested that the preferred sense of "vortices," as determined by the sign on the right of (4.15), should be different in northern and southern hemispheres, so that, if this is a necessary property of a realistic dynamo the model, (1.1) with (4.15) can only be expected to apply to the dynamo process within one hemisphere. Finally, the ordering of the magnetic Reynolds number $R = R_0\epsilon^{-\frac{1}{2}}$ utilized in our proof implies that the magnetic Reynolds number based on the characteristic length of the small-scale motion has been made small in order to achieve the dynamo effect. This is by no means a necessary property of kinematic dynamos, steady or nonsteady, nor is it an obvious feature of the geomagnetic dynamo. If local magnetic Reynolds numbers are taken to be *large*, there are severe analytic difficulties connected with the calculation of the small-scale magnetic field in a steady model, and, in this case, a scale separation based upon time appears to be more promising. However, in order to extend the present theory in any of these directions, it appears to be necessary to consider a comparison eigenvalue problem which is more complicated than (1.7). It is, in fact, the search for a simple self-adjoint comparison problem which originally led to the choice of (1.1).

ACKNOWLEDGMENTS

The author has benefited at various times throughout the course of this research from the helpful criticism and advice of Dr. J. B. Keller. Dr. H. Grad guided the present study in its early stages, and Dr. P. H. Roberts has, on a number of occasions, aided the author in the understanding of various parts of dynamo theory. The major portion of the work was sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grants AF-AFOSR-67-0124 and 67-815. A portion of the work was done while the author was associated with the Institut Henri Poincaré, University of Paris. The assistance at that time of Dr. P. Germain is gratefully acknowledged.

APPENDIX: CONSTRUCTION OF K AND E FOR THE UNIT SPHERE

Since the formulation (2.7) and (2.8) of the steady kinematic problem has not, to our knowledge, appeared before in the literature on dynamo theory, we give here explicit formulas for the Green's tensors $\mathbf{K} = \{K_{kl}\}$ and $\mathbf{E} = \{E_{kl}\} = \{\partial\Phi_k/\partial x_l\}$. In component form, the equations are

$$\epsilon_{ijk}\frac{\partial K_{kl}}{\partial x_j} = \frac{\partial\Phi_l}{\partial x_i} + \delta_{il}\delta(\mathbf{r}-\mathbf{r}'), \quad \mathbf{r},\mathbf{r}'\in D,$$
$$= 0, \quad \mathbf{r}\in V, \tag{A1}$$
$$\nabla^2\Phi_l = 0, \quad \mathbf{r}\in V, \tag{A2}$$
$$\frac{\partial K_{kl}}{\partial x_k} = 0, \tag{A3}$$
$$\Phi_l, K_{kl} \text{ continuous on } r = 1, \tag{A4}$$
$$r\Phi_l, r^3K_{kl} \text{ bounded in } V. \tag{A5}$$

To solve these, we note first that (A1) and (A2) imply

$$\nabla^2\Phi_l = \frac{\partial}{\partial x_l}\delta(\mathbf{r}-\mathbf{r}') = -\frac{\partial}{\partial x_l'}\delta(\mathbf{r}-\mathbf{r}') \tag{A6}$$

and that (2.5) provides the condition

$$\frac{\partial\Phi_l}{\partial r} = 0 \quad \text{on} \quad r = 1- \tag{A7}$$

From (A6) and (A7), we see that

$$E_{kl} = -\frac{\partial^2}{\partial x_k\partial x_l'}N(\mathbf{r},\mathbf{r}'), \tag{A8}$$

where N is a Neumann function for Laplace's equation

in the unit sphere,

$$N(\mathbf{r}, \mathbf{r}') = (4\pi)^{-1}[|\mathbf{r} - \mathbf{r}'|^{-1} + R^{-1} - \log(1 - \mathbf{r}\cdot\mathbf{r}' + R) + \log 2], \quad \text{(A9)}$$

where $R = [r^2r'^2 - 2\mathbf{r}\cdot\mathbf{r}' + 1]^{\frac{1}{2}}$.

Now, we consider $\mathbf{K}$ and set

$$K_{kl} = (4\pi)^{-1}\epsilon_{klm}(x_m - x_m')\,|\mathbf{r} - \mathbf{r}'|^{-3} + K_{kl}^*, \quad \text{(A10)}$$

$$N = (4\pi)^{-1}\,|\mathbf{r} - \mathbf{r}'|^{-1} + N^*. \quad \text{(A11)}$$

Then, from (A1), (A10), and (A11), we obtain, for $\mathbf{r} \in D$,

$$\epsilon_{ijk}\frac{\partial K_{kl}^*}{\partial x_j} = -\frac{\partial N^*}{\partial x_i \partial x_l'}. \quad \text{(A12)}$$

To solve (A12), we introduce the function

$$\psi = (4\pi)^{-1}[\log(1 - \mathbf{r}\cdot\mathbf{r}' + R) - 1]. \quad \text{(A13)}$$

From (A9) and the definition of N, we observe that ψ is harmonic with respect to $\mathbf{r}$ for $\mathbf{r} \neq \mathbf{r}'$, $\mathbf{r} \in D$, and that

$$\frac{\partial r\psi}{\partial r} = -N^* + \log 2.$$

Therefore, the vector field $\nabla \times \mathbf{r}\psi$ satisfies

$$\nabla \times (\nabla \times \mathbf{r}\psi) = -\nabla N^*$$

in D; (A12) thus has the general solution

$$K_{kl}^* = \epsilon_{kmn}\frac{\partial^2 x_n \psi}{\partial x_m \partial x_l'} + \frac{\partial \Psi_l'}{\partial x_k}, \quad \mathbf{r} \in D, \quad \text{(A14)}$$

where Ψ_l' is an arbitrary harmonic function of $\mathbf{r}$.

Now, considering (A4) and (A5), we introduce the decomposition (into harmonic poloidal and toroidal parts[3]) defined by

$$\frac{1}{4\pi}\epsilon_{klm}(x_m - x_m')\,|\mathbf{r} - \mathbf{r}'|^{-3} = \frac{\partial f_l}{\partial x_k} + \epsilon_{kmn}\frac{\partial x_n g_l}{\partial x_m}. \quad \text{(A15)}$$

Solving (A15) for f_l and g_l and using (A5), we obtain

$$f_l = -\frac{1}{4\pi}\mathbf{i}_l\cdot(\mathbf{r}\times\mathbf{r}')\,|\mathbf{r} - \mathbf{r}'|^{-1} \times [r^2 - \mathbf{r}\cdot\mathbf{r}' + r\,|\mathbf{r} - \mathbf{r}'|]^{-1}, \quad \text{(A16a)}$$

$$g_l = -\frac{1}{4\pi}\frac{\partial}{\partial x_l'} \times r^{-1}[\log(r^2 - \mathbf{r}\cdot\mathbf{r}' + r\,|\mathbf{r} - \mathbf{r}'|) - \log 2r^2]. \quad \text{(A16b)}$$

Both of these functions are harmonic with respect to $\mathbf{r}$ in D, except at $\mathbf{r}'$. Now, using (A13) and (A16), we obtain

$$\frac{\partial \psi}{\partial x_l'} + g_l = 0 \quad \text{for} \quad r = 1, \quad \text{(A17)}$$

so that the sum of the two terms containing ϵ_{kmn} in (A14) and (A15) vanishes on S. Therefore, we may set $\Psi_l' = \text{const}$ in (A14) and

$$K_{kl} = \frac{\partial f_l}{\partial x_k}, \quad \mathbf{r} \in V, \quad \text{(A18)}$$

$$K_{kl} = \frac{1}{4\pi}\epsilon_{klm}(x_m - x_m')\,|\mathbf{r} - \mathbf{r}'|^{-3} - \epsilon_{kmn}\frac{\partial x_n \psi}{\partial x_m \partial x_l'}, \quad r \in D. \quad \text{(A19)}$$

We note from (A19) that $\tilde{K}_{kl}(\mathbf{r}, \mathbf{r}') = \tilde{K}_{lk}(\mathbf{r}', \mathbf{r})$, where

$$\tilde{K}_{kl} = K_{kl} + \epsilon_{lmn}\frac{\partial^2 x_k' \psi}{\partial x_k \partial x_m'}$$

is obtained from K_{kl} by adding a gradient with respect to $\mathbf{r}$. This remark explicitly verifies the self-adjoint property of H as proved in Sec. 5.

[1] T. G. Cowling, *Magnetohydrodynamics* (Interscience, New York, 1957), Chap. 5, and P. H. Roberts, *An Introduction to Magnetohydrodynamics* (American Elsevier, New York, 1967), Chap. 3. In particular, see E. Bullard and H. Gellman, Phil. Trans. Roy. Soc. London **A247**, 213 (1954).

[2] W. M. Elsasser, Phys. Rev. **69**, 106 (1946).

[3] G. Backus, Ann. Phys. (N.Y.) **4**, 372 (1958).

[4] E. N. Parker, Astrophys. J. **122**, 293 (1955).

[5] A. Herzenberg, Phil. Trans. Roy. Soc. London **A250**, 543 (1958).

[6] S. Childress, Lecture Notes, University of Paris, 1968.

[7] S. Childress, Reports MF-53 and MF-54, Courant Institute of Mathematical Sciences (1967).

[8] S. Childress, in *Proceedings NATO Advance Study Institute, April, 1967* (Wiley, New York, 1969).

[9] In (2.1), we may set $\mathbf{h} = H^{-1}\mathbf{h}^*$, $\mathbf{e} = LH^{-1}\mathbf{e}^*$, $t = UL^{-1}t^*$, $\mathbf{q} = U^{-1}\mathbf{q}^*$, and $R = \sigma\mu UL$, where asterisked quantities are dimensionless and L is here the conductor radius.

[10] M. J. Lighthill, Phil. Trans. Roy. Soc. London **A252**, 397 (1960).

[11] The arbitrary constant may be absorbed into the definition of ϵ. The sign on the right of (4.15) is consistent with the restriction to positive comparison eigenvalues introduced in Sec. 1.

[12] That (5.3) defines a norm follows from the well-known inequality

$$\|\mathbf{h}\|_1 \geq \pi\,\|\mathbf{h}\|, \quad \mathbf{g} \in C_1.$$

[13] We note that if $\mathbf{f}$ is an eigenfunction of H corresponding to an eigenvalue $\nu \neq \mu_k$, then

$$\int_D \mathbf{f}_i\cdot\nabla\times\mathbf{f}\,d\mathbf{r} = \mu_k\int_D \nabla\times\mathbf{f}_i\cdot\nabla\times\mathbf{f}\,d\mathbf{r} = 0,$$

where the f_i are associated with μ_k.

[14] In differentiating the integrals with respect to r, the domain D_α may be regarded as fixed, relative to D, in which case integrals over ∂D_α do not arise from differentiation. This is possible because the total contribution from these surface integrals to $I_1 + I_2$ is zero.

[15] S. I. Braginskii, Zh. Eksp. Teor. Fiz. **47**, 1084, 2178 (1964) [Sov. Phys. JETP **20**, 726, 1462 (1965)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Multiplicities in the Classical Groups. I*

B. GRUBER
Department of Physics, Saint Louis University, St. Louis, Missouri 63103
and
Istituto di Fisica Teorica, Universita di Napoli, Naples, Italy†

(Received 1 December 1969)

A unified treatment of *all* three multiplicities, valid for *all* classical (compact, connected, simple Lie) groups, is described. The general theory is given and then applied to the rank-2 groups.

I. INTRODUCTION AND SUMMARY

The three multiplicities to be considered are[1]

(1) the multiplicity of weights,[2]

(2) the multiplicity of representations occurring in the (inner) direct product [Clebsch–Gordan (CG) series],[3]

(3) the branching multiplicity.[4]

The groups for which these multiplicities are studied are the simple, compact, and connected Lie groups. They are referred to as classical groups or simply as groups. Throughout this paper, by representations, we mean irreducible representations of the classical groups. For the third multiplicity, namely, the branching multiplicity, the subgroups to which we will restrict the groups are classical groups or direct products of classical groups (semisimple, compact, and connected Lie groups).

In Sec. II, a *method* of setting up "patterns" and "diagrams" is described for the classical groups up to rank 4 (but, in principle, valid for all classical groups). In Sec. III, *rules* are given, valid (again, in principle) for all classical groups, on how to apply the diagrams on the pattern in order to obtain the various multiplicities.

The particular form of a pattern depends on the group, while its size depends on the ambitions. The pattern is nothing but a systematic way "of writing down" a weight diagram. The more representations one wants to consider, the more the size of the pattern increases. On the other hand, the diagrams, one for each multiplicity (counting the restrictions to different subgroups as different multiplicities), are determined by properties of the classical group (and its subgroups) alone. They are related to the Weyl group of the classical group (and not representations of the group).

By setting up a pattern for a group and then applying the various diagrams according to generally valid rules on this pattern, the various multiplicities can be obtained from the pattern for all representations described by it. Thus, *all* multiplicities can be obtained from *one and the same* pattern according to rules which are valid for *all* classical groups. The number of representations for which this is true determines the size of the pattern.

In Sec. IV, the general methods developed in Secs. II and III are applied to the rank-2 groups, namely $SU(3)$, $SO(5) \sim Sp(4)$, and G_2. The group $SU(3)$ is trivial with respect to all multiplicities. It has, however, been included for reasons of completeness, and also as a convenient example due to its familiarity. [Moreover, due to its simplicity, the patterns and diagrams amount to a somewhat implicit tabulation of the multiplicities of $SU(3)$.] The groups $SO(5)$ and G_2 are already much less trivial, though for these groups also some well-known branching multiplicities have been included for reasons of completeness.

The following has been achieved:

$SU(3)$: Two patterns are given which allow the calculation of the multiplicities of weights for 113 representations as well as the branching of these 113 representations with respect to the subgroups $SU(2)$ and $SO(3)$. Moreover, two subpatterns are given which allow the calculation of the CG series of 52 representations with *any* representation of $SU(3)$.

$SO(5) \sim Sp(4)$: The multiplicity of weights for 132 representations, as well as the branching of these 132 representations with respect to the subgroups $SO(4)$, $SO(3)_1$, $SO(3)_2$, and $SO(3)_3$, can be obtained from two patterns. Two subpatterns give the CG series of 42 representations with any representation of $SO(5)$.

G_2: The multiplicity of weights for 36 representations, as well as their branching with respect to the subgroups $SU(3)$, $SU(2)_1 \times SU(2)_2$, $SU(2)_1$, $SU(2)_2$, $SO(3)_1$, and $SO(3)_2$ of the group G_2, can be obtained from one pattern. The CG series of any representation of G_2 with any of the 15 representations contained in a subpattern can be calculated.

Should the need arise, the patterns can easily be enlarged to include representations not contained in them (except for obvious limitations given by the size of the patterns).

The method developed in the present paper is applied in a subsequent paper to the rank-3 and rank-4 groups. It is the application to these groups which is of real interest. So far, no simple methods of obtaining the multiplicities for these groups exist—with a few exceptions—while the methods developed and described in the present paper allow a relatively easy handling of the multiplicities of representations with dimensionalities well in the tens of thousands. Moreover, the treatment is a "unified" one. *All* multiplicities are treated in the same manner for *all* the groups.

II. GENERAL THEORY

Three formulas are utilized to calculate the three multiplicities—one formula for each multiplicity. Before writing down these formulas, some definitions have to be made.

A highest weight is denoted by M, and m denotes an arbitrary weight. Then $m \in D(M)$ is a weight contained in the representation $D(M)$ with highest weight M. The symbol S denotes an element of the Weyl group W. The signature of the element S is denoted by δ_S: $\delta_S = -1$ if $S \in W$ is a reflection; $\delta_S = 1$ if $S \in W$ is not a reflection. The symbol $\delta_{x,y}$, where x and y are two (weight) vectors, is the Kronecker symbol; i.e., $\delta_{x,y} = 1$ if $x = y$, $\delta_{x,y} = 0$ if $x \neq y$. Adding a suffix r to one of the symbols, such as M_r, m_r, S_r, and W_r, denotes the same quantity in a subgroup of the group considered. Thus, M_r is a highest weight in a subgroup, etc. The mapping of a weight m of a group onto a weight m_r of a subgroup is denoted by Lm $(= m_r)$. The three multiplicities are denoted as follows:

$\gamma^M(m)$ or $\gamma(m)$: multiplicity of the weight m of the representation $D(M)$;

$\bar{\gamma}(M)$: multiplicity of the representation $D(M)$ contained in the decomposition of a direct product $D(M') \otimes D(M'')$;

$\tilde{\gamma}(M_r)$: multiplicity of the representation $D(M_r)$ (of a subgroup G_r) which is contained in a representation $D(M)$ of a group G under the restriction of G to the subgroup G_r.

The simple negative roots are denoted by β_i, $i = 1, 2, \cdots, l$, where l is the rank of the group. The sum over all positive roots of the group is denoted by $2R$.

The first formula, for the multiplicity of weights, is Racah's recursion formula.[5] In the notation described above, the formula is given as

$$\gamma(m) = -\sum_{\substack{S \in W \\ S \neq 1}} \delta_S \gamma(m + R - SR). \tag{1}$$

The second formula—for the multiplicity in the (inner) direct product—is also due to Racah.[5] If the direct product of two representations $D(M)$ and $D(M')$ is written as

$$D(M) \otimes D(M') = \oplus \sum_{\bar{M}} \bar{\gamma}(\bar{M}) D(\bar{M}),$$

then $\bar{\gamma}(\bar{M})$ is given as

$$\bar{\gamma}(M' + m) = \sum_{S \in W} \delta_S \gamma^M[m + M' + R - S(M' + R)], \quad \text{with} \quad m \in D(M). \tag{2}$$

Finally, the third formula—for the branching multiplicity—is due to Straumann[6] (it has also been independently obtained in Refs. 7). If the branching of some representation $D(M)$ is written as

$$D(M)|_{\text{restr.}} = \oplus \sum_{M_r} \tilde{\gamma}(M_r) D(M_r),$$

then

$$\tilde{\gamma}(M_r) = \sum_{S_r \in W_r} \sum_{m \in D(M)} \delta_{S_r} \gamma^M(m) \delta_{Lm, M_r + R_r - S_r R_r}. \tag{3}$$

It is easy to observe that all three multiplicities are related to weight diagrams (systems of weight vectors *without* their multiplicity). This makes it possible to obtain *all three* multiplicities from *one and the same* "pattern."

The Pattern

The pattern is nothing but a systematic way of writing down a weight diagram. If $D(M)$ is the representation to be considered, all its weights can be obtained from its highest weight M as

$$M + k_1\beta_1 + \cdots + k_l\beta_l, \tag{4}$$

where the k_i, $i = 1, 2, \cdots, l$, are nonnegative integers. If the dominant weights[3b] are found first—which is simple due to the conditions imposed on a weight through its dominancy—there is no problem in eliminating from the set of weights $M + k_1\beta_1 + \cdots + k_l\beta_l$, all possible nonnegative integers k_i, the weights which do not belong to $D(M)$. Namely, knowing the dominant weights of $D(M)$, we know all weights through the Weyl group W.

So far, only groups up to rank 4 have been treated by the methods described here. For them, the pattern is most conveniently obtained in the following manner.

The weights

$$M + k_1\beta_1 + k_2\beta_2 + 0 \cdot \beta_3 + 0 \cdot \beta_4 \tag{5}$$

are written down in matrix form. The position of the weight given by Eq. (5) for some values k_1 and k_2 is at the intersection of the $(k_1 + 1)$th line and $(k_2 + 1)$th column, lines and columns being counted from the *lower right* corner of the matrix. Positions in the matrix not corresponding to weights of the representation

are left empty (see Sec. IV). The value k_3 distinguishes different matrices—or blocks—one block being put on the top of the other (with increasing k_3). Thus, the weights

$$M + k_1\beta_1 + k_2\beta_2 + k_3\beta_3 + 0 \cdot \beta_4,$$

with k_1 and k_2 arbitrary (nonnegative integer) and k_3 some fixed value (nonnegative integer), form the $(k_3 + 1)$th block (matrix) counting from the bottom. (With k_1 and k_2 arbitrary we understand that outside some maximal values of k_1 and k_2—needed to describe the weights of the diagram—all entries of the matrix are zero anyway.) We call the set of blocks obtained in this manner a "column." The values of k_4 distinguish different columns, from *right to left.*

The weight vectors m of the pattern are written in the form ${}_q(m_1, m_2, \cdots, m_l)_p$ or the form $(m_1, m_2, \cdots, m_l)_p$, where m_i, $i = 1, \cdots, l$, are the components of the weight and the subscript p is the "dominant weight number" (d.w. number). The number p is the same for all members of a set of equivalent weights, while the dominant weight of this set is singled out by boldface **p**.[8] The suffix q gives the number of weights contained in the set of equivalent weights to which this weight belongs.

It should be noted that a pattern describes not just one representation but *all* those representations whose highest weights are contained in this representation.

The Diagram

The next things we consider are the diagrams corresponding to the different multiplicities. So far, the pattern—the weight diagram—has been set up on which the three equations [(1)–(3)] are based. The diagrams, determined by these equations, provide the tool to calculate the multiplicities from the pattern.

From Eq. (1), it follows that the multiplicity $\gamma(m)$ of the weight m is related to the multiplicity of several other weights. All these weights are entries of the pattern, and it is the diagram which determines for a given weight m the location (and signature δ_S) of the other weights to which the multiplicity $\gamma(m)$ is related. In other words, the relations

$$R - SR; \delta_S$$

have to be re-expressed in terms of the pattern language. This is done as follows:

In

$$R - SR = -(k_1\beta_1 + \cdots + k_l\beta_l),$$

the nonnegative integers k_i, $i = 1, 2, \cdots, l$, are determined for every $S \in W$ and, moreover, the value of δ_S is determined. The diagram is then defined to be the set of all elements

$$\begin{matrix} (k_1^s, k_2^s, \cdots, k_l^s); +1 \\ (k_1^{s'}, k_2^{s'}, \cdots, k_l^{s'}); -1 \\ \cdot \\ \cdot \\ \cdot \end{matrix} \tag{6}$$

A graphical representation of the diagram is obtained in the following manner: Along a horizontal line, the nonnegative integers 0, 1, 2, $\cdots$ are written from the left to the right. Then, for a given element

$$(k_1^s, k_2^s, \cdots, k_l^s); \delta_S, \tag{7}$$

the δ_S is written below the nonnegative integer k_1^s. The remaining k_i^s are written to the right of the diagram in the same line as the entry δ_S in the order

$$k_l^s, k_{l-1}^s, \cdots, k_2^s.$$

This is done for every element of the diagram. The ordering adopted is that, for two elements

$$(k_l, k_{l-1}, \cdots, k_u, k_{u-1}, \cdots, k_2),$$
$$(k_l, k_{l-1}, \cdots, k_u, k'_{u-1}, \cdots, k'_2)$$

having the same values $k_u, k_{u+1}, \cdots, k_l$, that element is entered first in the graphical representation of the diagram for which $k_{u-1} < k'_{u-1}$ holds.

For groups up to rank 4, the graphical representation of the diagram is of the form

0 1 2 3 4 5 6 $\cdots$	
-1	132
1	133
1	211
-1	220

where the four elements (4231); (−1), (5331); 1, (2112); 1, and (3022); (−1) have been represented graphically. Taking care of the conventions in which the pattern for the groups (up to rank 4) has been set up, we interpret a particular element (k_1, k_2, k_3, k_4); δ_S of the diagram in the following manner:

(1) Go horizontally k_1 columns to the right;

(2) from there go vertically k_2 lines down;

(3) from there go to the same position in the k_3th block below;

(4) from there go to the same position in the same block in the k_4th "block column" to the right. This is the position for δ_S.

Operations (1)–(4) are referred to as the "counting process."

Upon inspecting the two formulas Eqs. (1) and (2), it can be recognized that the diagram for the

multiplicity in the direct product is obtained in exactly the same way as the diagram for the multiplicity of weights. The only difference is that the positions of the values δ_S depend now on the particular direct product which is formed. Instead of depending on

$$R - SR; \delta_S,$$

Eq. (2) depends on

$$M' + R - S(M' + R); \delta_S.$$

This implies that the diagram depends on the representation $D(M')$ but remains the same for *all* representations $D(M)$. As in the case of the multiplicity of weights, the position of $M' + R - S(M' + R)$ in the pattern is obtained from

$$M' + R - S(M' + R) = -(k_1\beta_1 + \cdots + k_l\beta_l).$$

The system of numbers (the diagram)

$$(k_1, k_2, \cdots, k_l); \delta_S \tag{8}$$

has the same interpretation as in the case of the inner multiplicity. However, it depends not only on the element $S \in W$ but also on M' [for reasons of simplicity, this has not explicitly been indicated in Eq. (8)].

Finally, Eq. (3) is considered. For this case, the diagram is obtained as in the other two cases. However, some care has to be taken due to the fact that the diagram has to be applied in the *subgroup*. The pattern for the group has to be projected into the subgroup by the map L. It is after this projection that the diagram is to be applied. The position of the weight $M_r + R_r - S_rR_r$, in the projected pattern, is obtained from

$$M_r + R_r' - S_rR_r = -(k_1\beta_1' + \cdots + k_t\beta_t'), \quad t \leq l,$$

where t is the rank of the subgroup and $\beta_1', \beta_2', \cdots, \beta_t'$ are the simple roots which span the projected pattern. Thus, $\beta_1', \cdots, \beta_t'$ may or may not be the simple negative roots of the subgroup. [The two cases are, for instance, represented by $SO(5) \downarrow SO(3)$ and $SO(5) \downarrow SO(4)$, respectively. See below, Sec. IV.]

The interpretation of the set of numbers

$$(k_1, k_2, \cdots, k_t); \delta_{S_r}$$

in the *projected* pattern is the same as in the above two cases.

III. RULES

In this section, the rules for obtaining the three multiplicities by applying the diagrams to the patterns are given. The rules can be seen to be valid for all classical groups. For convenience, the rules are given here for the groups up to rank 4. It is a matter of simple generalization to obtain the rules for the general case.

A. The Counting Process

To begin the counting process at some weight m of the pattern means the following:

(1) Go from the weight m horizontally to the weight k_1 columns to the right of it.

(2) Go from the weight reached by (1) vertically to the weight k_2 lines below it.

(3) The weight reached by (1) and (2) is at some position in the block containing it. Go vertically to the k_3th block below the block containing the weight m and take the weight which is at the same position in this block as the weight reached by (1) and (2).

(4) From the "block column" containing the weight m go horizontally k_4 "block columns" to the right. In the block column thus reached take the weight m' which has the same position as the weight reached by (1), (2), and (3) in the "block column" containing m.

An element $(k_1, k_2, k_3, k_4); \delta_S$ can lead to an empty space in the pattern or even out of the pattern. Those elements of the diagram can be ignored. They do not contribute.

B. Multiplicity of Weights: $\gamma(m)$

The multiplicities of the dominant weights have to be calculated only. (Sets of equivalent weights have the same multiplicity.) This is done successively, beginning with the highest weight of the representation. Each boldface d.w. number corresponds to the highest weight of an (irreducible) representation; *all* d.w. numbers equal or smaller than a given d.w. number correspond to weights of that representation.[8] Thus, beginning with some d.w. number $\mathbf{p}$, corresponding to some representation $D(M)$, the multiplicities of *all* weights corresponding to the d.w. numbers $\mathbf{p}, \mathbf{p-1}, \mathbf{p-2}, \cdots, \mathbf{1}$ have to be calculated successively. This is done as follows (m_p as well as p is used to denote the weights):

(B.1) Let the highest weight M of the representation $D(M)$ have the d.w. number p. Then all weights of the pattern having d.w. numbers $p, p-1, \cdots, 1$ belong to $D(M)$.[9]

(B.2) $\gamma(\mathbf{p}) = 1$ (multiplicity of highest weight).

(B.3) $\gamma(\mathbf{p}-1)$:

(b.1) Begin the counting process (Sec. IIIA) at the weight $m_{\mathbf{p}-1}$.

(b.2) Multiply the multiplicity $\gamma(m')$ of the weight $m' \in D(M)$ reached by (b.1) by δ_S.

(b.3) Do this for all elements $(k_1, k_2, k_3, k_4); \delta_S$ of the diagram.

(b.4) Add all multiplicities $\gamma(m')$ obtained in this fashion; the resulting number is the multiplicity $\gamma(p-1)$.

(B.4) Perform step (B.3) successively for the weights

$$m_{\mathbf{p}-2}, m_{\mathbf{p}-3}, \cdots, m_1.$$

The resulting numbers are the multiplicities

$$\gamma(p-2), \gamma(p-3), \cdots, \gamma(1).$$

The following two observations can be made:

(i) If only the multiplicity of weights is of interest, the weights in the pattern can be substituted by their d.w. number. No property other than the d.w. number labeling is used. (The same is true for the calculation of the branching multiplicity.)

(ii) If the multiplicity of dominant weights only is calculated (which is, in fact, the simplest choice), the pattern can be made smaller. The last weight to be included is the weight corresponding to d.w. number **1**.

C. Multiplicity in the Direct Product $D(M) \otimes D(M')$

The multiplicities $\bar{\gamma}(\bar{M})$ in the direct product $D(M) \otimes D(M')$ are obtained as follows:

(C.1) If possible, choose $D(M')$ to be the larger of the two representations (it simplifies the calculation). Then determine the diagram for the weight M'.

(C.2) A certain d.w. number corresponds to the weight M of $D(M)$ in the pattern. All weights with this or a lower d.w. number are weights of $D(M)$. The multiplicity of these weights has to be calculated by the procedure of Sec. IIIB.

(C.3) The multiplicity $\bar{\gamma}(M'+m)$, $m \in D(M)$, of the representation $D(M'+m)$ (i.e., $M'+m$ is supposed to be a dominant weight) is obtained by applying the counting process of Sec. IIIA with the diagram obtained in (C.1) on the weight m. The calculation of the multiplicity $\bar{\gamma}(M'+m)$ is identical to the calculation of $\gamma(\mathbf{p}-1)$ by steps (b.1) to (b.4) in (B.3) except that the diagram is different. [The resulting number in step (b.4) is, of course, the multiplicity $\bar{\gamma}(M'+m)$.]

(C.4) The calculation of the $\bar{\gamma}(M'+m)$ has to proceed successively, beginning with the weight M, moving from the right to the left and from the bottom to the top of the pattern.

Observation: The explicit weight diagram is needed in order to determine the argument $M'+m$ of $D(M'+m)$. The d.w. numbers do not distinguish equivalent weights, and thus, in the case of this multiplicity, the pattern cannot be reduced to a pattern of d.w. numbers.

D. Branching Multiplicity

The branching multiplicity $\tilde{\gamma}(M_r)$ for a representation $D(M)$ is obtained by applying the counting process (by means of the diagrams for branching) on the weights M_r of the *projected* weight diagram $LD(M)$.

(D.1) To the highest weight M of the representation $D(M)$ corresponds some d.w. number $\mathbf{p}$. All weights of the pattern with d.w. numbers $\leq p$ belong to $D(M)$. Determine the multiplicity of weights for the representation $D(M)$ by (B).

(D.2) $\tilde{\gamma}(M_r)$ is obtained as follows:

- (d.1) Apply the counting process (Sec IIIA) (for an element of the diagram for branching) on the highest weight M_r in the projected pattern. Some weight m_r is reached.
- (d.2) Sum the multiplicities of all weights m of $D(M)$ in the original pattern which are mapped by L on the weight m_r.
- (d.3) Multiply the number obtained by rule (d.2) by the δ_S of the corresponding element of the diagram.
- (d.4) Do this for all elements of the diagram.
- (d.5) Add all numbers obtained by (d.1) to (d.4). The resulting number is the branching multiplicity $\tilde{\gamma}(M_r)$.

(D.3) As in the previous two cases (B) and (C) the multiplicities $\tilde{\gamma}$ are calculated successively, beginning with the highest possible M_r. The succession is, as in case (C), from right to left and from bottom to top of the pattern.

IV. THE RANK-2 GROUPS

In the following patterns and diagrams are given for the rank-2 groups $SU(3)$, $SO(5) \sim Sp(4)$, and G_2.

In order to keep the size of the patterns small, two sorts of patterns are given. Namely, patterns containing d.w. numbers only and patterns in which the weights themselves are given. The former patterns are good for multiplicity of weights and branching multiplicity only. They have the advantage of being able to accommodate many more representations than a pattern of the second kind of equal size. Patterns of the second kind are obviously good for all three multiplicities. That these patterns of the second kind contain fewer representations is of no concern. The only multiplicity which would be affected is the multiplicity in the direct product $D(M) \otimes D(M')$ for which fewer representations $D(M)$ are available. This is more than balanced by the fact that $D(M')$ may be *any* representation.

A. $SU(3)$

Simple roots:

$$\beta_1 = (-1, 1, 0), \quad \beta_2 = (0, -1, 1),$$
$$R = (1, 0, -1).$$

Weights:

$$m = (m_1, m_2, m_3), \quad m_i = \tfrac{1}{3}k, \quad k \text{ integer},$$
$$m_1 + m_2 + m_3 = 0.$$

In order to minimize the size of the patterns, the (p, q) notation is used:

$$m_1 = \tfrac{1}{3}(2p + q), \quad m_2 = \tfrac{1}{3}(-p + q),$$
$$m_3 = \tfrac{1}{3}(-p - 2q),$$

where

$$p = m_1 - m_2, \quad q = m_2 - m_3,$$

and

$$\dim (p, q) = \tfrac{1}{2}(p + 1)(q + 1)(p + q + 2),$$

with

$$p \geq 0, \quad q \geq 0 \text{ (d.w. condition).}$$

Diagram for multiplicity of weights γ:

0	1	2	
	1		0
1		−1	1
	−1	1	2

Diagrams for branching multiplicity $\tilde{\gamma}$:

(1) Subgroup $SU(2)$, $L(p, q) = \frac{1}{2}p$:

0	1	
1	−1	0

(2) Subgroup $SO(3)$, $L(p, q) = p + q$:

0	1	
1	−1	0

Diagram for multiplicity $\tilde{\gamma}$:

0 1 2 ⋯ $p'+1$ ⋯ $p'+q'+2$		
1	−1	0
−1	1	$q'+1$
	1 −1	$p'+q'+2$

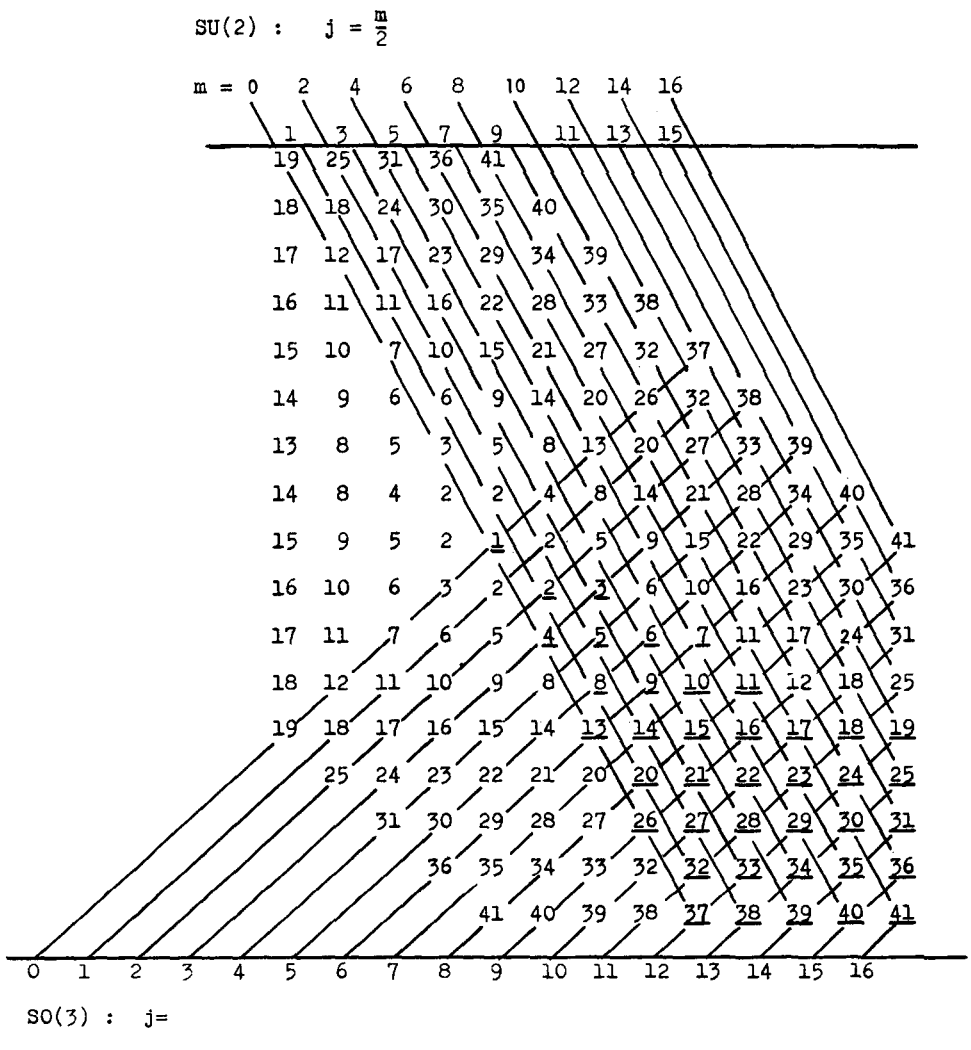

FIG. 1. Dominant-weight-number pattern for $S\mathcal{U}(3)$. The representations corresponding to the d.w. numbers are given in the text. The (dominant) weights j of the subgroups $SU(2)$ and $SO(3)$ are given on top and on bottom of the diagram, respectively. The straight lines in the d.w. number pattern indicate the projection $L(p, q) = j$ of $SU(3)$ weights onto the weight j of the subgroup. All weights with d.w. numbers lying along such a straight line are projected onto the subgroup weight j which lies on the same straight line.

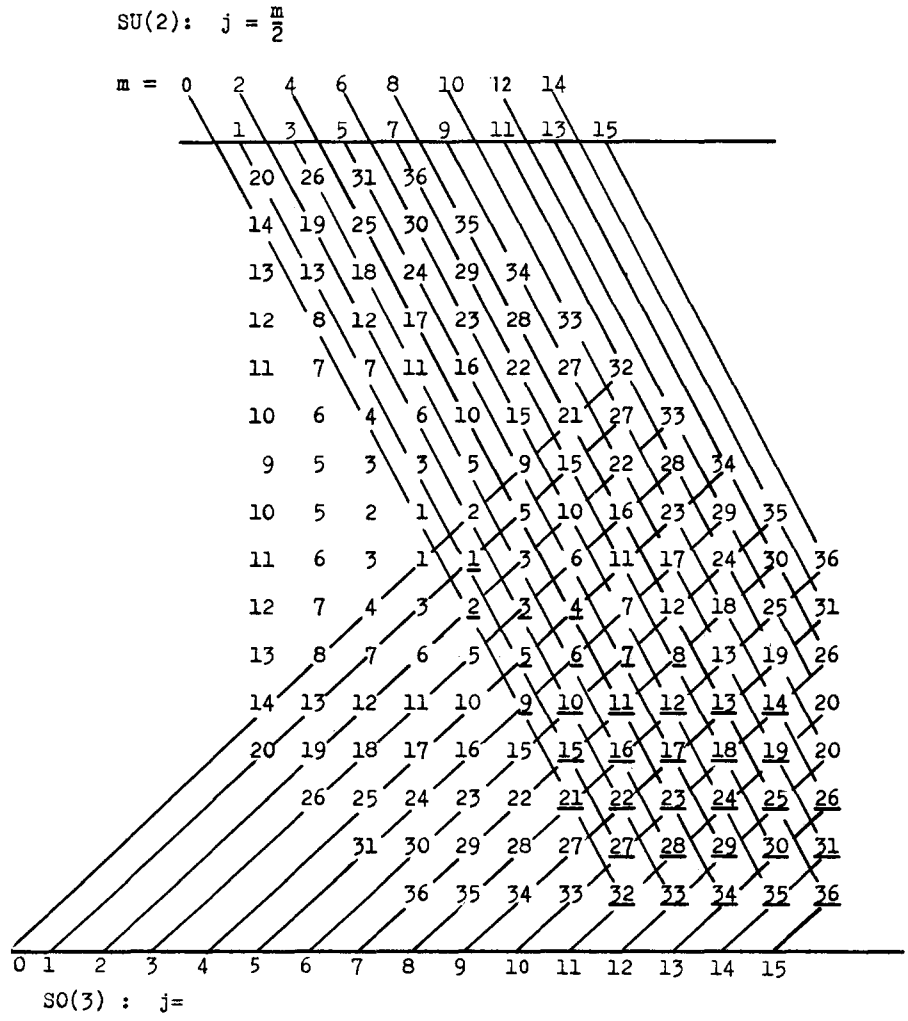

FIG. 2. As in Fig. 1.

Representations:

For the pattern Fig. 1, the representations are

$(8,8)_{41}$, $(6,9)_{40}$, $(4,10)_{39}$, $(2,11)_{38}$, $(0,12)_{37}$, $(9,6)_{36}$, $(7,7)_{35}$, $(5,8)_{34}$, $(3,9)_{33}$, $(1,10)_{32}$, $(10,4)_{31}$, $(8,5)_{30}$, $(6,6)_{29}$, $(4,7)_{28}$, $(2,8)_{27}$, $(0,9)_{26}$, $(11,2)_{25}$, $(9,3)_{24}$, $(7,4)_{23}$, $(5,5)_{22}$, $(3,6)_{21}$, $(1,7)_{20}$, $(12,0)_{19}$, $(10,1)_{18}$, $(8,2)_{17}$, $(6,3)_{16}$, $(4,4)_{15}$, $(2,5)_{14}$, $(0,4)_{13}$, $(9,0)_{12}$, $(7,1)_{11}$, $(5,2)_{10}$, $(3,3)_9$, $(1,4)_8$, $(6,0)_7$, $(4,1)_6$, $(2,2)_5$, $(0,3)_4$, $(3,0)_3$, $(1,1)_2$, $(0,0)_1$.

For the pattern Fig. 2, the representations are

$(8,7)_{36}$, $(6,8)_{35}$, $(4,9)_{34}$, $(2,10)_{33}$, $(0,11)_{32}$, $(9,5)_{31}$, $(7,6)_{30}$, $(5,7)_{29}$, $(3,8)_{28}$, $(1,9)_{27}$, $(10,3)_{26}$, $(8,4)_{25}$, $(6,5)_{24}$, $(4,6)_{23}$, $(2,7)_{22}$, $(0,8)_{21}$, $(11,1)_{20}$, $(9,2)_{19}$, $(7,3)_{18}$, $(5,4)_{17}$, $(3,5)_{16}$, $(1,6)_{15}$, $(10,0)_{14}$, $(8,1)_{13}$, $(6,2)_{12}$, $(4,3)_{11}$, $(2,4)_{10}$, $(0,5)_9$, $(7,0)_8$, $(5,1)_7$, $(3,2)_6$, $(1,3)_5$, $(4,0)_4$, $(2,1)_3$, $(0,2)_2$, $(1,0)_1$.

Sets of equivalent weights:

For the dominant weights given above the following holds:

(a) ${}_6(p,q)$, if $p \neq 0$ and $q \neq 0$,
(b) ${}_3(p,q)$, if $p = 0$ or $q = 0$,
(c) ${}_1(p,q)$, if $p = q = 0$.

Conjugate representations:

The two representations

$$D(p,q) \quad \text{and} \quad D(q,p)$$

are conjugate to each other. Conjugate representations have the same multiplicity structure. Thus, if the multiplicities of the representation $D(p,q)$ are known, the multiplicities of the representation $D(q,p)$ are given by the mapping

$$(p',q') \leftrightarrow (q',p') \tag{9}$$

of the weights of $D(p,q)$ onto the weights of $D(q,p)$.

When the direct product is formed, it should be noted that

$$D(p,q) \otimes D(p',q')$$

and

$$D(q,p) \otimes D(q',p')$$

$(-5,5)_{22}$, $(-3,6)_{16}$, $(-1,7)_{11}$, $(1,-8)_{11}$, $(3,-9)_{16}$, $(5,-10)_{22}$

$(-6,-3)_{21}$, $(-4,-4)_{15}$, $(-2,-5)_{10}$, $(0,-6)_7$, $(2,-7)_{10}$, $(4,-8)_{15}$, $(6,-9)_{21}$

$(-7,-1)_{20}$, $(-5,-2)_{14}$, $(-3,-3)_9$, $(-1,-4)_6$, $(1,-5)_6$, $(3,-6)_9$, $(5,-7)_{14}$, $(7,-8)_{20}$

$(-8,1)_{20}$, $(-6,0)_{13}$, $(-4,-1)_8$, $(-2,-2)_5$, $(0,-3)_3$, $(2,-4)_5$, $(4,-5)_8$, $(6,-6)_{13}$, $(8,-7)_{20}$

$(-9,3)_{21}$, $(-7,2)_{14}$, $(-5,1)_8$, $(-3,0)_4$, $(-1,-1)_2$, $(1,-2)_2$, $(3,-3)_4$, $(5,-4)_8$, $(7,-5)_{14}$, $(9,-6)_{21}$

$(-10,5)_{22}$, $(-8,4)_{15}$, $(-6,3)_9$, $(-4,2)_5$, $(-2,1)_2$, $(0,0)_{\underline{1}}$, $(2,-1)_2$, $(4,-2)_5$, $(6,-3)_9$, $(8,-4)_{15}$, $(10,-5)_{22}$

$(-9,6)_{16}$, $(-7,5)_{10}$, $(-5,4)_6$, $(-3,3)_3$, $(-1,2)_2$, $(1,1)_{\underline{2}}$, $(3,0)_{\underline{3}}$, $(5,-1)_6$, $(7,-2)_{10}$, $(9,-3)_{16}$

$(-8,7)_{11}$, $(-6,6)_7$, $(-4,5)_6$, $(-2,4)_5$, $(0,3)_{\underline{4}}$, $(2,2)_{\underline{5}}$, $(4,1)_{\underline{6}}$, $(6,0)_{\underline{7}}$, $(8,-1)_{11}$

$(-7,8)_{11}$, $(-5,7)_{10}$, $(-3,6)_9$, $(-1,5)_8$, $(1,4)_{\underline{8}}$, $(3,3)_{\underline{9}}$, $(5,2)_{\underline{10}}$, $(7,1)_{\underline{11}}$

$(-6,9)_{16}$, $(-4,8)_{15}$, $(-2,7)_{14}$, $(0,6)_{\underline{13}}$, $(2,5)_{\underline{14}}$, $(4,4)_{\underline{15}}$, $(6,3)_{\underline{16}}$

$(-5,10)_{22}$, $(-3,9)_{21}$, $(-1,8)_{20}$, $(1,7)_{\underline{20}}$, $(3,6)_{\underline{21}}$, $(5,5)_{\underline{22}}$

FIG. 3. Subpattern of pattern in Fig. 1. In this subpattern the weights are given, not only their d.w. numbers. From this subpattern all three multiplicities can be obtained, in particular the multiplicity $\bar{\gamma}$ of the direct product $D(M) \otimes D(M')$. The representations $D(M)$ are limited to those contained in the subpattern while the representation $D(M')$ can be chosen completely arbitrary.

$(-3,-7)_{18}$, $(-1,-8)_{13}$, $(1,-9)_{13}$, $(3,-10)_{18}$

$(-4,-5)_{17}$, $(-2,-6)_{12}$, $(0,-7)_8$, $(2,-8)_{12}$, $(4,-9)_{17}$

$(-5,-3)_{16}$, $(-3,-4)_{11}$, $(-1,-5)_7$, $(1,-6)_7$, $(3,-7)_{11}$, $(5,-8)_{16}$

$(-6,-1)_{15}$, $(-4,-2)_{10}$, $(-2,-3)_6$, $(0,-4)_4$, $(2,-5)_6$, $(4,-6)_{10}$, $(6,-7)_{15}$

$(-7,1)_{15}$, $(-5,0)_9$, $(-3,-1)_5$, $(-1,-2)_3$, $(1,-3)_3$, $(3,-4)_5$, $(5,-5)_9$, $(7,-6)_{15}$

$(-8,3)_{16}$, $(-6,2)_{10}$, $(-4,1)_5$, $(-2,0)_2$, $(0,-1)_1$, $(2,-2)_2$, $(4,-3)_5$, $(6,-4)_{10}$, $(8,-5)_{16}$

$(-9,5)_{17}$, $(-7,4)_{11}$, $(-5,3)_6$, $(-3,2)_3$, $(-1,1)_1$, $(1,0)_{\underline{1}}$, $(3,-1)_3$, $(5,-2)_6$, $(7,-3)_{11}$, $(9,-4)_{17}$

$(-10,7)_{18}$, $(-8,6)_{12}$, $(-6,5)_7$, $(-4,4)_4$, $(-2,3)_3$, $(0,2)_{\underline{2}}$, $(2,1)_{\underline{3}}$, $(4,0)_{\underline{4}}$, $(6,-1)_7$, $(8,-2)_{12}$, $(10,-3)_{18}$

$(-9,8)_{13}$, $(-7,7)_8$, $(-5,6)_7$, $(-3,5)_6$, $(-1,4)_5$, $(1,3)_{\underline{5}}$, $(3,2)_{\underline{6}}$, $(5,1)_{\underline{7}}$, $(7,0)_{\underline{8}}$, $(9,-1)_{13}$

$(-8,9)_{13}$, $(-6,8)_{12}$, $(-4,7)_{11}$, $(-2,6)_{10}$, $(0,5)_{\underline{9}}$, $(2,4)_{\underline{10}}$, $(4,3)_{\underline{11}}$, $(6,2)_{\underline{12}}$, $(8,1)_{\underline{13}}$

$(-7,10)_{18}$, $(-5,9)_{17}$, $(-3,8)_{16}$, $(-1,7)_{15}$, $(1,6)_{\underline{15}}$, $(3,5)_{\underline{16}}$, $(5,4)_{\underline{17}}$, $(7,3)_{\underline{18}}$

FIG. 4. Subpattern of pattern Fig. 2. The rest as in Fig. 3.

are related to each other by the mapping (9).

B. $SO(5)$

Simple roots:

$$\beta_1 = (0, -1), \quad \beta_2 = (-1, 1),$$
$$R = \tfrac{1}{2}(3, 1).$$

Weights:

$m = (m_1, m_2)$, $m_1, m_2 = k$ or $m_1, m_2 = \frac{1}{2}k$, k integer,

$\dim(M) = \frac{1}{6}(2M_1 + 3)(2M_2 + 1)(M_1 + M_2 + 2) \times (M_1 - M_2 + 1)$

with

$M_1 \geq M_2 \geq 0$ (d.w. condition).

Diagram for multiplicity of weights γ:

0	1	2	3	4	
	1				0
1			-1		1
	-1			1	2
			1	-1	3

Diagram for multiplicity $\bar{\gamma}$:

0	1	2	$\cdots$	$2M_2+1$	$\cdots$	$2M_1+3$	$\cdots$	$2(M_1+M_2+2)$	
1				-1					0
-1						1			$M_1 - M_2 + 1$
				1				-1	$M_1 + M_2 + 2$
						-1		1	$2M_1 + 3$

Diagrams for branching multiplicity $\tilde{\gamma}$:

(1) Subgroup $SO(4)$, $Lm = m$:

0	1	2	
	1		0
−1		−1	1
		1	2

For a dominant weight of $SO(4)$ we have

$$m = (m_1, m_2), \quad m_1 \geq |m_2|.$$

(2) Subgroup $SO(3)_1$, $Lm = \frac{1}{2}(m_1 + m_2)$,
subgroup $SO(3)_2$, $Lm = m_1$,
subgroup $SO(3)_3$, $Lm = 2m_1 + m_2$:

0	1	
1	−1	0

Representations:

The representations of the pattern Fig. 5 are

$(10, 10)_{66}$, $(10, 9)_{65}$, $(10, 8)_{64}$, $(10, 7)_{63}$, $(10, 6)_{62}$, $(10, 5)_{61}$, $(10, 5)_{61}$, $(10, 4)_{60}$, $(10, 3)_{59}$, $(10, 2)_{58}$, $(10, 1)_{57}$, $(10, 0)_{56}$, $(9, 9)_{55}$, $(9, 8)_{54}$, $(9, 7)_{53}$, $(9, 6)_{52}$, $(9, 5)_{51}$, $(9, 4)_{50}$, $(9, 3)_{49}$, $(9, 2)_{48}$, $(9, 1)_{47}$, $(9, 0)_{46}$, $(8, 8)_{45}$, $(8, 7)_{44}$, $(8, 6)_{43}$, $(8, 5)_{42}$, $(8, 4)_{41}$, $(8, 3)_{40}$, $(8, 2)_{39}$, $(8, 1)_{38}$, $(8, 0)_{37}$, $(7, 7)_{36}$, $(7, 6)_{35}$, $(7, 5)_{34}$, $(7, 4)_{33}$, $(7, 3)_{32}$, $(7, 2)_{31}$, $(7, 1)_{30}$, $(7, 0)_{29}$, $(6, 6)_{28}$, $(6, 5)_{27}$, $(6, 4)_{26}$, $(6, 3)_{25}$, $(6, 2)_{24}$, $(6, 1)_{23}$, $(6, 0)_{22}$, $(5, 5)_{21}$, $(5, 4)_{20}$, $(5, 3)_{19}$, $(5, 2)_{18}$, $(5, 1)_{17}$, $(5, 0)_{16}$, $(4, 4)_{15}$, $(4, 3)_{14}$, $(4, 2)_{13}$, $(4, 1)_{12}$, $(4, 0)_{11}$, $(3, 3)_{10}$, $(3, 2)_9$, $(3, 1)_8$, $(3, 0)_7$, $(2, 2)_6$, $(2, 1)_5$, $(2, 0)_4$, $(1, 1)_3$, $(1, 0)_2$, $(0, 0)_1$.

The representations of the pattern Fig. 6 are (it is understood that each weight given below has to be multiplied by the factor $\frac{1}{2}$):

$(21, 21)_{66}$, $(21, 19)_{65}$, $(21, 17)_{64}$, $(21, 15)_{63}$, $(21, 13)_{62}$, $(21, 11)_{61}$, $(21, 9)_{60}$, $(21, 7)_{59}$, $(21, 5)_{58}$, $(21, 3)_{57}$, $(21, 1)_{56}$, $(19, 19)_{55}$, $(19, 17)_{54}$, $(19, 15)_{53}$, $(19, 13)_{52}$, $(19, 11)_{51}$, $(19, 9)_{50}$, $(19, 7)_{49}$, $(19, 5)_{48}$, $(19, 3)_{47}$, $(19, 1)_{46}$, $(17, 17)_{45}$, $(17, 15)_{44}$, $(17, 13)_{43}$, $(17, 11)_{42}$, $(17, 9)_{41}$, $(17, 7)_{40}$, $(17, 5)_{39}$, $(17, 3)_{38}$, $(17, 1)_{37}$, $(15, 15)_{36}$, $(15, 13)_{35}$, $(15, 11)_{34}$, $(15, 9)_{33}$, $(15, 7)_{32}$, $(15, 5)_{31}$, $(15, 3)_{30}$, $(15, 1)_{29}$, $(13, 13)_{28}$, $(13, 11)_{27}$, $(13, 9)_{26}$, $(13, 7)_{25}$, $(13, 5)_{24}$, $(13, 3)_{23}$, $(13, 1)_{22}$, $(11, 11)_{21}$, $(11, 9)_{20}$, $(11, 7)_{19}$, $(11, 5)_{18}$, $(11, 3)_{17}$, $(11, 1)_{16}$, $(9, 9)_{15}$, $(9, 7)_{14}$, $(9, 5)_{13}$, $(9, 3)_{12}$, $(9, 1)_{11}$, $(7, 7)_{10}$, $(7, 5)_9$, $(7, 3)_8$, $(7, 1)_7$, $(5, 5)_6$, $(5, 3)_5$, $(5, 1)_4$, $(3, 3)_3$, $(3, 1)_2$, $(1, 1)_1$. (Also see Figs. 7–9 for preceding diagrams.)

Sets of equivalent weights:

For the dominant weights given above there holds

(a) ${}_8(m_1, m_2)$, if $m_1 \neq m_2$, $m_2 \neq 0$,
(b) ${}_4(m_1, m_2)$, if $m_1 = m_2 \neq 0$ or $m_2 = 0$,
(c) ${}_1(m_1, m_2)$, if $m_1 = m_2 = 0$.

C. G_2

Simple roots:

$$\beta_1 = (0, -1, 1), \quad \beta_2 = (-1, 2, -1),$$
$$R = (3, -1, -2).$$

Weights:

$$m = (m_1, m_2, m_3), \quad m_1 + m_2 + m_3 = 0,$$
$$m_i \text{ integers.}$$

Again, as in the case of $SU(3)$, the patterns are given in the (p, q) notation,

$$\dim (p, q) = \tfrac{1}{1\,2\,0}(p + 4)(q + 1)(p + q + 5)\tfrac{1}{3}(2p + q + 9) \times \tfrac{1}{3}(p - q + 3)\tfrac{1}{3}(p + 2q + 6)$$

with

$$p \geq 0, \quad q \geq 0, \quad p \geq q \text{ (d.w. condition).}$$

[It should be remarked that, while to every (p, q)—p and q some nonnegative integer—there corresponds a representation in $SU(3)$, this is not the case for G_2. There are values p and q satisfying the d.w. condition in G_2 which do not correspond to a representation.]

Diagram for multiplicity of weights γ:

0	1	2	3	4	5	6	7	8	9	10	
	1										0
1			−1								1
	−1				1						2
			1					−1			4
					−1					1	5
									1	−1	6

Diagrams for branching multiplicity $\tilde{\gamma}$:

(1) Subgroup $SU(3)$: $L(p, q) = \frac{1}{3}(p - q, p + 2q)$:

0	1	2	3	4	5	6	
1							0
−1			−1				1
			1			1	3
						−1	4

(2) Subgroup $SU(2)_1$, $L(p, q) = \frac{1}{2}(p + q)$,
subgroup $SU(2)_2$, $L(p, q) = \frac{1}{6}(-p + q)$,
subgroup $SO(3)_1$, $L(p, q) = \frac{1}{3}(p - q)$,
subgroup $SO(3)_2$, $L(p, q) = \frac{1}{3}(4p + 5q)$.

Their diagram is (j); 1, $(j + 1)$; −1, or equivalently

0	1	
1	−1	0

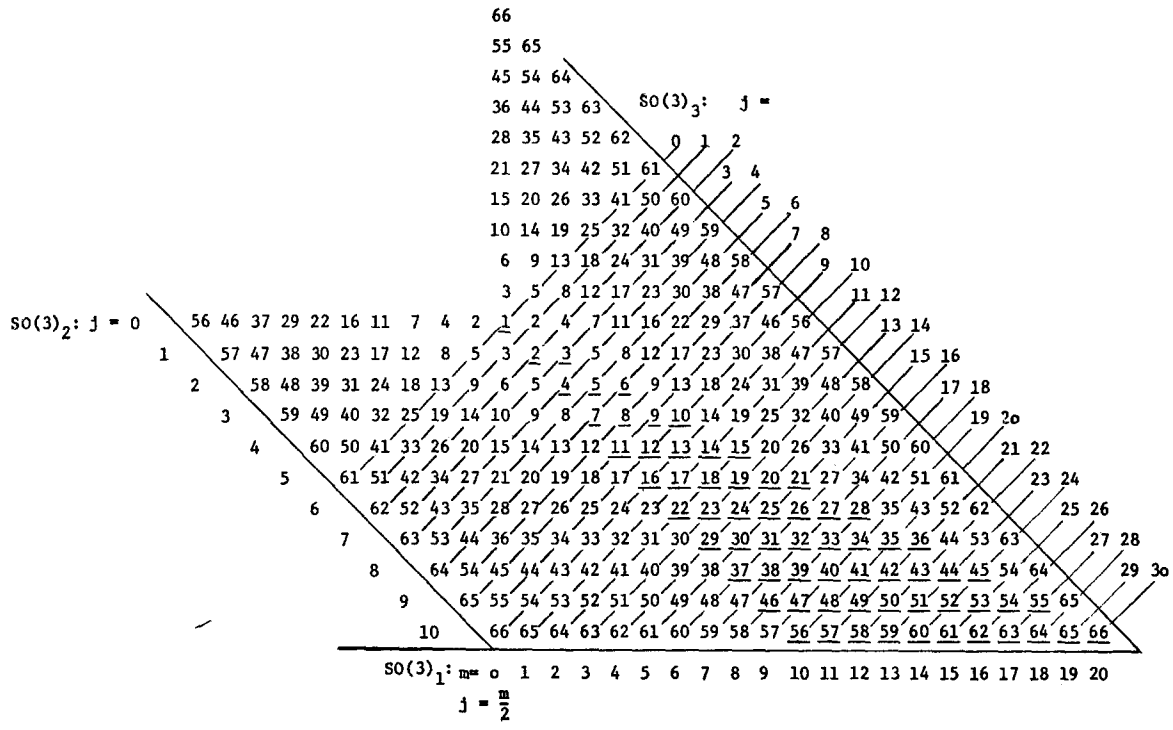

FIG. 5. Dominant weight number pattern for $SO(5)$. The representations $D(m_1, m_2)$, m_1, m_2 integer, are given in the text. The dominant weights j of the three subgroups $SO_1(3)$, $SO(3)_2$, and $SO(3)_3$ are given on the sides of the pattern. For the subgroup $SO(3)_2$ all weights of a line are projected by $L(m_1, m_2)$ onto the weight j which stands in the same line. For the subgroup $SO(3)_1$ all weights of a column are projected onto the weight j standing in that column. For the subgroup $SO(3)_3$ all weights along a straight line—as drawn in the pattern— are projected onto the weight j lying on that line.

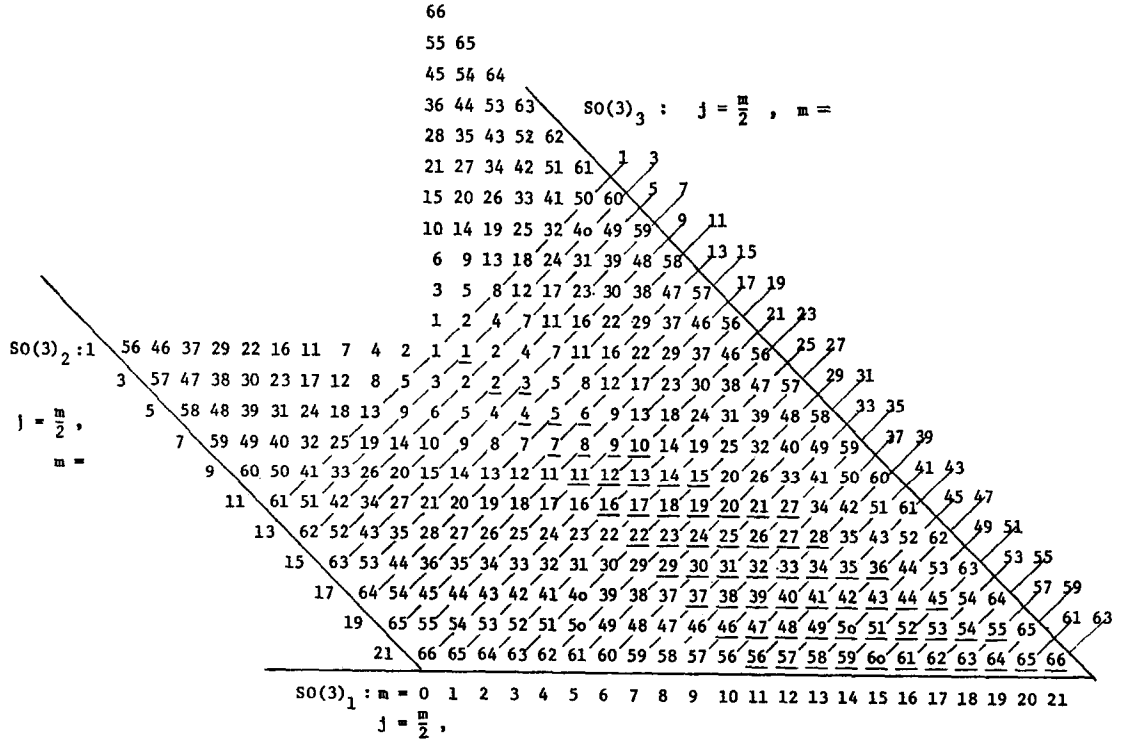

FIG. 6. As in Fig. 5. The representations are given in the text and correspond to $D(m_1, m_2)$ with m_1, m_2 half-integer.

$(0,0)_1$

$(1,-1)_3$, $(1,0)_2$, $(1,1)_3$

$(2,-2)_6$, $(2,-1)_5$, $(2,0)_4$, $(2,1)_5$, $(2,2)_6$

$(3,-3)_{10}$, $(3,-2)_9$, $(3,-1)_8$, $(3,0)_7$, $(3,1)_8$, $(3,2)_9$, $(3,3)_{10}$

$(4,-4)_{15}$, $(4,-3)_{14}$,$(4,-2)_{13}$,$(4,-1)_{12}$,$(4,0)_{11}$, $(4,1)_{12}$, $(4,2)_{13}$, $(4,3)_{14}$, $(4,4)_{15}$

$(5,-5)_{21}$, $(5,-4)_{20}$,$(5,-3)_{19}$,$(5,-2)_{18}$,$(5,-1)_{17}$,$(5,0)_{16}$, $(5,1)_{17}$, $(5,2)_{18}$, $(5,3)_{19}$, $(5,4)_{20}$, $(5,5)_{21}$

$(6,-6)_{28}$, $(6,-5)_{27}$,$(6,-4)_{26}$,$(6,-3)_{25}$,$(6,-2)_{24}$,$(6,-1)_{23}$,$(6,0)_{22}$, $(6,1)_{23}$, $(6,2)_{24}$, $(6,3)_{25}$, $(6,4)_{26}$, $(6,5)_{27}$, $(6,6)_{28}$

$(7,-7)_{36}$, $(7,-6)_{35}$,$(7,-5)_{34}$,$(7,-4)_{33}$,$(7,-3)_{32}$,$(7,-2)_{31}$,$(7,-1)_{30}$, $(7,0)_{29}$, $(7,1)_{30}$, $(7,2)_{31}$, $(7,3)_{32}$, $(7,4)_{33}$, $(7,5)_{34}$ $(7,6)_{35}$, $(7,7)_{36}$

$(8,-8)_{45}$, $(8,-7)_{44}$,$(8,-6)_{43}$,$(8,-5)_{42}$,$(8,-4)_{41}$,$(8,-3)_{40}$,$(8,-2)_{39}$, $(8,-1)_{38}$, $(8,0)_{37}$, $(8,1)_{38}$, $(8,2)_{39}$, $(8,3)_{40}$, $(8,4)_{41}$ $(8,5)_{42}$, $(8,6)_{43}$, $(8,7)_{44}$, $(8,8)_{45}$

$(9,-9)_{55}$, $(9,-8)_{54}$,$(9,-7)_{53}$,$(9,-6)_{52}$,$(9,-5)_{51}$,$(9,-4)_{50}$,$(9,-3)_{49}$, $(9,-2)_{48}$, $(9,-1)_{47}$, $(9,0)_{46}$, $(9,1)_{47}$, $(9,2)_{48}$, $(9,3)_{49}$ $(9,4)_{50}$, $(9,5)_{51}$, $(9,6)_{52}$, $(9,7)_{53}$, $(9,8)_{54}$, $(9,9)_{55}$

$(10,-10)_{66}$$(10,-9)_{65}$$(10-8)_{64}$,$(10,-7)_{63}$$(10,-6)_{62}$$(10,-5)_{61}$$(10,-4)_{60}$,$(10,-3)_{59}$,$(10,-2)_{58}$,$(10,-1)_{57}$,$(10,0)_{56}$, $(10,1)_{57}$,$(10,2)_{58}$ $(10,3)_{59}$, $(10,4)_{60}$, $(10,5)_{61}$, $(10,6)_{62}$, $(10,7)_{63}$, $(10,8)_{64}$, $(10,9)_{65}$, $(10,10)_{66}$

FIG. 7. This figure represents the mapping $L(m)$ of Fig. 5 onto the dominant weights of the $SO(4)$ subgroup of $SO(5)$. The dominant weight numbers, however, refer to $SO(5)$ [while the weights are those of $SO(4)$]. The branching law corresponding to the restriction of $SO(5)$ to its $SO(4)$ subgroup is very simple as well as very familiar. Figure 7 has been given for reasons of illustration and to achieve some kind of completeness. See also Fig. 11.

$(-5,-5)_{21}$, $(-5,-4)_{20}$, $(-5,-3)_{19}$, $(-5,-2)_{18}$, $(-5,-1)_{17}$, $(-5,0)_{16}$, $(-5,1)_{17}$, $(-5,2)_{18}$, $(-5,3)_{19}$, $(-5,4)_{20}$, $(-5,5)_{21}$

$(-4,-5)_{20}$, $(-4,-4)_{15}$, $(-4,-3)_{14}$, $(-4,-2)_{13}$, $(-4,-1)_{12}$, $(-4,0)_{11}$, $(-4,1)_{12}$, $(-4,2)_{13}$, $(-4,3)_{14}$, $(-4,4)_{15}$ $(-4,5)_{20}$

$(-3,-5)_{19}$, $(-3,-4)_{14}$, $(-3,-3)_{10}$, $(-3,-2)_9$, $(-3,-1)_8$, $(-3,0)_7$, $(-3,1)_8$, $(-3,2)_9$, $(-3,3)_{10}$ $(-3,4)_{14}$, $(-3,5)_{19}$

$(-2,-5)_{18}$, $(-2,-4)_{13}$, $(-2,-3)_9$, $(-2,-2)_6$, $(-2,-1)_5$, $(-2,0)_4$, $(-2,1)_5$, $(-2,2)_6$ $(-2,3)_9$, $(-2,4)_{13}$, $(-2,5)_{18}$

$(-1,-5)_{17}$, $(-1,-4)_{12}$, $(-1,-3)_8$, $(-1,-2)_5$, $(-1,-1)_3$, $(-1,0)_2$, $(-1,1)_3$ $(-1,2)_5$, $(-1,3)_8$, $(-1,4)_{12}$, $(-1,5)_{17}$

$(0,-5)_{16}$, $(0,-4)_{11}$, $(0,-3)_7$, $(0,-2)_4$, $(0,-1)_2$, $(0,0)_{\underline{1}}$ $(0,1)_2$, $(0,2)_4$, $(0,3)_7$, $(0,4)_{11}$, $(0,5)_{16}$

$(1,-5)_{17}$, $(1,-4)_{12}$, $(1,-3)_8$, $(1,-2)_5$, $(1,-1)_3$ $(1,0)_{\underline{2}}$, $(1,1)_{\underline{3}}$, $(1,2)_5$, $(1,3)_8$, $(1,4)_{12}$, $(1,5)_{17}$

$(2,-5)_{18}$, $(2,-4)_{13}$, $(2,-3)_9$, $(2,-2)_6$ $(2,-1)_5$, $(2,0)_{\underline{4}}$, $(2,1)_{\underline{5}}$, $(2,2)_{\underline{6}}$, $(2,3)_9$, $(2,4)_{13}$, $(2,5)_{18}$

$(3,-5)_{19}$, $(3,-4)_{14}$,$(3,-3)_{10}$ $(3,-2)_9$, $(3,-1)_8$, $(3,0)_{\underline{7}}$, $(3,1)_{\underline{8}}$, $(3,2)_{\underline{9}}$, $(3,3)_{\underline{10}}$, $(3,4)_{14}$, $(3,5)_{19}$

$(4,-5)_{20}$,$(4,-4)_{15}$ $(4,-3)_{14}$, $(4,-2)_{13}$, $(4,-1)_{12}$, $(4,0)_{\underline{11}}$, $(4,1)_{\underline{12}}$, $(4,2)_{\underline{13}}$, $(4,3)_{\underline{14}}$, $(4,4)_{\underline{15}}$, $(4,5)_{20}$

$(5,-5)_{21}$ $(5,-4)_{20}$, $(5,-3)_{19}$, $(5,-2)_{18}$, $(5,-1)_{17}$, $(5,0)_{16}$, $(5,1)_{17}$, $(5,2)_{18}$, $(5,3)_{19}$, $(5,4)_{20}$, $(5,5)_{21}$

FIG. 8. Subpattern of d.w. pattern given in Fig. 5. See Fig. 3.

$(-11,-11)_{21}$, $(-11,-9)_{20}$, $(-11,-7)_{19}$, $(-11,-5)_{18}$, $(-11,-3)_{17}$, $(-11,-1)_{16}$, $(-11,1)_{16}$, $(-11,3)_{17}$, $(-11,5)_{18}$, $(-11,7)_{19}$, $(-11,9)_{20}$ $(-11,11)_{21}$

$(-9,-11)_{20}$, $(-9,-9)_{15}$, $(-9,-7)_{14}$, $(-9,-5)_{13}$, $(-9,-3)_{12}$, $(-9,-1)_{11}$, $(-9,1)_{11}$, $(-9,3)_{12}$, $(-9,5)_{13}$, $(-9,7)_{14}$ $(-9,9)_{15}$, $(-9,11)_{20}$

$(-7,-11)_{19}$, $(-7,-9)_{14}$, $(-7,-7)_{10}$, $(-7,-5)_{9}$, $(-7,-3)_{8}$, $(-7,-1)_{7}$, $(-7,1)_{7}$, $(-7,3)_{8}$, $(-7,5)_{9}$ $(-7,7)_{10}$, $(-7,9)_{14}$, $(-7,11)_{19}$

$(-5,-11)_{18}$, $(-5,-9)_{13}$, $(-5,-7)_{9}$, $(-5,-5)_{6}$, $(-5,-3)_{5}$, $(-5,-1)_{4}$, $(-5,1)_{4}$, $(-5,3)_{5}$ $(-5,5)_{6}$, $(-5,7)_{9}$, $(-5,9)_{13}$, $(-5,11)_{18}$

$(-3,-11)_{17}$, $(-3,-9)_{12}$, $(-3,-7)_{8}$, $(-3,-5)_{5}$, $(-3,-3)_{3}$, $(-3,-1)_{2}$, $(-3,1)_{2}$ $(-3,3)_{3}$, $(-3,5)_{5}$, $(-3,7)_{8}$, $(-3,9)_{12}$, $(-3,11)_{17}$

$(-1,-11)_{16}$, $(-1,-9)_{11}$, $(-1,-7)_{7}$, $(-1,-5)_{4}$, $(-1,-3)_{2}$, $(-1,-1)_{1}$ $(-1,1)_{1}$, $(-1,3)_{2}$, $(-1,5)_{4}$, $(-1,7)_{7}$, $(-1,9)_{11}$, $(-1,11)_{16}$

$(1,-11)_{16}$, $(1,-9)_{11}$, $(1,-7)_{7}$, $(1,-5)_{4}$, $(1,-3)_{2}$ $(1,-1)_{1}$, $(1,1)_{\underline{1}}$, $(1,3)_{2}$, $(1,5)_{4}$, $(1,7)_{7}$, $(1,9)_{11}$, $(1,11)_{16}$

$(3,-11)_{17}$, $(3,-9)_{12}$, $(3,-7)_{8}$, $(3,-5)_{5}$ $(3,-3)_{3}$, $(3,-1)_{2}$, $(3,1)_{\underline{2}}$, $(3,3)_{\underline{3}}$, $(3,5)_{5}$, $(3,7)_{8}$, $(3,9)_{12}$, $(3,11)_{17}$

$(5,-11)_{18}$, $(5,-9)_{13}$, $(5,-7)_{9}$ $(5,-5)_{6}$, $(5,-3)_{5}$, $(5,-1)_{4}$, $(5,1)_{\underline{4}}$, $(5,3)_{\underline{5}}$, $(5,5)_{\underline{6}}$, $(5,7)_{9}$, $(5,9)_{13}$, $(5,11)_{18}$,

$(7,-11)_{19}$, $(7,-9)_{14}$ $(7,-7)_{10}$, $(7,-5)_{9}$, $(7,-3)_{8}$, $(7,-1)_{7}$, $(7,1)_{\underline{7}}$, $(7,3)_{\underline{8}}$, $(7,5)_{\underline{9}}$, $(7,7)_{\underline{10}}$, $(7,9)_{14}$, $(7,11)_{19}$.

$(9,-11)_{20}$ $(9,-9)_{15}$, $(9,-7)_{14}$, $(9,-5)_{13}$, $(9,-3)_{12}$, $(9,-1)_{11}$, $(9,1)_{\underline{11}}$, $(9,3)_{\underline{12}}$, $(9,5)_{\underline{13}}$, $(9,7)_{\underline{14}}$, $(9,9)_{\underline{15}}$, $(9,11)_{20}$

$(11,-11)_{21}$ $(11,-9)_{20}$ $(11,-7)_{19}$, $(11,-5)_{18}$, $(11,-3)_{17}$ $(11,-1)_{16}$ $(11,1)_{\underline{16}}$, $(11,3)_{\underline{17}}$ $(11,5)_{\underline{18}}$, $(11,7)_{\underline{19}}$, $(11,9)_{\underline{20}}$, $(11,11)_{\underline{21}}$

FIG. 9. Subpattern of d.w. pattern given in Fig. 6. See Fig. 3.

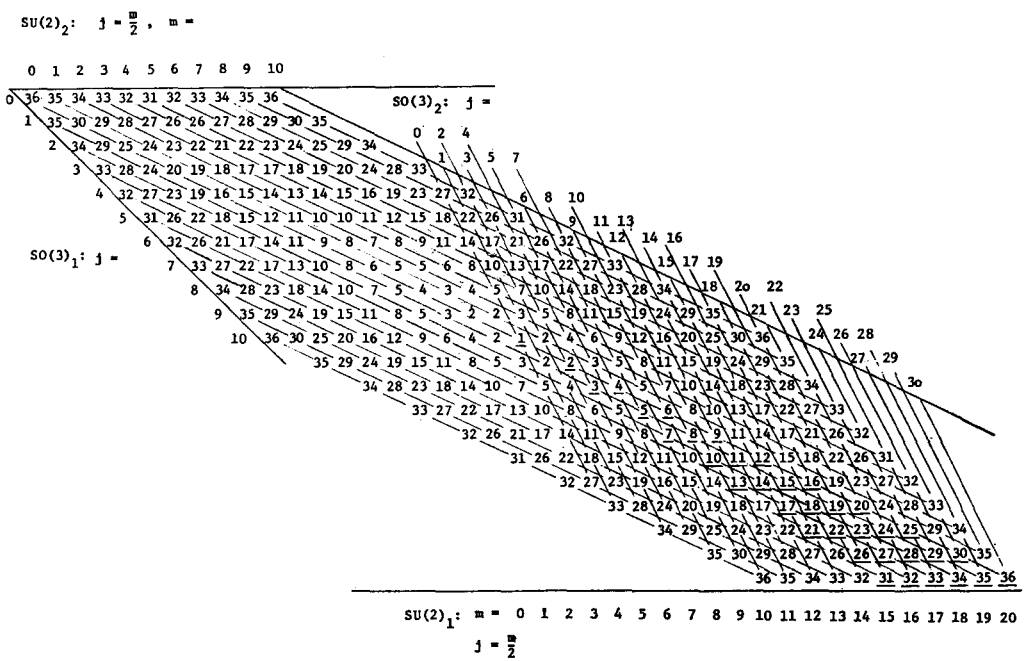

FIG. 10. Dominant weight number pattern for G_2. The representations $D(p, q)$ corresponding to the d.w. numbers are given in the text. The dominant weights of the four subgroups $SO(3)_1$, $SO(3)_2$, $SU(2)_1$ and $SU(2)_2$ are given on the sides of the pattern. For $SU(2)_1$ all weights (corresponding to d.w. numbers) of a column are projected by L onto the weight j of $SU(2)$, standing in the same column. For the other subgroups the projection is as indicated by the straight lines drawn in the pattern. (All weights along such a straight line are projected by L onto the weight j of the subgroup under consideration which lies on that straight line.)

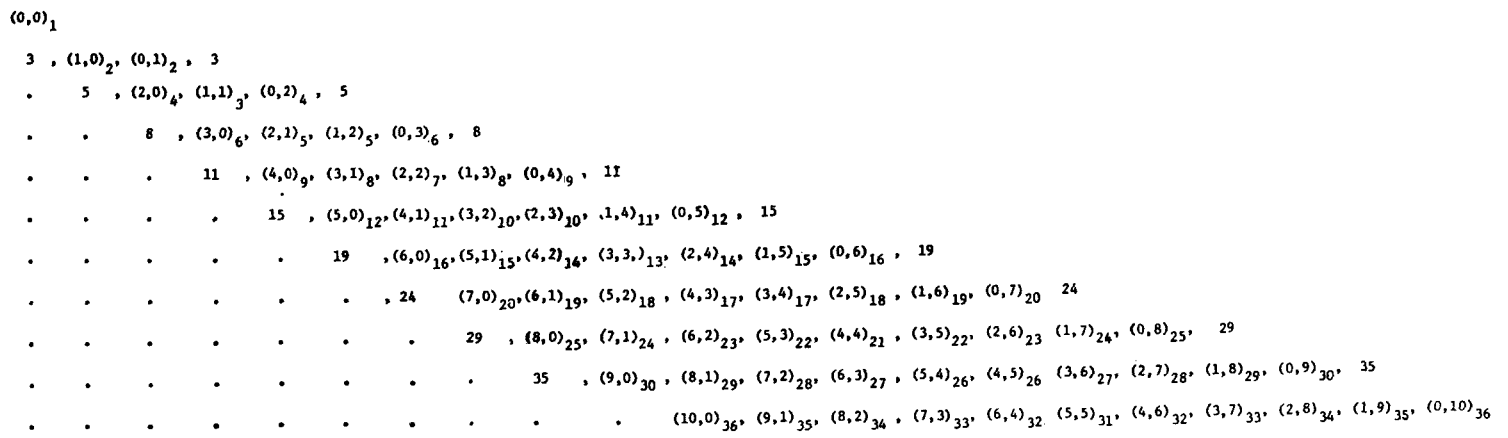

FIG. 11. This figure represents the mapping L of pattern Fig. 10 onto weights of the subgroup $SU(3)$. The weights of this pattern are weights of $SU(3)$, the d.w. numbers, however, refer to group G_2. It should be noted, that in distinction to Fig. 7, weights of G_2 contribute which are mapped onto nondominant weights of the subgroup $SU(3)$. (Namely, those weights of G_2 whose image in this pattern is characterized solely by the d.w. number of the mapped G_2 weight.)

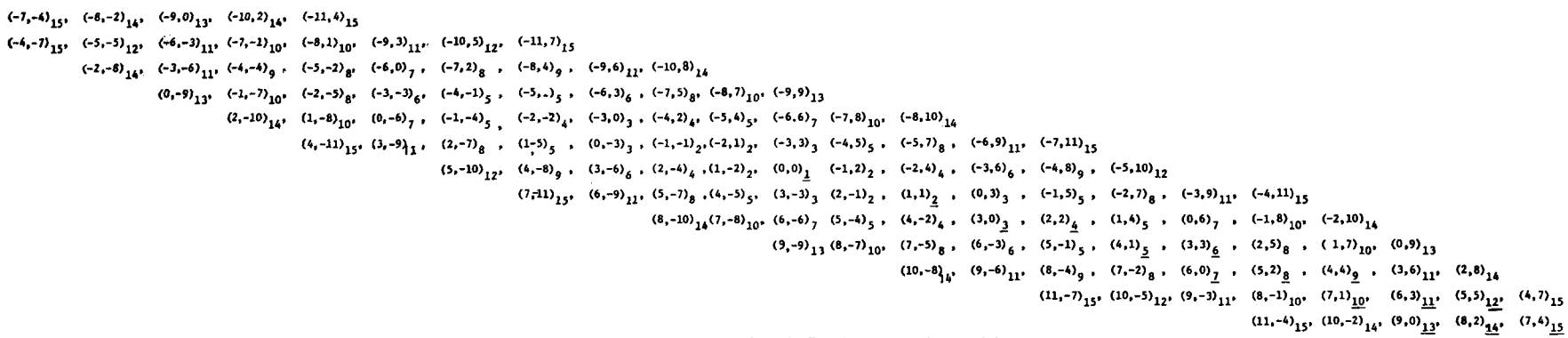

FIG. 12. Subpattern of Fig. 10. See figure caption of Fig. 3.

(3) Subgroup $SU(2)_1 \times SU(2)_2$:

$$L(p,q) = [\tfrac{1}{2}(p+q); \tfrac{1}{6}(-p+q)].$$

The diagram is given as

$$(j,j');1,\quad (j+1;j');-1,\quad (j;j'+1);-1,\quad (j+1;j'+1);1,$$

where j and j' refer to the two subgroups, respectively.

Multiplicity $\bar{\gamma}$:

0	$q+1$	$p+4$	$p+2q+6$	$2p+q+9$	$2p+2q+10$	
1	−1					0
−1		1				$\frac{1}{3}(p-q)+1$
	1		−1			$\frac{1}{3}(p+2q)+2$
		−1		1		$p+4$
			1		−1	$p+q+5$
				−1	1	$\frac{2}{3}(2p+q)+6$

Representations:

The representations of the pattern Fig. 11 are

$(10,10)_{36}$, $(11,8)_{35}$, $(12,6)_{34}$, $(13,4)_{33}$, $(14,2)_{32}$, $(15,0)_{31}$, $(9,9)_{30}$, $(10,7)_{29}$, $(11,5)_{28}$, $(12,3)_{27}$, $(13,1)_{26}$, $(8,8)_{25}$, $(9,6)_{24}$, $(10,4)_{23}$, $(11,2)_{22}$, $(12,0)_{21}$, $(7,7)_{20}$, $(8,5)_{19}$, $(9,3)_{18}$, $(10,1)_{17}$, $(6,6)_{16}$, $(7,4)_{15}$, $(8,2)_{14}$, $(9,0)_{13}$, $(5,5)_{12}$, $(6,3)_{11}$, $(7,1)_{10}$, $(4,4)_9$, $(5,2)_8$, $(6,0)_7$, $(3,3)_6$, $(4,1)_5$, $(2,2)_4$, $(3,0)_3$, $(1,1)_2$, $(0,0)_1$.

Sets of equivalent weights:

For the dominant weights given above there holds:

(a) ${}_{12}(p,q)$ if $p \neq 0, q \neq 0$, and $p \neq q$,
(b) ${}_{6}(p,q)$ if $p \neq 0, q \neq 0$, and $p = q$,
if $p \neq 0, q = 0$,
(c) ${}_{1}(p,q)$ if $p = q = 0$.

* Supported in part by the National Science Foundation, Washington, D.C. under NSF Grant GP-9623.

† Visiting member.

[1] The literature on the subject is very extensive, and it is rather doubtful whether any kind of completeness could be achieved by the author. We therefore prefer to quote a few publications on the subject which together cover a rather wide range of references.

[2] J.-P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226, 1560 (1965); J. McConnel, Proc. Roy. Irish Acad. **65A**, No. 1 (1966).

[3] (a) B. Preziosi, A. Simoni, and B. Vitale, Nuovo Cimento **34**, 1101 (1964); (b) A. J. Macfarlane, L. O'Raifeartaigh, and P. S. Rao, J. Math. Phys. **8**, 536 (1967).

[4] A. P. Stone, J. Math. Phys. **11**, 29 (1970).

[5] G. Racah, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach, New York, 1962).

[6] N. Straumann, Helv. Phys. Acta **38**, 481 (1966).

[7] (a) A. U. Klymyk, Academy of Sciences, Ukrainian SSR, Institute for Theoretical Physics, Kiev, No. 67-17, 1967; (b) R. M. Delaney and B. Gruber, J. Math. Phys. **10**, 252 (1969).

[8] For technical reasons, the boldface numerals are represented in the patterns by underlined numerals.

[9] Some may have multiplicity zero. The d.w. numbers greater than p obviously do not belong to this representation and are therefore to be ignored. The weights corresponding to these d.w. numbers may also be looked upon as having multiplicity zero.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Linear Integral Equations for a Certain Class of H-Functions Applicable to the Theory of Neutron Transport and Radiative Transfer*

E. E. BURNISTON
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607
AND
C. E. SIEWERT
Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina 27607

(Received 16 April 1970)

A matrix version of the classical Riemann–Hilbert problem defined on an open contour is discussed. The problem is reduced to a quasiregular integral equation for cases where the sufficient Hölder continuity condition is satisfied and the component indices are nonnegative. As an illustration of this procedure, linear integral equations (rather than the usual nonlinear forms), for Chandrasekhar's functions $H_l(\mu)$ and $H_r(\mu)$ are established in a form amenable to solution by numerical iteration.

I. INTRODUCTION

In general, the use of the singular eigenfunction expansion technique, introduced by Case[1] for treating problems in neutron transport theory to solve boundary value problems in "particle" transport analysis, requires solutions to singular integral equations in order to establish the various expansion coefficients.[2–4] Once the appropriate singular equations are developed, the methods of Muskhelishvili[5] can be used to convert the boundary value problem to an equivalent Riemann–Hilbert problem, and in many cases[2] closed-form solutions may be obtained.

Following Case's original paper[1] on the subject of singular eigenfunction expansions, the method has been extended to include many different models in neutron transport theory and radiative transfer. For example, the degenerate kernel model for energy transfer has been discussed by Mika,[6] the case of anisotropic scattering in 1-speed neutron transport theory has been thoroughly investigated by McCormick and Kuščer,[3] and several studies in "multigroup" theory have been reported by Siewert and Zweifel,[4,7] Siewert and Fraley,[8] Mourad and Siewert,[9] and Shultis.[10]

The usual procedure,[1,2] once the normal modes of the considered equation of transfer are established, is first to attempt the proof of a full-range expansion theorem. In developing this proof, the singular integral equations encountered can normally[2] be reduced to a special case of the inhomogeneous Riemann–Hilbert problem which can then be solved straightforwardly even for the case of matrices.[9]

The considerably more interesting half-range expansion theorem[2] cannot, in general, be established quite so readily; in fact, no constructive proofs for the multigroup or matrix models considered by Shultis[10] or Mourad and Siewert,[9] for example, have been reported. Although for some cases the proof of the half-range theorem applicable to matrix models has been converted to the need to solve systems of regular integral equations,[10,11] there have been no rigorous proofs of the existence of solutions to these equations; in some instances, however, this approach has been shown to be feasible computationally for nonmultiplying media.[12] Another approach used for half-range applications has been exhibited by Metcalf and Zweifel[13] and Mourad,[14] who have shown it possible to solve by numerical iteration the *singular* integral equations encountered in two different matrix problems.

A more direct method for solving half-range problems with the singular eigenfunction expansion technique is to pursue the homogeneous Riemann–Hilbert problem.[5] However, as Leonard and Ferziger[15] and Kuščer[16] have illustrated, multigroup models normally lead to a matrix form of the Riemann–Hilbert problem, and closed-form solutions are not generally available. We should thus like to discuss the analysis required to reduce these analytically formidable problems to forms computationally more feasible.

II. GENERAL ANALYSIS

For multigroup application of Case's method of normal modes,[1] the proof of the required half-range expansion theorem reduces to the need to solve a homogeneous Riemann–Hilbert problem for the normally called $\mathbf{X}$ matrix.[15] Here a matrix $\mathbf{X}(z)$, holomorphic in the complex plane cut from zero to one along the real line, is sought such that the boundary values from above (+) and below (−) the cut are related by

$$\mathbf{X}^{+}(\mu) = \mathbf{G}(\mu)\mathbf{X}^{-}(\mu), \quad \mu \in [0, 1], \tag{1}$$

where $\mathbf{G}(\mu)$ is a given matrix. We seek here the fundamental solution to Eq. (1), and thus $\mathbf{X}(z)$ and $\mathbf{X}^{\pm}(\mu)$ are required to be nonsingular in the finite plane.[5]

Since, in general, there exists no analytical solution to Eq. (1), we wish to make use of Muskhelishvili's theory[5] to convert Eq. (1) to a quasiregular linear integral equation for $\mathbf{X}^{-}(\mu)$, or alternatively for Chandrasekhar's **H**-matrix equivalent.[17]

If we now stipulate that $\mathbf{G}(\mu)$ obeys the Hölder condition[5] on the interval $[0, 1]$ and further that $\mathbf{G}(0) = \mathbf{G}(1) = \mathbf{I}$, **I** being the unit matrix, then we can write [see Ref. 5, p. 386, Eq. (126.5)]

$$\mathbf{X}^{-}(\mu) = \mathbf{X}_{\infty}(\mu) + \frac{1}{2\pi i}\int_0^1 [\mathbf{G}^{-1}(\mu)\mathbf{G}(\nu) - \mathbf{I}]\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu} + \frac{1}{2\pi i}[\mathbf{G}^{-1}(\mu) - \mathbf{I}]\int_{C_1}\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu}, \quad (2)$$

where the arbitrary arc C_1 has been added (with the proviso that there be a continuously turning tangent) to the real-line segment $[0, 1]$ to yield a closed contour C, as depicted in Fig. 1. Further, we have denoted the principal part of $\mathbf{X}(z)$ at infinity by $\mathbf{X}_{\infty}(z)$.

In order to establish Eq. (2), we have also defined $\mathbf{G}(\mu) \triangleq \mathbf{I}$ for $\mu \in C_1$. A similar procedure for closing the contour has been used by Leonard and Ferziger[15] and Kuščer,[16] though in the latter case a term due to the integral on C_1 appears to be missing [see Ref. 16, p. 267, Eq. (113)].

We note that Muskhelishvili's derivation[5] of the equation equivalent to our Eq. (2) was based on the proposition that the matrix $\mathbf{G}(\mu)$ was Hölder continuous on C, and that assumption is maintained here. Clearly, the fact that $\mathbf{G}(\mu)$ is taken to be a Hölder matrix is sufficient to ensure that Eq. (2) is quasiregular[5]; however, Leonard and Ferziger[15] applied Muskhelishvili's analysis without modification to a multigroup problem where the **G** matrix is not of the Hölder class, and the assertion that their equivalent to Eq. (2) is quasiregular is not immediately justified. [Clearly, simple continuity, as opposed to Hölder continuity, is not sufficient to ensure that Eq. (2) is quasiregular.]

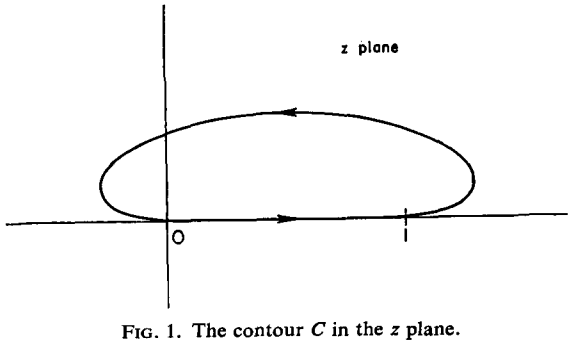

FIG. 1. The contour C in the z plane.

Since we are seeking to develop an integral equation for $\mathbf{X}^{-}(\mu)$ only on the real-line segment $\mu \in [0, 1]$, we rewrite Eq. (2) for the two cases $\mu \in [0, 1]$ and $\mu \in C_1$:

$$\mathbf{X}^{-}(\mu) = \mathbf{X}_{\infty}(\mu) + \frac{1}{2\pi i}\int_0^1 [\mathbf{G}^{-1}(\mu)\mathbf{G}(\nu) - \mathbf{I}]\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu} + \frac{1}{2\pi i}[\mathbf{G}^{-1}(\mu) - \mathbf{I}]\int_{C_1}\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu}, \quad \mu \in [0, 1], \quad (3a)$$

and

$$\mathbf{X}^{-}(\mu) = \mathbf{X}_{\infty}(\mu) + \frac{1}{2\pi i}\int_0^1 [\mathbf{G}(\nu) - \mathbf{I}]\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu}, \quad \mu \in C_1. \quad (3b)$$

Equation (3b) is clearly an explicit expression which relates $\mathbf{X}^{-}(\mu)$ for μ on C_1 to $\mathbf{X}_{\infty}(\mu)$ and values of $\mathbf{X}^{-}(\nu)$, where ν is confined to the real-line segment $[0, 1]$. This equation can thus be entered into the last term of Eq. (3a) to yield

$$\mathbf{X}^{-}(\mu) = \mathbf{X}_{\infty}(\mu) + \frac{1}{2\pi i}\int_0^1 [\mathbf{G}^{-1}(\mu)\mathbf{G}(\nu) - \mathbf{I}]\mathbf{X}^{-}(\nu)\frac{d\nu}{\nu - \mu} + \frac{1}{2\pi i}[\mathbf{G}^{-1}(\mu) - \mathbf{I}]\int_{C_1}\left(\mathbf{X}_{\infty}(\nu) + \frac{1}{2\pi i}\int_0^1 [\mathbf{G}(\nu') - \mathbf{I}]\mathbf{X}^{-}(\nu')\frac{d\nu'}{\nu' - \nu}\right)\frac{d\nu}{\nu - \mu}, \quad \mu \in [0, 1]. \quad (4)$$

We should now like to consider the repeated integral in the above equation and therefore introduce the definition

$$\mathbf{I}(\mu) \triangleq \int_{C_1}\left\{\int_0^1 [\mathbf{G}(\nu') - \mathbf{I}]\mathbf{X}^{-}(\nu')\frac{d\nu'}{\nu' - \nu}\right\}\frac{d\nu}{\nu - \mu}, \quad \mu \in (0, 1). \quad (5)$$

Since the inner integral in Eq. (5) is nonsingular [because **G** is Hölder on C and $\mathbf{G}(0) = \mathbf{G}(1) = \mathbf{I}$], we invert the order of integration to obtain

$$\mathbf{I}(\mu) = \int_0^1 \left([\mathbf{G}(\nu') - \mathbf{I}]\mathbf{X}^{-}(\nu')\int_{C_1}\frac{1}{\nu' - \nu}\frac{1}{\nu - \mu}d\nu\right)d\nu', \quad \mu \in (0, 1). \quad (6)$$

Performing now the integration over ν in Eq. (6), we find

$$\mathbf{I}(\mu) = \int_0^1 [\mathbf{G}(\nu') - \mathbf{I}]\left[\ln\left(\frac{1}{\nu'} - 1\right) - \ln\left(\frac{1}{\mu} - 1\right)\right]\mathbf{X}^{-}(\nu')\frac{d\nu'}{\nu' - \mu}, \quad \mu \in (0, 1). \quad (7)$$

Equation (7) may now be substituted for the repeated integral in Eq. (4) to yield

$$\mathbf{X}^-(\mu) = \mathbf{X}_\infty(\mu) + \frac{1}{2\pi i}[\mathbf{G}^{-1}(\mu) - \mathbf{I}]\int_{C_1} \mathbf{X}_\infty(\nu)\frac{d\nu}{\nu - \mu} + \frac{1}{2\pi i}\int_0^1 \mathbf{K}(\nu, \mu)\mathbf{X}^-(\nu)\frac{d\nu}{\nu - \mu}, \quad \mu \in [0, 1], \tag{8}$$

where $\mathbf{K}(\nu, \mu)$ is given by

$$\mathbf{K}(\nu, \mu) = \mathbf{G}^{-1}(\mu)\mathbf{G}(\nu) - \mathbf{I} + \frac{1}{2\pi i}[\mathbf{G}^{-1}(\mu) - \mathbf{I}] \times [\mathbf{G}(\nu) - \mathbf{I}]\left[\ln\left(\frac{1}{\nu} - 1\right) - \ln\left(\frac{1}{\mu} - 1\right)\right]. \tag{9}$$

Finally, the integral defined on C_1 in Eq. (8) may be written as

$$\int_{C_1} \mathbf{X}_\infty(\nu)\frac{d\nu}{\nu - \mu} = -P\int_0^1 \mathbf{X}_\infty(\nu)\frac{d\nu}{\nu - \mu} + \pi i\mathbf{X}_\infty(\mu), \tag{10}$$

so that the desired equation in terms only of variables on the line segment [0, 1] is obtained:

$$\mathbf{X}^-(\mu) = \mathbf{X}_\infty(\mu) + [\mathbf{G}^{-1}(\mu) - \mathbf{I}] \times \left(\tfrac{1}{2}\mathbf{X}_\infty(\mu) - \frac{1}{2\pi i}P\int_0^1 \mathbf{X}_\infty(\nu)\frac{d\nu}{\nu - \mu}\right) + \frac{1}{2\pi i}\int_0^1 \mathbf{K}(\nu, \mu)\mathbf{X}^-(\nu)\frac{d\nu}{\nu - \mu}, \quad \mu \in [0, 1]. \tag{11}$$

Equation (11) represents the basic version of the classical result [see Ref. 5, p. 386, Eq. (126.5)] modified for an open contour and is based on the proposition that the given G-matrix is Hölder continuous on the interval [0, 1] and further that $\mathbf{G}(0) = \mathbf{G}(1) = \mathbf{I}$.

It is clear that once $\mathbf{X}^-(\mu)$ is determined, as say from Eq. (11), $\mathbf{X}(z)$ follows immediately through the appropriate Cauchy integral; however, there remains the task of proving the equivalence between the original problem and Eq. (11), the ordinary integral equation for $\mathbf{X}^-(\mu)$. Furthermore, the solubility of Eq. (11) needs to be established in order to ensure that a solution to the original problem exists. It is stated by Muskhelishvili (Ref. 5, p. 389) that conditions sufficient for proving the required equivalence and solubility are that neither the accompanying problem,

$$\mathbf{\Psi}^+(\mu) = \mathbf{G}^{-1}(\mu)\mathbf{\Psi}^-(\mu), \quad \mu \in C, \tag{12}$$

nor the associate problem,

$$\mathbf{\Phi}^+(\mu) = [\mathbf{G}^{\mathrm{T}}(\mu)]^{-1}\mathbf{\Phi}^-(\mu), \quad \mu \in C, \tag{13}$$

has a nontrivial solution vanishing at infinity.[5] Here the transpose of $\mathbf{G}(\mu)$ is denoted by $\mathbf{G}^{\mathrm{T}}(\mu)$.

Although there seems to be no general method for calculating the so-called component indices[5] for the original Riemann–Hilbert problem given by Eq. (1), it appears that, for problems normally encountered in neutron transport theory and radiative transfer, the G matrix is such that its indices are nonnegative.[16] We thus restrict our attention to those problems for which the G matrix leads to nonnegative component indices. It now follows that the boundary value problems defined by $\mathbf{G}^{-1}$ and $[\mathbf{G}^T]^{-1}$ will have nonpositive component indices, and therefore the only solution of the accompanying or associate problems which vanishes at infinity must be the trivial solution.

III. QUASIREGULAR FREDHOLM EQUATIONS FOR $H_l(\mu)$ AND $H_r(\mu)$

We should like to apply the analysis of the previous section to two special cases pertinent to the study of polarized light in a free-electron atmosphere.[8,17] We thus seek solutions $X_1(z)$ and $X_2(z)$ of the Riemann–Hilbert problem given by Eq. (1) for the two scalar cases:

$$G_\alpha(\mu) = \frac{\Lambda_\alpha^+(\mu)}{\Lambda_\alpha^-(\mu)}, \quad \mu \in [0, 1], \quad \alpha = 1 \text{ or } 2, \tag{14}$$

with $\Lambda_\alpha^\pm(\mu)$ being the boundary values of the function

$$\Lambda_\alpha(z) = (-1)^\alpha + 3(1 - z^2)\left(1 + \tfrac{1}{2}z\int_{-1}^1 \frac{d\mu}{\mu - z}\right). \tag{15}$$

Since we require here the canonical solutions, i.e., solutions which are nonvanishing in the finite plane and which yield nonvanishing boundary values $X_\alpha^\pm(\mu)$, $\alpha = 1$ or 2, on the cut $\mu \in [0, 1]$, we follow Muskhelishvili[5] and first calculate the required indices $\aleph_1$ and $\aleph_2$ of the two problems:

$$\aleph_\alpha = \frac{1}{2\pi i}[\arg G_\alpha(\mu)]_C, \quad \alpha = 1 \text{ or } 2, \tag{16}$$

where $[\ ]_C$ is used to denote the increase of the function in brackets as the contour C is traversed in the positive direction. It is a simple matter[8] to show for the functions $G_\alpha(\mu)$ considered here that

$$\aleph_1 = 1 \tag{17a}$$

and

$$\aleph_2 = 0. \tag{17b}$$

Consequently, from the remarks of the previous section, it follows that the integral equation for $X_\alpha^-(\mu)$ is equivalent to Eq. (1) and, furthermore, that Eq. (11) is soluble, the solution being unique to within a multiplicative constant.

It follows from Eq. (17a) that $X_1(z)$ will have a simple zero at infinity, and thus $X_{1\infty}(\nu) \equiv 0$. In a similar manner, $X_{2\infty}(\nu)$ must be a constant, which we arbitrarily choose equal to unity in order to normalize our results in the established manner.[8] Since the principal parts of the functions $X_\alpha(\mu)$ have been determined, we can now write the forms of Eq. (11) appropriate here:

$$X_1^-(\mu) = \frac{1}{2\pi i}\int_0^1 K_1(\nu,\mu)X_1^-(\nu)\frac{d\nu}{\nu-\mu}, \quad \mu \in [0,1], \tag{18a}$$

and

$$X_2^-(\mu) = 1 + \frac{G_2^{-1}(\mu)-1}{2\pi i}\left[\pi i - \ln\left(\frac{1}{\mu}-1\right)\right] + \frac{1}{2\pi i}\int_0^1 K_2(\nu,\mu)X_2^-(\nu)\frac{d\nu}{\nu-\mu}, \quad \mu \in [0,1], \tag{18b}$$

where

$$K_\alpha(\nu,\mu) = G_\alpha^{-1}(\mu)G_\alpha(\nu) - 1 + \frac{1}{2\pi i}[G_\alpha^{-1}(\mu)-1] \times [G_\alpha(\nu)-1]\left[\ln\left(\frac{1}{\nu}-1\right) - \ln\left(\frac{1}{\mu}-1\right)\right]. \tag{19}$$

Equation (18a) is clearly a homogeneous equation for $X_1^-(\mu)$, and thus we wish to select a normalization consistent with that used previously, since the desired canonical solution is fixed only to within an arbitrary multiplicative constant. In the process of establishing the exact form of $X_1(z)$, Siewert and Fraley[8] normalized their solution such that

$$X_1(0) = \sqrt{5}. \tag{20}$$

Now setting $\mu = 0$ in Eq. (18a), we find

$$X_1^-(0) = X_1(0) = \frac{1}{2\pi i}\int_0^1 [G_1(\nu)-1]X_1^-(\nu)\frac{d\nu}{\nu}, \tag{21}$$

which, when the explicit forms of $G_1(\nu)$ and Eq. (20) are used, yields the identity

$$\int_0^1 \Psi_l(\nu)\frac{X_1^-(\nu)}{\Lambda_1^-(\nu)}\,d\nu = \frac{\sqrt{5}}{2}, \tag{22}$$

where we have introduced Chandrasekhar's characteristic function[17]

$$\Psi_l(\nu) = \tfrac{3}{4}(1-\nu^2). \tag{23}$$

We thus seek a solution to the homogeneous Eq. (18a) such that $X_1(0) = \sqrt{5}$ or, alternatively, such that it is subject to the integral constraint given by Eq. (22).

Rather than pursue the analysis for $X_1^-(\mu)$, we prefer to write our equations in terms of Chandrasekhar's function[17] $H_l(\mu)$,

$$H_l(\mu) = \frac{2\sqrt{5}}{5}\frac{X_1^-(\mu)}{\Lambda_1^-(\mu)}, \quad \mu \in [0,1], \tag{24}$$

and thus convert Eq. (18a) to the equivalent

$$H_l(\mu) = \frac{1}{2\pi i}\int_0^1 \frac{\Lambda_1^-(\nu)}{\Lambda_1^-(\mu)}K_1(\nu,\mu)H_l(\nu)\frac{d\nu}{\nu-\mu}. \tag{25}$$

Furthermore, the normalization $H_l(0) = 1$ follows from Eqs. (15), (20), and (24), whereas the alternative integral constraint follows from Eq. (22):

$$\int_0^1 \Psi_l(\nu)H_l(\nu)\,d\nu = 1. \tag{26}$$

In the process of simplifying the algebra once the explicit form of $\Lambda_1(z)$ is used in Eq. (25), it becomes possible to split off a term proportional to the integral given in Eq. (26), and thus by using that identity we are able to convert the homogeneous integral equation for $H_l(\mu)$ to an inhomogeneous form, the solution of which necessarily is normalized in the desired manner. We find finally that $H_l(\mu)$ is the solution to the Fredholm equation

$$H_l(\mu) = 2g_1(\mu)(2-3\mu^2) + \frac{9g_1(\mu)}{4}\int_0^1 \left(\frac{\nu\mu(1-\nu^2)(1-\mu^2)\Delta(\mu,\nu)}{\nu-\mu} + \frac{2\mu(\nu+\mu)}{3}\right)H_l(\nu)\,d\nu, \tag{27}$$

where

$$g_\alpha(\mu) = [\Lambda_\alpha^+(\mu)\Lambda_\alpha^-(\mu)]^{-1}, \quad \alpha = 1 \text{ or } 2, \tag{28a}$$

$$\Lambda_\alpha^\pm(\mu) = (-1)^\alpha + 3(1-\mu^2)(1-\mu\tanh^{-1}\mu) \pm \tfrac{3}{2}\pi i\mu(1-\mu^2), \quad \alpha = 1 \text{ or } 2, \tag{28b}$$

and

$$\Delta(\mu,\nu) = \ln\left(\frac{(1-\mu)^2}{\mu(1+\mu)}\right) - \ln\left(\frac{(1-\nu)^2}{\nu(1+\nu)}\right). \tag{28c}$$

For the sake of brevity, we state simply that the expressions

$$X_{2\infty}(\nu) = 1, \tag{29a}$$

$$X_2(0) = \sqrt{2}, \tag{29b}$$

$$H_r(\nu) = 2\sqrt{2}\frac{X_2^+(\nu)}{\Lambda_2^+(\nu)}, \tag{29c}$$

and

$$\int_0^1 \Psi_r(\nu)H_r(\nu)\,d\nu = 1 - \tfrac{1}{2}\sqrt{2}, \tag{29d}$$

with

$$\Psi_r(\nu) = \tfrac{3}{8}(1-\nu^2), \tag{29e}$$

can be used in a manner analogous to that used for the $\alpha = 1$ case to find the resulting version of Eq. (18b) for $H_r(\mu)$:

$$H_r(\mu) = g_2(\mu)\left[3\sqrt{2}\,\mu(1-\mu^2)\ln\left(\frac{(1-\mu)^2}{\mu(1+\mu)}\right) + 4(4-3\mu^2)\right] + \frac{9g_2(\mu)}{4}\int_0^1\left(\frac{\nu\mu(1-\nu^2)(1-\mu^2)\Delta(\mu,\nu)}{\nu-\mu} - \frac{2\mu(\nu+\mu)}{3}\right)H_r(\nu)\,d\nu. \quad (30)$$

Equations (27) and (30) clearly are not so concise as Chandrasekhar's nonlinear equations[17]:

$$H_\alpha(\mu) = 1 + \mu H_\alpha(\mu)\int_0^1 \Psi_\alpha(\nu)H_\alpha(\nu)\frac{d\nu}{\nu+\mu}, \qquad \alpha = l \text{ or } r. \quad (31)$$

Furthermore, for the case of scalar Riemann–Hilbert problems, exact analytical solutions are available; however, the extension to matrices cannot be made analytically, whereas it is felt that the method employed here may be used to advantage for certain classes of matrix Riemann–Hilbert problems.

ACKNOWLEDGMENT

The authors wish to acknowledge the contribution made by T. W. Mullikin of Purdue University to this work. Appreciation is also extended to J. T. Kriese who has established that the developed equations for $H_l(\mu)$ and $H_r(\mu)$ can be solved numerically to yield known results.

* This work was supported in part by the National Science Foundation through Grant GK 11935 and the Joint Services Advisory Group through Grant AFOSR-69-1779.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[3] N. J. McCormick and I. Kuščer, J. Math. Phys. **6**, 1939 (1965).
[4] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[5] N. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[6] J. Mika, Nucl. Sci. Eng. **22**, 235 (1965).
[7] C. E. Siewert and P. F. Zweifel, J. Math. Phys. **7**, 2092 (1966).
[8] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[9] S. A. Mourad and C. E. Siewert, Astrophys. J. **155**, 555 (1969).
[10] J. K. Shultis, Nucl. Sci. Eng. **38**, 83 (1969).
[11] R. Zelazny and A. Kuszell, Ann. Phys. (N.Y.) **16**, 81 (1961).
[12] J. K. Shultis, Ph.D. thesis, University of Michigan, 1968.
[13] D. R. Metcalf and P. F. Zweifel, Nucl. Sci. Eng. **33**, 307 (1968).
[14] S. A. Mourad, Ph.D. thesis, North Carolina State University, 1970.
[15] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 181 (1966).
[16] I. Kuščer, in *Developments In Transport Theory*, edited by E. Inönü and P. F. Zweifel (Academic, New York, 1967).
[17] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Exact Recursion Relation for $2 \times N$ Arrays of Dumbbells

R. B. McQuistan and S. J. Lichtman*†
Physics Department and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

(Received 9 March 1970)

It is shown that $A(q, N)$, the number of ways of arranging q indistinguishable dumbbells on a $2 \times N$ rectangular array of compartments, is exactly described by the recursion relation

$$A(q, N) = A(q, N-1) + 2A(q-1, N-1) + A(q-1, N-2) - A(q-3, N-3).$$

For large values of N the normalization of the distribution generated by this recursion relation is found to be $0.665(3.214)^N$ and the maximum number of arrangements occurs when the array is approximately 61% occupied.

I. INTRODUCTION

There are important aspects of a number of physical phenomena, e.g., magnetism, adsorption, crystallization, which can be treated by considering the occupation statistics of a regular space lattice or array. One of the interesting problems arising from this approach is that of determining the number of possible arrangements of dumbbells on a lattice space. Here the lattice space is considered to be a rectangular array of compartments and the dumbbells occupying two adjacent compartments. A 2-dimensional form of this problem is encountered in the theory of adsorption of diatomic molecules.[1]

As is generally true for problems of this nature, exact solutions have been found for the 1-dimensional case,[2–5] but exact solutions for spaces of higher-order

can be used in a manner analogous to that used for the $\alpha = 1$ case to find the resulting version of Eq. (18b) for $H_r(\mu)$:

$$H_r(\mu) = g_2(\mu)\left[3\sqrt{2}\,\mu(1-\mu^2)\ln\left(\frac{(1-\mu)^2}{\mu(1+\mu)}\right) + 4(4-3\mu^2)\right] + \frac{9g_2(\mu)}{4}\int_0^1\left(\frac{\nu\mu(1-\nu^2)(1-\mu^2)\Delta(\mu,\nu)}{\nu-\mu} - \frac{2\mu(\nu+\mu)}{3}\right)H_r(\nu)\,d\nu. \quad (30)$$

Equations (27) and (30) clearly are not so concise as Chandrasekhar's nonlinear equations[17]:

$$H_\alpha(\mu) = 1 + \mu H_\alpha(\mu)\int_0^1 \Psi_\alpha(\nu)H_\alpha(\nu)\frac{d\nu}{\nu+\mu}, \qquad \alpha = l \text{ or } r. \quad (31)$$

Furthermore, for the case of scalar Riemann–Hilbert problems, exact analytical solutions are available; however, the extension to matrices cannot be made analytically, whereas it is felt that the method employed here may be used to advantage for certain classes of matrix Riemann–Hilbert problems.

ACKNOWLEDGMENT

The authors wish to acknowledge the contribution made by T. W. Mullikin of Purdue University to this work. Appreciation is also extended to J. T. Kriese who has established that the developed equations for $H_l(\mu)$ and $H_r(\mu)$ can be solved numerically to yield known results.

* This work was supported in part by the National Science Foundation through Grant GK 11935 and the Joint Services Advisory Group through Grant AFOSR-69-1779.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[3] N. J. McCormick and I. Kuščer, J. Math. Phys. **6**, 1939 (1965).
[4] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[5] N. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[6] J. Mika, Nucl. Sci. Eng. **22**, 235 (1965).
[7] C. E. Siewert and P. F. Zweifel, J. Math. Phys. **7**, 2092 (1966).
[8] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[9] S. A. Mourad and C. E. Siewert, Astrophys. J. **155**, 555 (1969).
[10] J. K. Shultis, Nucl. Sci. Eng. **38**, 83 (1969).
[11] R. Zelazny and A. Kuszell, Ann. Phys. (N.Y.) **16**, 81 (1961).
[12] J. K. Shultis, Ph.D. thesis, University of Michigan, 1968.
[13] D. R. Metcalf and P. F. Zweifel, Nucl. Sci. Eng. **33**, 307 (1968).
[14] S. A. Mourad, Ph.D. thesis, North Carolina State University, 1970.
[15] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 181 (1966).
[16] I. Kuščer, in *Developments In Transport Theory*, edited by E. Inönü and P. F. Zweifel (Academic, New York, 1967).
[17] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Exact Recursion Relation for $2 \times N$ Arrays of Dumbbells

R. B. McQuistan and S. J. Lichtman*†
Physics Department and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

(Received 9 March 1970)

It is shown that $A(q, N)$, the number of ways of arranging q indistinguishable dumbbells on a $2 \times N$ rectangular array of compartments, is exactly described by the recursion relation

$$A(q, N) = A(q, N-1) + 2A(q-1, N-1) + A(q-1, N-2) - A(q-3, N-3).$$

For large values of N the normalization of the distribution generated by this recursion relation is found to be $0.665(3.214)^N$ and the maximum number of arrangements occurs when the array is approximately 61% occupied.

I. INTRODUCTION

There are important aspects of a number of physical phenomena, e.g., magnetism, adsorption, crystallization, which can be treated by considering the occupation statistics of a regular space lattice or array. One of the interesting problems arising from this approach is that of determining the number of possible arrangements of dumbbells on a lattice space. Here the lattice space is considered to be a rectangular array of compartments and the dumbbells occupying two adjacent compartments. A 2-dimensional form of this problem is encountered in the theory of adsorption of diatomic molecules.[1]

As is generally true for problems of this nature, exact solutions have been found for the 1-dimensional case,[2–5] but exact solutions for spaces of higher-order

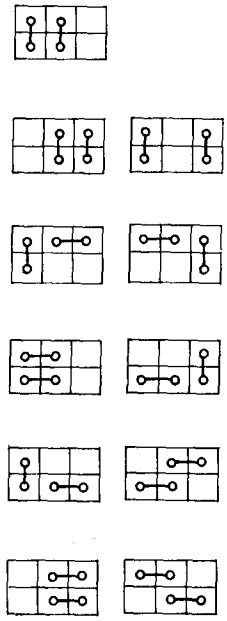

FIG. 1. There are 11 possible ways of arranging two indistinguishable dumbbells on a 2×3 array.

dimensionality have been obtained only for very special cases,[6,7] i.e., a 2-dimensional array completely covered with dumbbells. Consequently, approximation methods have been utilized[8–10] to attack this problem.

In the present paper we are concerned with the occupation statistics for dumbbells on a pseudo-2-dimensional rectangular array—the $2 \times N$ array. We wish to determine the number of ways of arranging dumbbells on a $2 \times N$ rectangular array of compartments (see Fig. 1).

II. DERIVATION OF RECURSION RELATION

In this section we seek to determine a recursion relation for $A(q, N)$, the number of arrangements of q indistinguishable dumbbells on a $2 \times N$ rectangular array of compartments. For the purposes of this calculation we shall define the following arrays: (i) An $\alpha(N)$ array (see Fig. 2a) is defined to be an array of compartments arranged in two adjacent, aligned rows of N compartments each; (ii) a $\beta(N)$ array (see Fig. 2b) is one in which the compartments are arranged in two adjacent, aligned rows; one row of N compartments and one row of $N+1$ compartments.

$A(q, N)$ is then the number of ways of arranging q indistinguishable dumbbells on an $\alpha(N)$ array, and we define $B(q, N)$ as the number of ways q indistinguishable dumbbells may be arranged on a $\beta(N)$ array.

Theorem 1:

$$B(q, N) = A(q, N) + B(q-1, N-1). \quad (1)$$

Proof: Let $b(q, N)$ be the set of all possible arrangements of q indistinguishable dumbbells on a $\beta(N)$ array; $c(q, N)$ is the subset of $b(q, N)$ in which the extra compartment is vacant, and $d(q, N)$ is the subset of $b(q, N)$ in which the extra compartment is occupied. Then every arrangement in $c(q, N)$ differs from every arrangement in $d(q, N)$ by the condition of occupation of the extra compartment, i.e., $c(q, N) \cap d(q, N) = \phi$, a null set. In addition, every member of $b(q, N)$ will be found either in $c(q, N)$ or $d(q, N)$, i.e., $c(q, N) \cup d(q, N) = b(q, N)$.

We conclude that $\#b(q, N)$, the number of members of the set $b(q, N)$, is given by

$$\#b(q, N) = \#c(q, N) + \#d(q, N) \equiv B(q, N). \quad (2)$$

The extra compartment of the β array is unoccupied in the set $c(q, N)$ so that by definition $\#c(q, N) \equiv A(q, N)$. If the extra compartment is occupied, then the adjacent compartment in the same row is also occupied. Hence, all other possible arrangements must involve the remaining $q-1$ dumbbells on the remainder of the array, which is a $\beta(N-1)$ array. The number of elements in $d(q, N)$ is therefore $B(q-1, N-1)$, i.e., $\#d(q, N) \equiv B(q-1, N-1)$. The theorem then follows from Eq. (2).

Corollary 1:

$$B(q, N) = \sum_{j=0}^{q} A(q-j, N-j). \quad (3)$$

Proof: Use Theorem 1 to evaluate $B(q-1, N-1)$, i.e.,

$$B(q-1, N-1) = A(q-1, N-1) + B(q-2, N-2). \quad (4)$$

Substitution of this into the theorem yields

$$B(q, N) = A(q, N) + A(q-1, N-1) + B(q-2, N-2).$$

Repeated use of Eq. (4) gives

$$\begin{aligned} B(q, N) &= A(q, N) + A(q-1, N-1) + \cdots + A(0, N-q) \\ &= \sum_{j=0}^{q} A(q-j, N-j). \end{aligned} \quad (5)$$

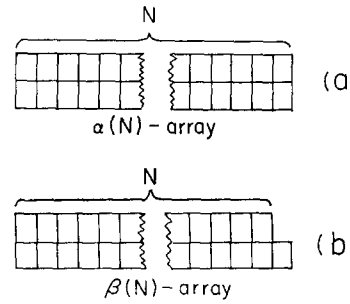

FIG. 2 (a) An $\alpha(N)$ array; (b) A $\beta(N)$ array.

Theorem 2:

$$A(q, N) = A(q, N-1) + 2B(q-1, N-2) + A(q-1, N-1) + A(q-2, N-2). \quad (6)$$

Proof: Let $a(q, N)$ be the set of all possible arrangements of q dumbbells on an $\alpha(N)$ array and let $e_1(q, N)$, $e_2(q, N), \cdots, e_5(q, N)$ be subsets of $a(q, N)$ in which the Nth column of the array is occupied in a manner shown in Fig. 3. In other words, the $e_k(q, N)$ are defined on the basis of the manner in which the two compartments forming the Nth column are occupied. Since every member of $e_k(q, N)$ differs from any and every member of $e_j(q, N)(k \neq j)$, we state that $e_j(q, N) \cap e_k(q, N) = \phi$, $k \neq j$. In addition, these five configurations are clearly the only possible Nth column configurations; thus

$$\bigcup_{k=1}^{5} e_k(q, N) = a(q, N).$$

We conclude that

$$\#a(q, N) = \#e_1(q, N) + \#e_2(q, N) + \#e_3(q, N) + \#e_4(q, N) + \#e_5(q, N) \equiv A(q, N). \quad (7)$$

The set $e_1(q, N)$ contains only those arrangements in which the Nth column is unoccupied. All q dumbbells are then arranged on the remaining $\alpha(N-1)$ array; hence $\#e_1(q, N) \equiv A(q, N-1)$.

Both sets $e_2(q, N)$ and $e_3(q, N)$ have one compartment of the Nth column occupied and one compartment empty. Necessarily, this implies that one compartment of the $(N-1)$th column is also occupied. In each of these sets the remaining $q-1$ dumbbells are arranged on an array composed of the original array minus the three precluded compartments, i.e., on a $\beta(N-2)$ array. We may then write $\#e_2(q, N) = \#e_3(q, N) \equiv B(q-1, N-2)$.

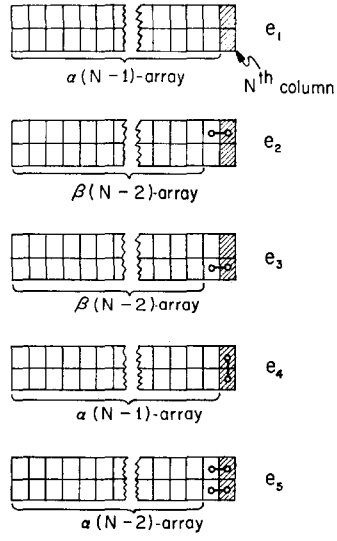

FIG. 3. The state of occupation of the Nth column defines the subsets e_1, e_2, e_3, e_4, and e_5.

One dumbbell covers both compartments of the Nth column in the set $e_4(q, N)$. Thus, the remaining $q-1$ dumbbells can be arranged on a $\alpha(N-1)$ array, i.e., $\#e_4(q, N) \equiv A(q-1, N-1)$.

In the set $e_5(q, N)$ two dumbbells cover the Nth column [and also the $(N-1)$th column]. The remaining $q-2$ dumbbells are arranged on the $\alpha(N-2)$ array; thus $\#e_5(q, N) \equiv A(q-2, N-2)$. Application of Eq. 7 yields Theorem 2.

Corollary 2:

$$A(q, N) = A(q, N-1) + 2A(q-1, N-1) + A(q-1, N-2) - A(q-3, N-3). \quad (8)$$

Proof: Use Corollary 1 to evaluate $B(q-1, N-2)$ in Theorem 2, i.e.,

$$A(q, N) = A(q, N-1) + 2\sum_{j=0}^{q-1} A(q-1-j, N-2-j) + A(q-1, N-1) + A(q-2, N-2). \quad (9)$$

If Eq. (9) is used to evaluate $A(q-1, N-1)$ and we form the difference $A(q, N) - A(q-1, N-1)$, we obtain Corollary 2 by noting that

$$A(q-1, N-2) = \sum_{j=0}^{q-1} A(q-1-j, N-2-j) - \sum_{j=0}^{q-2} A(q-2-j, N-3-j). \quad (10)$$

Corollary 3:

$$A(q, q) = f_q, \quad (11)$$

where f_q is the qth Fibonacci number.

Proof: Since, if $q > N$, no arrangements are possible, i.e., $A(q, N) = 0$, the special case in which the array is completely filled has the recursion

$$A(q, N) = 2A(q-1, N-1) - A(q-3, N-3)$$

or

$$A(q, q) = 2A(q-1, q-1) - A(q-3, q-3). \quad (12)$$

The initial conditions $A(0, 0) = A(1, 1) = 1$ and the use of Eq. (12) yield the Fibonacci sequence. Thus, the number of arrangements possible for $q = \frac{1}{2}N$

N \ q	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1									
2	1	4	2								
3	1	7	11	3							
4	1	10	29	26	5						
5	1	13	56	94	56	8					
6	1	16	92	234	263	114	13				
7	1	19	137	473	815	667	223	21			
8	1	22	191	838	1982	2504	1577	424	34		
9	1	25	254	1356	4115	7191	7018	3538	789	55	
10	1	28	326	2054	7646	17266	23431	18336	7622	1444	89

FIG. 4. The number of arrangements when indistinguishable dumbbells are placed on a $2 \times N$ array for N and q in the range 0–10.

dumbbells on a filled array is the qth Fibonacci number.

Figure 4 shows the number of arrangements of q indistinguishable dumbbells on a $2 \times N$ array for q and N in the range 0–10, according to Eq. (8).

III. NORMALIZATION

In this section we attempt to determine

$$\Delta_N \equiv \sum_{q=0}^{N} A(q, N), \tag{13}$$

the normalization of the statistics defined by the recursion given in Eq. (8). By Corollary 2 we find the recursion relation for Δ_N to be

$$\Delta_N = 3\Delta_{N-1} + \Delta_{N-2} - \Delta_{N-3} \tag{14}$$

with the initial conditions $\Delta_0 = 1$ and $\Delta_1 = 2$. Utilizing the results of Zeitlin,[11] we find the generating function of Δ_N to be

$$\frac{1 - x}{1 - 3x - x^2 + x^3} = \sum_{N=0}^{\infty} \Delta_N x^N. \tag{15}$$

This generating function may be rewritten as

$$\frac{1 - x}{1 - 3x - x^2 + x^3} = \frac{k_1}{1 - S_1 x} + \frac{k_2}{1 - S_2 x} + \frac{k_3}{1 - S_3 x}, \tag{16}$$

where the k's are constants and the S's are the zeros of $x^3 - 3x^2 - x + 1$, i.e.,

$$S_1 = 3.214320, \quad S_2 = 0.460811, \quad S_3 = -0.675131.$$

Thus, the equivalent generating function may be expanded as

$$\frac{k_1}{1 - S_1 x} + \frac{k_2}{1 - S_2 x} + \frac{k_3}{1 - S_3 x} = \sum_{N=0}^{\infty} (k_1 S_1^N + k_2 S_2^N + k_3 S_3^N) x^N. \tag{17}$$

Since the absolute values of S_2 and S_3 are less than unity, S_2^N and S_3^N approach zero as $N \to \infty$. Thus in the limit we have

$$\lim_{N \to \infty} \Delta_N = k_1 S_1^N \tag{18}$$

As $x \to S_1^{-1}$, only the first term on the right-hand side of Eq. (16) is important, so that we may write

$$\lim_{x \to S_1^{-1}} \left(\frac{1 - x}{1 - 3x - x^2 + x^3} - \frac{k_1}{1 - S_1 x} \right) = 0. \tag{19}$$

Utilizing L'Hospital's rule, we may determine k_1:

$$k_1 = \left(1 - \frac{1}{S_1}\right) \left(\frac{S_1}{\frac{3}{S_1^2} - \frac{2}{S_1} - 3} \right) \simeq 0.665. \tag{20}$$

Thus we may write

$$\lim_{N \to \infty} \Delta_N = k_1 S_1^N \simeq 0.665(3.214)^N. \tag{21}$$

This normalization may be compared with that obtained for single-particle statistics, 2^N, and with 1-dimensional dumbbell statistics[3]

$$\sum_{q=0}^{N/2} \binom{N - q}{q} = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^N - \left(\frac{1 - \sqrt{5}}{2} \right)^N \right],$$

which for large N becomes $0.447\gamma^N$, where γ is the golden mean (1.618).

IV. EVALUATION OF THE MEAN VALUE (EXPECTATION) FOR LARGE N

In this section we attempt to determine θ, the coverage $[\theta \equiv q/N]$ at which the distribution is a maximum. Here we define the mean value μ_N to be

$$\mu_N \equiv \sum_{q=0}^{N} qA(q, N) \Big/ \sum_{q=0}^{N} A(q, N) \tag{22}$$

and assume that

$$\lim_{N \to \infty} \mu_N = k_0 N \lim_{N \to \infty} \Delta_N = k_0 N k_1 (S_1)^N, \tag{23}$$

where k_0 is a constant.

By Corollary 2 we have

$$\begin{aligned} \sum_{q=0}^{N} qA(q, N) &= \sum_{q=0}^{N-1} qA(q, N-1) \\ &\quad + 2\sum_{q=0}^{N-1} (q+1)A(q, N-1) \\ &\quad + \sum_{q=0}^{N-2} (q+1)A(q, N-2) \\ &\quad - \sum_{q=0}^{N-3} (q+3)A(q-3, N-3) \\ &= 3\mu_{N-1} + \mu_{N-2} - \mu_{N-3} \\ &\quad + 2\Delta_{N-1} + \Delta_{N-2} - 3\Delta_{N-3}. \end{aligned} \tag{24}$$

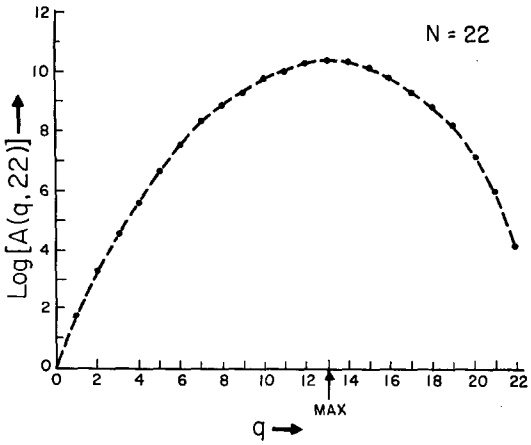

FIG. 5. A graph showing the logarithm of the number arrangements of q dumbbells on a 2×22 for all possible q. There are approximately 1.9×10^{10} ways of arranging 13 dumbbells on a 2×22 array.

For large values of N, Eq. (24) yields

$$k_0 N S_1^N = 3k_0(N-1)S_1^{N-1} + k_0(N-2)S_1^{N-2} - k_0(N-3)S_1^{N-3} + 2S_1^{N-1} + S_1^{N-2} - 3S_1^{N-3}. \quad (25)$$

We may divide this by S_1^N and, noting that

$$S_1^N = 3S_1^{N-1} + S_1^{N-2} - S_1^{N-3}, \quad (26)$$

we obtain

$$\frac{1}{S_1}(2 - 3k_0) + \frac{1}{S_1^2}(1 - 2k_0) - \frac{1}{S_1^3}(3 - 3k_0) = 0 \quad (27)$$

or

$$k_0 = 1 - \frac{S_1^2 + S_1}{3S_1^2 + 2S_1 - 3} \simeq 0.6064927. \quad (28)$$

Thus, the distribution reaches a maximum when the dumbbells occupy 61% of the compartments. Figure 5 shows $A(q, 22)$ as a function of q. In this case the maximum occurs at $q = 13$ or $\theta = 0.59$.

ACKNOWLEDGMENTS

The authors wish to express appreciation to Professor L. Parker for many helpful discussions and to R. Ristic for drawing the figures.

* Present address: University of Wisconsin–Madison, Madison, Wisconsin.

† Supported by the National Science Foundation (GY 6109).

[1] J. K. Roberts, Proc. Roy. Soc. (London) **A161**, 141 (1937).
[2] E. G. D. Cohn, J. DeBoer, and Z. W. Salsburg, Physica **21**, 137 (1955).
[3] D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
[4] R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).
[5] R. B. McQuistan, J. Math. Phys. **10**, 2205 (1969).
[6] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961).
[7] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[8] H. A. Bethe, Proc. Roy. Soc. (London) **A150**, 552 (1935).
[9] R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).
[10] J. K. Roberts and A. R. Miller, Proc. Cambridge Phil. Soc. **35**, 293 (1939).
[11] D. Zeitlin, Am. Math. Monthly **74**, 235 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Extension of the Riemann–Lebesgue Lemma*

PAUL B. KANTOR
Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106

(Received 23 May 1969; Revised Manuscript Received 20 March 1970)

We show that, in the limit of large λ, integrals of the form

$$H(\lambda) \equiv \int_a^b \frac{f(x)\,dx}{u(x) + e^{i\lambda x}}$$

are essentially given by $\int_{R'} [f(x)/u(x)]\,dx$ where the region R' is the union of all those subintervals in which $|u| \geq 1$. The corrections to this expression are of two kinds: terms $O(1/\lambda)$ which depend on the details of averaging to remove logarithmic singularities in $H(\lambda)$ and terms $O[(\ln \lambda)/\lambda]$. Some examples are given. If $|u| \leq 1$, the leading term in H vanishes and $H(\lambda)$ is bounded by $(\ln \lambda)/\lambda$.

I. INTRODUCTION

It is frequently useful to be able to bound integrals of the form

$$G(\lambda) = \int_a^b f(x)e^{i\lambda x}\,dx$$

for large values of the parameter λ. Typical physical problems would be the behavior of wavepackets at large times and the behavior of the Born approximation for scattering amplitudes at large energy.[1] The mathematical tool is, of course, the Riemann–Lebesgue lemma, which ensures that

$$|G(\lambda)| \leq \text{const}\, v(f)\lambda^{-1}.$$

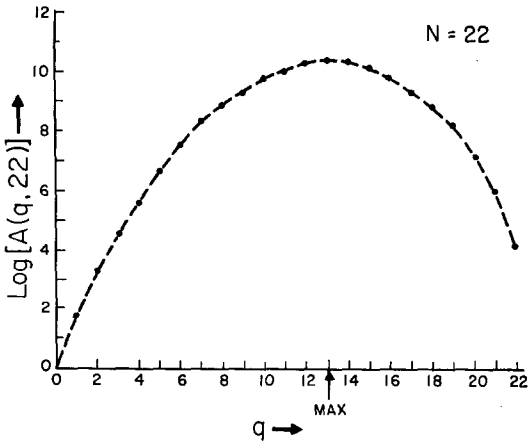

FIG. 5. A graph showing the logarithm of the number arrangements of q dumbbells on a 2×22 for all possible q. There are approximately 1.9×10^{10} ways of arranging 13 dumbbells on a 2×22 array.

For large values of N, Eq. (24) yields

$$k_0 N S_1^N = 3k_0(N-1)S_1^{N-1} + k_0(N-2)S_1^{N-2} - k_0(N-3)S_1^{N-3} + 2S_1^{N-1} + S_1^{N-2} - 3S_1^{N-3}. \quad (25)$$

We may divide this by S_1^N and, noting that

$$S_1^N = 3S_1^{N-1} + S_1^{N-2} - S_1^{N-3}, \quad (26)$$

we obtain

$$\frac{1}{S_1}(2 - 3k_0) + \frac{1}{S_1^2}(1 - 2k_0) - \frac{1}{S_1^3}(3 - 3k_0) = 0 \quad (27)$$

or

$$k_0 = 1 - \frac{S_1^2 + S_1}{3S_1^2 + 2S_1 - 3} \simeq 0.6064927. \quad (28)$$

Thus, the distribution reaches a maximum when the dumbbells occupy 61% of the compartments. Figure 5 shows $A(q, 22)$ as a function of q. In this case the maximum occurs at $q = 13$ or $\theta = 0.59$.

ACKNOWLEDGMENTS

The authors wish to express appreciation to Professor L. Parker for many helpful discussions and to R. Ristic for drawing the figures.

* Present address: University of Wisconsin–Madison, Madison, Wisconsin.

† Supported by the National Science Foundation (GY 6109).

[1] J. K. Roberts, Proc. Roy. Soc. (London) **A161**, 141 (1937).
[2] E. G. D. Cohn, J. DeBoer, and Z. W. Salsburg, Physica **21**, 137 (1955).
[3] D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
[4] R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).
[5] R. B. McQuistan, J. Math. Phys. **10**, 2205 (1969).
[6] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961).
[7] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[8] H. A. Bethe, Proc. Roy. Soc. (London) **A150**, 552 (1935).
[9] R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).
[10] J. K. Roberts and A. R. Miller, Proc. Cambridge Phil. Soc. **35**, 293 (1939).
[11] D. Zeitlin, Am. Math. Monthly **74**, 235 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Extension of the Riemann–Lebesgue Lemma*

PAUL B. KANTOR
Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106

(Received 23 May 1969; Revised Manuscript Received 20 March 1970)

We show that, in the limit of large λ, integrals of the form

$$H(\lambda) \equiv \int_a^b \frac{f(x)\,dx}{u(x) + e^{i\lambda x}}$$

are essentially given by $\int_{R'} [f(x)/u(x)]\,dx$ where the region R' is the union of all those subintervals in which $|u| \geq 1$. The corrections to this expression are of two kinds: terms $O(1/\lambda)$ which depend on the details of averaging to remove logarithmic singularities in $H(\lambda)$ and terms $O[(\ln \lambda)/\lambda]$. Some examples are given. If $|u| \leq 1$, the leading term in H vanishes and $H(\lambda)$ is bounded by $(\ln \lambda)/\lambda$.

I. INTRODUCTION

It is frequently useful to be able to bound integrals of the form

$$G(\lambda) = \int_a^b f(x)e^{i\lambda x}\,dx$$

for large values of the parameter λ. Typical physical problems would be the behavior of wavepackets at large times and the behavior of the Born approximation for scattering amplitudes at large energy.[1] The mathematical tool is, of course, the Riemann–Lebesgue lemma, which ensures that

$$|G(\lambda)| \leq \text{const}\, v(f)\lambda^{-1}.$$

Here $v(f)$ denotes the total variation[2] of the function f in the interval $[a, b]$.

The study of multiple scattering from several particles, using separable potentials and the closure approximation,[3] leads to expressions of the form

$$\int_R \frac{f(x)}{u(x) + e^{i\lambda x}}\, dx = H(\lambda).$$

The purpose of the present paper is to show that $H(\lambda)$ is "essentially" given by

$$\hat{H} = \int_{R'} \frac{f(x)}{u(x)}\, dx,$$

where R' is the union of all those subintervals of R in which $|u| \geq 1$.

There are four complications which we encounter.

(i) The function $H(\lambda)$ may have recurring logarithmic singularities as λ grows without bound. We deal with this by defining an averaged function of λ:

$$H_\Delta(\lambda) = \Delta^{-1} \int_{\lambda - \frac{1}{2}\Delta}^{\lambda + \frac{1}{2}\Delta} H(\lambda')\, d\lambda'.$$

Δ must be small enough so that only one singularity appears in the integrand for any value of λ. Note that the limit of $H(\lambda)$ as $\lambda \to \infty$ is independent of Δ, provided that R' is not empty.

(ii) Again because $u(x)$ may have unit modulus, the corrections to H are not quite bounded by $1/\lambda$. The best bound we can obtain is (for $\Delta < 1$)

$$|H_\Delta(\lambda) - \hat{H}| \leq C(\ln \lambda)/\lambda\, |\ln \Delta - 1|.$$

(iii) The function f must be of bounded variation, as in the Riemann–Lebesgue lemma. Our proof requires that $|u|$ satisfy the stronger condition of piecewise monotonicity. This insures that the variation of u^n for any n can be bounded *uniformly with respect to n.*

(iv) For purely technical reasons we have had to make the following additional assumptions. We do not know whether they are necessary.

For every x_0, such that $|u(x_0)| = 1$, there is a neighborhood N of x_0 contained in $[a, b]$ in which

(a) u is differentiable,

(b) $\exists\, B_2 \ni \left| \frac{u'(x) - u'(x_0)}{u(x) + e^{i\lambda x}} \right| < B_2,$

(c) $\exists\, D > 0 \ni \left| \frac{|u(x)| - |u(x_0)|}{x - x_0} \right| > D,$

(d) $\exists\, B_1 \ni \left| \frac{f(x) - f(x_0)}{u(x) + e^{i\lambda x}} \right| < B_1.$

Using condition (iii), we can divide $[a, b]$ into a finite number of subintervals of three different kinds:

(A) $|u| \leq R < 1$,

(B) $|u| \geq S > 1$,

(C) $R \leq |u| \leq S$ and $|u(x)| = 1$ has exactly *one* root.

In Sec. II we find the limiting form of $H(\lambda)$ for each of these cases. In Sec. III we combine these results. Finally, in Sec. IV we discuss examples and possible extensions of this technique.

Before presenting the proof we remark that the *complex* nature of $e^{i\lambda x}$ is essential for the simple form of $\hat{H}$. For example, with the real function $\sin \lambda x$, in case (B) one may show that

$$\int \frac{1}{u(x) + \sin \lambda x}\, dx = \int \frac{1}{[u(x)^2 - 1]^{\frac{1}{2}}}\, dx + O\left(\frac{1}{\lambda}\right). \quad (1)$$

II. DISCUSSION OF THREE CASES

Case A: We have the uniformly convergent expansion

$$[u(x) + e^{i\lambda x}]^{-1} = e^{-i\lambda x} \sum_{n=0}^{\infty} [u(x)e^{-i\lambda x}]^n (-1)^n.$$

Let $v(u)$ be the variation of $|u|$. Since $|u|$ is piecewise monotonic (we denote the number of "pieces" by N),

$$v(u) \leq 2NR$$

and, because of the monotonicity,

$$v(u^n) \leq 2NR^n$$

and

$$\sup |u^n| \leq R^n.$$

Hence, the function $u^n f$ is of bounded variation, and there is a constant K such that[4]

$$\begin{aligned} v(f \cdot u^n) &\leq 2NR^n \max f + R^n v(f) \\ &\leq KR^n. \end{aligned}$$

We now use the expansion of the denominator in the integral defining $H(\lambda)$ and apply the Riemann–Lebesgue lemma to every term in the sum:

$$\begin{aligned} \int u(x)^n f(x) e^{-\lambda x(n+1)}\, dx \\ &\leq \frac{4}{\lambda(n+1)} [v(fu^n) + \sup |u^n f|] \\ &\leq \frac{4}{\lambda}\frac{R^n}{n+1} [v(f) + (2N+1) \sup |f|] \\ &\equiv \frac{4}{\lambda}\frac{R^n}{n+1} D(f, N). \end{aligned}$$

Thus

$$f(x)[u(x) + e^{i\lambda x}]^{-1}\, dx \leq \frac{4D(f, N)}{\lambda} \frac{1}{R} |\ln (1 - R)|.$$

This completes the discussion of intervals of type A. They do not contribute to the limiting value at large λ.

Case B: Suppose $|u(x)| \geq S > 1$. Then we may write

$$f(x)[u(x) + e^{i\lambda x}]^{-1} = \frac{f(x)}{u(x)}\left(1 - \frac{1/u(x)}{[1/u(x)] + e^{-i\lambda x}}\right).$$

The first term in brackets is independent of λ. The second satisfies all the hypotheses of Case A with R replaced by $1/S$. Thus

$$\int \frac{f(x)}{u(x) + e^{i\lambda x}}\,dx = \int \frac{f(x)}{u(x)}\,dx + O\left(\frac{1}{\lambda}\right).$$

Case C: We now consider the case where $|u|$ assumes the value 1 exactly once in the interval of interest. The phase is unimportant in our argument; so we assume $u(x_1) = 1$. Now we choose an $\epsilon > 0$ such that, when

$$|x - x_1| < \epsilon,$$

x lies in the neighborhood N specified in condition (iv) of Sec. I. We have

$$\left|\frac{f(x) - f(x_1)}{u(x) + e^{i\lambda x}}\right| < B_1$$

and

$$\left|\frac{u'(x) - u'(x_1)}{u(x) + e^{i\lambda x}}\right| < B_2.$$

Since the integral exhibits logarithmic singularities at $\lambda = (\pi/x_1)(2n + 1)$, we extract them with the aid of the following two relations:

$$\frac{f(x)}{u(x) + e^{i\lambda x}} = \frac{f(x) - f(x_1)}{u(x) + e^{i\lambda x}} + f(x_1)\frac{1}{u(x) + e^{i\lambda x}},$$

$$\frac{1}{u(x) + e^{i\lambda x}} = \frac{1}{u'(x) - i\lambda}\left(\frac{d}{dx}\ln\,[u(x) + e^{i\lambda x}]\right.$$
$$\left. = \frac{u'(x) - u'(x_1)}{u(x) + e^{i\lambda x}} - i\lambda\frac{1 + e^{i\lambda x}}{u(x) + e^{i\lambda x}}\right).$$

Now choose $\delta \leq \min\{\epsilon, 1/\lambda\}$. Note that the definition of ϵ does not depend on λ so that, as $\lambda \to \infty$, δ will eventually be $1/\lambda$. We will show that, with suitable averaging over λ,

$$\int_{x_1 - \frac{1}{2}\delta}^{x_1 + \frac{1}{2}\delta} f(x)[u(x) + e^{i\lambda x}]^{-1}\,dx = O\left(\frac{1}{\lambda}\right).$$

Using the first relation above, we have

$$\int \frac{f(x)}{u(x) + e^{i\lambda x}}\,dx \leq \int \frac{|f(x) - f(x_1)|}{|u(x) + e^{i\lambda x}|}\,dx$$
$$+ |f(x_1)|\left|\int_{x_1 - \frac{1}{2}\delta}^{x_1 + \frac{1}{2}\delta} \frac{1}{u(x) + e^{i\lambda x}}\right|dx.$$

The first term on the right-hand side is bounded by B_1/λ. By elementary geometry we have

$$\frac{1 + e^{i\lambda x}}{u(x) + e^{i\lambda x}} \leq \frac{2}{1 + u(x)} \leq 2.$$

Thus

$$\int_{x_1 - \frac{1}{2}\delta}^{x_1 + \frac{1}{2}\delta} \frac{f(x)}{u(x) + e^{i\lambda x}}\,dx$$
$$\leq \frac{B_1}{\lambda} + \frac{1}{\lambda}\left(\frac{B_2}{\lambda} + 2\right)|f(x_1)|$$
$$+ \frac{1}{\lambda}|f(x_1)|\left|\int \frac{d}{dx}\ln\,[u(x) + e^{i\lambda x}]\,dx\right|.$$

This last integral exhibits the logarithmic singularity. Since the product $\lambda\delta$ is less than 1, the imaginary part of the integral is bounded by 1. To eliminate the divergence, we average over $\lambda - \frac{1}{2}\Delta \leq \lambda' \leq \lambda + \frac{1}{2}\Delta$:

$$\text{av}\int \frac{d}{dx}\ln\,[u(x) + e^{i\lambda x}]\,dx \leq C(|\ln\Delta| + 1).$$

Hence the integral in these dangerous regions is bounded by $1/\lambda$, multiplied by a factor which depends in a well-defined way on the averaging over λ.

III. PROOF OF THE THEOREM

Since u has only a finite number of oscillations, we can cover the range of integration with a finite number of intervals of type A, B, or C. Since the intervals of type C are of length $1/\lambda$, the values of R and S for the intervals of type A and B will be quite close to 1. However, for sufficiently large λ, we can be sure that $1 - R$ and $S - 1$ are at least as large as D/λ, where D is defined by condition (c). Inserting this in the bounds for case A, we find

$$H_\Delta(\lambda) = \int_{\text{type B}} \frac{f(x)}{u(x)}\,dx + K_1\frac{\ln\lambda}{\lambda} + K_2\frac{1 + |\ln\Delta|}{\lambda}.$$

The first correction term is due to intervals of type A and B, while the second is due to intervals of type C.

IV. EXAMPLES

Integrals of the form discussed here do not arise frequently in the mathematical literature, where expressions of the form e^{ix} usually arise from contour integration. An example arising in the physical situation discussed in Sec. I will be presented elsewhere.[3]

Examples for which the limiting form of the integral is known by other means can be found by considering Fermi–Dirac integrals.

Let

$$H(\beta) = \int_0^\infty \frac{f(\epsilon)}{e^{\beta(\epsilon - \mu)} + 1}\,d\epsilon.$$

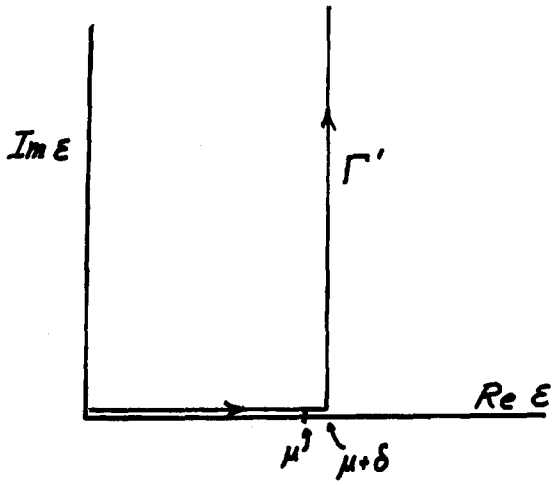

FIG. 1. The deformed contour Γ' in the complex ϵ plane.

Of course, this is not yet the form in which we are interested.

However, the integrand is regular except at the singularities of f and at $\beta(\epsilon - \mu) = (2n + 1)\pi i$. Since the integrand falls as $e^{-\beta \operatorname{Re}\epsilon}$ at infinity, we can almost deform the contour of integration to the axis $\mu + iy$. We can apply our technique as follows: Let the contour Γ' be as shown in Fig. 1. Assume that f has no singularities in the first quadrant. Then

$$\int_0^\infty \frac{f(\epsilon)\,d\epsilon}{e^{(\epsilon-\mu)\beta}+1} = \int_{\Gamma'} \frac{f(\epsilon)\,d\epsilon}{e^{(\epsilon-\mu)\beta}+1}$$

$$= \int_0^{\mu+\delta} \frac{f(\epsilon)\,d\epsilon}{e^{(\epsilon-\mu)\beta}+1} + i\int_0^\infty \frac{f(iy)\,dy}{e^{(\delta+iy)\beta}+1}.$$

Assuming f to be of bounded variation on Γ', we can apply our lemma, using $u(y) = e^{-\beta\delta} < 1$. We can maintain this condition with $\delta \to 0$; for example, let $\delta = 1/\beta$. Thus

$$\int_0^\infty \frac{f(\epsilon)\,d\epsilon}{e^{(\epsilon-\mu)\beta}+1} = \int_0^\mu \frac{f(\epsilon)\,d\epsilon}{e^{(\epsilon-\mu)\beta}+1} + O\left(\frac{1}{\beta}\right),$$

which is the familiar result.

Further examples can be constructed by considering the integral of the exact derivative

$$\frac{d}{dx}\{g(x)\ln[u(x)+e^{i\lambda x}]\},$$

which is easily seen to be equal to

$$\frac{-i\lambda g(x)u(x)}{u(x)+e^{i\lambda x}} + i\lambda g(x) + ig'(x)\arg[u(x)+e^{i\lambda x}]$$

$$+ \frac{g(x)u'(x)}{u(x)+e^{i\lambda x}} + g'(x)\ln|u(x)+e^{i\lambda x}|.$$

If we set $g(x) = f(x)/u(x)$, the first term is seen to be an integral of the desired type multiplied by λ. Thus we examine the rest of this equation for terms of order λ. When $|u(x)| > 1$, there are none but $i\lambda g(x)$, and the integral is given by

$$\int \frac{f(x)}{u(x)+e^{i\lambda x}}\,dx \to \int g(x).$$

When $u(x) < 1$, there are several terms of order λ, with only the last two being negligible. For large λ and sufficiently smooth g, we can replace

$$\int_a^b \frac{d}{dx}\{g(x)\ln[u(x)+e^{i\lambda x}]\} \quad \text{by} \quad i\lambda g(x)x\Big|_a^b$$

and

$$\arg[u(x)+e^{i\lambda x}] \quad \text{by} \quad i\lambda x.$$

A single partial integration then yields

$$i\lambda\int_a^b \frac{g(x)u(x)}{u(x)+e^{ikx}}\,dx$$

$$= i\lambda\int_a^b g(x)\,dx$$

$$+ i\lambda\int_a^b xg'(x)\,dx + O(1) - i\lambda g(x)x\Big|_a^b$$

$$= 0 + O(1).$$

Of course, when $u(x)$ can assume the value 1, the logarithmic correction enters in several terms.

This "example" serves as an analog to the usual heuristic proof of the Riemann–Lebesgue lemma based on partial integration. It would be nice to find some g and u for which every term can be integrated in terms of elementary functions, but I have been unable to do so.

V. DISCUSSION

By extending the Riemann–Lebesgue lemma to cases where the oscillating factor appears as the argument of a rational function, we can discuss the high-energy limit of certain very simple multiple scattering problems. Subject to possible convergence difficulties, the discussion can be extended to meromorphic functions of $e^{i\lambda x}$. Unfortunately, we have not found a relation of these ideas to the calculus of residues and are unable to extend the argument to holomorphic functions of $e^{i\lambda x}$. This last extension would be particularly useful in more realistic multiple scattering problems, where operators of the form $(1 - V_1G_0V_2)$ must be inverted.[3]

ACKNOWLEDGMENTS

I have enjoyed discussions of this question with several colleagues at Case Western Reserve University, particularly Professor L. Foldy and Dr. P. Johnson, who has given a simple proof of Eq. (1). This work arose from a consideration of the high-energy limit of multiple scattering, stimulated by an article by Professor L. Foldy and Professor D. Walecka.[5] I thank them for showing me their manuscript prior to publication.

* Work supported in part by the U.S. Atomic Energy Commission AEC Contract No. AT (11-1)-1753.

[1] A good reference for all these matters is M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1965), esp. Chaps. 3 and 6.

[2] E.T. Whittaker and G.N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1952), expecially Secs. 3.64 and 9.41.

[3] The results of that analysis will be published elsewhere. They are contained in P. B. Kantor, "Scattering From A Composite System; High Energy Limit of the Closure Approximation," Case Western Reserve University, Cleveland, Ohio, Preprint, 1970.

[4] To see this, simply apply the inequality $V_{fg} \leq V_g \sup|f| + V_f \sup|g|$ to each subinterval in which u is monotonic, and use induction on n.

[5] L. L. Foldy and J. D. Walecka, Ann. Phys. (N.Y.) **54**, 447 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Diagrammatic Perturbation Expansion for Ensembles of Random Matrices

J. F. McDONALD*
Department of Mathematics, University of Windsor, Windsor, Ontario, Canada
AND
L. D. FAVRO
Department of Physics, Wayne State University, Detroit, Michigan

(Received 22 September 1969; Revised Manuscript Received 23 April 1970)

A method for obtaining a perturbation expansion of various eigenvalue distributions corresponding to a certain class of perturbed ensembles of random matrices is given. The terms in the expansion can be written down immediately as diagrams analogous to those used in other kinds of perturbation theory. Further, part of the expansion can be summed explicitly, and the result of the summation read off the diagrams. In addition, a new perturbing ensemble is introduced. It has the advantage that the number of matrix elements which are perturbed simultaneously and the size of the perturbation can be varied independently. The expansion given is an expansion in the number of perturbed matrix elements rather than the usual expansion in the size of the elements. Finally, the conditions for convergence of the expansion are discussed.

1. INTRODUCTION

There has been recent interest in the problem of how a small perturbation to the Hamiltonian of a complex system effects the statistical properties of the energy levels.[1-6] The primary purpose of such work is to ascertain the possibility of determining whether or not a particular quantity is an exact invariant or only an approximate invariant by measurement of these statistical properties. Of particular concern is the question of how a small time-reversal invariant term in a Hamiltonian would manifest itself in the statistical properties of the energy spectrum.[2,4,6]

The ensemble which has received the most attention is a simple generalization of the Gaussian ensemble. That is, the unperturbed ensemble is assumed to be Gaussian, say of width $\alpha^{-\frac{1}{2}}$, and the perturbing ensemble is also assumed to be Gaussian, say of width $\gamma^{-\frac{1}{2}}$. Then it is assumed that the relative strength of the perturbation is given by $(\alpha/\gamma)^{\frac{1}{2}}$. One is, of course, interested in the limit where α/γ is small.

In studying this ensemble one encounters mathematical difficulties which seem to be inherent in it. In particular, if one considers the case where the unperturbed distribution is orthogonal (time-reversal invariant) and the perturbing distribution is unitary (not time-reversal invariant), it appears that the results will be purely unitary or orthogonal in the $\lim N \to +\infty$ (N is the dimensionality of the matrices) unless $\gamma \to +\infty$ with N in exactly the right way.[2,6] This results from the fact that the number of nonvanishing off-diagonal matrix elements is of order N^2. The convergence of a perturbation expansion in general depends not only on the size of the off-diagonal elements, but also on the number of such elements, since each succeeding term in the expansion involves another summation.

However, the physical interpretation of letting γ

ACKNOWLEDGMENTS

I have enjoyed discussions of this question with several colleagues at Case Western Reserve University, particularly Professor L. Foldy and Dr. P. Johnson, who has given a simple proof of Eq. (1). This work arose from a consideration of the high-energy limit of multiple scattering, stimulated by an article by Professor L. Foldy and Professor D. Walecka.[5] I thank them for showing me their manuscript prior to publication.

* Work supported in part by the U.S. Atomic Energy Commission AEC Contract No. AT (11-1)-1753.

[1] A good reference for all these matters is M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1965), esp. Chaps. 3 and 6.

[2] E.T. Whittaker and G.N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1952), expecially Secs. 3.64 and 9.41.

[3] The results of that analysis will be published elsewhere. They are contained in P. B. Kantor, "Scattering From A Composite System; High Energy Limit of the Closure Approximation," Case Western Reserve University, Cleveland, Ohio, Preprint, 1970.

[4] To see this, simply apply the inequality $V_{fg} \leq V_g \sup|f| + V_f \sup|g|$ to each subinterval in which u is monotonic, and use induction on n.

[5] L. L. Foldy and J. D. Walecka, Ann. Phys. (N.Y.) **54**, 447 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Diagrammatic Perturbation Expansion for Ensembles of Random Matrices

J. F. McDONALD*
Department of Mathematics, University of Windsor, Windsor, Ontario, Canada
AND
L. D. FAVRO
Department of Physics, Wayne State University, Detroit, Michigan

(Received 22 September 1969; Revised Manuscript Received 23 April 1970)

A method for obtaining a perturbation expansion of various eigenvalue distributions corresponding to a certain class of perturbed ensembles of random matrices is given. The terms in the expansion can be written down immediately as diagrams analogous to those used in other kinds of perturbation theory. Further, part of the expansion can be summed explicitly, and the result of the summation read off the diagrams. In addition, a new perturbing ensemble is introduced. It has the advantage that the number of matrix elements which are perturbed simultaneously and the size of the perturbation can be varied independently. The expansion given is an expansion in the number of perturbed matrix elements rather than the usual expansion in the size of the elements. Finally, the conditions for convergence of the expansion are discussed.

1. INTRODUCTION

There has been recent interest in the problem of how a small perturbation to the Hamiltonian of a complex system effects the statistical properties of the energy levels.[1-6] The primary purpose of such work is to ascertain the possibility of determining whether or not a particular quantity is an exact invariant or only an approximate invariant by measurement of these statistical properties. Of particular concern is the question of how a small time-reversal invariant term in a Hamiltonian would manifest itself in the statistical properties of the energy spectrum.[2,4,6]

The ensemble which has received the most attention is a simple generalization of the Gaussian ensemble. That is, the unperturbed ensemble is assumed to be Gaussian, say of width $\alpha^{-\frac{1}{2}}$, and the perturbing ensemble is also assumed to be Gaussian, say of width $\gamma^{-\frac{1}{2}}$. Then it is assumed that the relative strength of the perturbation is given by $(\alpha/\gamma)^{\frac{1}{2}}$. One is, of course, interested in the limit where α/γ is small.

In studying this ensemble one encounters mathematical difficulties which seem to be inherent in it. In particular, if one considers the case where the unperturbed distribution is orthogonal (time-reversal invariant) and the perturbing distribution is unitary (not time-reversal invariant), it appears that the results will be purely unitary or orthogonal in the $\lim N \to +\infty$ (N is the dimensionality of the matrices) unless $\gamma \to +\infty$ with N in exactly the right way.[2,6] This results from the fact that the number of nonvanishing off-diagonal matrix elements is of order N^2. The convergence of a perturbation expansion in general depends not only on the size of the off-diagonal elements, but also on the number of such elements, since each succeeding term in the expansion involves another summation.

However, the physical interpretation of letting γ

approach infinity as N goes to infinity is somewhat obscure. Since this results in making all the matrix elements of the perturbation go to zero, it might appear that the result of this would be to eliminate the effect of the perturbation entirely. This is not true because the total *number* of elements increases as each of them goes to zero. The problem with this ensemble is that the ratio α/γ is not simply a measure of the relative strength of the perturbation. It is also, in some sense, a measure of how many matrix elements are simultaneously perturbed on the average. Therefore, in the next section we introduce a 2-parameter distribution function in which one can vary the size and the number of perturbing matrix elements independently. This distribution function (and others of the same form) has the additional advantage that it lends itself readily to diagrammatic summation techniques.

2. THE DISTRIBUTION

Two approaches to the problem of perturbing a random ensemble of matrices have been used in previous work.[6-9] In one case it is assumed that the unperturbed matrix element distribution is completely known. In the other case it is assumed that the only thing that is known about the unperturbed matrix element distribution is its eigenvalue distribution.[7] Here we choose the latter approach because of its simplicity.

We will denote the unperturbed joint eigenvalue distribution by $f_N(\boldsymbol{\lambda}) = f_N(\lambda_1, \cdots, \lambda_N)$, where $\boldsymbol{\lambda}$ represents the N eigenvalues. The only assumption we make about $f_N(\boldsymbol{\lambda})$ is that it is invariant with respect to the labeling of the λ_i, and it is normalized to unity.

The perturbed distribution for the matrix elements will then be given by

$$P(\mathbf{H}) = \int d^N\boldsymbol{\lambda} p(\mathbf{H}, \boldsymbol{\lambda}) f_N(\boldsymbol{\lambda}), \tag{1}$$

where $p(\mathbf{H}, \boldsymbol{\lambda})$ is the perturbing distribution which in previous work was taken to be Gaussian; i.e.,

$$p(\mathbf{H}, \boldsymbol{\lambda}) \propto \exp\,[-\gamma\ \mathrm{Tr}\ (\mathbf{H} - \boldsymbol{\lambda})^2]. \tag{2}$$

From the work of Porter and Rosenzweig as corrected by Leff, it is known that this Gaussian form of the perturbing distribution is the only one for which the matrix elements are statistically independent and which is representationally invariant.[8] There are two simple extensions of that distribution which might form the basis of a more rapidly converging perturbation expansion. Each has a finite probability for the vanishing of the off-diagonal matrix elements. The first is a distribution in which the requirement of statistical independence of the matrix elements has been relaxed[9]:

$$p(\mathbf{H}, \boldsymbol{\lambda}) = (1 + \beta)^{-1}\Big[\Big(\prod_i \delta(H_{ii} - \lambda_i)\Big)\Big(\prod_{i>j} \delta(H_{ij})\Big) + \beta 2^{\frac{1}{4}N(N-1)}(\gamma/\pi)^{\frac{1}{4}N(N+1)} \times \exp\,[-\gamma\ \mathrm{Tr}\,(\mathbf{H} - \boldsymbol{\lambda})^2]\Big]. \tag{3}$$

This distribution function is representationally invariant in the sense that the product of δ functions in it may be regarded as a limit of a Gaussian distribution and hence is only a function of Tr $[(\mathbf{H} - \boldsymbol{\lambda})^2]$. The second term in the distribution simply adds a fraction β of a Gaussian perturbation to the unperturbed distribution.

The second distribution is one in which one maintains statistical independence but not representational invariance:

$$p(\mathbf{H}, \boldsymbol{\lambda}) = (1 + \beta)^{-N(N+1)/2} \times \prod_i \{\delta(H_{ii} - \lambda_i) + \beta(\gamma/\pi)^{\frac{1}{2}} \times \exp\,[-\gamma(H_{ii} - \lambda_i)^2]\} \times \prod_{j>k} [\delta(H_{jk}) + \beta(2\gamma/\pi)^{\frac{1}{2}} \exp\,(-2\gamma H_{jk}^2)]. \tag{4}$$

In this ensemble the distribution of *each* matrix element consists of a broad Gaussian with a sharp spike in the middle.

The parameters in these two distributions have been chosen such that in either distribution the probability for the vanishing of an off-diagonal element, say H_{21}, is $(1 + \beta)^{-1}$ and the mean square value $\langle H_{21}^2 \rangle$ is $\beta/[4\gamma(1 + \beta)]$.[10] The first one, however, has the disadvantage that the vanishing of one off-diagonal element implies the vanishing of all of them (i.e., the probability that one element is zero is the same as the probability that all elements are zero simultaneously). This follows from the fact that the integral over the product of δ functions is independent of the range of integration so long as the origin is included. For the second distribution the average number of diagonal matrix elements which are nonzero is $N(N - 1)\beta/[2(1 + \beta)]$.[11] Furthermore, it has the advantage that it leads itself readily to a perturbation treatment by simply multiplying out the terms in (4) explicitly. It is this second distribution which we shall study here.

As a measure of the strength of the perturbation, we shall take the mean square value of H_{ij}, $i > j$, divided by the parts of the H_{ij} which are simultaneously nonzero. This may be interpreted as the mean

square size of one of the nonzero elements, i.e.,

$$\langle H_{ij}^2\rangle_{\neq 0} = \frac{\langle H_{ij}^2\rangle}{\beta/(1+\beta)} = \frac{1}{4\gamma}. \tag{5}$$

The *number* of nonzero elements is independent of this and may be varied by changing β.

3. THE n-LEVEL CORRELATION FUNCTION AND DIAGRAMS

In this section we illustrate some techniques which can be used with the ensemble (4). For simplicity the discussion will be restricted to the calculation of the n-level correlation function (i.e., the joint distribution for n eigenvalues).

We first multiply out the product in the perturbing distribution explicitly. This will give a sum of terms. The general term in the sum for the perturbed matrix element distribution will be[12]

$$\begin{aligned} P_{\mathbf{i}_l,\mathbf{j}_m,\mathbf{k}_m} = {} & [\sigma(\gamma/\pi)^{\frac{1}{2}}]^l[\beta(2\gamma/\pi)^{\frac{1}{2}}]^m \\ & \times \int d^N\boldsymbol{\lambda} f_N(\boldsymbol{\lambda}) \cdot \exp\left[-\gamma(H_{i_1i_1} - \lambda_{i_1})^2\right] \cdots \\ & \times \exp\left[-\gamma(H_{i_li_l} - \lambda_{i_l})^2\right] \exp\left(-2\gamma H_{j_1k_1}^2\right) \cdots \\ & \times \exp\left(-2\gamma H_{j_mk_m}^2\right) \cdot \delta^{\mathbf{i}_l,\mathbf{j}_m,\mathbf{k}_m(\mathbf{H}-\boldsymbol{\lambda})}, \end{aligned} \tag{6}$$

where $\mathbf{i}_l = (i_1, \cdots, i_l)$, etc., and

$$\delta^{\mathbf{i}_l,\mathbf{j}_m,\mathbf{k}_m(\mathbf{H}-\boldsymbol{\lambda})} = \prod_{i\neq \mathbf{i}_l} \delta(H_{ii} - \lambda_i) \prod_{\substack{j>k \\ (j,k)\neq(\mathbf{j}_m,\mathbf{k}_m)}} \delta(H_{jk}). \tag{7}$$

In this term the set $H_{i_1i_1}$ through $H_{i_li_l}$ of the diagonal elements are perturbed (represented by factors like $\exp\left[-\gamma(H_{i_1i_1} - \lambda_{i_1})^2\right]$) and the set $H_{j_1k_1}$ through $H_{j_mk_m}$ of the off-diagonal matrix elements are perturbed (represented by factors like $\exp\left[-2\gamma H_{j_1k_1}^2\right]$). All of the other matrix elements are unperturbed (represented by δ functions). The perturbed distribution is then given by

$$p(\mathbf{H}) = C_N \sum_{l=0}^{N} \sum_{m=0}^{\kappa_N} \sum_{i_1>i_2>\cdots i_l} \sum_{j_1>k_1} \cdots \sum_{j_m>k_m} p_{\mathbf{i}_l,\mathbf{j}_m,\mathbf{k}_m}(\mathbf{H}), \tag{8}$$

where

$$C_N^{-1} = (1+\alpha)^N(1+\beta)^{\kappa_N}, \tag{9}$$

with

$$\kappa_N = N(N-1)/2. \tag{10}$$

The sum over the j's and k's includes only terms with all (j_τ, k_τ), $\tau = 1, \cdots, m$, different, and further includes distinct terms once and only once.

In order to obtain the n-level correlation function, one must make a change of variables from the matrix elements H_{ij}, $i \geq j$, to the eigenvalues E_i, $i = 1, \cdots, N$, and κ_N other variables $\boldsymbol{\Phi}$. Since $f_N(\boldsymbol{\lambda})$ has been assumed to be symmetric in the E_i, the result will be independent (except for labeling, of course) of which E_i remain. For definiteness, let $E_1, \cdots, E_n$ be the remaining eigenvalues, and let $f_N^n(E_1, \cdots, E_n)$ denote the n-level correlation function corresponding to $f_N(\mathbf{E})$.

In the following discussion it will be useful to think of the individual unperturbed eigenvalues as "particles" with perturbing off-diagonal matrix elements causing interactions between them. Thus H_{21}, for instance, may be thought of as causing an interaction between λ_1 and λ_2. A perturbing diagonal element causes a "self-interaction" of the particle with itself. We may continue this particle analogy further by drawing diagrams in which the vertices (indicated by symbols such as ○, ●, and △) are particles (eigenvalues) and the lines connecting them are interactions (perturbing matrix elements). This use of diagrams simplifies the bookkeeping associated with summing terms in the perturbation expansion.

Consider first the case $n = 1$ (i.e., single eigenvalue distribution). If one examines (6), it is clear that the only matrix elements which influence the functional dependence on E_1 are those H_{i1} which are perturbed (i.e., have a Gaussian distribution) and those H_{ij}, $i \geq j$, which are indirectly connected to H_{11} by perturbed matrix elements. That is, if zeros are put in the Hamiltonian matrix for the unperturbed off-diagonal matrix elements (i.e., those with a δ-function distribution) and the Hamiltonian is put in reduced form, it is only the block containing H_{11} which is relevant. This is because for a particular term in the expansion the matrix elements are statistically independent and thus the disjoint blocks are statistically independent.

In terms of diagrams, the situation may be thought of as a series of unconnected clusters of particles. Each cluster represents one of the blocks. The cluster of interest is the one which contains the particle of interest, namely E_1. The statistical independence of the matrix elements and the nonconnected nature of the diagrams allow the variables occurring in the clusters not containing E_1 to be integrated out immediately. The problem then is to find the single eigenvalue distribution in the usual way for the remaining submatrix (cluster). Let us denote the dimensionality of this submatrix by s.

Consider first those $p_{\mathbf{i}_l,\mathbf{j}_m,\mathbf{k}_m}$ such that $s = 1$ (i.e., those terms corresponding to matrices having no nonzero off-diagonal elements in the first row or column). There are two classes of such terms. In the first class H_{11} is not perturbed. The contribution to the single eigenvalue distribution from such a term is simply $\sigma^l\beta^m f_N^1(E_1)$. The distribution $f_N^1(E_1)$ is the

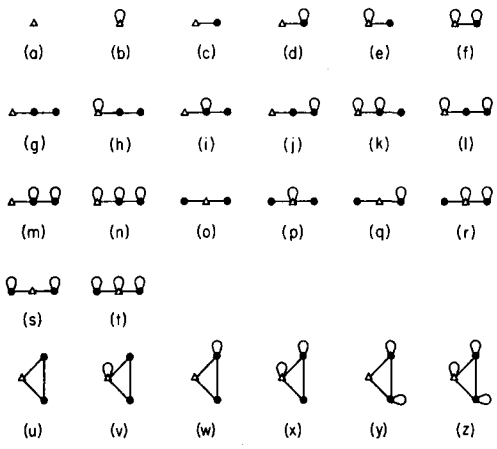

FIG. 1. All possible diagrams for the perturbation expansion of the single eigenvalue distribution to order $s = 3$.

single eigenvalue distribution for the unperturbed ensemble. Since this result does not depend on the particular values of $(\mathbf{i}_l, \mathbf{j}_m, \mathbf{k}_m)$, the calculation of the contribution from all such terms, with l and m fixed, simply involves counting the number of ways such a contribution will arise. It is easily seen that this number is

$$\binom{N-1}{l}\binom{\kappa_N - 1}{m},$$

where κ_N is defined by (10) and where $\binom{P}{Q}$ denotes the binomial coefficient. We can now sum over l and m to obtain the total contribution:

$$C_N(1+\sigma)^{N-1}(1+\beta)^{\kappa(N-1)}f_N^1(E_1) = (1+\sigma)^{-1}(1+\beta)^{-N+1}f_N^1(E_1). \quad (11)$$

This summation amounts to adding the contributions from all possible diagrams in the clusters not containing E_1. The diagram for $f_N^1(E_1)$ will simply be denoted by $\triangle$.[13]

The calculation in the second case, i.e., the case in which H_{11} is perturbed, goes through in an analogous manner. The only differences are that the above coefficient is multiplied by σ and that $f_N^1(E)$ is replaced by

$$(\gamma/\pi)^{\frac{1}{2}}\int d\lambda_1 f_N^1(\lambda_1)\exp[-\gamma(E_1-\lambda_1)^2].$$

This last expression represents the "particle" E_1 interacting with itself via H_{11}, and will be denoted by the diagram in Fig. 1(b).

Next we consider clusters with $s = 2$. There are four cases. First, we can have the case in which there are two unperturbed diagonal elements H_{11} and, say, H_{ii} which are connected by a perturbed off-diagonal element (in this case H_{i1}). The contribution from such a term is the single eigenvalue distribution corresponding to the 2×2 matrix-element distribution

$$(2\gamma/\pi)^{\frac{1}{2}}f_N^2(H_{11}, H_{ii})\exp(-2\gamma H_{i1}^2).$$

This contribution will be denoted by the diagram in Fig. 1(c). Since the result will not depend on i, the total contribution can again be found by counting the number of ways it occurs for fixed l and m and then summing over l and m. This coefficient is found to be

$$C_N(N-1)\beta(1+\sigma)^{N-2}(1+\beta)^{\kappa_{N-2}} = \beta(N-1)(1+\sigma)^{-2}(1+\beta)^{-2(N-3)}. \quad (12)$$

In the second case H_{ii} is also perturbed. The contribution from such a term is the single eigenvalue distribution corresponding to

$$\sqrt{2}\,(\gamma/\pi)\int d\lambda_i f_N^2(H_{11}, \lambda_i) \times \exp(-2\gamma H_{i1}^2)\exp[-\gamma(H_{ii}-\lambda_i)^2].$$

This is indicated by Fig. 1(d). Again, since the result does not depend on i, the necessary sums are easily carried out. The result can be found in Table I. The

TABLE I. Coefficients and matrix-element distributions corresponding to the diagrams given in Fig. 1. Note that

$$B_{ii} \equiv (\gamma/\pi)^{\frac{1}{2}}\exp[-\gamma(H_{ii}-\lambda_i)^2]$$

and

$$B_{ij} \equiv (2\gamma/\pi)^{\frac{1}{2}}\exp[-\gamma H_{ij}^2], \quad i > j.$$

Coefficient		Matrix-element distribution
a	A_1^1	$f_N^1(H_{11})$
b	σA_1^1	$\int d\lambda_1 f_N^1(\lambda_1)B_{11}$
c	βA_2^1	$f_N^2(H_{11}, H_{22})B_{21}$
d	$\sigma\beta A_2^1$	$\int d\lambda_2 f_N^2(H_{11}, \lambda_2)B_{22}B_{21}$
e	$\sigma\beta A_2^1$	$\int d\lambda_1 f_N^2(\lambda_1, H_{22})B_{11}B_{21}$
f	$\sigma^2\beta A_2^1$	$\int d\lambda_1\, d\lambda_2 f_N^2(\lambda_1, \lambda_2)B_{11}B_{22}B_{21}$
g	$\beta^2 A_3^1$	$f_N^3(H_{11}, H_{22}, H_{33})B_{21}B_{32}\delta(H_{31})$
h	$\sigma\beta^2 A_3^1$	$\int d\lambda_1 f_N^3(\lambda_1, H_{22}, H_{33})B_{11}B_{21}B_{32}\delta(H_{31})$
i	$\sigma\beta^2 A_3^1$	$\int d\lambda_2 f_N^3(H_{11}, \lambda_2, H_{33})B_{22}B_{21}B_{32}\delta(H_{31})$
j	$\sigma\beta^2 A_3^1$	$\int d\lambda_3 f_N^3(H_{11}, H_{22}, \lambda_3)B_{33}B_{21}B_{32}\delta(H_{31})$
k	$\sigma^2\beta^2 A_3^1$	$\int d\lambda_1\, d\lambda_2 f_N^3(\lambda_1, \lambda_2, H_{33})B_{11}B_{22}B_{21}B_{32}\delta(H_{31})$
l	$\sigma^2\beta^2 A_3^1$	$\int d\lambda_1\, d\lambda_3 f_N^3(\lambda_1, H_{22}, \lambda_3)B_{11}B_{33}B_{21}B_{32}\delta(H_{31})$
m	$\sigma^2\beta^2 A_3^1$	$\int d\lambda_2\, d\lambda_3 f_N^3(H_{11}, \lambda_2, \lambda_3)B_{22}B_{33}B_{21}B_{32}\delta(H_{31})$
n	$\sigma^3\beta^2 A_3^1$	$\int d\lambda_1\, d\lambda_2\, d\lambda_3 f_N^3(\lambda_1, \lambda_2, \lambda_3)B_{11}B_{22}B_{33}B_{21}B_{32}\delta(H_{31})$
o	$\frac{1}{2}\beta^2 A_3^1$	$f_N^3(H_{11}, H_{22}, H_{33})B_{21}B_{31}\delta(H_{32})$
p	$\frac{1}{2}\sigma\beta^2 A_3^1$	$\int d\lambda_1 f_N^3(\lambda_1, H_{22}, H_{33})B_{11}B_{21}B_{31}\delta(H_{32})$
q	$\sigma\beta^2 A_3^1$	$\int d\lambda_2 f_N^3(H_{11}, \lambda_2, H_{33})B_{22}B_{21}B_{31}\delta(H_{32})$
r	$\sigma^2\beta^2 A_3^1$	$\int d\lambda_1\, d\lambda_2 f_N^3(\lambda_1, \lambda_2, H_{33})B_{11}B_{22}B_{21}B_{31}\delta(H_{32})$
s	$\frac{1}{2}\sigma^2\beta^2 A_3^1$	$\int d\lambda_2\, d\lambda_3 f_N^3(H_{11}, \lambda_2, \lambda_3)B_{22}B_{33}B_{21}B_{31}\delta(H_{32})$
t	$\frac{1}{2}\sigma^3\beta^2 A_3^1$	$\int d\lambda_1\, d\lambda_2\, d\lambda_3 f_N^3(\lambda_1, \lambda_2, \lambda_3)B_{11}B_{22}B_{33}B_{21}B_{31}\delta(H_{32})$
u	$\frac{1}{2}\beta^3 A_3^1$	$f_N^3(H_{11}, H_{22}, H_{33})B_{21}B_{31}B_{32}$
v	$\frac{1}{2}\sigma\beta^3 A_3^1$	$\int d\lambda_1 f_N^3(\lambda_1, H_{22}, H_{33})B_{11}B_{21}B_{31}B_{32}$
w	$\sigma\beta^3 A_3^1$	$\int d\lambda_2 f_N^3(H_{11}, \lambda_2, H_{33})B_{22}B_{21}B_{31}B_{32}$
x	$\sigma^2\beta^3 A_3^1$	$\int d\lambda_1\, d\lambda_2 f_N^3(\lambda_1, \lambda_2, H_{33})B_{11}B_{22}B_{21}B_{31}B_{32}$
y	$\frac{1}{2}\sigma^2\beta^3 A_3^1$	$\int d\lambda_2\, d\lambda_3 f_N^3(H_{11}, \lambda_2, \lambda_3)B_{22}B_{33}B_{21}B_{31}B_{32}$
z	$\frac{1}{2}\sigma^3\beta^3 A_3^1$	$\int d\lambda_1\, d\lambda_2\, d\lambda_3 f_N^3(\lambda_1, \lambda_2, \lambda_3)B_{11}B_{22}B_{33}B_{21}B_{31}B_{32}$

diagrams and coefficients for the other two cases (i.e., H_{11} perturbed but H_{ii} not perturbed, and finally both H_{11} and H_{ii} perturbed) are given in Fig. 1 and Table I, respectively.

If one next considers $s = 3$, there are 20 distinct cases. The procedure is the same, and the results are given in Fig. 1 and Table I. One can continue this procedure for any value of s. All possible diagrams for any value of s can be written down immediately. The matrix-element distribution from which the contribution arises then can immediately be written down. These are given for $s = 1, 2, 3$ in Table I. In addition, the coefficient corresponding to a given diagram is given by the expression

$$p^{-1}A_s^1\sigma^a\beta^b, \tag{13}$$

where

$$A_s^n \equiv \frac{(N-n)!}{(N-s)!}(1+\sigma)^{-s}(1+\beta)^{-s(2N-s-1)/2}, \tag{14}$$

s is the number of vertices in the diagram (i.e., dimensionality of the submatrix involved), a is the number of self-interactions (represented by closed loops attached to the vertices), b is the number of interacting pairs (represented by the connecting lines in the diagram), and p is the number of permutations of the vertices which leave the connectivity of the diagram unchanged.[14] It should be noted that factor $p^{-1}(N-1)!/(N-s)!$ occurring in (13) is simply the number of ways that an s-fold cluster containing E_1 can occur for a particular set of clusters not involving E_1. The coefficients for $s = 1, 2, 3$ are given in Table I. The total expansion to order $s = 3$ is the sum of all the diagrams in Fig. 1 with the corresponding coefficients given in Table I.

The problem for the 2-level correlation function can be treated in an analogous manner. Now two eigenvalues are not integrated out. In the diagrams we let $\triangle$ correspond to E_1 and $\circ$ to E_2. The existence

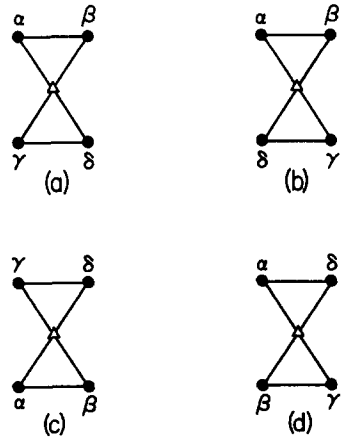

FIG. 2. Sample diagrams for single eigenvalue distribution with $s = 5$, illustrating the definition of "connectivity," as defined in Footnote 14.

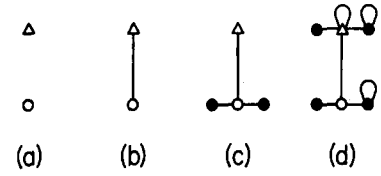

FIG. 3. Some sample diagrams for the 2-level correlation function.

of two unique eigenvalues produces one difference from the single-level case. The matrix elements H_{11} and H_{22} may be in the same cluster or different clusters. However, the counting problem is analogous, and the coefficient for a diagram is given by the expression

$$p^{-1}\sigma^a\beta^b A_s^2, \tag{15}$$

where the symbols are defined exactly as before if H_{11} and H_{22} are in the same cluster while, if H_{11} and H_{22} are in different clusters, s is the number of vertices in both clusters (i.e., the sum of dimensions of the two matrices), p is the product of the values of p for each cluster, as defined above, a is the number of self-interactions in both diagrams, and b is the total number of interactions between pairs in both diagrams. Some sample diagrams are given in Fig. 3, and the matrix-element distribution arises as well as the coefficient are given in Table II.

At this point the generalization of the above results to the n-level correlation function should be clear. One first writes down all possible diagrams. In order to obtain the matrix-element distribution corresponding to a particular diagram, one first integrates out all the variables in the clusters which do not contain $\lambda_1, \cdots, \lambda_n$. The matrix-element distribution is then

$$\int d\lambda_1 \cdots d\lambda_s f_N^s(\lambda_1, \cdots, \lambda_s) \times \left(\prod_i (\gamma/\pi)^{\frac{1}{2}} \exp\left[-\gamma(H_{ii}-\lambda_i)^2\right]\right) \times \left(\prod_{i>j}(2\gamma/\pi)^{\frac{1}{2}} \cdot \exp\left[-2\gamma H_{ij}^2\right]\right)\left(\prod_{h>l}\delta(H_{hl})\right), \tag{16}$$

where s is the number of vertices, $\prod_i$ is over the self-interacting vertices, $\prod_{i>j}$ is over the vertices connected

TABLE II. Coefficients and matrix-element distributions corresponding to the diagrams given in Fig. 3.

	Coefficient	Matrix-element distribution
a	A_2^2	$f_N^2(H_{11}, H_{22})\delta(H_{21})$
b	βA_2^2	$f_N^2(H_{11}, H_{22})E_{21}$
c	$\frac{1}{2}\beta^3 A_4^2$	$f_N^4(H_{11}, H_{22}, H_{33}, H_{44})E_{21}E_{32}E_{42}\delta(H_{31})\delta(H_{41})\delta(H_{43})$
d	$\sigma^3\beta^5 A_6^2$	$\int d\lambda_1\, d\lambda_3\, d\lambda_6 f_N^6(\lambda_1, H_{22}, \lambda_3, H_{44}, H_{55}, \lambda_6)$ $\times E_{11}E_{33}E_{66}E_{21}E_{31}E_{41}E_{52}E_{62}\delta(H_{51})\delta(H_{61})\delta(H_{32})$ $\times \delta(H_{42})\delta(H_{43})\delta(H_{53})\delta(H_{63})\delta(H_{54})\delta(H_{64})\delta(H_{65})$

in pairs, and $\prod_{h>l}$ is over those vertices which are free (i.e., not connected to any other). The contribution from the diagram is then the n-level correlation function corresponding to this matrix-element distribution (which can be found in the usual way) times the coefficient

$$p^{-1}\sigma^a\beta^b A_s^n, \tag{17}$$

where, of course, $s \geq n$ and where the symbols are just the obvious generalizations of those given above for $n = 2$.

It should be noted that the above perturbation techniques are not restricted to the ensemble given by (4). In principle they can be used for any distribution for which the matrix elements are all statistically independent and for which the distribution for each matrix element can be written as a sum of an unperturbed distribution and a perturbing distribution. For example, the Gaussian ensemble can be treated analogously if one notes that (2) can be written in the form

$$\prod_i (\delta(H_{ii} - \lambda_i) + \{(\gamma/\pi)^{\frac{1}{2}} \exp[-\gamma(H_{ii} - \lambda_i)^2] - \delta(H_{ii} - \lambda_i)\}) \times \prod_{j>k} (\delta(H_{jk}) + \{(2\gamma/\pi)^{\frac{1}{2}} \exp[-2\gamma H_{jk}^2] - \delta(H_{jk})\}). \tag{18}$$

Here the unperturbed distribution for each matrix element is a δ function, while the perturbing distribution is the difference of an exponential and a δ function. The details of the calculation, as well as the validity of the expansion (i.e., convergence in the limit of large N), of course depend on the particular distribution involved. In the next section we examine the convergence of the expansion for the distribution given by (4).

4. CONVERGENCE OF THE PERTURBATION EXPANSION IN THE LIMIT OF LARGE N

In the limit in which the dimensionality N of the matrices goes to infinity, the finite sums in our perturbation expansion become infinite series. These infinite series are partially summed series in powers of the parameter β which measures the number of matrix elements which are simultaneously perturbed. Each term in the series contains a coefficient which depends on β, N, n, and s. Each also contains a function of the eigenvalues E_i, which is essentially an average over the unperturbed distribution. In some sense these eigenvalue-dependent functions are "well-behaved" functions, since the integral of each is unity. We will discuss the convergence of the series when each of them is replaced by unity.[15] One then only need consider the asymptotic behavior of the coefficient for large N.

When n and s are small compared to N, the coefficient is given approximately by

$$p^{-1}\beta^{a+b}N^{s-n}(1+p)^{-s(2N-s+1)/2}. \tag{19}$$

This asymptotic coefficient has some properties which at first sight seem to be contradictory.

First, consider the case in which β goes as N^{-2} (which corresponds to having only a finite number of matrix elements perturbed simultaneously). In this case all of the coefficients except the first one (with $s = n$ and $a + b = 0$) vanish, and one is left with the unperturbed distribution. This is reasonable, since one does not expect a finite number of matrix elements to make any substantial changes in the eigenvalue distributions for an infinite matrix.

Next consider the case in which β behaves as N^{-1} (which corresponds to the number of perturbed elements being of order N). In this case only those diagrams with $a + b = s - n$ contribute, and those have coefficients given by

$$p^{-1}\bar{\beta}^{s-n}e^{-\beta s}, \tag{20}$$

where $\bar{\beta} = N\beta$. This will clearly converge in s provided $\bar{\beta} > 0$. Thus, making the number of perturbed elements larger has produced a finite, but converging, correction to the unperturbed distribution. One is then led to ask what happens when the number of terms is made still larger, say of order N^2. This corresponds to making β independent of N. In the limit N going to infinity, we then find that (19) goes to zero. This is a surprising result because the previous discussion (and common sense) seem to indicate that increasing the number of perturbed matrix elements causes the series to converge more slowly. What has happened is that infinitely many terms in the series are contributing equally (roughly speaking) and hence each of them individually must vanish. An interesting sidelight is that in $\lim \beta \to +\infty$ the series converges again, but it is the last term in the series [with $s = N$ and $a + b = N(N-1)/2$] which gives the entire contribution.

The discussion above shows that the only nontrivial large dimensionality limit of our expansion is in the case with β proportional to N^{-1}. The perturbation expansion simplifies considerably in this case. The only terms which contribute are those with $a = 0$ and $b = s - n$. These correspond to a particularly simple class of diagrams. First of all, there are no loops, i.e., self-interactions. Next there are n disjoint diagrams with one and only one vertex corresponding to $E_1, E_2, \cdots, E_n$ in each diagram. Finally, each

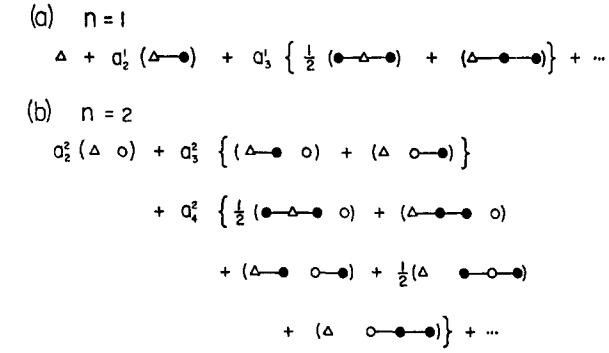

FIG. 4. The perturbation expansion for $n = 1, 2$ to order $s = 3$ for the case $\beta \sim 1/N$ in the limit of large N. Note that

$$a_s^n = \bar{\beta}^{s-n} e^{-s\bar{\beta}}.$$

disjoint diagram is a simple chain; i.e., if the vertices are arranged in a straight line, the only interactions are only those between nearest neighbors. For a particular value of s, all the diagrams can be obtained by taking all possible (distinct) permutations of the vertices in each disjoint diagram.[16] The perturbation expansions for $n = 1, 2$ to order $s = 3$ is shown in Fig. 4.

5. A SAMPLE CALCULATION

To illustrate the above results we shall calculate the first correction to the n-level correlation function. For simplicity we restrict our discussion to $f_N(\boldsymbol{\lambda})$ of the form

$$f_N(\boldsymbol{\lambda}) = \prod_{i=1}^{N} f_N^1(\lambda_i). \tag{21}$$

In addition we assume the $\beta = \bar{\beta}/N$ and consider only $\lim N \to +\infty$.

In this case there will be a total of n diagrams in the first correction term. In each of these, $(n-1)$ of the particles (levels) have no self-interaction or interaction with other particles. The other level interacts with only one other. The diagrams differ only in the fact that in each one a different one of the λ_i interacts. Thus, we need only consider one such diagram in detail. The others can be treated by simply relabeling the λ_i.

Consider the diagram where λ_1 is the eigenvalue with the interaction. This diagram corresponds to the perturbed matrix-element distribution

$$\int d^N\boldsymbol{\lambda} \left(\prod_{i=1}^{N} f_N^1(\lambda_i)\right) \delta^{n+1,1}(\mathbf{H} - \boldsymbol{\lambda})(2\gamma/\pi)^{\frac{1}{2}} \times \exp(-2\gamma H_{n+1,1}^2), \tag{22}$$

where

$$\delta^{n+1,1}(\mathbf{H} - \boldsymbol{\lambda}) \equiv \prod_{i=1}^{N} \delta(H_{ii} - \lambda_i) \prod_{j>k} \delta(H_{jk}) \tag{23}$$

with $(j, k) \neq (n+1, 1)$. It follows trivially that the contribution to the n-level correlation function is

$$P_{N,n,1}(E) \equiv p(E_1) \prod_{i=2}^{n} f_N^1(E_i), \tag{24}$$

where $p(E_1)$ is the single eigenvalue distribution corresponding to the ensemble

$$P(H_{11}, H_{n+1}, H_{n+1,1}) = f_N^1(H_{n+1,n+1})(2\gamma/\pi)^{\frac{1}{2}} \exp(-2\gamma H_{n+1,1}^2). \tag{25}$$

In the usual manner[8] it follows that

$$p(E_1) = \tfrac{1}{4} \int_0^{\pi} d\Phi \int_{-\infty}^{+\infty} dx \times P(H_{11}, H_{n+1,n+1}, H_{n+1,1}) |E_1 - x|, \tag{26}$$

where

$$H_{11} = E_1 \cos^2\Phi + x \sin^2\Phi, \tag{27}$$

$$H_{n+1,n+1} = E_1 \sin^2\Phi + x \cos^2\Phi, \tag{28}$$

and

$$H_{n+1,1} = (x - E_1) \cos\Phi \sin\Phi. \tag{29}$$

The total contribution from all diagrams in this correction term will be simply

$$\sum_{i=1}^{n} P_{N,n,i}(E),$$

where i denotes the perturbed level. The coefficient for this term is, from (20), $\bar{\beta} \exp[-(n+1)\bar{\beta}]$.

In general the integrals that are required to evaluate $p(E_i)$ explicitly [i.e., (26)] are not trivial. Even in the seemingly simple case when

$$f_N^1(\lambda) = (\sigma/\pi)^{\frac{1}{2}} \exp(-\sigma\lambda^2), \tag{30}$$

the integrals cannot be done in closed form.[8] Hence, since this calculation is just a sample to indicate the nature of the terms in the expansion and does not have direct physical interest, we will not carry the calculation any further here. Calculations of physical interest will undoubtedly require numerical computation.

APPENDIX

Here we show that for the ensemble given by (4) the average number of off-diagonal matrix elements which are nonzero is $N(N-1)\beta/[2(1+\beta)]$.

The probability that n particular H_{jk}, $j > k$, are zero, while all the others are nonzero is defined by

$$\lim_{\epsilon \to 0^+} \left(\prod_{i=1} \int_{-\infty}^{+\infty} dH_{ii}\right) \left(\prod_{\alpha=1}^{n} \int_{-\epsilon}^{+\epsilon} dH_{j_\alpha k_\alpha}\right) \times \prod_{j>k}{}' \left(\int_{-\infty}^{-\epsilon} dH_{jk} + \int_{\epsilon}^{+\infty} dH_{jk}\right) P(\mathbf{H}),$$

where the last product includes the $\kappa_N - n$ off-diagonal elements not included in $\prod_\alpha$. Here $P(\mathbf{H})$ is given in terms of $p(\mathbf{H}, \boldsymbol{\lambda})$ by (1) and $\kappa_N \equiv N(N-1)/2$ is the total number of off-diagonal elements H_{jk}, $j > k$. If one inserts the distribution given by (4) into the above definition, the integrals are easily calculated, and one finds the given probability to be

$$[\beta/(1+\beta)]^{\kappa_N}\beta^{-n}.$$

Since this result is independent of which H_{jk} are zero and which are nonzero, the probability that *any* n are zero and the remaining are nonzero is

$$p(n=0) = \binom{\kappa_N}{n}\left(\frac{\beta}{1+\beta}\right)^{\kappa_N}\beta^{-n}.$$

Thus, the average number which are zero is given by

$$\langle n\rangle_0 = \sum_{n=0}^{\kappa_N} np(n=0).$$

The required sum is easily evaluated. The result is

$$\langle n\rangle_0 = \kappa_N/(1+\beta).$$

Thus, the average number of off-diagonal elements which are not zero is

$$\kappa_N - \langle n=0\rangle = \beta\kappa_N/(1+\beta).$$

* Supported in part by an National Research Council grant and in part by a University Affairs Grant (University of Windsor).

[1] N. Rosenzweig and C. E. Porter, Phys. Rev. **120**, 1968 (1960).
[2] N. Rosenzweig, J. E. Monahan, and M. L. Mehta, Nucl. Phys. **A109**, 437 (1968).
[3] M. L. Mehta and N. Rosenzweig, Nucl. Phys. **A109**, 449 (1968).
[4] L. D. Favro and J. F. McDonald, Phys. Rev. Letters **19**, 1254 (1967).
[5] L. D. Favro and J. F. McDonald, J. Math. Phys. **9**, 1429 (1968).
[6] J. F. McDonald, J. Math. Phys. **10**, 1191 (1969).

[7] If one is interested in the eigenvalues and not the eigenvectors, the two approaches can be shown to be equivalent under certain conditions.

[8] C. E. Porter and N. Rosenzweig, Ann. Acad. Sci. Fennicae, Ser. AVI, **44** (1960).

[9] Here, for simplicity, we have assumed that the perturbing distribution is characterized by the orthogonal group.

[10] The probability that H_{21} is zero is defined as $\lim_{\epsilon\to 0+} p_{21}(\epsilon)$, where $p_{21}(\epsilon)$ is the probability that H_{21} is in the interval $(-\epsilon, \epsilon)$ regardless of the values of the other H_{ij}.

[11] See Appendix.

[12] Note that we have replaced the β for the diagonal elements with a third parameter σ. This controls the number of diagonal elements which are zero. This can be treated as an independent parameter and set equal to β in the end if so desired. We introduce it here to facilitate the bookkeeping in the perturbation expansion.

[13] A triangle will be used to indicate a vertex corresponding to the unique eigenvalue E_1.

[14] The "connectivity" of a diagram is defined in the following way: Label each vertex of the diagram with a symbol such as α, β, γ, etc. The connectivity of two diagrams is the same if in each of them each of the vertices is connected to the same set of symbols as in the other. For example, in Fig.2(a),α, β, γ, and δ are all connected to the starred vertex and $\alpha\beta$ and $\gamma\delta$ are connected in pairs. Thus, Figs. 2(b) and 2(c) have the same connectivity as Fig. 2(a), but Fig. 2(d) does not. The value of p for this diagram is 8.

[15] It is understood that, for sufficiently pathological unperturbed distributions, this replacement may not be appropriate.

[16] If one does not let $\sigma = \beta$, the situation is only slightly more complicated. The parameter σ can be taken to be independent of N (i.e., of order N nonzero diagonal elements). Furthermore, the diagrams which appear will be those which are present for $\sigma = \beta$ with all possible combinations of loops. Of course, the coefficients will be different.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Use of the WKB Quantization Condition for Calculating Expectation Values

E. SIEBERT AND J. B. KRIEGER
Department of Physics, Polytechnic Institute of Brooklyn, Brooklyn, New York 11201

(Received 16 February 1970)

A systematic technique is developed for the calculation of expectation values for 1-dimensional problems by a direct use of the WKB quantization rule including the higher-order integrals without the introduction of wavefunctions. As an example of the technique we calculate the moments $\langle r^p \rangle$ for the Coulomb potential for $2 \geq p \geq -3$, for all bound states of the system. Our results are exact except for $p = -3$, in which case the inclusion of the higher-order integrals substantially improves the accuracy of the calculation.

I. INTRODUCTION

One of the constantly recurring calculations in quantum mechanics is the evaluation of expectation values and sum rules for the various states of a system with a given Hamiltonian. When the wavefunction of the state is known, the calculation of expectation values is in principle straightforward, but the evaluation of sums involving off-diagonal elements of some operator, such as appear in second-order perturbation theory for the energy, continues to be difficult in practice. Furthermore, when the wavefunction is not exactly known, it is not generally possible to calculate directly the expectation values, although some progress has recently been made in this area through the development of variational principles for expectation values other than the energy.[1] Moreover, variational techniques have also been developed for the calculation of sums involving off-diagonal matrix elements.[2]

In this communication we examine the application of the WKB method to this problem. The standard application of this technique to the calculation is well known, i.e., one merely makes use of the WKB wavefunctions to approximately calculate expectation values.[3] However, this technique does not readily lead to a simple systematic method of improving the calculation because the higher-order WKB contributions to the wavefunction become progressively more singular near the classical turning points, and thus a more detailed analysis of the wavefunction in the neighborhood of these points is required.

On the other hand, the WKB method has long been used to calculate the approximate eigenvalues of the 1-dimensional Schrödinger equation. The accuracy of the method has been improved by the inclusion of the higher-order correction terms obtained by Dunham.[4] The inclusion of these terms in the quantization condition has led to the calculation of eigenvalues with considerable precision even for the ground state of the system.[5-7] However, little use has been made of the WKB method as a means of calculating expectation values, moments, and other quantities of interest.[8]

In Sec. II we show that a systematic technique can be developed for the calculation of expectation values and sum rules for 1-dimensional problems by a direct use of the WKB quantization rule without the introduction of wavefunctions. As an example of the technique, we calculate, in Sec. III, the moments $\langle r^p \rangle$ for the Coulomb potential for $2 \geq p \geq -3$, for all bound states of the system. Finally, in Sec. IV we conclude with some remarks about the applicability of the technique to other systems.

II. DERIVATION OF THE TECHNIQUE

It is well known from perturbation theory that, if a small perturbation $\lambda V'$ is added to a potential, then the energy of the perturbed state can be written as

$$\epsilon_n(\lambda) = \epsilon_n(0) + \sum_{m=1}^{\infty} \epsilon_{nm} \lambda^m, \tag{1}$$

where

$$\epsilon_{n1} = \langle n | V' | n \rangle, \tag{2}$$

$$\epsilon_{n2} = \sum_s{}' \frac{\langle n | V' | s \rangle \langle s | V' | n \rangle}{E_n - E_s}, \tag{3}$$

and the higher-order coefficients are given in the standard textbook accounts.[9] It then follows from Eq. (1) that

$$\left. \frac{\partial \epsilon_n(\lambda)}{\partial \lambda} \right|_{\lambda=0} = \langle n | V' | n \rangle \tag{4}$$

and

$$\left. \frac{1}{2} \frac{\partial^2 \epsilon_n}{\partial \lambda^2} \right|_{\lambda=0} = \sum_s{}' \frac{\langle n | V' | s \rangle \langle s | V' | n \rangle}{E_n - E_s}, \tag{5}$$

with similar relations being valid for the higher-order derivatives. It is interesting to note that Eqs. (4) and (5) are valid even if the perturbation expansion given

by Eq. (1) does not exist (either has no radius of convergence or has a divergent higher-order coefficient) provided only that the right-hand sides of these equations are finite.[10]

Hence the expectation value of any function V' can be calculated by adding $\lambda V'$ to the potential and finding the change in the energy to first order in λ. Similarly, the sum in Eq. (5) can be found by evaluating the change in the energy to second order in λ. In both cases all that is required is an independent method of calculating the perturbed energy of the state.

The WKB method offers just such an independent means of calculation for potentials that are separable. Suppose we write the quantization condition for a 1-dimensional Schrödinger equation as

$$\oint Q_t(\epsilon_n, V(x))\,dx = n + \tfrac{1}{2}, \tag{6}$$

where Q_t is an expression obtained from summing the first t integrals in the quantization condition[4] and the path of integration is taken to enclose the zeros of Q_1 and no other singularities of Q_1. Then the equation for $\epsilon_n(\lambda)$ is

$$\oint Q_t(\epsilon_n(\lambda), V(x) + \lambda V'(x))\,dx = n + \tfrac{1}{2}. \tag{7}$$

Expanding Eq. (7) around $\lambda = 0$ gives

$$\left.\frac{\partial\epsilon_n(\lambda)}{\partial\lambda}\right|_{\lambda=0} = \left.\frac{-(\partial/\partial\lambda)\oint Q_t(\epsilon, V+\lambda V')\,dx}{(\partial/\partial\epsilon)\oint Q_t(\epsilon, V)\,dx}\right|_{\lambda=0}, \tag{8}$$

where ϵ is determined by Eq. (6). Similarly, expressions can be derived for the higher derivatives of $\epsilon_n(\lambda)$. It is of practical importance that the higher-order WKB integrals generally decrease rapidly,[5-7] so that we expect the right-hand side of Eq. (8) to be a rapidly converging expression as t, the number of integrals included in the quantization condition, is increased. As an example we calculate below the moments for the Coulomb potential.

III. MOMENTS FOR COULOMB POTENTIAL

The WKB quantization condition for spherically symmetric potentials has been found by Langer[11] and the first two higher-order integrals have been correctly derived.[12] The quantization condition, correct to the third order in $\hbar^2$, is[12]

$$\frac{2\pi(N_r+\tfrac{1}{2})\hbar}{(2m)^{\frac{1}{2}}} = \oint q\frac{dr}{r} - \frac{\hbar^2}{64m}\oint\frac{(Dq^2)^2}{q^5}\frac{dr}{r} - \frac{\hbar^4}{8192m^2}\oint\left(\frac{49(Dq^2)^4}{q^{11}} - \frac{16(Dq^2)(D^3q^2)}{q^7}\right)\frac{dr}{r}, \tag{9}$$

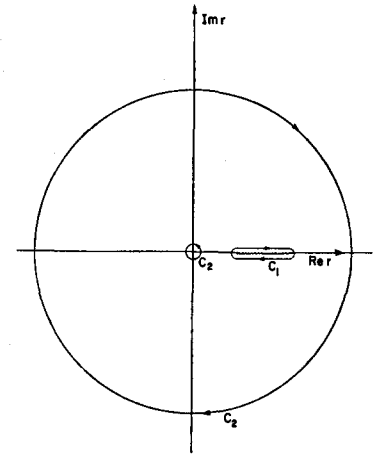

FIG. 1. The contour C_1 in the complex r plane encloses the branch points corresponding to the zeros of q. The contour C_2 consists of two parts and is equivalent to C_1.

where

$$q = \left(r^2E - r^2V(r) - \frac{(l+\tfrac{1}{2})^2\hbar^2}{2m}\right)^{\frac{1}{2}},$$

$$D^nq^2 \equiv \left(r\frac{d}{dr}\right)^n q^2,$$

N_r is an integer ≥ 0, and $E < 0$ is the energy of a bound state. The integrals are all taken about a contour enclosing the zeros of q and no other singularities.

For the Coulomb potential, $V(r) = -ze^2/r$, the only zeros of q are on the positive real axis. We integrate around the contour C_1 as shown in Fig. 1, where the integrands have been made single valued by a cut connecting the zeros of q and with phases chosen so that $q > 0$ just above the cut. Since the only singularities of the integrands occur at the branch points and at $r = 0$ and infinity, we may deform C_1 into the equivalent contour C_2 in which case we see that the only possible contributions to the integrals occur at the origin and infinity. We evaluate these integrals by expanding the integrands about the origin and infinity, keeping only the terms having the $1/r$ dependence necessary for calculating the residues.

A. Coulomb Eigenvalues

It is well known that the first-order integral gives the exact eigenvalues for the Coulomb potential and that the higher-order integrals in Eq. (9) are zero.[12] Evaluation of the first-order integral gives

$$\frac{2\pi(N_r+\tfrac{1}{2})\hbar}{(2m)^{\frac{1}{2}}} = \oint_\infty q\frac{dr}{r} - \oint_0 q\frac{dr}{r} = 2\pi\left(\frac{ze^2}{2(-E)^{\frac{1}{2}}} - \frac{\hbar(l+\tfrac{1}{2})}{(2m)^{\frac{1}{2}}}\right).$$

Solving for the energy gives the well-known eigenvalues

$$E_N = -\frac{z^2e^2}{2a_0N^2}, \tag{10}$$

where $N = N_r + l + 1$ and $a_0 = \hbar^2/me^2$.

B. First-Order Calculation of $\langle r^n\rangle$

We add a term λr^n to the potential. Using Eqs. (4) and (8) and retaining only terms from the first-order integral, we have

$$\langle r^n\rangle = \frac{(-E_N)^{\frac{3}{2}}}{\pi ze^2}\oint \frac{r^{n+2}}{q}\frac{dr}{r}. \tag{11}$$

Evaluation of the residue at the origin gives

$$\begin{aligned}\oint_0 \frac{r^{n+2}}{q}\frac{dr}{r} &= \frac{2\pi i}{(-n-2)!}\left(\frac{d}{dr}\right)^{-n-2} q^{-1}\Big|_{r=0}\\ &= \frac{2\pi i}{(-n-2)!}\left(\frac{d}{dr}\right)^{-n-2}\\ &\quad\times\left(-\frac{(l+\frac{1}{2})^2}{2m}\hbar^2 + ze^2r + E_Nr^2\right)^{-\frac{1}{2}}\Big|_{r=0}, \qquad n \le -2,\\ &= 0, \qquad n > -2.\end{aligned} \tag{12}$$

Evaluation of the residue at infinity gives

$$\begin{aligned}\oint_\infty \frac{r^{n+1}}{(q/r)}\frac{dr}{r} &= \frac{-2\pi i}{(n+1)!}\left(\frac{d}{d(1/r)}\right)^{n+1}\left(\frac{q}{r}\right)^{-1}\Big|_{1/r=0}\\ &= \frac{-2\pi i}{(n+1)!}\left(\frac{d}{ds}\right)^{n+1}\\ &\quad\times\left(E_N + ze^2s - \frac{(l+\frac{1}{2})^2\hbar^2s^2}{2m}\right)^{-\frac{1}{2}}\Big|_{s=1/r=0}, \qquad n \ge -1,\\ &= 0, \qquad n < -1.\end{aligned} \tag{13}$$

The derivatives can be evaluated with the aid of the formula[13]

$$\begin{aligned}&\frac{1}{m!}\left(\frac{d}{dx}\right)^m (a + bx + cx^2)^l\Big|_{x=0}\\ &\quad = a^l\left(\frac{b}{a}\right)^m \sum_{k=0}^{m}\binom{l}{m-k}\binom{m-k}{k}\left(\frac{ac}{b^2}\right)^k.\end{aligned} \tag{14}$$

Using the above formula, we express $\langle r^n\rangle$ in terms of the physical parameters, i.e.,

$$\begin{aligned}\langle r^n\rangle &= a_0^n\frac{z}{N^2}\left(\frac{-2N^2}{z}\right)^{n+1}\\ &\quad\times\sum_{k=0}^{n+1}\binom{-\frac{1}{2}}{n+1-k}\binom{n+1-k}{k}\left(\frac{l+\frac{1}{2}}{2N}\right)^{2k}, \qquad n \ge -1,\end{aligned} \tag{15}$$

$$\begin{aligned}\langle r^n\rangle &= \frac{a_0^nz^2}{(l+\frac{1}{2})N^3}\left(\frac{-2z}{(l+\frac{1}{2})^2}\right)^{-n-2}\\ &\quad\times\sum_{k=0}^{-n-2}\binom{-\frac{1}{2}}{-n-2-k}\binom{-n-2-k}{k}\left(\frac{l+\frac{1}{2}}{2N}\right)^{2k}, \qquad n \le -2.\end{aligned}$$

While these moments are, in general, only approximations, the approximations are good for large N and l and exact[14] for n equal to 0, -1, and -2. Furthermore, one cannot help but notice the ease with which all the moments for all n, N, and l have been calculated by doing only one contour integral.

The moments generated by the first-order integral for $2 \ge n \ge -3$ are

$$\begin{aligned}\langle r^2\rangle &= a_0^2\left(\frac{5N^4 - 3N^2(l+\frac{1}{2})^2}{2z^2}\right),\\ \langle r\rangle &= a_0\left(\frac{3N^2 - (l+\frac{1}{2})^2}{2z}\right),\\ \langle r^0\rangle &= 1,\\ \langle r^{-1}\rangle &= a_0^{-1}\left(\frac{z}{N^2}\right),\\ \langle r^{-2}\rangle &= a_0^{-2}\left(\frac{z^2}{N^3(l+\frac{1}{2})}\right),\\ \langle r^{-3}\rangle &= a_0^{-3}\left(\frac{z^3}{N^3(l+\frac{1}{2})^3}\right).\end{aligned} \tag{16}$$

C. Second-Order Calculation of $\langle r^n\rangle$

The accuracy of the calculated moments can be improved by the use of the higher-order integrals in Eq. (9). Expansion of the second integral to the first power in λ gives, from Eq. (8), the second-order correction to Eq. (11) as

$$\begin{aligned}\Delta\langle r^n\rangle &= \frac{2(-E_N)^{\frac{3}{2}}}{\pi ze^2}\frac{\hbar^2}{64m}\\ &\quad\times\oint\frac{Dq^2}{q^5}\left(2(n+2) - \frac{5}{2}\frac{Dq^2}{q^2}\right)r^{n+2}\frac{dr}{r}.\end{aligned}$$

Evaluation of the residue at the origin gives

$$\frac{1}{2\pi i}\oint_0 \frac{Dq^2}{q^5}\left(2(n+2)-\frac{5}{2}\frac{Dq^2}{q^2}\right)r^{n+2}\frac{dr}{r}$$
$$=\frac{2(n+2)}{2\pi i}\oint\left(\frac{ze^2r^{n+3}+2E_Nr^{n+4}}{q^5}\right)\frac{dr}{r}$$
$$-\frac{5}{4\pi i}\oint\left(\frac{z^2e^4r^{n+4}+4ze^2E_Nr^{n+5}+4E_N^2r^{n+6}}{q^7}\right)\frac{dr}{r}$$
$$=\frac{2(n+2)}{(-n-3)!}ze^2\left(\frac{d}{dr}\right)^{-n-3}q^{-5}\Big|_{r=0}$$
$$+\frac{4(n+2)E_N}{(-n-4)!}\left(\frac{d}{dr}\right)^{-n-4}q^{-5}\Big|_{r=0}$$
$$-\frac{5}{2}\frac{z^2e^4}{(-n-4)!}\left(\frac{d}{dr}\right)^{-n-4}q^{-7}\Big|_{r=0}$$
$$-\frac{10E_Nze^2}{(-n-5)!}\left(\frac{d}{dr}\right)^{-n-5}q^{-7}\Big|_{r=0}$$
$$-\frac{10E_N^2}{(-n-6)!}\left(\frac{d}{dr}\right)^{-n-6}q^{-7}\Big|_{r=0}.$$

Evaluating the residue at infinity gives

$$-\frac{1}{2\pi i}\oint_\infty \frac{Dq^2}{q^5}\left(2(n+2)-\frac{5}{2}\frac{Dq^2}{q^2}\right)r^{n+2}\frac{dr}{r}$$
$$=\frac{2(n+2)}{2\pi i}\oint\left(\frac{2E_Nr^{n-1}+ze^2r^{n-2}}{q^5}\right)\frac{dr}{r}$$
$$-\frac{5}{4\pi i}\oint\left(\frac{4E_N^2r^{n-1}+4E_Nze^2r^{n-2}+z^2e^4r^{n-3}}{q^7}\right)\frac{dr}{r}$$
$$=\frac{4(n+2)}{(n-1)!}E_N\left(\frac{d}{ds}\right)^{n-1}(sq)^{-5}\Big|_{s=0}$$
$$+\frac{2(n+2)}{(n-2)!}ze^2\left(\frac{d}{ds}\right)^{n-2}(sq)^{-5}\Big|_{s=0}$$
$$-\frac{10E_N^2}{(n-1)!}\left(\frac{d}{ds}\right)^{n-1}(sq)^{-7}\Big|_{s=0}$$
$$-\frac{10E_Nze^2}{(n-2)!}\left(\frac{d}{ds}\right)^{n-2}(sq)^{-7}\Big|_{s=0}$$
$$-\frac{5z^2e^4}{2(n-3)!}\left(\frac{d}{ds}\right)^{n-3}(sq)^{-7}\Big|_{s=0}.$$

The use of Eq. (14) permits the derivatives to be expressed in terms of the physical parameters. It is found that the five terms on the right-hand side of the above two equations all have the same functional dependence on these parameters. The corrections to the first-order moments are thus found to be

$$\Delta\langle r^2\rangle = 7a_0^2N^2/2z^2,$$
$$\Delta\langle r\rangle = a_0/8z,$$
$$\Delta\langle r^{-3}\rangle = a_0^{-3}z^3/4N^3(l+\tfrac{1}{2})^5.$$

The moments for n equal to 1 and 2 become exact with the addition of the second-order corrections, while the moment for $n=-3$ is still inexact. In all cases the second-order corrections provide a means for estimating the accuracy of the calculated moments. If further precision is required, the third-order correction term can be evaluated.

D. Third-Order Calculation of $\langle r^n\rangle$

The inclusion of the third-order integral in Eq. (8) gives

$$\Delta^2\langle r^n\rangle = \frac{2(-E_N)^{\frac{3}{2}}}{\pi ze^2}\frac{\hbar^4}{8192m^2}$$
$$\times\left[\oint\frac{(49)(Dq^2)^4}{q^{11}}\left(\frac{4(n+2)}{Dq^2}-\frac{11}{2q^2}\right)r^{n+2}\frac{dr}{r}\right.$$
$$\left.-\oint\frac{16(Dq^2)(D^3q^2)}{q^7}\times\left(\frac{(n+2)}{Dq^2}+\frac{(n+2)^3}{D^3q^2}-\frac{7}{2q^2}\right)r^{n+2}\frac{dr}{r}\right].$$

Since the second-order correction gave exact results for $-2\le n\le+2$, we expect the third-order correction for these moments will be zero, which is precisely true (as can be easily verified by performing the contour integrals above). For $n=-3$, the last integral in the above equation has a residue at the origin, the other integrals being zero. Evaluating this residue for the third-order correction yields

$$\Delta^2\langle r^{-3}\rangle = \frac{a_0^{-3}z^3}{16N^3(l+\frac{1}{2})^7}.$$

The moment for $n=-3$, correct to third order, is then

$$\langle r^{-3}\rangle = \frac{a_0^{-3}z^3}{N^3(l+\frac{1}{2})^3}\left(1+\frac{1}{(2l+1)^2}+\frac{1}{(2l+1)^4}\right).$$

While the moment is still inexact, it is the correct series expansion for the exact moment, which can be written as

$$\langle r^{-3}\rangle = \frac{a_0^{-3}z^3}{N^3(l+\frac{1}{2})l(l+1)}$$
$$=\frac{a_0^{-3}z^3}{N^3(l+\frac{1}{2})^3\{1-[1/(2l+1)^2]\}}$$
$$=\frac{a_0^{-3}z^3}{N^3(l+\frac{1}{2})^3}\left(1+\frac{1}{(2l+1)^2}+\frac{1}{(2l+1)^4}+\frac{1}{(2l+1)^6}+\cdots\right).$$

It is seen that the difference between the calculated third-order moment and the exact moment is small even for low values of l. For $l=1$ the error is 0.14%,

and for $l = 2$ the error is 0.0064%. For $l = 0$ one could infer that the moment is infinite from the fact that the higher-order corrections do not appear to be converging.

IV. DISCUSSION

The above calculation demonstrates the feasibility of obtaining expectation values from the WKB approximation without the necessity of first obtaining the WKB wavefunction. This is particularly convenient since it is often difficult to evaluate the integrals that represent the space dependence of the wavefunction. Furthermore, the higher-order corrections to the wavefunction involve integrals which become increasingly singular near the classical turning points of the effective Schrödinger equation, and hence would require a more exact approach to determine the wavefunction in these regions.

The technique yields a systematic way of improving the accuracy of the calculation by including the higher-order integrals. Thus, for the Coulomb potential, $\langle r^n \rangle$ for $-2 \geq n \leq 2$ is given exactly when the first two integrals are included, the third-order integral then giving a zero contribution. Moreover, even when the inclusion of the third-order integral does not lead to the exact result as in the calculation of the very singular term $\langle r^{-3} \rangle$, contributions from the higher-order integrals rapidly decrease giving not only excellent results even for low quantum numbers, but a means to estimate the accuracy of the calculation. Since the relative smallness of the higher-order integrals appears to be a general property of the WKB expansion, we anticipate that similar results will obtain for other potentials. Furthermore, by performing coordinate transformations that preserve the boundary conditions, it is possible to make the contributions of the higher-order integrals smaller[15] and thus improve the accuracy of the calculation.

Finally, we note that, when the relevant integrals can be analytically evaluated, this technique is capable of yielding analytic expressions for expectation values for all quantum states by performing only a trivial calculation, as is evident in the case of the moments for the Coulomb potential.

[1] C. L. Delves, Nucl. Phys. **41**, 497 (1963); S. Aranoff and J. K. Percus, Phys. Rev. **166**, 1255 (1968).
[2] S. Borowitz and M. O. Vassel, J. Quant. Spectry. Rad. Trans. **4**, 663 (1964); T. K. Rebane, Opt. Spektrosk. **21**, 118 (1966) [Opt. Spectrosc. **21**, 66 (1966)]; M. N. Adamov, B. I. Orlov, and T. K. Rebane, Opt. Spektrosk. **24**, 657 (1968) [Opt. Spectrosc. **24**, 353 (1968)].
[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, London, 1958), p. 178.
[4] J. L. Dunham, Phys. Rev. **41**, 713 (1932).
[5] J. E. Killpatrick and M. F. Killpatrick, J. Chem. Phys. **16**, 781 (1948).
[6] J. B. Krieger, M. L. Lewis, and C. Rosenzweig, J. Chem. Phys. **47**, 2942 (1967).
[7] J. D. Bekenstein and J. B. Krieger, Phys. Rev. **188**, 130 (1969).
[8] J. D. Bekenstein and J. B. Krieger, J. Math. Phys. **11**, 2921 (1970).
[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 1004.
[10] J. B. Krieger, J. Math. Phys. **9**, 432 (1968).
[11] R. E. Langer, Phys. Rev. **51**, 669 (1937).
[12] J. B. Krieger and C. Rosenzweig, Phys. Rev. **164**, 171 (1967).
[13] This follows directly from a double binomial expansion.
[14] A. S. Davydov, *Quantum Mechanics* (Pergamon, London, 1965), p. 138.
[15] J. B. Krieger, J. Math. Phys. **10**, 1455 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 10 OCTOBER 1970

Three-Colorings of the Square Lattice: A Hard Squares Model

R. J. BAXTER
Mathematics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 16 February 1970)

The evaluation of the partition function of the 2-dimensional ice model is equivalent to counting the number of ways of coloring the faces of the square lattice with three colors so that no two adjacent faces are colored alike. In this paper we solve a generalized problem in which activities are associated with the colors. If one of the colors is regarded as a particle and the others as forming a background, then the model is reminiscent of the hard-square lattice gas. It is found to undergo a phase transition with infinite compressibility at the density $\rho = 1/3$.

1. INTRODUCTION

The 2-dimensional ice model was solved by Lieb[1,2] in 1967. Lenard (see Ref. 2) has pointed out that the model is equivalent to counting the number of ways of coloring the faces of the square lattice with three colors, so that no two adjacent faces are colored alike. Calling these colors 1, 2, and 3, it follows that the partition function for a lattice with N_t faces can be written as

$$Z = \sum G(N_1, N_2, N_3), \tag{1.1}$$

where the summation is over all nonnegative integers N_1, N_2, and N_3 such that $N_1 + N_2 + N_3 = N_t$ and $G(N_1, N_2, N_3)$ is the number of allowed ways of coloring the faces so that N_1 are colored 1, N_2 are colored 2, and N_3 are colored 3.

Clearly, these colors can be regarded as three types of particles, with an infinitely repulsive force between nearest neighbors of the same type. This suggests associating activities z_1, z_2, z_3 with the three colors and generalizing (1.1) to

$$Z = \sum z_1^{N_1} z_2^{N_2} z_3^{N_3} G(N_1, N_2, N_3), \tag{1.2}$$

where the summation is as before. Z is then the grand canonical partition function of the particle system, evaluated at close packing of the lattice.

In this paper we solve this problem exactly in the thermodynamic limit when N_t becomes large. We find that

$$Z \sim W^{N_t}, \tag{1.3}$$

where

$$W = (z_1 z_2 z_3)^{\frac{1}{3}} W_{\mathrm{D}} \tag{1.4}$$

and W_{D} is a function only of the dimensionless parameter

$$B = (z_2 z_3 + z_3 z_1 + z_1 z_2)/[3(z_1 z_2 z_3)^{\frac{2}{3}}]. \tag{1.5}$$

In fact, we find that W_{D} is simply an algebraic function of B. If t is the root of the equation

$$B^3 = (1 - 3t^2)^3/(1 - 9t^2) \tag{1.6}$$

such that $0 \leq t < \frac{1}{3}$ (since from its definition $B \geq 1$ there is always one and only one such root), then W_{D} is given by

$$W_{\mathrm{D}}^2 = 64(1 - 9t^2)^{\frac{2}{3}}/[27(1 + t)^3(1 - 3t)]. \tag{1.7}$$

From (1.6) and (1.7) we can see that W_{D} is an analytic function of B in the interval $1 < B < \infty$, while at $B = 1$ it has a branch point. Since, for real positive activities, $B = 1$ only when $z_1 = z_2 = z_3$, it follows that W is an analytic function of z_1, z_2, and z_3, except when they are all equal.

When $z_1 = z_2 = z_3 = 1$, the partition function (1.2) reduces to (1.1), and we regain the ice model. In this case, $B = 1$ and $t = 0$, so that (1.4) and (1.7) give

$$W = (\tfrac{4}{3})^{\frac{3}{2}}, \tag{1.8}$$

which is Lieb's result.

Another special case that is particularly interesting is when

$$z_1 = z, \quad z_2 = z_3 = 1. \tag{1.9}$$

Colors 1 can then be regarded as particles and colors 2 and 3 as forming a background. Since no two such particles can occupy adjacent faces of the lattice, the model is reminiscent of the hard-square lattice gas.[3] In fact, any allowed coloring of the lattice corresponds to a unique hard square configuration, and any hard square configuration corresponds to at least two colorings, since the empty squares can always be colored 2 and 3 in one of two possible checkerboard fashions.

Unfortunately, the correspondence is not simply 2-to-1, since the hard squares may enclose "lagoons" of empty squares such that no empty square inside the lagoon borders one outside. Such a lagoon can be colored in a checkerboard fashion independently of that used for the other empty squares. For instance, the empty squares in the hard square configuration shown in Fig. 1 can be colored 2 and 3 in four ways.

Nevertheless, this case of the coloring problem resembles the hard squares gas in that it has only

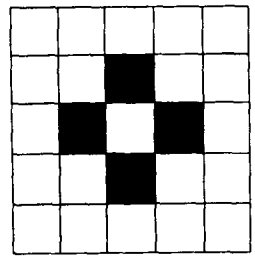

FIG. 1. A hard square configuration that corresponds to four possible 3-colorings of the lattice (the hard squares being colored 1).

infinitely repulsive nearest-neighbor forces. Further, since the probability of occurrence of lagoons such as that of Fig. 1 is quite small at low densities, the two models should agree fairly closely in this regime. In particular, as four particles are needed to form a lagoon, the first three virial coefficients of both models should be the same.

Making the substitutions (1.9) into the previous equations, we see that the density ρ of particles is given by

$$\rho = z \frac{d \log W}{dz} \tag{1.10}$$

and lies between 0 and the close-packed value $\frac{1}{2}$. It is found that W is a continuous function of ρ, monotonic increasing except at $\rho = \frac{1}{3}$, where its derivative vanishes and $z = 1$. Neglecting the density-independent Boltzmann factor, we see that the compressibility K_T is given by

$$K_T^{-1} = \rho \frac{d \log W}{d\rho} \tag{1.11}$$

and near $\rho = \frac{1}{3}$ is found to behave as

$$K_T \sim 2/[9\,|\rho - \tfrac{1}{3}|]. \tag{1.12}$$

Thus the system undergoes a phase transition with continuous density and infinite compressibility, the values of ρ, z, and W at the critical point being

$$\rho_c = \tfrac{1}{3}, \quad z_c = 1, \quad W_c = (\tfrac{4}{3})^{\frac{3}{2}}. \tag{1.13}$$

Despite the simplicity of the result (1.7), its derivation is quite lengthy, involving the theory of elliptic functions, and is given in Secs. 2–7. In Sec. 8 the pseudo hard-squares model mentioned above is discussed in more detail and graphs of $\log W$ and $d(\log W)/d\rho$ given. It is found that the results do approximately agree with those of the true hard-square lattice gas at low densities (say $\rho < 0.2$), but there are considerable differences in the transition region and above.

2. THE TRANSFER MATRIX

Consider a square lattice of M rows and N columns, and suppose it to be wound on a torus so that column N is followed by column 1 and row M by row 1. Then the total number of faces of the lattice is

$$N_t = MN. \tag{2.1}$$

Suppose all the faces are colored 1, 2, or 3 so that no two adjacent faces are colored alike. It is convenient, though not essential, to describe such a coloring in ice model terms. To do this, we order the colors cyclically (so that 2 follows 1, 3 follows 2, and 1 follows 3) and place arrows on all the bonds of the lattice so that an observer facing along an arrow sees the color on his right as following the one on his left. Then the four arrows at a vertex are found to satisfy the ice condition, namely that there are two pointing in and two pointing out.

It can then be seen that the number of down (or up) arrows in each row of vertical bonds is the same. Suppose there are n down arrows in each row. Then the coloring of the faces of a row can be specified by the color σ of the extreme left-hand face and the positions $x_1, x_2, \cdots, x_n$ of the down arrows.

The general coloring of a row is shown in Fig. 2. The color numbers increase from left to right across an up arrow and decrease across a down arrow, and are to be interpreted mod 3. For example, colors 1, 4, 7, 10, etc., are the same. The cyclic boundary condition implies that colors at the extreme left and right must be the same, and so N and n must satisfy the relation

$$N - 2n = 3J, \tag{2.2}$$

where J is an integer.

Noting that the jth down arrow has the effect of repeating colors $\sigma + x_j - 2j$ and $\sigma + x_j - 2j + 1$ in the otherwise increasing sequence, we can see that the product of the activities of the colors in the row is

$$D(X) = \xi \prod_{j=1}^{n} \zeta(\sigma + x_j + j), \tag{2.3}$$

where

$$\xi = (z_1 z_1 z_3)^{\frac{1}{3}N}, \tag{2.4}$$

$$\zeta(\sigma) = z_\sigma z_{\sigma+1}/(z_1 z_2 z_3)^{\frac{2}{3}}, \tag{2.5}$$

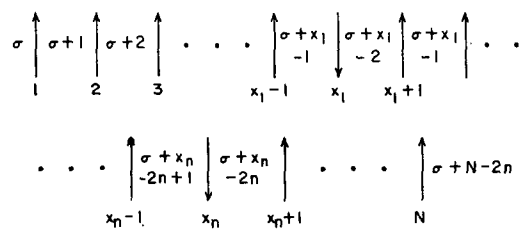

FIG. 2. Coloring of a row of the lattice; the arrows lie on vertical bonds and the numbers between are the colors of the faces. (The second line is to be regarded as to the right of the first.)

and the mod 3 convention, i.e.,

$$\zeta(\sigma + 3) \equiv \zeta(\sigma), \tag{2.6}$$

has been used.

The transfer matrix can now be constructed in the same way as for the ice model,[2] except that it is multiplied by a diagonal matrix with elements $D(X)$ and it is necessary to keep track of the color of the left-hand face of each row. Let $f_\sigma(x_1, x_2, \cdots, x_n)$ be the elements (corresponding to the row-coloring of Fig. 2) of the eigenvector of the transfer matrix. Then f and the corresponding eigenvalue λ are given by

$$\lambda f_\sigma(x_1, \cdots, x_n) = D(X)\Big(\sum_{y_1=1}^{x_1}\sum_{y_2=x_1}^{x_2}\cdots\sum_{y_n=x_{n-1}}^{x_n} f_{\sigma+1}(y_1, \cdots, y_n) + \sum_{y_1=x_1}^{x_2}\sum_{y_2=x_2}^{x_3}\cdots\sum_{y_n=x_n}^{N} f_{\sigma+2}(y_1, \cdots, y_n)\Big), \tag{2.7}$$

where $1 \leq x_1 < x_2 < \cdots < x_n \leq N$, $f_{\sigma+3} \equiv f_\sigma$, and f is to be replaced by zero on the right-hand side of (2.7) when two successive y's become equal (e.g., $y_1 = x_1 = y_2$).

In the limit of M large, the partition function Z is given by

$$Z \sim \lambda^M, \tag{2.8}$$

where λ is the largest eigenvalue of Eq. (2.7). Thus, from (1.3) and (2.1),

$$\lambda \sim W^N \tag{2.9}$$

when N becomes large.

3. DIAGONALIZATION OF THE TRANSFER MATRIX

By analogy with the method of the ice[1,2] and similar[4] models, we now attempt to solve the eigenvalue equation (2.7) by trying a solution for the eigenvector f which is the sum of products of single-particle wavefunctions. Some consideration of the translation symmetries of the problem suggests the ansatz

$$f_\sigma(x_1, \cdots, x_n) = \sum_P A_{\alpha_1,\cdots,\alpha_n} \prod_{j=1}^{n} G_{\alpha_j}(x_j'), \tag{3.1}$$

where

$$x_j' = x_j + \sigma + j \tag{3.2}$$

and the summation is over all $n!$ permutations $\alpha_1, \cdots, \alpha_n$ of the integers $1, \cdots, n$. Thus there are n single-particle functions $G_1(x'), \cdots, G_2(x')$ to be determined, together with the $n!$ coefficients A.

The condition $f_{\sigma+3} = f_\sigma$ can be satisfied by requiring that for each function $G_\alpha(x')$ there exists p_α such that

$$G_\alpha(x' + 3) = e^{3ip_\alpha} G_\alpha(x') \tag{3.3}$$

for all integers x' and that

$$p_1 + p_2 + \cdots + p_n = 0. \tag{3.4}$$

Thus the single-particle functions are plane waves mod 3 with wavenumbers $p_1, \cdots, p_n$, and f is translation invariant in the sense that it is unchanged by replacing each x_j by $x_j + 3$.

As a first step, we insert the single product $f = G_1(x_1')G_2(x_2') \cdots G_n(x_n')$ into Eq. (2.7). Using (2.3), we see that the right-hand side becomes

$$\xi \prod_{j=1}^{n} \{\zeta(x_j')[H_j(x_{j-1}') - H_j(x_j')]\} + \xi \prod_{j=1}^{n} \{\zeta(x_j')[H_j(x_j') - H_j(x_{j+1}')]\}, \tag{3.5}$$

where $x_0' = \sigma + 1$, $x_{n+1}' = N + n + \sigma + 1$, and

$$H_\alpha(x) = \sum_{y=x+2}^{N} G_\alpha(y) + \text{const.} \tag{3.6}$$

However, (3.5) spuriously includes terms which arise when two y's are equal in (2.7). These must be subtracted; for instance, the terms coming from the case $y_1 = x_1 = y_2$ in the first set of summations give a correction

$$-\xi\Big(\prod_{j=1}^{n} \zeta(x_j')\Big) G_1(x_1' + 1)G_2(x_1' + 2) \times \prod_{j=3}^{n} [H_j(x_{j-1}') - H_j(x_j')] \tag{3.7}$$

to (3.5).

Expanding each product in (3.5) gives 2^n terms, of which all but one contain either x_0', x_{n+1}', or two functions H of the same variable. If n is even, the two other terms (one from each product) reinforce one another to give a contribution

$$2\xi \prod_{j=1}^{n} [\zeta(x_j')H_j(x_j')] \tag{3.8}$$

to the right-hand side of (2.7). This is also a product of single-particle functions and can be made to equal the left-hand side of (2.7) by requiring that

$$\mu_\alpha G_\alpha(x) = \zeta(x)H_\alpha(x), \tag{3.9}$$

for $\alpha = 1, \cdots, n$ and all integers x, and that

$$\lambda = 2\xi\mu_1\mu_2 \cdots \mu_n. \tag{3.10}$$

The constant in Eq. (3.6) may depend on α, but not on x. By comparing (2.6), (3.3), and (3.9), it is apparent that this constant must be chosen so that

$$H_\alpha(x + 3) = e^{3ip_\alpha} H_\alpha(x), \tag{3.11}$$

i.e., $H_\alpha(x)$ is also a plane wave mod 3. Equations (3.6) and (3.9) then reduce to a cubic eigenvalue equation determining μ_α and the single-particle function $G_\alpha(x)$ in terms of p_α.

From (3.10) it is clear that λ is unchanged by permuting the single-particle functions and hence $\mu_1, \cdots, \mu_n$. Thus one can try the general form (3.1) in Eq. (2.7) and attempt to choose $p_1, \cdots, p_n$ and the coefficients A so as to cancel all the remaining unwanted terms on the right-hand side. Since the correction terms such as (3.7) are of the same type as the terms in the expansion of (3.5) which contain two functions H of the same variable, these can be made to cancel by requiring that

$$s_{\alpha,\beta}(x)A_{\cdots,\alpha,\beta,\cdots} + s_{\beta,\alpha}(x)A_{\cdots,\beta,\alpha,\cdots} = 0 \quad (3.12)$$

for all integers x, where

$$s_{\alpha,\beta}(x) = H_\alpha(x)H_\beta(x) + G_\alpha(x+1)G_\beta(x+2). \quad (3.13)$$

The coefficients A in (3.12) differ only in that the two successive indices α and β are interchanged.

It is also necessary to cancel the terms in (3.5) containing x_0' and x_{n+1}'. This can be done by performing a single cyclic shift of $\alpha_1, \cdots, \alpha_n$ and noting that the terms containing x_0' in the original permutation are the same as those containing x_{n+1}' in the second, which leads to the condition

$$H_{\alpha_1}(\sigma+1)A_{\alpha_1,\cdots,\alpha_n} = H_{\alpha_1}(N+n+\sigma+1)A_{\alpha_2,\cdots,\alpha_n,\alpha_1} \quad (3.14)$$

for $\sigma = 1, 2, 3$. It can be shown that the conditions (3.12) and (3.14) are, in fact, sufficient to ensure that all the unwanted terms in the transfer matrix equation cancel.

It remains to show that $p_1, \cdots, p_n$ and the A's can be chosen to satisfy these conditions. To do this, it is first necessary to solve Eqs. (3.6), (3.9), and (3.11) for the single-particle functions. From (3.6) it can be seen that

$$G_\alpha(x) = H_\alpha(x-2) - H_\alpha(x-1). \quad (3.15)$$

Substituting this result into (3.9), using (3.11), and setting

$$\nu_\alpha = -\mu_\alpha e^{-3ip_\alpha}, \quad (3.16)$$

we find that three successive values of $H_\alpha(x)$ can be written as

$$\begin{aligned} H_\alpha(x-1) &= \mu_\alpha\nu_\alpha + \nu_\alpha\zeta(x+1), \\ H_\alpha(x) &= \mu_\alpha\nu_\alpha - \zeta(x-1)\zeta(x+1), \\ H_\alpha(x+1) &= \mu_\alpha\nu_\alpha + \mu_\alpha\zeta(x-1), \end{aligned} \quad (3.17)$$

while μ_α and ν_α must satisfy the relation

$$\mu_\alpha^2\nu_\alpha + \mu_\alpha\nu_\alpha^2 + 3B\mu_\alpha\nu_\alpha - 1 = 0, \quad (3.18)$$

where

$$3B = \zeta(1) + \zeta(2) + \zeta(3), \quad (3.19)$$

i.e., B is given by (1.5). We have used the relation

$$\zeta(1)\zeta(2)\zeta(3) = 1, \quad (3.20)$$

which follows from (2.5).

Substituting (3.15) and (3.17) into the definition (3.13) of $s_{\alpha,\beta}(x)$ and using (3.20) give

$$\zeta(x)s_{\alpha,\beta}(x) = u_{\alpha,\beta} + \zeta(x-1)v_{\alpha,\beta} + \zeta(x)w_{\alpha,\beta}, \quad (3.21)$$

where $u_{\alpha,\beta}$, $v_{\alpha,\beta}$, and $w_{\alpha,\beta}$ are independent of x, being given by

$$\begin{aligned} u_{\alpha,\beta} &= -3B\nu_\alpha - \mu_\alpha\nu_\alpha - \mu_\beta\nu_\beta - \nu_\alpha\mu_\beta, \\ v_{\alpha,\beta} &= \nu_\alpha - \mu_\beta, \\ w_{\alpha,\beta} &= \mu_\alpha\nu_\alpha\mu_\beta\nu_\beta + \nu_\alpha. \end{aligned} \quad (3.22)$$

Using (3.18), we now find an essential and otherwise not obvious result, namely, that

$$u_{\alpha,\beta}/u_{\beta,\alpha} = v_{\alpha,\beta}/v_{\beta,\alpha} = w_{\alpha,\beta}/w_{\beta,\alpha}. \quad (3.23)$$

Thus the ratio $s_{\alpha,\beta}(x)/s_{\beta,\alpha}(x)$ is also equal to the common ratio in (3.23). Selecting the simplest expression in (3.22), we can say that

$$s_{\alpha,\beta}(x)/s_{\beta,\alpha}(x) = v_{\alpha,\beta}/v_{\beta,\alpha}. \quad (3.24)$$

The condition (3.12) is therefore independent of x and can be satisfied by choosing the A's so that

$$A_{\alpha_1,\cdots,\alpha_n} = \epsilon_P \prod_{1\le j<k\le n} v_{\alpha_k,\alpha_j}, \quad (3.25)$$

where $\epsilon_P = \pm 1$ is the sign of the permutation $\alpha_1, \cdots, \alpha_n$.

Finally, the condition (3.14) must be satisfied. From (2.2) it is apparent that $\sigma + 1$ and $N + n + \sigma + 1$ differ by an integer multiple of 3; thus, using (3.11), we obtain

$$A_{\alpha_1,\cdots,\alpha_n} = \exp[i(N+n)p_{\alpha_1}]A_{\alpha_2,\cdots,\alpha_n,\alpha_1}. \quad (3.26)$$

By substituting (3.25) into (3.26) and using (3.22), it follows that the condition (3.14) is satisfied provided that

$$e^{i(N+n)p_\alpha} = (-)^{n-1}\prod_{\beta=1}^{n}\frac{\nu_\beta - \mu_\alpha}{\nu_\alpha - \mu_\beta} \quad (3.27)$$

for $\alpha = 1, 2, \cdots, n$.

The results of this section can now be summarized: Equations (3.16) and (3.18) define μ_α and ν_α as functions of p_α. Thus (3.27) is a set of n equations for the n unknown wavenumbers $p_1, \cdots, p_n$. Once these are satisfied, the ansatz does indeed give a solution of the transfer matrix equations, with the eigenvalue λ given by (3.10).

It should be noted that (3.27) has solutions where two or more of the wavenumbers p_α are equal. However, such solutions should be ignored since all the elements of the eigenvector f can then be seen to vanish.

4. EQUATIONS FOR THE MAXIMUM EIGENVALUE

Equations (3.16), (3.18), (3.27), and (3.10) can be written more explicitly by defining a set of quantities $g_1, \cdots, g_n$ such that

$$\mu_\alpha = ig_\alpha^{-1} e^{\frac{3}{2}ip_\alpha}. \tag{4.1}$$

Equation (3.16) then becomes

$$v_\alpha = -ig_\alpha^{-1} e^{-\frac{3}{2}ip_\alpha}, \tag{4.2}$$

and from (3.18) it follows that $g_\alpha = g(p_\alpha)$, where the function $g \equiv g(p)$ is defined by

$$g^3 - 3Bg + 2\sin(\tfrac{3}{2}p) = 0. \tag{4.3}$$

Using (4.1), (4.2), and the condition (3.4), we can write Eqs. (3.27) as

$$e^{iNp_\alpha} = (-)^{n-1} \prod_{\beta=1}^{n} e^{-i\theta_{\alpha,\beta}} \tag{4.4}$$

for $\alpha = 1, \cdots, n$, where

$$e^{-i\theta_{\alpha,\beta}} = F_{\beta,\alpha}/F_{\alpha,\beta}, \tag{4.5}$$

and

$$F_{\alpha,\beta} = e^{i[\frac{1}{2}p_\alpha + p_\beta]} g_\alpha + e^{-i[p_\alpha + \frac{1}{2}p_\beta]} g_\beta. \tag{4.6}$$

Lastly, using (4.1), we see that Eq. (3.10) for λ becomes

$$\lambda = 2(-)^{\frac{1}{2}n} \xi/(g_1 g_2 \cdots g_n). \tag{4.7}$$

The problem is now to find the solution $p_1, \cdots, p_n$ of (4.3)–(4.6) which gives the largest value of λ. Equation (4.7) suggests that this solution is obtained when the function g of p is chosen to be the numerically smallest root of the cubic equation (4.3). Since B can be seen from its definition (1.5) to be not less than one, the smallest root g of (4.3) is a real function of p, of period $\frac{4}{3}\pi$, and odd. By taking the appropriate logarithm of both sides of (4.5) and using (4.6), it follows that $\theta_{\alpha,\beta} \equiv \theta(p_\alpha, p_\beta)$ is a real function of p_α and p_β such that

$$\theta(p_\alpha, p_\beta) = -\theta(p_\beta, p_\alpha) = -\theta(-p_\alpha, -p_\beta). \tag{4.8}$$

Using these properties, we find that (4.4) has solutions such that the wavenumbers $p_1, \cdots, p_n$ are real and occur in pairs p, $-p$. By analogy with the ice[1,2] and similar[4] models (or by considering the case when $N \gg n$), we expect the maximum eigenvalue λ to be given by taking the logarithms of both sides of (4.4) so that

$$Np_\alpha = \pi(2\alpha - n - 1) - \sum_{\beta=1}^{n} \theta_{\alpha,\beta} \tag{4.9}$$

for $\alpha = 1, \cdots, n$. This choice of the logarithms ensures that

$$p_1 < p_2 < \cdots < p_n, \tag{4.10}$$

$$p_{n+1-\alpha} = -p_\alpha, \tag{4.11}$$

and gives the solution of (4.4) which has the most closely packed distribution of distinct p_α about the origin.

5. TRANSFORMATION TO DIFFERENCE KERNEL FORM

In all the ice-type models[1,2,4] it is found that a transformation

$$u_\alpha \equiv u(p_\alpha) \tag{5.1}$$

exists such that $\theta_{\alpha,\beta}$ depends only on the difference between u_α and u_β, i.e.,

$$\theta_{\alpha,\beta} \equiv \theta(u_\alpha - u_\beta). \tag{5.2}$$

In this case, it is by no means obvious that such a transformation exists, but let us suppose that it does and consider the consequences.

Equation (5.2) is equivalent to asserting that the functions $\theta(p_\alpha, p_\beta)$ and $u(p_\alpha) - u(p_\beta)$ of the two variables p_α and p_β are functionally dependent, so that their Jacobian must vanish identically, i.e.,

$$\begin{vmatrix} u'(p) & -u'(q) \\ \dfrac{\partial\theta(p,q)}{\partial p} & \dfrac{\partial\theta(p,q)}{\partial q} \end{vmatrix} \equiv 0, \tag{5.3}$$

where p_α and p_β have been replaced by p and q and $u'(p)$ is the derivative of the function $u(p)$.

By taking the limit of $q \to 0$ in (5.3) and using (4.3)–(4.6), it follows that

$$u'(p) = L/[B - g^2(p)], \tag{5.4}$$

where L is an arbitrary constant. Substituting this form for the functions $u'(p)$ and $u'(q)$ back into (5.3) and using (4.3)–(4.6), we find that (5.3) is indeed satisfied. This establishes the required identity (5.2).

When $p_\beta = 0$, it also follows from (4.3)–(4.6) that $g_\beta = 0$ and $\theta_{\alpha,\beta} = p_\alpha$. By inverting the relation between p and u so as to regard p as a function $p(u)$ of u, it follows that

$$\theta_{\alpha,\beta} \equiv p(u_\alpha - u_\beta). \tag{5.5}$$

It therefore becomes necessary to use (4.3) and (5.4) to express g and p as functions of u. Equation (4.3) can be used to evaluate dg/dp in terms of g; dividing the result into (5.4) gives

$$\frac{du}{dg} = 2L/[4 - g^2(3B - g^2)^2]^{\frac{1}{2}}. \tag{5.6}$$

Changing to the variable $x = g^2$, we can integrate Eq. (5.6) by using elliptic functions (Sec. 3.147.2 and 8.11–8.15 of Ref. 5). Let a, b, and c be the three roots of the cubic equation

$$x(3B - x)^2 - 4 = 0. \tag{5.7}$$

Since $B \geq 1$, these roots are real and positive and can be ordered so that

$$a > b \geq c > 0. \tag{5.8}$$

Choosing

$$L = \tfrac{1}{2}[(a - c)b]^{\frac{1}{2}} \tag{5.9}$$

and u to be zero when p and g are zero, we find that g is given as a function $g(u)$ of u by

$$g^2(u) = ac\,\mathrm{sn}^2(u)/[a - c + c\,\mathrm{sn}^2(u)], \tag{5.10}$$

where sn (u) is the usual elliptic sn function, with modulus

$$k = \left(\frac{(a - b)c}{(a - c)b}\right)^{\frac{1}{2}}. \tag{5.11}$$

p is now given as a function of u by Eqs. (4.3) and (5.10). Alternatively, Eq. (5.4) can be inverted to give

$$\frac{dp}{du} = h(u), \tag{5.12}$$

where

$$h(u) = [B - g^2(u)]/L. \tag{5.13}$$

6. THE THERMODYNAMIC LIMIT

Using (5.5), we can write Eq. (4.9) in terms of the new variables $u_1, \cdots, u_n$ as

$$Np(u_\alpha) = \pi(2\alpha - n - 1) - \sum_{\beta=1}^{n} p(u_\alpha - u_\beta) \tag{6.1}$$

for $\alpha = 1, \cdots, n$. It can be seen from the above that p is a monotonic increasing odd function of u; thus, from (4.10) and (4.11), $u_1, \cdots, u_n$ are arranged in increasing order and are distributed symmetrically about the value zero.

We are interested in the thermodynamic limit when the width of the lattice becomes large, so that n, $N \to \infty$, the ratio n/N remaining fixed. As with the ice-type models,[1,2,4] we expect $u_1, \cdots, u_n$ to approach a continuous distribution along the real axis in some interval $(-Q, Q)$, so that a distribution function $\rho(u)$ can be defined such that $N\rho(u)\,du$ is the number of u_α lying between u and $u + du$. Thus the total number of u_α is

$$n = N\int_{-Q}^{Q} \rho(u)\,du, \tag{6.2}$$

while the value of α corresponding to $u_\alpha = u$ is

$$\alpha = N\int_{-Q}^{u} \rho(u')\,du'. \tag{6.3}$$

In this limit, (6.1) therefore becomes the integral equation

$$p(u) = 2\pi\int_0^u \rho(u')\,du' - \int_{-Q}^{Q} p(u - u')\rho(u')\,du' \tag{6.4}$$

[by using the fact that $\rho(u)$ must be an even function].

The eigenvalue λ and hence the partition function can be obtained from (4.7). First note that from (1.4), (2.4), and (2.9)

$$N^{-1}\log(\lambda/2\xi) = \log W_{\mathrm{D}} \tag{6.5}$$

(N large). Since the g_α occur in pairs g, $-g$, it follows that, in the thermodynamic limit, (4.7) becomes

$$\log W_{\mathrm{D}} = -\tfrac{1}{2}\int_{-Q}^{Q} \rho(u)\log g^2(u)\,du. \tag{6.6}$$

Equation (6.4) determines $\rho(u)$ in terms of the known function $p(u)$ and the constant Q. Q can then be determined in terms of n/N from (6.2) and W_{D} evaluated from (6.6).

Differentiating (6.4) and using (5.12) give

$$h(u) = 2\pi\rho(u) - \int_{-Q}^{Q} h(u - u')\rho(u')\,du'. \tag{6.7}$$

From (5.10), (5.13), and the theory of elliptic functions,[5] it can be seen that $h(u)$ is an even function of u, of period $2K$, where K is the complete elliptic integral of the first kind of modulus k. Thus, when $Q = K$, Eq. (6.7) can be solved by Fourier series.

$h(u)$ is also periodic with period $2iK'$, where K' is the complete elliptic integral of the first kind of modulus

$$k' = (1 - k^2)^{\frac{1}{2}} = \left(\frac{(b - c)a}{(a - c)b}\right)^{\frac{1}{2}} \tag{6.8}$$

[from Eq. (5.11)]. The only singularities of $h(u)$ in the complex u plane are simple poles at

$$u = \pm i\tau + \text{integral multiples of } 2K \text{ and } 2iK', \tag{6.9}$$

where

$$\mathrm{sn}(i\tau) = i[(a - c)/c]^{\frac{1}{2}}. \tag{6.10}$$

From the cubic equation (5.7), it follows that a, b, and c satisfy the dimensionless relation

$$a^{\frac{1}{2}} = b^{\frac{1}{2}} + c^{\frac{1}{2}}. \tag{6.11}$$

Using this result together with (5.11), (6.10), and the addition theorems for elliptic functions (8.156) and (8.157) of Ref. 5, we find that

$$\tau = \tfrac{2}{3}K'. \tag{6.12}$$

Also, using (5.9)–(5.11) and the differentiation formulas for elliptic functions, we find the residues of $g^2(u)$ at $\pm i\tau$ to be

$$\mathrm{Res}\,[\pm i\tau \,|\, g^2(u)] = \pm iL. \tag{6.13}$$

The coefficients of the Fourier series for $h(u)$ can now be found[6] by integrating around a rectangle in the complex u plane with vertices at $-K$, K, $K+2iK'$, and $-K+2iK'$ and by using Cauchy's residue theorem. This gives

$$h(u)=\frac{\pi}{K}\sum_{m=-\infty}^{\infty}\frac{e^{i\pi mu/K}}{r^{-m}+1+r^{m}}, \tag{6.14}$$

where

$$r=e^{-2\pi K'/3K}. \tag{6.15}$$

Equation (6.7) can now be solved by Fourier series to give

$$\rho(u)=\frac{1}{2K}\sum_{m=-\infty}^{\infty}\frac{e^{i\pi mu/K}}{r^{m}+r^{-m}}. \tag{6.16}$$

Substituting the expression (6.16) into (6.2), we find that

$$n=\tfrac{1}{2}N \tag{6.17}$$

when $Q=K$. Thus there are as many up arrows as down arrows in each row of the lattice. Intuitively, we expect this to be the case which gives the maximum eigenvalue λ of the transfer matrix and, hence, the correct values of W and W_{D}.

To evaluate W_{D}, we first obtain the coefficients of the Fourier series for $\log g^2(u)$ by integrating around the above mentioned rectangle in the complex u plane and taking account of the branch cuts on the segments $[0, i\tau]$ and $[2i\tau, 2iK']$ of the imaginary axis.[7] This gives

$$\log g^2(u)=-\frac{4\pi K'}{9K}-2\sum_{m=1}^{\infty}\frac{(1-r^{2m})\cos(\pi mu/K)}{m(1+r^{m}+r^{2m})}. \tag{6.18}$$

Using (6.16) and (6.18) in Eq. (6.6), we find that

$$\log W_{\mathrm{D}}=\frac{\pi K'}{9K}+\sum_{m=1}^{\infty}\frac{r^{m}-r^{3m}}{m(1+r^{2m})(1+r^{m}+r^{2m})}. \tag{6.19}$$

Equation (6.19) is the essential result of this and the four previous sections. Together with (5.7), (5.8), (5.11), and (6.15) it defines W_{D} as a function of the dimensionless parameter B introduced in (1.5).

7. ELIMINATION OF ELLIPTIC FUNCTIONS

The above working leans heavily on the theory of elliptic functions. It is remarkable that the series in (6.19) can be summed so as to completely eliminate these functions and integrals, giving simply an algebraic expression for W_{D} in terms of the roots a, b, and c of the cubic equation (5.7).

To show this, we generalize (6.19) by multiplying the summand of the series by $\cos(\pi mu/K)$. The summand can then be written as the sum of two terms to give

$$\log W_{\mathrm{D}}=\lim_{u\to 0}[S_1(u)+S_2(u)], \tag{7.1}$$

where

$$S_1(u)=\frac{\pi K'}{3K}+\sum_{m=1}^{\infty}\frac{(1-r^{2m})\cos(\pi mu/K)}{m(1+r^{2m})}, \tag{7.2}$$

$$S_2(u)=-\frac{2\pi K'}{9K}-\sum_{m=1}^{\infty}\frac{(1-r^{2m})\cos(\pi mu/K)}{m(1+r^{m}+r^{2m})}. \tag{7.3}$$

By comparing (6.18) and (7.3), it is apparent that

$$S_2(u)=\tfrac{1}{2}\log g^2(u). \tag{7.4}$$

Also, using the series expansions for $\log \mathrm{sn}^2(u)$ (8.146.20/23 of Ref. 5), one finds that

$$S_1(u)=-6\log k-\tfrac{1}{2}\sum_{\alpha=0}^{3}\sum_{\beta=0}^{2}\log \mathrm{sn}^2[\tfrac{1}{4}u+\tfrac{1}{2}\alpha K+\tfrac{2}{3}\beta iK']. \tag{7.5}$$

By substituting the expressions (7.4) and (7.5) into Eq. (7.1) and using (5.10) and the periodicity and symmetry properties of sn (u), it follows that

$$W_{\mathrm{D}}=\frac{4[ac/(a-c)]^{\frac{1}{2}}}{k^{6}|\eta_1\eta_2\eta_3\eta_4^2|^2}, \tag{7.6}$$

where

$$\eta_1=\mathrm{sn}(i\tau),\quad \eta_2=\mathrm{sn}(K+i\tau),$$
$$\eta_3=\mathrm{sn}(\tfrac{1}{2}K),\quad \eta_4=\mathrm{sn}(\tfrac{1}{2}K+i\tau). \tag{7.7}$$

The numbers $\eta_1,\cdots,\eta_4$ can be evaluated by using (5.11), (6.10), and the addition theorems for elliptic functions, giving

$$\eta_1=i[(a-c)/c]^{\frac{1}{2}},\quad \eta_2=(b/c)^{\frac{1}{2}},$$
$$\eta_3=(1-k')^{\frac{1}{2}}/k,\quad |\eta_4|^2=(\eta_3^2-\eta_1^2)/(1-k^2\eta_1^2\eta_3^2). \tag{7.8}$$

Using (5.11), (6.8), and (6.11), we can simplify these expressions to give

$$W_{\mathrm{D}}=\frac{4abc^{\frac{1}{2}}}{[(a-c)^3a]^{\frac{1}{2}}-[(b-c)^3b]^{\frac{1}{2}}}. \tag{7.9}$$

The formula (1.7) quoted in the introduction can now be established by noting that, if B is parameterized according to (1.6), the roots of Eq. (5.7) are

$$c=(1-9t^2)^{\frac{2}{3}},$$
$$a=\tfrac{1}{4}c(3y+1)^2, \tag{7.10}$$
$$b=\tfrac{1}{4}c(3y-1)^2,$$

where

$$y=\left(\frac{1-t^2}{1-9t^2}\right)^{\frac{1}{2}}. \tag{7.11}$$

Substituting these expressions into (7.9), we obtain the result (1.7).

8. HARD SQUARES MODEL

Having evaluated the partition function for the general 3-coloring problem, let us now specialize to the case when color 1 is regarded as a particle and colors 2, 3 as forming a background. Thus we set

$$z_1 = z, \quad z_2 = z_3 = 1, \tag{8.1}$$

where z is the activity of the particles. As was discussed in the introduction, this system is reminiscent of the hard-square lattice gas.

From (1.4),

$$W = z^{\frac{1}{3}} W_{\mathrm{D}}, \tag{8.2}$$

while from (1.5) B is now given by

$$B = (1 + 2z)/(3z^{\frac{2}{3}}). \tag{8.3}$$

It is easy to see that one of the roots a, b, and c of Eq. (5.7) is $z^{-\frac{2}{3}}$, the others therefore being given by a quadratic equation. Owing to the requirement that a, b, and c be ordered so that $a > b > c$, two cases arise which must be discussed separately.

A. $0 < z \leq 1, 0 < \rho \leq \frac{1}{3}$

When $z \leq 1$, it is found that the roots of (5.7), in their correct ordering, are

$$\begin{aligned} a &= z^{-\frac{2}{3}} \tfrac{1}{4}(\epsilon + 1)^2, \\ b &= z^{-\frac{2}{3}}, \\ c &= z^{-\frac{2}{3}} \tfrac{1}{4}(\epsilon - 1)^2, \end{aligned} \tag{8.4}$$

where

$$\epsilon = (1 + 8z)^{\frac{1}{2}}. \tag{8.5}$$

Using these results in (7.9) and (8.2) gives

$$W = (\epsilon^2 - 1)/\{\epsilon^{\frac{1}{2}}[\epsilon - (3 - \epsilon)X]\}, \tag{8.6}$$

where

$$X = \tfrac{1}{4}[(3 - \epsilon)(1 + \epsilon)/\epsilon]^{\frac{1}{2}}, \tag{8.7}$$

Since $Z \sim W^{N_t}$ is the grand-partition function of this system, $\log W$ is the grand potential $P/k_{\mathrm{B}}T$ (P being the pressure, k_{B} Boltzmann's constant, and T the temperature). Thus the density ρ of particles per lattice site (or face) is

$$\rho = z \frac{d \log W}{dz}. \tag{8.8}$$

Substituting the expression (8.6) into (8.8) and using (8.5) and (8.7), we find that

$$\rho = (\epsilon - 1)/[2\epsilon(1 + 2X)]. \tag{8.9}$$

By inspection of these equations, it can be seen that, as z increases from 0 to 1, ρ varies from 0 to $\frac{1}{3}$, while W goes from 1 to the ice-model value $(\frac{4}{3})^{\frac{3}{2}}$.

It is of some interest to consider W as a function of ρ. Eliminating ϵ and X from Eqs. (8.5)–(8.9), we find that W^2 is the positive root of the equation

$$W^4 - \mu^2(4 + 8\mu - 27\mu^4)W^2 - 16\mu^6 = 0, \tag{8.10}$$

where

$$\mu = 1 - \rho. \tag{8.11}$$

From this result the virial expansion for the grand potential can be obtained, namely

$$\begin{aligned} \log W = \rho + 2\tfrac{1}{2}\rho^2 + 4\tfrac{1}{3}\rho^3 + 1\tfrac{1}{4}\rho^4 \\ - 31\tfrac{4}{5}\rho^5 - 179\tfrac{1}{6}\rho^6 - \cdots. \end{aligned} \tag{8.12}$$

Comparing this with the corresponding expansion for the hard-square lattice gas,[3] we find that the first three coefficients are indeed the same, as expected.

From (8.10) it can be seen that the only singularities in the complex ρ plane of the function W^2 are branch points at the four points

$$\begin{aligned} \rho = \tfrac{1}{3}[(\sqrt{2}) + 1][(\sqrt{2}) \pm i], \\ \tfrac{1}{3}[(\sqrt{2}) - 1][(\sqrt{2}) \pm i], \end{aligned} \tag{8.13}$$

while the only zero is at $\rho = 1$. Thus, $\log W$ is analytic in a circle, with center at the origin, of radius $[(\sqrt{2}) - 1]/\sqrt{3} \simeq 0.239$. This radius is therefore the radius of convergence of the virial expansion (8.12).

B. $z \geq 1, \frac{1}{3} \leq \rho < \frac{1}{2}$

When $z > 1$, the roots of (5.7) are still given by (8.4), except that now b and c must be interchanged to ensure the correct ordering $a > b > c$. Thus, $c = z^{-\frac{2}{3}}$, and from (7.10) the parameter t is the positive root of the equation

$$z = (1 - 9t^2)^{-1}. \tag{8.14}$$

From (1.7) and (8.2) it follows that

$$W^2 = 64/[27(1 + t)^3(1 - 3t)], \tag{8.15}$$

and (8.8) now gives

$$\rho = (1 + 3t)/(3 + 3t). \tag{8.16}$$

Thus, in this case, z and W^2 are rational functions of t and ρ. In particular, eliminating t from (8.15) and (8.16) gives

$$W^2 = 4(1 - \rho)^4/(1 - 2\rho). \tag{8.17}$$

From these results it can be seen that as z increases from 1 to ∞, t goes from 0 to $\frac{1}{3}$, ρ from $\frac{1}{3}$ to its close-packed value $\frac{1}{2}$, and W from $(\frac{4}{3})^{\frac{3}{2}}$ to ∞.

C. Nature of the Transition

In Figs. 3 and 4 the grand potential $\log W$ and its derivative with respect to density are plotted as functions of ρ for the physically permissible range $0 \leq \rho < \frac{1}{2}$. It can be seen that $\log W$ is a continuous function of ρ, monotonic increasing except at the point $\rho = \frac{1}{3}$, where the algebraic form of W changes from (8.10) to (8.17). Expanding both forms in Taylor series about $\rho = \frac{1}{3}$, we find that

$$\log W = \tfrac{3}{2} \log \tfrac{4}{3} - \tfrac{3}{4}\theta(3\rho - 1)^2 + O\{(3\rho - 1)^3\}, \tag{8.18}$$

where

$$\theta = 1 \quad \text{if} \quad \rho < \tfrac{1}{3},$$
$$\theta = -1 \quad \text{if} \quad \rho > \tfrac{1}{3}. \tag{8.19}$$

Thus the derivative of $\log W$ vanishes at the critical point $\rho = \frac{1}{3}$, and the isothermal compressibility, given by (1.11), becomes infinite, being given to first order by (1.12).

Also shown in Fig. 3 are the numerical estimates of $\log W$ for the hard-square lattice gas.[3] It can be seen that the two models agree at low densities as expected,

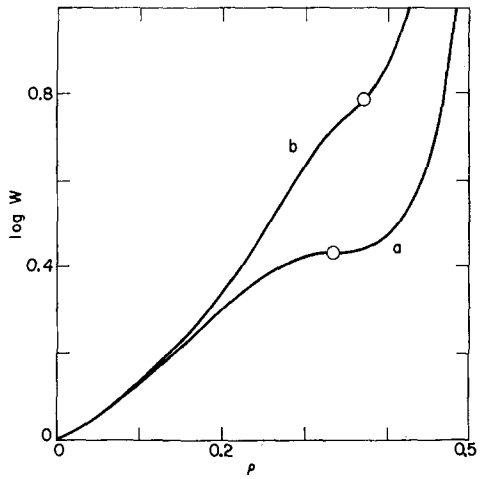

FIG. 3. Equation of state of (a) the hard square model of this paper and (b) the true hard-square lattice gas.[3] The circles indicate critical points.

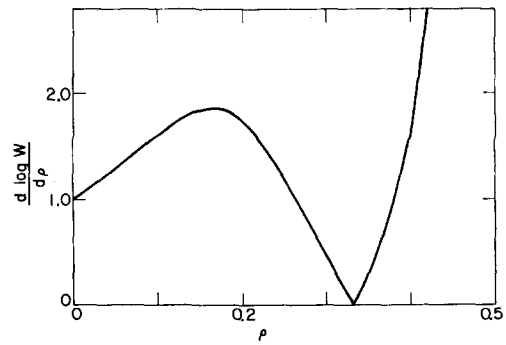

FIG. 4. Plot of $d(\log W)/d\rho$, which is proportional to the inverse compressibility, for the hard square model of this paper.

but that there are considerable differences in the transition region and above.

It is interesting to note that both (8.10) and (8.17) give functions W of density that are analytic throughout the entire permissible range $0 \leq \rho < \frac{1}{2}$. However, each applies only in its appropriate interval, $[0, \frac{1}{3}]$ and $[\frac{1}{3}, \frac{1}{2}]$, respectively, and one is not the analytic continuation of the other.

ACKNOWLEDGMENTS

The author thanks Professor E. H. Lieb for his interest and encouragement in this work and Dr. D. S. Gaunt and Professor M. E. Fisher for making available graphs of their results for the hard-square lattice gas.

[1] E. H. Lieb, Phys. Rev. Letters **18**, 692 (1967).
[2] E. H. Lieb, Phys. Rev. **162**, 162 (1967).
[3] D. S. Gaunt and M. E. Fisher, J. Chem. Phys. **43**, 2840 (1965); L. K. Runnels and L. L. Combs, *ibid.* **45**, 2482 (1966); M. E. Fisher, J. Math. Phys. **4**, 278 (1963).
[4] C. N. Yang and C. P. Yang, Phys. Rev. **150**, 321 (1966); **150**, 327 (1966); E. H. Lieb, Phys. Rev. Letters **18**, 1046 (1967); **19**, 108 (1967); R. J. Baxter, J. Math. Phys. **10**, 1211 (1969); **11**, 784 (1970).
[5] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965).
[6] This method fails for $m = 0$, but in this case we can use Eq. (5.12) and the relation $p(u + 2K) - p(u) = \frac{2}{3}\pi$.
[7] Again, the method fails for $m = 0$, but, once the other coefficients are found, it can be established that $g(u)g(u + i\tau)g(u - i\tau) =$ const. Taking the limit $u \to 0$, we find that this constant is unity, from which the $m = 0$ coefficient can be obtained.