JOURNAL OF MATHEMATICAL PHYSICS

VOLUME 13, NUMBER 12 DECEMBER 1972

Coordinate systems associated with asymptotically shear-free null congruences*

B. Aronson
Geauga Regional Campus, Kent State University, Chardon, Ohio

E.T.Newman
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 23 June 1972)

We present here the asymptotic coordinate transformation between a coordinate system associated with null hypersurfaces and one associated with an asymptotically shear-free (but twisting) null congruence. The general asymptotically flat metric is expressed in this new coordinate system. Special cases of this are the algebraically special metrics in their "natural" coordinate system.

1. INTRODUCTION

In the study of asymptotically flat solutions of the Einstein (or Einstein–Maxwell) equations two types of "null coordinate" systems have been commonly employed. The first of these (referred to as Type I) and by far the most frequently used is based on null hypersurfaces at infinity.[1–5] By now a reasonable understanding of the physical meaning (energy–momentum, angular momentum, multipole moments, etc.) of many of the geometric quantities involved, has been acquired. The second of these coordinate systems[6–10] (referred to as Type II) is based on twisting, asymptotically shear-free null geodesics. It has been most commonly associated with the class of solutions of the Einstein equations known as the algebraically special, twisting metrics. The physical meaning of the geometric quantities involved in these metrics has been obscure. It is one purpose of this paper to express asymptotically these solutions in the first coordinate system (and thereby clarify their meaning). This is accomplished by solving the broader problem of determining the general coordinate (and associated tetrad) transformation between the Type I and II systems.

Asymptotically flat spaces have been investigated from several points of view. Most earlier investigations were based on reasonable guesses for the behavior of the metric tensor at spatial infinity. Major work by Bondi[1] and Sachs[2] improved the situation greatly by utilizing characteristic surfaces and deriving from very simple assumptions the asymptotic behavior of the metric tensor and the Riemann tensor. In the spin coefficient formalism developed and applied to asymptotically flat space in NP[4] and NU[5] the emphasis is shifted from the metric tensor to the empty space Riemann (Weyl) tensor. This point of view is adopted in this work. It is assumed that the reader is familiar with this formalism.

2. PRELIMINARIES

The coordinatization of asymptotically flat empty spaces is most easily approached by considering future null infinity $\mathscr{J}^+$ of Penrose.[11] (Past null infinite $\mathscr{J}^-$ could just as easily have been considered.) If for descriptive purpose we consider only the conformal structure of space-time, then $\mathscr{J}^+$ can be treated as an ordinary three-dimensional boundary to a four-dimensional region.

We coordinatize $\mathscr{J}^+$, which is a null surface, by first introducing arbitrary, nonintersecting, spacelike cuts which can be labeled by $x^0 = u =$ const. Since $\mathscr{J}^+$ is $S^2 \times R$, its generators can be labeled by the complex stereographic coordinates of a sphere, ζ and $\bar{\zeta}$.

There are several ways to coordinatize an interior neighborhood of $\mathscr{J}^+$ once we have chosen a particular coordinate system on $\mathscr{J}^+$. Two types of coordinate systems are of particular interest to us. To develop a Type I coordinate system, we choose null geodesics, with tangent vector l^μ, from the interior that are *orthogonal* to the $u =$ const cuts of $\mathscr{J}^+$ at every point on the cut. Each of these geodesics is identified by the u and the ζ and $\bar{\zeta}$ of its intersection with $\mathscr{J}^+$. The affine parameter along each geodesic serves as the radial coordinate $x^1 = r$. Because of the hypersurface orthogonal character of the geodesics we can choose $l_\mu = u_{,\mu} = \delta^0_\mu$ and $l^\mu = \delta^\mu_1$. The freedom in the choice of the affine parameter permits us to choose the expansion ($\rho = -\frac{1}{2} l^\mu_{;\mu}$) of the congruence to be of the form

$$\rho = -r^{-1} + O(r^{-3}). \tag{2.1}$$

The complex shear σ of the congruence is found to be of the form:

$$\sigma = \sigma^0 r^{-2} + O(r^{-4}). \quad (2.2)$$

A tetrad system of the type used in NU can be adapted to this coordinate system. The proof of these assertions can be found in NU.

To develop a Type II coordinate system, we choose null geodesics, with tangent vector $\tilde{l}^\mu$, from the interior that are not orthogonal to the u = const cuts. As will be shown later, we can always choose from such geodesics, three parameter families which are *asymptotically shear free*, though in general twisting. Though these families do not generate null surfaces they still induce a coordinate system in the neighborhood of J^+ in a manner similar to the way Type I was introduced. Each geodesic is labeled by the u, ζ, and $\bar{\zeta}$ of its intersection with J^+, with the affine length $\tilde{r}$ serving as the radial coordinate, so that $\tilde{l}^\mu = \delta^\mu_1$. The freedom in the choice of the origin of $\tilde{r}$ will enable us to choose the expansion, $\tilde{\rho}$ to be of the form $\tilde{\rho} = -(\tilde{r} + i\Sigma)^{-1} + O(\tilde{r}^{-3})$ or

$$\tilde{\rho} = -\tilde{r}^{-1} + i\Sigma\tilde{r}^{-2} + O(\tilde{r}^{-3}), \quad (2.3)$$

where Σ, the asymptotic twist, is real. The shear vanishes asymptotically so that

$$\tilde{\sigma} = O(r^{-3}). \quad (2.4)$$

A tetrad system of the type used by Talbot[9] can be adapted to this coordinate system.

In each of the above cases (i.e., Types I and II) the tangent vector l^μ to the geodesics is chosen as the first tetrad vector. The remaining three tetrad vectors are restricted by the condition that they be parallel propagated along each of the geodesics; i.e., the spin coefficients κ, ϵ, and π must vanish. The tetrad vectors are expressed in the form

$$l^\mu = \delta^\mu_1, \quad n^\mu = U\delta^\mu_1 + X^a\delta^\mu_a, \quad m^\mu = \omega\delta^\mu_1 + \xi^a\delta^\mu_a, \quad (2.5)$$

where the index assumes the values 0, 2, 3. In a Type I system $X^0 = 1$ and $\xi^0 = 0$.

In order to find the coordinate transformation from a Type I to II system, it is most convenient to first find the associated tetrad transformation where all the vectors are described in the I coordinate system.

The tetrad transformations used are the three two-parameter tetrad transformations[12] which are equivalent to the six-parameter restricted Lorentz group. Though the notation does not show it, these transformations are to be performed consecutively. We first consider the null rotation about n^μ, with complex parameter b, given by

$$\tilde{l}^\mu = l^\mu + b\bar{m}^\mu + \bar{b}m^\mu + b\bar{b}n^\mu, \quad \tilde{n}^\mu = n^\mu,$$
$$\tilde{m}^\mu = m^\mu + bn^\mu. \quad (2.6)$$

This will be followed by the null rotation about l^μ with complex parameter a, given by

$$\tilde{l}^\mu = l^\mu, \quad \tilde{n}^\mu = n^\mu + a\bar{m}^\mu + \bar{a}m^\mu + a\bar{a}l^\mu,$$
$$\tilde{m}^\mu = m^\mu + al^\mu. \quad (2.7)$$

Finally the Lorentz transformation in the l^μ, n^μ plane coupled with the spatial rotation in the $m, \bar{m}^\mu$ plane with real parameters G and H is given by

$$\tilde{l}^\mu = Gl^\mu, \quad \tilde{n}^\mu = G^{-1}n^\mu, \quad \tilde{m}^\mu = e^{iH}m^\mu. \quad (2.8)$$

3. THE TRANSFORMATIONS

In NU the NP equations were solved asymptotically in a Type I coordinate system with its associated tetrad, under the assumption that $\psi_0 = \psi_0^0 r^{-5} + O(r^{-6})$. (This is probably the most general asymptotically flat solution.) The entire solution in the form used here is presented elsewhere.[13]

Starting with this solution (in Type I coordinates) we will utilize asymptotically vanishing tetrad transformations to obtain a tetrad associated with a Type II system; i.e., one with $\tilde{\kappa} = \tilde{\pi} = \tilde{\epsilon} = 0$ and *asymptotically vanishing shear*. We first use Eqs. (2.6), the null rotation about n^μ, to introduce an $\tilde{l}^\mu$ with twist, i.e., cause the new $\tilde{\rho} \neq \bar{\tilde{\rho}}$. The complex parameter b of the transformation is assumed to be of the form

$$b = -L(u,\zeta,\bar{\zeta})r^{-1} + M(u,\zeta,\bar{\zeta})r^{-2} + O(r^{-3}). \quad (3.1)$$

The new shear $\tilde{\sigma}$ then takes the form (Appendix):

$$\tilde{\sigma} = (\sigma^0 - L\dot{L} - \eth L)r^{-2} + (\eth M + \dot{L}M + L\dot{M} - L^3\eth\dot{P}/P - 2L^2\eth\bar{\eth}\ln P + L\bar{\eth}\sigma^0 + \sigma^0\bar{\eth}L - LM\dot{P}/P)r^{-3} + O(r^{-4}), \quad (3.2)$$

where the dot denotes $\partial/\partial u$. For a Type II system the leading term of $\tilde{\sigma}$ must vanish by definition; therefore, L is given by a solution of the differential equation

$$\sigma^0 = \eth L + L\dot{L}. \quad (3.3)$$

The arbitrariness in the solution corresponds to the freedom in the choice of the Type II coordinates.

The spin coefficient $\tilde{\kappa}$ is given by

$$\tilde{\kappa} = (M - \bar{L}\sigma^0 + \bar{L}\eth L + L\bar{\eth}L + L\dot{L}\bar{L})r^{-3} + O(r^{-4}). \quad (3.4)$$

With Eq. (3.3) we see that the leading term of $\tilde{\kappa}$ can be made to vanish by choosing

$$M = -L\bar{\eth}L. \quad (3.5)$$

(This choice of M causes in addition the coefficient of r^{-3} in $\tilde{\sigma}$ to vanish.) The vanishing of $\tilde{\kappa}$ is even stronger than we have indicated; we can make $\tilde{\kappa}$ vanish to any order of r^{-1} by specifying b to further orders. This "vanishing" of $\tilde{\kappa}$ is not changed by the remaining tetrad transformations (see Appendix), and so we can consider $\tilde{\kappa}$ to actually vanish and $\tilde{l}^\mu$ to actually be tangent to a geodesic congruence, also a necessary condition for a Type II system.

We now have a tetrad system for asymptotically flat space whose l^μ field is tangent to a geodesic and which is asymptotically shear free but twisting. To introduce a Type II coordinate system, we wish to have further a tetrad for which $\tilde{\pi}$ and $\tilde{\epsilon}$ vanish as well. We saw that for $\tilde{\kappa}$ to vanish we had to, in principle, specify that the higher order terms of b cancel the higher order terms of $\tilde{\kappa}$. In order to have $\tilde{\pi}$ and $\tilde{\epsilon}$ vanish we will have to make similar conditions on the other transformation parameters a, G, and H of (2.7) and (2.8). After a rather tedious calculation one finds

$$a = \dot{L} + (L\bar{\eth}\dot{L} - \dot{L}^2\bar{L} + L\eth\bar{\eth}\ln P)r^{-1} + O(r^{-2}), \quad (3.6)$$

$$G = 1 - L\bar{L}(\dot{P}/P)r^{-1} + O(r^{-2}), \quad (3.7a)$$

$$H = i(\bar{L}\eth \ln P - L\bar{\eth} \ln P)r^{-1} + O(r^{-2}). \quad (3.7b)$$

The transformations (2.6), (2.7), and (2.8), with (3.1), (3.3), (3.5), (3.6), and (3.7), applied consecutively to a Type I tetrad yield the Type II tetrad.

In order to obtain the Type I to II coordinate transformation, we assume that the coordinates of J^+ are the same in the two systems, i.e., the transformation should be asymptotically the identity transformation and have the form

$$u' = u + a_1 r^{-1} + a_2 r^{-2} + O(r^{-3}), \quad (3.8a)$$

$$r' = r - b_0 + b_1 r^{-1} + O(r^{-2}), \quad (3.8b)$$

$$\zeta' = \zeta + c_1 r^{-1} + c_2 r^{-2} + O(r^{-3}). \quad (3.8c)$$

Since $\tilde{l}^\mu$ (the tangent vector to the Type II geodesics) is known (from the tetrad transformation) but with the components expressed in a I coordinate system, the condition (2.5) (namely that $\tilde{l}^\mu = \delta^\mu_1$, in a II coordinate system) severely restricts the transformation (3.8). In fact it determines all the parameters except b_0, yielding

$$a_1 = L\bar{L}, \quad (3.9a)$$

$$a_2 = \tfrac{1}{2}L\bar{L}(L\dot{\bar{L}} + \bar{L}\dot{L} - L\bar{L}\dot{P}/P + 2\eth\bar{L} + 2\bar{\eth}L) + \tfrac{1}{2}\bar{L}^2\eth L + \tfrac{1}{2}L^2\bar{\eth}L, \quad (3.9b)$$

$$b_1 = \dot{b}_0 L\bar{L} + L\bar{\eth}b_0 + \bar{L}\eth b_0, \quad (3.9c)$$

$$c_1 = 2\bar{L}P, \quad (3.9d)$$

$$c_2 = 2P\bar{L}\eth\bar{L} + P\bar{L}^2\eth \ln P. \quad (3.9e)$$

The b_0 is determined by (2.3), i.e., by the condition that the r'^{-2} term in $\tilde{\rho}$ be imaginary. One thus obtains

$$b_0 = -L\bar{L}\dot{P}/P + \tfrac{1}{2}(\bar{\eth}L + \eth\bar{L} + L\dot{\bar{L}} + \bar{L}\dot{L}) \quad (3.10)$$

and

$$2i\Sigma = \eth\bar{L} - \bar{\eth}L + L\dot{\bar{L}} - \bar{L}\dot{L}. \quad (3.11)$$

4. ASYMPTOTICALLY FLAT SPACE IN A TYPE II COORDINATE SYSTEM

In this section we present a summary of the results of the combined coordinate and tetrad transformation on the NU solutions. All quantities are expressed in a Type II coordinate and tetrad system. These results represent a solution to the general spin coefficient (NP) equations in a Type II coordinate system in asymptotically flat space. As such they are presented independently of the coordinate transformation, without any indication of the transformation such as primes and tildes.

I. The tetrad vectors:

$$l^\mu = \delta^\alpha_1, \quad (4.1a)$$

$$n^\alpha = U\delta^\alpha_1 + X^a\delta^\alpha_a, \quad (4.1b)$$

$$m^\alpha = \omega\delta^\alpha_1 + \xi^a\delta^\alpha_a. \quad (4.1c)$$

II. The metric variables:

$$U = (\dot{P}/P)r + U^0 + O(r^{-1}), \quad (4.2a)$$

$$X^0 = 1 + O(r^{-2}), \quad (4.2b)$$

$$X^2 = X^3 = O(r^{-3}), \quad (4.2c)$$

$$\omega = \dot{L} - L\dot{P}/P + \omega^0 r^{-1} + O(r^{-2}), \quad (4.2d)$$

$$\xi^0 = -Lr^{-1} - iL\Sigma r^{-2} + O(r^{-3}), \quad (4.2e)$$

$$\xi^2 = Pr^{-1} + i\Sigma P r^{-2} + O(r^{-3}), \quad (4.2f)$$

$$\xi^3 = iPr^{-1} - \Sigma P r^{-2} + O(r^{-3}), \quad (4.2g)$$

with

$$U^0 = -\eth\bar{\eth} \ln P - \tfrac{1}{2}[L\eth\dot{P}/P + \bar{L}\bar{\eth}\dot{P}/P + L\ddot{\bar{L}} + \bar{L}\ddot{L} - (\dot{P}/P)(L\dot{\bar{L}} + \bar{L}\dot{L}) + \eth\dot{\bar{L}} + \bar{\eth}\dot{L}], \quad (4.2h)$$

$$\omega^0 = 2i\Sigma(\dot{L} - L\dot{P}/P) + i(L\dot{\Sigma} + \eth\Sigma), \quad (4.2i)$$

$$2i\Sigma = \eth\bar{L} - \bar{\eth}L + L\dot{\bar{L}} - \bar{L}\dot{L}. \quad (4.2j)$$

III. The spin coefficients:

$$\kappa = \epsilon = \pi = 0, \quad (4.3a)$$

$$\sigma = O(r^{-4}), \quad (4.3b)$$

$$\tau = \lambda = O(r^{-3}), \quad (4.3c)$$

$$\rho = -r^{-1} + i\Sigma r^{-2} + O(r^{-3}), \quad (4.3d)$$

$$\mu = -(L\bar{\eth}\dot{P}/P + \eth\bar{\eth} \ln P - L\dot{\bar{L}}\dot{P}/P + L\ddot{\bar{L}} + \eth\dot{\bar{L}})r^{-1} + O(r^{-2}), \quad (4.3e)$$

$$\nu = \bar{\eth}\dot{P}/P - \dot{\bar{L}}\dot{P}/P + \ddot{\bar{L}} + O(r^{-1}), \quad (4.3f)$$

$$\alpha = \tfrac{1}{2}(\bar{L}\dot{P}/P - 2\dot{\bar{L}} - \bar{\eth} \ln P)r^{-1} - \tfrac{1}{2}i\Sigma(\bar{L}\dot{P}/P - 2\dot{\bar{L}} - \bar{\eth} \ln P)r^{-2} + O(r^{-3}), \quad (4.3g)$$

$$\beta = \tfrac{1}{2}(L\dot{P}/P + \eth \ln P)r^{-1} + O(r^{-2}), \quad (4.3h)$$

$$\gamma = -\tfrac{1}{2}\dot{P}/P + O(r^{-2}). \quad (4.3i)$$

IV. The tetrad components of the Weyl tensor:

$$\psi_0 = \psi_0^0 r^{-5} + O(r^{-6}), \quad (4.4a)$$

$$\psi_1 = \psi_1^0 r^{-4} + O(r^{-5}), \quad (4.4b)$$

$$\psi_2 = \psi_2^0 r^{-3} + O(r^{-4}), \quad (4.4c)$$

$$\psi_3 = \psi_3^0 r^{-2} + O(r^{-3}), \quad (4.4d)$$

$$\psi_4 = \psi_4^0 r^{-1} + O(r^{-2}). \quad (4.4e)$$

V. With

$$\psi_2^0 - \bar{\psi}_2^0 = -2i\,\mathrm{Re}(\eth W + L\dot{W} + \dot{L}W - 2LW\dot{P}/P) + 4i\Sigma U^0, \quad (4.5a)$$

$$\psi_3^0 = -\eth R - L\dot{R} + 2LR\dot{P}/P, \quad (4.5b)$$

$$\psi_4^0 = -\dot{R} + 2R\dot{P}/P, \quad (4.5c)$$

with

$$W = \bar{\eth}\Sigma + P(\bar{L}\Sigma P^{-1})^{\cdot}, \quad (4.5d)$$

$$R = \bar{\eth}N + \bar{L}\dot{N} - \bar{L}N\dot{P}/P + N^2 - 2N\bar{\eth} \ln P, \quad (4.5e)$$

$$N = \dot{\bar{L}} - \bar{\eth}\ \ln P. \tag{4.5f}$$

VI. The differential equations relating the ψ_A^0 (these equations are found directly from the Bianchi identities, Eqs (4.5) of NP]:

$$\eth\psi_1^0 = 3(\dot{P}/P)\psi_0^0 - \dot{\psi}_0^0 + L\dot{\psi}_1^0 - 3L(\dot{P}/P)\psi_1^0 + 4\dot{L}\psi_1^0, \tag{4.6a}$$

$$\eth\psi_2^0 = 3(\dot{P}/P)\psi_1^0 - \dot{\psi}_1^0 + L\dot{\psi}_2^0 - 3L(\dot{P}/P)\psi_2^0 + 3\dot{L}\psi_2^0, \tag{4.6b}$$

$$\eth\psi_3^0 = 3(\dot{P}/P)\psi_2^0 - \dot{\psi}_2^0 + L\dot{\psi}_3^0 - 3L(\dot{P}/P)\psi_3^0 + 2\dot{L}\psi_3^0, \tag{4.6c}$$

$$\eth\psi_4^0 = 3(\dot{P}/P)\psi_3^0 - \dot{\psi}_3^0 + L\dot{\psi}_4^0 - 3L(\dot{P}/P)\psi_4^0 + \dot{L}\psi_4^0. \tag{4.6d}$$

[Equation (4.6d) is identically satisfied with the use of Eqs. (4.5b) and (4.5c).]

The variables $\psi_0^0, \psi_1^0, \psi_2^0, \psi_3^0$, and ψ_4^0 are easily related to their Type I counterparts (symbolized by I) by

$$\psi_0^0 = \overset{\mathrm{I}}{\psi}{}_0^0 - 4L\overset{\mathrm{I}}{\psi}{}_1^0 + 6L^2\overset{\mathrm{I}}{\psi}{}_2^0 - 4L^3\overset{\mathrm{I}}{\psi}{}_3^0 + L^4\overset{\mathrm{I}}{\psi}{}_4^0, \tag{4.7a}$$

$$\psi_1^0 = \overset{\mathrm{I}}{\psi}{}_1^0 - 3L\overset{\mathrm{I}}{\psi}{}_2^0 + 3L^2\overset{\mathrm{I}}{\psi}{}_3^0 - L^3\overset{\mathrm{I}}{\psi}{}_4^0, \tag{4.7b}$$

$$\psi_2^0 = \overset{\mathrm{I}}{\psi}{}_2^0 - 2L\overset{\mathrm{I}}{\psi}{}_3^0 + L^2\overset{\mathrm{I}}{\psi}{}_4^0, \tag{4.7c}$$

$$\psi_3^0 = \overset{\mathrm{I}}{\psi}{}_3^0 - L\overset{\mathrm{I}}{\psi}{}_4^0, \tag{4.7d}$$

$$\psi_4^0 = \overset{\mathrm{I}}{\psi}{}_4^0. \tag{4.7e}$$

5. DISCUSSION

A result of this study of Type I and Type II coordinate systems in asymptotically flat space is a solution to the general spin coefficient equations (which are equivalent to the Einstein field equations) based on an asymptotically shear free but twisting congruence of null geodesics. This solution is obviously equivalent to the solution one would obtain by actually integrating the spin coefficient equations in a Type II system in asymptotically flat space. The condition for asymptotic flatness used by NU for a Type I system also applies for the Type II system; i.e.,

$$\psi_0 = \psi_0^0 r^{-5} + O(r^{-6}), \tag{5.1}$$

where the order symbols do not change when differentiated with respect to the nonradial coordinates.[4] With this point of view the variable L is interpreted as the variable of integration associated with the r^{-1} part of the metric variable ξ^0 [Eq. (4.2d)]. The solution is given in terms of the basic variables L and P and the variables ψ_0^0, ψ_1^0, and $\psi_2^0 + \bar{\psi}_2^0$ which satisfy differential equations (4.6). Comparison with NU shows that L takes the place of σ^0 as a basic variable. We interpret this to mean that the information or news carried by σ^0 in the NU solution is carried by L for the same solution in a Type II coordinate system.

We can define a subset of the asymptotically flat spaces as being asymptotically algebraically special if there exists a Type II system such that

$$\psi_0^0 = \psi_1^0 = 0. \tag{5.2}$$

[Note that this is more than the condition that $\psi_0^0 = 0$ in (4.7a) have a repeated root for L. L must also satisfy (3.3).] Algebraically special metrics when looked at asymptotically obviously have this form (though having this form is no guarantee of algebraic specialness) and thus satisfy (4.6) with (5.2). More important, we now have the tool for taking the geometric quantities associated with the algebraically special metrics and interpreting them physically by studying them in a Type I system. This work is now in progress.

APPENDIX

We can find the transformation law of the spin coefficients by substituting the transformed tetrad, in terms of the original, into the definition of the spin coefficient; e.g., the transformation of σ under (2.6) is given by

$$\begin{aligned}\tilde{\sigma} &= \tilde{l}_{\mu;\nu}\tilde{m}^\mu \tilde{m}^\nu \\ &= (l_\mu + b\bar{m}_\mu + \bar{b}m_\mu + b\bar{b}n_\mu)_{;\nu}(m^\mu + bn^\mu)(m^\nu + bn^\nu).\end{aligned}$$

Simplifying and using the definitions of the original spin coefficients, we obtain

$$\sigma = \sigma + b(\tau + 2\beta) + b^2(\mu + 2\gamma) + b^3\nu - \delta b - b\Delta b.$$

where

$$D = l^\mu \frac{\partial}{\partial x^\mu}, \quad \Delta = n^\mu \frac{\partial}{\partial x^\mu}, \quad \delta = m^\mu \frac{\partial}{\partial x^\mu}. \tag{A1}$$

The behavior of the spin coefficients and the ψ_A under (2.6) (the null rotation about n^μ) is given by

$$\tilde{\rho} = \rho + 2b\alpha + \bar{b}\tau + 2b\bar{b}\gamma + b^2\lambda + b^2\bar{b}\nu - \bar{\delta}b - \bar{b}\Delta b, \tag{A2a}$$

$$\tilde{\sigma} = \sigma + b(\tau + 2\beta) + b^2(\mu + 2\gamma) + b^3\nu - \delta b - b\Delta b, \tag{A2b}$$

$$\begin{aligned}\tilde{\kappa} = \kappa &+ b(\rho + 2\epsilon) + \bar{b}\sigma + b\bar{b}(\tau + 2\beta) + b^2(\pi + 2\alpha) \\ &+ b^2\bar{b}(\mu + 2\gamma) + b^3\lambda + b^3\bar{b}\nu - Db - \bar{b}\delta b \\ &- b\bar{\delta}b - b\bar{b}\Delta b,\end{aligned} \tag{A2c}$$

$$\tilde{\tau} = \tau + 2b\gamma + b^2\nu - \Delta b, \tag{A2d}$$

$$\tilde{\mu} = \mu + b\nu, \tag{A2e}$$

$$\tilde{\lambda} = \lambda + \bar{b}\nu, \tag{A2f}$$

$$\tilde{\nu} = \nu, \tag{A2g}$$

$$\tilde{\pi} = \pi + \bar{b}\mu + b\lambda + b\bar{b}\nu, \tag{A2h}$$

$$\tilde{\alpha} = \alpha + b\lambda + \bar{b}\gamma + b\bar{b}\nu, \tag{A2i}$$

$$\tilde{\beta} = \beta + b(\mu + \gamma) + b^2\nu, \tag{A2j}$$

$$\tilde{\gamma} = \gamma + b\nu, \tag{A2k}$$

$$\tilde{\epsilon} = \epsilon + b(\alpha + \pi) + \bar{b}\beta + b\bar{b}(\mu + \gamma) + b^2\lambda + b^2\bar{b}\nu \tag{A2l}$$

and

$$\tilde{\psi}_0 = \psi_0 + 4b\psi_1 + 6b^2\psi_2 + 4b^3\psi_3 + b^4\psi_4, \tag{A3a}$$

$$\tilde{\psi}_1 = \psi_1 + 3b\psi_2 + 3b^2\psi_3 + b^3\psi_4, \tag{A3b}$$

$$\tilde{\psi}_2 = \psi_2 + 2b\psi_3 + b^2\psi_4, \tag{A3c}$$

$$\tilde{\psi}_3 = \psi_3 + b\psi_4, \tag{A3d}$$

$$\tilde{\psi}_4 = \psi_4. \tag{A3e}$$

The behavior of the spin coefficients and the ψ_A under (2.7) (the null rotation about l^μ) is given by:

$$\tilde{\rho} = \rho + \bar{a}\kappa, \tag{A4a}$$

$$\tilde{\sigma} = \sigma + a\kappa, \tag{A4b}$$

$$\tilde{\kappa} = \kappa, \tag{A4c}$$

$$\tilde{\tau} = \tau + \bar{a}\sigma + a\rho + a\bar{a}\kappa, \tag{A4d}$$

$$\tilde{\mu} = \mu + a\pi + 2\bar{a}\beta + 2a\bar{a}\epsilon + \bar{a}^2\sigma + \bar{a}^2 a\kappa + \delta\bar{a} + aD\bar{a} \tag{A4e}$$

$$\tilde{\lambda} = \lambda + \bar{a}(\pi + 2\alpha) + \bar{a}^2(\rho + 2\epsilon) + \bar{a}^3\kappa + \bar{\delta}\bar{a} + \bar{a}D\bar{a} \tag{A4f}$$

$$\begin{aligned}\tilde{\nu} = \nu &+ a\lambda + \bar{a}(\mu + 2\gamma) + a\bar{a}(\pi + 2\alpha) \\ &+ \bar{a}^2(\tau + 2\beta) + a\bar{a}^2(\rho + 2\epsilon) \\ &+ \bar{a}^3\sigma + a\bar{a}^3\kappa + \Delta\bar{a} + \bar{a}\delta\bar{a} + a\bar{\delta}\bar{a} + a\bar{a}D\bar{a},\end{aligned} \tag{A4g}$$

$$\tilde{\pi} = \pi + 2\bar{a}\epsilon + \bar{a}^2\kappa + D\bar{a}, \tag{A4h}$$

$$\tilde{\alpha} = \alpha + \bar{a}(\rho + \epsilon) + \bar{a}^2\kappa, \tag{A4i}$$

$$\tilde{\beta} = \beta + a\epsilon + \bar{a}\sigma + a\bar{a}\kappa, \tag{A4j}$$

$$\tilde{\gamma} = \gamma + a\alpha + \bar{a}(\beta + \tau) + a\bar{a}(\rho + \epsilon) + \bar{a}^2\sigma + a\bar{a}^2\kappa, \tag{A4k}$$

$$\tilde{\epsilon} = \epsilon + \bar{a}\kappa \tag{A4l}$$

and

$$\tilde{\psi}_0 = \psi_0, \tag{A5a}$$

$$\tilde{\psi}_1 = \psi_1 + \bar{a}\psi_0, \tag{A5b}$$

$$\tilde{\psi}_2 = \psi_2 + 2\bar{a}\psi_1 + \bar{a}^2\psi_0, \tag{A5c}$$

$$\tilde{\psi}_3 = \psi_3 + 3\bar{a}\psi_2 + 3\bar{a}^2\psi_1 + \bar{a}^3\psi, \tag{A5d}$$

$$\tilde{\psi}_4 = \psi_4 + 4\bar{a}\psi_3 + 6\bar{a}^2\psi_2 + 4\bar{a}^3\psi_1 + \bar{a}^4\psi_4. \tag{A5e}$$

The behavior of the spin coefficients and the ψ_A under (2.8) (the combined Lorentz transformation in the l^μ, n^μ plane and spatial rotation in the $m^\mu, \bar{m}^\mu$ plane) is given by

$$\tilde{\rho} = G\rho, \tag{A6a}$$

$$\tilde{\sigma} = Ge^{2iH}\sigma, \tag{A6b}$$

$$\tilde{\kappa} = G^2e^{iH}\kappa, \tag{A6c}$$

$$\tilde{\tau} = e^{iH}\tau, \tag{A6d}$$

$$\tilde{\mu} = G^{-1}\mu, \tag{A6e}$$

$$\tilde{\lambda} = G^{-1}e^{-2iH}\lambda, \tag{A6f}$$

$$\tilde{\nu} = G^{-2}e^{-iH}\nu, \tag{A6g}$$

$$\tilde{\pi} = e^{-iH}, \tag{A6h}$$

$$\tilde{\alpha} = e^{-iH}\alpha + \tfrac{1}{2}e^{-iH}(G^{-1}\bar{\delta}G + i\bar{\delta}H), \tag{A6i}$$

$$\tilde{\beta} = e^{iH}\beta + \tfrac{1}{2}e^{iH}(G^{-1}\delta G + i\delta H), \tag{A6j}$$

$$\tilde{\gamma} = G^{-1}\gamma + \tfrac{1}{2}G^{-1}(G^{-1}\Delta G + i\Delta H), \tag{A6k}$$

$$\tilde{\epsilon} = G\epsilon + \tfrac{1}{2}G(G^{-1}DG + iDH) \tag{A6l}$$

and

$$\tilde{\psi}_0 = G^2e^{2iH}\psi_0, \tag{A7a}$$

$$\tilde{\psi}_1 = Ge^{iH}\psi_1, \tag{A7b}$$

$$\tilde{\psi}_2 = \psi_2, \tag{A7c}$$

$$\tilde{\psi}_3 = G^{-1}e^{-H}\psi_4 \tag{A7d}$$

$$\tilde{\psi}_4 = G^{-2}e^{-2iH}\psi_4. \tag{A7e}$$

*This research was supported in part by the NSF under grant GP22789.
[1] H. Bondi *et al.*, Proc. R. Soc. A **269**, 21 (1962).
[2] R. K. Sachs, Proc. R. Soc. A **270**, 103 (1962).
[3] I. Robinson and A. Trautman, Proc. R. Soc. A **265**, 463 (1962).
[4] E. T. Newman and R. Penrose, J. Math. Phys. (N.Y.) **3**, 566 (1962).
[5] E. T. Newman and T. Unti, J. Math. Phys. (N.Y.) **3**, 891 (1962).
[6] E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. (N.Y.) **4**, 915 (1963).
[7] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[8] G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. (N.Y.) **10**, 1842 (1969).
[9] C. J. Talbot, Commun. Math. Phys. **13**, 45 (1969).
[10] R. W. Lind, Ph.D. thesis (University of Pittsburgh, 1970).
[11] R. Penrose, in *Relativity, Groups and Topology*, edited by C. Dewitt and B. Dewitt (Gordon and Breach, New York, 1964).
[12] A. Janis and E. T. Newman, J. Math. Phys. (N.Y.) **6**, 915 (1965).
[13] B. Aronson, Ph.D. thesis (University of Pittsburgh, 1971). Also, R. Lind, J. Messmer, and E. T. Newman, J. Math. Phys. (N.Y.) **13**, 1884 (1972).

Relativistic dynamical symmetries and dilatations

P. Roman, J. J. Aghassi, and P. L. Huddleston
Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 8 May 1972)

Characterizing the state of a relativistic particle by a pair (x_μ, ξ_μ) of 4-vectors, we are led, in a natural way, to a group $\mathcal{K}_5$ of canonical transformations which includes the Poincaré group and dilatations. The structure of the group and its induced irreducible unitary representations are explored. It is shown that $\mathcal{K}_5$ has a semisimple noncompact subgroup which permits a systematic treatment of exact and of broken dilatation symmetry. The relevance of these ideas to scale dimension and to a new symmetry, scale conjugation, is discussed. As an application, a mass formula is derived from broken dilatation symmetry.

1. INTRODUCTION

The role of group theory and of algebraic methods in the description and exploitation of kinematical symmetries has a long history, both in the nonrelativistic and in the relativistic domain. On the other hand, it is hardly more than a decade since the importance of dynamical symmetries and the usefulness of algebraic methods in this field has been realized. The prototype and ideal of a dynamical group is the nonrelativistic Galilei group which, apart from accounting for the kinematical symmetries and the associated conservation laws, contains a full statement of the nonrelativistic dynamical law.

There are many reasons which make it desirable to have a relativistic analog of the Galilei group or, even more generally, to construct a relativistic group which incorporates dynamical space–time symmetries. In our opinion, the recent evidence that in the high energy region, especially in inelastic collisions which probe the internal structure of hadrons, approximate dilatation symmetry is found, lends added impetus to the search for a relativistic dynamical space–time symmetry group. Broken dilatation invariance is clearly a dynamical symmetry, closely tied to space–time, yet not of a kinematical character. The major purpose of this paper is to show that it is possible to extend the (kinematical) Poincaré group in a manner which leads to a dynamical group that contains, in an essentially unavoidable way, dilatations.

We first review the standard process by which one arrives at a nonrelativistic dynamical group. In nonrelativistic physics, an "event" is labeled by the coordinates q_k, $k = 1, 2, 3$, and a *universal time* t. We may represent the state of a particle by q_k, which then is to be considered as a function of t. The dynamical development of states is precisely given by specifying $q_k(t)$. The kinematics is described by the Euclidean group[1] $SO(3) \otimes T_3^P$, acting on the q_k coordinate space. In order to have a dynamical group, we adjoin time translations (generated by H) and the nonkinematical velocity transformations (generated by Galilean boosts $\mathbf{G}$). The latter connect the kinematical coordinates q_k and the time t. In this manner we obtain the Galilei group

$$\mathcal{G}_4 = (SO(3) \times T_1^H) \otimes (T_3^P \times T_3^G).$$

It is not difficult to adjoin further, nonkinematical transformations so as to obtain a bigger dynamical group, in fact one which contains dilatations. This group $\mathcal{K}_4$, or rather its central extension $\tilde{\mathcal{K}}_4$, has interesting quantum mechanical applications.[2]

The situation is very different in relativistic physics. Here, an "event" is labeled by the world coordinates x_μ, $\mu = 0, 1, 2, 3$, *but there is no relativistic universal time*. At first sight it therefore appears that we cannot have a group which acts on a manifold larger than the geometrical Minkowski space. The corresponding group of motions is the Poincaré group $SO(3,1) \otimes T_4^P$, which is a purely kinematical group.

However, closer inspection of the nonrelativistic case suggests a way to enlarge the Poincaré group to a dynamical group, without reliance on some analog of universal time. As well known, nonrelativistic dynamics may be formulated without reference to universal time if we adopt a "phase space" approach. In essence, this means that we characterize the states of a (noninteracting and nonconstrained) particle not by the values of the function $q_k(t)$, but rather, we define

$$x_k = q_k(0), \qquad \xi_k = \left(\frac{dq_k}{dt}\right)_{t=0}$$

and label every possible state by the pair (x_k, ξ_k). Thus, a state corresponds to a point of the six-dimensional phase space. We than define for any pair $A(x, \xi)$ and $B(x, \xi)$ of dynamical observables of Poisson bracket, setting

$$[A,B]_P = \frac{\partial A}{\partial x_j}\frac{\partial B}{\partial \xi_j} - \frac{\partial A}{\partial \xi_j}\frac{\partial B}{\partial x_j} \quad \text{(summation over } j). \tag{1.1}$$

In particular, we have

$$[x_k, x_l]_P = 0, \qquad [\xi_k, \xi_l]_P = 0, \qquad [x_k, \xi_l]_P = \delta_{kl}. \tag{1.2}$$

The search for a dynamical group can now be formulated in the following fashion: We look for a (linear inhomogeneous) group of *canonical transformations*

$$x_k \to x_k'(x, \xi), \qquad \xi_k \to \xi_k'(x, \xi),$$

which leaves (1.2) invariant and which contains the kinematical transformations $x_k \to x_k' = R_{kj} x_j$ and $x_k \to x_k' = x_k + a_k$. In Appendix A we show that the smallest dynamical group so defined is precisely the standard Galilei group which, when realized on the phase space manifold (x, ξ), assumes the form

$$\mathbf{J}: \begin{cases} x_k' = R_{kj} x_j, \\ \xi_k' = R_{kj} \xi_j, \end{cases} \tag{1.3a}$$

$$\mathbf{P}: \begin{cases} \xi_k' = x_k + a_k, \\ \xi_k' = \xi_k, \end{cases} \tag{1.3b}$$

$$\mathbf{G}: \begin{cases} x_k' = x_k, \\ \xi_k' = \xi_k + v_k \end{cases} \tag{1.3c}$$

$$\mathbf{H}: \begin{cases} x_k' = x_k - \tau \xi_k, \\ \xi_k' = \xi_k, \end{cases} \tag{1.3d}$$

where R_{kj}, a_k, v_k, τ are the usual parameters of the respective subgroups of $\mathcal{G}_4$.

These observations suggest an attempt for the formulation of a relativistic dynamical group.

We base our future work on the following:

Postulate: The state of relativistic particle is characterized by a pair (x_μ, ξ_μ) of independent 4-vectors.

This means that the state of a particle is represented by a point in an eight-dimensional "phase space." Correspondingly, the dynamical development is represented by a curve in the (x, ξ) space.

We may paraphrase this postulate by saying that *the full description of a state requires two Minkowski frames,* $E_{3,1}(x)$ and $E_{3,1}(\xi)$. For brevity, we may refer to the first as the "external frame" and to the second as the "internal frame." In this language, then, x summarizes the data of the external state and ξ subsumes the information on the internal state. In a sense, we have a version of a bilocal theory, inasmuch as when we wish to formulate field theory in our $E_{3,1}(x) \times E_{3,1}(\xi)$ background, the field functions $\psi(x, \xi)$ will have to depend on both sets of coordinates.

An alternative and equivalent viewpoint is to say that we look upon a particle as a vector field over a Minkowski space. That is, a particle is characterized by a world point x_μ and an *attached* vector ξ_μ. In a sense, this characterization of a particle is not much more outlandish than the accustomed picture of visualizing a particle by a point in ordinary space to which a "spin vector" is attached. Accordingly, it may be permitted to think of ξ_μ an an "internal state variable."

We are now prepared to construct a dynamical group, following the nonrelativistic analogy. We define, for a pair $A(x, \xi)$ and $B(x, \xi)$ of dynamical observables, the *relativistic Poisson brackets*

$$[A,B]_P = \frac{\partial A}{\partial x^\rho}\frac{\partial B}{\partial \xi^\rho} - \frac{\partial A}{\partial \xi^\rho}\frac{\partial B}{\partial x^\rho}, \tag{1.4}$$

and obtain[3] for x and ξ,

$$[x_\mu, x_\nu]_P = 0, \quad [\xi_\mu, \xi_\nu]_P = 0, \quad [x_\mu, \xi_\nu]_P = g_{\mu\nu}. \tag{1.5}$$

The dynamical group will consist of certain canonical transformations

$$x_\mu \to x'_\mu(x, \xi), \quad \xi_\mu \to \xi'_\mu(x, \xi),$$

which leave (1.5) *unchanged.*

2. THE DYNAMICAL GROUP $\mathcal{K}_5$

A. Construction of the group

To start with, we wish to include Lorentz transformations $x_\mu \to x'_\mu = \Lambda_{\mu\rho} x^\rho$. But these leave (1.5) invariant if and only if they are accompanied by a Lorentz transformation $\xi_\mu \to \xi'_\mu = \Lambda_{\mu\rho}\xi^\rho$ with the *same* set of parameters. The necessity of this pair of equal Lorentz transformations is also evident from the fact that, because of our basic postulate, ξ_μ is a vector field over the Minkowski space.[4] We denote the generators of this Lorentz subgroup by $J_{\mu\rho}$.

Next, we consider translations $x_\mu \to x'_\mu = x_\mu + a_\mu$. Together with $\xi_\mu \to \xi'_\mu = \xi_\mu$, these form canonical transformations. The corresponding generators will be denoted by P_μ.

At this point, we have nothing more than the Poincaré group $SO(3,1) \otimes T_4^P$. We are perfectly free to adjoin independent translations for the internal reference frame: the set $x_\mu \to x'_\mu = x_\mu$, $\xi_\mu \to \xi'_\mu = \xi_\mu + b_\mu$ of transformations leaves (1.5) invariant. If we denote the generators of these new transformations by Π_μ, the structure of our group is $SO(3,1) \otimes (T_4^P \times T_4^\Pi)$. This group is purely kinematical because the effect of both P_μ and Π_μ is to merely shift the curve of dynamical development in the (x, ξ) space parallel to itself [see Figs. 1a and 1b]. Another way to express this is to note that the Casimir invariants of this group are

$$\mathcal{C}_1 = P_\mu P^\mu, \quad \mathcal{C}_2 = \Pi_\mu \Pi^\mu, \quad \mathcal{C}_3 = P_\mu \Pi^\mu,$$
$$\mathcal{C}_4 = W_\mu W^\mu, \quad \mathcal{C}_5 = V_\mu V^\mu, \quad \mathcal{C}_6 = W_\mu V^\mu,$$

where W_μ is the usual Pauli–Lubanski vector

$$W_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\nu\rho} P^\sigma$$

and V_μ is its analog with P replaced by Π, i.e.,

$$V_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\nu\rho} \Pi^\sigma.$$

It then follows that the state functions (or, in a field theory, the field functions) are separable products, $\psi = \psi_1(p)\psi_2(\pi)$, so that the system is essentially trivial: the internal and external state variables are unrelated.

Thus, we now look for dynamical canonical transformations which will have to mix the x and ξ variables, so as to give an intrinsic change in the dynamical development curve in the (x, ξ) space. The simplest such one-parameter transformation is given by $x_\mu \to x'_\mu = x_\mu - \sigma\xi_\mu$ which, together with $\xi_\mu \to \xi'_\mu = \xi_\mu$, is indeed canonical. By denoting the corresponding generator by S, the group structure becomes

$$\mathcal{G}_5 = (SO(3,1) \times T_1^S) \otimes (T_4^P \times T_4^\Pi).$$

This, the smallest relativistic dynamical group, is a complete analog of the nonrelativistic Galilei group. It has been introduced by us earlier,[5–8] based on a different

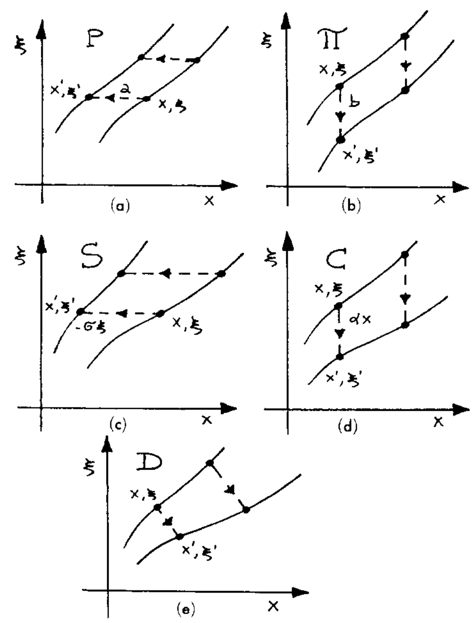


FIG. 1. The action of transformations in the (x, ξ) space.

line of argument and it was realized on a different carrier space.[9] We showed[7] that, remarkably enough, the Casimir invariants of $\mathcal{G}_5$ are

$$\mathcal{C}_1 = P_\mu P^\mu, \quad \mathcal{C}_2 = W_\mu W^\mu,$$

i.e., precisely those of the Poincaré group.

At this point we observe that there is no reason to single out x over ξ, so that we are naturally led to consider the analogous transformations $x_\mu \to x'_\mu = x_\mu$, $\xi_\mu \to \xi'_\mu = \xi_\mu + \alpha x_\mu$. This one-parameter group (whose generator we denote by C) is indeed a canonical transformation. The effect of the S and C transformations on the dynamical state development curve is illustrated in Figs. 1c and 1d. However, *the set of canonical transformations generated* by $J_{\mu\rho}, P_\mu, \Pi_\mu, S,$ *and C does not show closure:* We do not have a group. It is not difficult to prove that, in order to complete the group structure, we must add yet another one-parameter set of canonical transformations which is given by $x_\mu \to x'_\mu = e^\lambda x_\mu$ and $\xi_\mu \to \xi'_\mu = e^{-\lambda}\xi_\mu$. We realize that these transformations (whose generator will be denoted by D) are precisely dilatations on the external variable x, accompanied by corresponding "contractions" on the internal variable ξ. It is striking how the demand of having a closed group of canonical transformations *leads, in a natural and unavoidable manner, to the inclusion of dilatations*. The effect of D in the (x, ξ) space is illustrated in Fig. 1e.

B. Structure and basic properties of $\mathcal{H}_5$

We now summarize the features of the group of canonical transformations which we arrived at in the above manner.

The carrier space[10] is $E_{3,1}(x) \times E_{3,1}(\xi)$ and the defining transformations are

$$J_{\mu\rho}: \begin{cases} x'_\mu = \Lambda_{\mu\rho} x^\rho, \\ \xi'_\mu = \Lambda_{\mu\rho}\xi^\rho, \end{cases} \tag{2.1a}$$

$$P_\mu: \begin{cases} x'_\mu = x_\mu + a_\mu, \\ \xi'_\mu = \xi_\mu, \end{cases} \tag{2.1b}$$

$$\Pi_\mu: \begin{cases} x'_\mu = x_\mu, \\ \xi'_\mu = \xi_\mu + b_\mu, \end{cases} \tag{2.1c}$$

$$S: \begin{cases} x'_\mu = x_\mu - \sigma\xi_\mu, \\ \xi'_\mu = \xi_\mu, \end{cases} \tag{2.1d}$$

$$C: \begin{cases} x'_\mu = x_\mu, \\ \xi'_\mu = \xi_\mu + \alpha x_\mu, \end{cases} \tag{2.1e}$$

$$D: \begin{cases} x'_\mu = e^\lambda x_\mu, \\ \xi'_\mu = e^{-\lambda}\xi_\mu. \end{cases} \tag{2.1f}$$

Performing the transformations in the order J, Π, P, S, C, D, these formulas yield, in a condensed form,

$$x' = e^\lambda\{\Lambda(x - \sigma\xi) + a - \sigma b\}, \tag{2.2a}$$

$$\xi' = e^{-\lambda}\{\Lambda[(1-\alpha\sigma)\xi + \alpha x] + (1-\alpha\sigma)b + \alpha a\}. \tag{2.2b}$$

Denoting such an arbitrary transformation by the symbol[11]

$$g = (\lambda, \alpha, \sigma, a, b, \Lambda),$$

we easily find the composition law

$$\begin{aligned} g'g &= (\lambda', \alpha', \sigma', a', b', \Lambda')(\lambda, \alpha, \sigma, a, b, \Lambda) \\ &= \Big(\lambda' + \lambda + \log(1 - e^{-2\lambda}\alpha\sigma'), \\ &\quad (1 - e^{-2\lambda}\alpha\sigma')(\alpha + e^{2\lambda}\alpha' - \alpha\alpha'\sigma'), \\ &\quad (\sigma + e^{-2\lambda}\sigma' - e^{-2\lambda}\alpha\sigma\sigma')/(1 - e^{-2\lambda}\alpha\sigma'), \\ &\quad \Lambda' a + e^\lambda b'\sigma + e^{-\lambda}a' - e^{-\lambda}\alpha\sigma a', \\ &\quad \Lambda' b + e^\lambda b' - e^{-\lambda}\alpha a', \Lambda'\Lambda\Big). \end{aligned} \tag{2.3}$$

The inverse element of g is then

$$\begin{aligned} g^{-1} = (&-\lambda + \log(1-\alpha\sigma), -e^{-2\lambda}\alpha(1-\alpha\sigma), \\ &-e^{2\lambda}\sigma(1-\alpha\sigma)^{-1}, -e^{-\lambda}\Lambda^{-1}a + e^\lambda\sigma\Lambda^{-1}b, \\ &-e^{-\lambda}(1-\alpha\sigma)\Lambda^{-1}b - e^{-\lambda}\alpha\Lambda^{-1}a, \Lambda^{-1}). \end{aligned} \tag{2.4}$$

Thus, the transformations indeed close to form a group which we shall denote by $\mathcal{H}_5$. It is interesting to note (see Appendix B) that this group is the natural relativistic generalization of Hagen's group $\mathcal{H}_4$ whose central extension $\tilde{\mathcal{H}}_4$ he called the "conformal Galilei group" (cf. Ref. 2).

From the composition law (2.3) we obtain, by standard methods, the following Lie algebra:

$$[J_{\mu\nu}, J_{\rho\sigma}] = i(g_{\nu\rho}J_{\mu\sigma} - g_{\mu\rho}J_{\nu\sigma} - g_{\mu\sigma}J_{\rho\nu} + g_{\nu\sigma}J_{\rho\mu}), \tag{2.5a}$$

$$[J_{\rho\sigma}, P_\mu] = i(g_{\mu\sigma}P_\rho - g_{\mu\rho}P_\sigma), \tag{2.5b}$$

$$[J_{\rho\sigma}, \Pi_\mu] = i(g_{\mu\sigma}\Pi_\rho - g_{\mu\rho}\Pi_\sigma), \tag{2.5c}$$

$$[P_\mu, \Pi_\nu] = 0, \tag{2.5d}$$

$$[P_\mu, P_\nu] = [\Pi_\mu, \Pi_\nu] = 0, \tag{2.5e}$$

$$[S, P_\mu] = 0, \quad [S, \Pi_\mu] = iP_\mu, \tag{2.5f}$$

$$[C, P_\mu] = -i\Pi_\mu, \quad [C, \Pi_\mu] = 0, \tag{2.5g}$$

$$[D, P_\mu] = -iP_\mu, \quad [D, \Pi_\mu] = i\Pi_\mu, \tag{2.5h}$$

$$[S, C] = iD, \quad [S, D] = 2iS, \quad [D, C] = 2iC, \tag{2.5i}$$

$$[J_{\mu\nu}, S] = [J_{\mu\nu}, C] = [J_{\mu\nu}, D] = 0. \tag{2.5j}$$

At this point we observe that if we introduce the linear combinations

$$I_1 = \tfrac{1}{2}D, \quad I_2 = \tfrac{1}{2}(C + S), \quad I_3 = \tfrac{1}{2}(C - S), \tag{2.6}$$

then (2.5i) can be rewritten as

$$[I_1, I_2] = iI_3, \quad [I_2, I_3] = iI_1, \quad [I_3, I_1] = -iI_2. \tag{2.7}$$

Thus, the dynamical canonical transformations generated by S, C, D form a noncompact $SO(2,1)$ [or, equivalently, $SU(1,1)$] subalgebra.[12]

We can now exhibit the structure of the (covering of) the dynamical group $\mathcal{H}_5$ as

$$\mathcal{H}_5 = (SL(2,C)_J \times SU(1,1)_I) \otimes (T_4^P \times T_4^\Pi). \tag{2.8}$$

Thus, the maximal Abelian subgroup (radical) is the direct product of the "external" and "internal" translation group, and the other factor in the semidirect product decomposition is the direct product of a standard kinematical Lorentz and a purely dynamical $SU(1,1)$ group. The latter contains the dilatations. The aesthetic aspects of this structure are indeed pleasing.[13]

In order to write down the *Casimir invariants* of $\mathcal{H}_5$, it is useful to introduce the antisymmetric Lorentz tensor $R_{\mu\nu}$ defined as

$$R_{\mu\nu} = P_\mu \Pi_\nu - P_\nu \Pi_\mu. \tag{2.9}$$

It is then found that there are two Casimir invariants which can be written in the form

$$\mathcal{C}_1 = -\tfrac{1}{2} R_{\mu\nu} R^{\mu\nu}, \tag{2.10a}$$

$$\mathcal{C}_2 = \tfrac{1}{2} J_{\mu\nu} R^{\mu\nu} + S\Pi_\mu \Pi^\mu + C P_\mu P^\mu - D P_\mu \Pi^\mu. \tag{2.10b}$$

We may say that (2.10a) gives the equation of the orbits (see next subsection) in the carrier space of the radical. Explicitly, we have

$$\mathcal{C}_1 = (P_\mu \Pi^\mu)^2 - (P_\mu P^\mu)(\Pi_\nu \Pi^\nu). \tag{2.11}$$

C. Representations of $\mathcal{H}_5$

In order to classify the irreducible unitary representations of our group, we use the method of induced representations.[14]

The semidirect product structure (2.8) is realized by the automorphisms

$$n \to t_h(n) = hnh^{-1},$$

where

$$n \in N \equiv T_4^P \times T_4^\Pi, \quad h \in H \equiv SL(2,C)_J \times SU(1,1)_I.$$

The irreducible unitary representations of N are, of course, one dimensional and we denote them by

$$(a, b\,|\,p, \pi) = \exp[i(ap + b\pi)],$$

where the pair p, π of vectors is referred to as the character of the representation. The set $(ab\,|\,p\pi)$ of all representations is the character group $\hat{N}$. For each $h \in H$, the automorphism t_h defines a one-to-one mapping of $\hat{N}$ onto itself. If, under t_h, $a \to a'$ and $b \to b'$, then, using (2.3), we find that

$$(a'b'\,|\,p\pi) = (ab\,|\,p'\pi'),$$

where

$$p' = e^{\lambda}\Lambda^{-1}p + e^{-\lambda}\alpha\Lambda^{-1}\pi,$$
$$\pi' = e^{-\lambda}(1 - \alpha\sigma)\Lambda^{-1}\pi - e^{\lambda}\sigma\Lambda^{-1}p. \tag{2.12}$$

One easily checks that

$$(p'\pi')^2 - p'^2\pi'^2 = (p\pi)^2 - p^2\pi^2,$$

so that [by comparison with (2.11)] we verified that t_h defines the orbits in N. The *little groups* of $\mathcal{H}_5$ are those subgroups of H which leave a given point of the orbit fixed, i.e., for which in (2.12) we obtain $p' = p$, $\pi' = \pi$. To find the possible little groups, we have to consider five special cases.[15]

Case 1: $p = 0, \pi = 0$. The little group is $SL(2,C)_J \times SU(1,1)_I$ itself. For a maximal set of commuting operators we may choose the union of the standard maximal set for $SL(2,C)$ given by[16]

$$\mathbf{T}^2 - \mathbf{N}^2 \equiv \tfrac{1}{2} J_{\mu\nu} J^{\mu\nu}, \quad \mathbf{TN} \equiv \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\mu\nu} J^{\rho\sigma}, \quad \mathbf{T}^2, \quad \text{and} \quad \mathbf{T}_3 \tag{2.13}$$

and of the standard maximal set for $SU(1,1)$ given by[17]

$$I^2 \equiv I_1^2 - I_2^2 + I_3^2 \quad \text{and} \quad I_2.$$

Thus, the maximal set is $\Sigma = \{\mathbf{T}^2 - \mathbf{N}^2, \mathbf{TN}, \mathbf{T}^2, T_3, I^2, I_2\}$ and the canonical basis functions are labeled as

$$\psi = \psi^{\mathbf{T}^2-\mathbf{N}^2, \mathbf{TN}, I^2}_{\mathbf{T}^2, T_3, I_2}, \tag{2.14}$$

where the upper labels fix the representation and the lower ones are state labels.[18] Note that $\mathcal{C}_1 = 0, \mathcal{C}_2 = 0$.

Case 2: $p = 0, \pi \neq 0$. We have to further distinguish the subcases where $\pi^2 > 0$, $\pi^2 < 0$, $\pi^2 = 0$. We then find the little groups

$$SO(1,1)_{I^-} \times \begin{cases} SU(2) & \text{if } \pi \text{ is timelike,} \\ SU(1,1) & \text{if } \pi \text{ is spacelike,} \\ \overline{E}(2) & \text{if } \pi \text{ is lightlike.} \end{cases}$$

Here $SU(2), SU(1,1), \overline{E}(2)$ are the familiar little groups that occur for the Poincaré group[19] and $SO(1,1)_{I^-}$ is the (noncompact) one-parameter group generated by $I^- \equiv I_2 - I_3 = S$. If we define

$$V_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\nu\rho}\Pi^\sigma, \tag{2.15}$$

then $V_\mu V^\mu$ is an effective Casimir operator for the Poincaré factor and V_3 generates the $SO(2)$ subgroup of the relevant Poincaré little groups. Furthermore (since $p = 0$), Eqs. (2.10a), (2.10b) tells us that now the Casimir operator $\mathcal{C}_1 = 0$ and

$$\mathcal{C}_2 \psi = S\Pi^2\psi. \tag{2.16}$$

For the maximal commuting set we can now choose $\Sigma = \{\mathcal{C}_2, V^2, V_3, \Pi_\mu\}$ and the canonical basis is

$$\psi = \psi^{\mathcal{C}_2}_{V^2, V_3}(\pi). \tag{2.17}$$

Case 3: $p \neq 0, \pi = 0$. Similarly to the previous case, we find the little groups

$$SO(1,1)_{I^+} \times \begin{cases} SU(2) & \text{if } p \text{ is timelike,} \\ SU(1,1) & \text{if } p \text{ is spacelike,} \\ \overline{E}(2) & \text{if } p \text{ is lightlike,} \end{cases}$$

where $SO(1,1)_{I^+}$ is generated by $I^+ \equiv I_2 + I_3 = C$. With the usual definition

$$W_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\nu\rho} P^\sigma \tag{2.18}$$

of the Pauli–Lubanski vector, $W_\mu W^\mu$ and W_3 become state labels. Furthermore, $\mathcal{C}_1 = 0$ and

$$\mathcal{C}_2 \psi = CP^2\psi. \tag{2.19}$$

We choose the maximal set $\Sigma = \{\mathcal{C}_2, W^2, W_3, P_\mu\}$, and the canonical basis is

$$\psi = \psi^{\;2}_{W^2, W_3}(p). \tag{2.20}$$

Case 4: $p \neq 0, \pi \neq 0$ but p and π are parallel. We write $\pi_\mu = dp_\mu$ and so π can be suppressed in the following. The little groups are

$$SO(1,1)_Y \times \begin{cases} SU(2) & \text{if } p \text{ and } \pi \text{ are timelike,} \\ SU(1,1) & \text{if } p \text{ and } \pi \text{ are spacelike,} \\ \overline{E}(2) & \text{if } p \text{ and } \pi \text{ are lightlike,} \end{cases}$$

where now $SO(1,1)_Y$ is generated by

$$Y \equiv 2I_1 - \frac{(1+d^2)}{d} I_2 - \frac{(1-d^2)}{d} I_3.$$

For state labels, we can choose either W^2 and W_3 or V^2 and V_3 (because p and π are parallel); for definiteness we take the former. We have $\mathcal{C}_1 = 0$ and

$$\mathcal{C}_2 \psi = P^2(d^2 S + C - dD)\psi = (S\Pi^2 + CP^2 - DP\Pi)\psi. \tag{2.21}$$

We choose the maximal set $\Sigma = \{\mathcal{C}_2, W^2, W_3, P_\mu\}$, and have the canonical basis

$$\psi = \psi^{\mathcal{C}_2}_{W^2, W_3}(p). \tag{2.22}$$

Case 5: $p \neq 0, \pi \neq 0$, and p and π are not parallel. The little group is the trivial identity group $e_J \times e_I = e$. Now $\mathcal{C}_1 \neq 0, \mathcal{C}_2 \neq 0$ and the maximal set may be taken to be $\Sigma = \{\mathcal{C}_1, \mathcal{C}_2, P_\mu, \Pi_\mu\}$, to which corresponds the canonical basis

$$\psi = \psi^{\mathcal{C}_1 \cdot \mathcal{C}_2}(p, \pi). \tag{2.23}$$

We systematically obtained in the above manner all five possible cases of induced representations. However, from the mathematical point of view, the Cases 2, 3, and 4 are equivalent. This is formally evidenced by the fact that the little groups for these cases are isomorphic to each other. The reason for this equivalence is that, as seen from (2.12), a pure C-transformation ($\Lambda = 1$, $\lambda = \sigma = 0$, $\alpha \neq 0$) can transform a pair ($p = 0, \pi \neq 0$) into a nonvanishing but parallel pair ($p = \alpha\pi, \pi$) and vice versa. Similarly, a pure S transformation can transform a pair ($p \neq 0, \pi = 0$) into the pair ($p, \pi = \sigma p$) and vice versa. Thus, in effect, we have three classes of irreducible unitary representations[20]:

Class I : Case 1 : $\mathcal{C}_1 = 0, \quad \mathcal{C}_2 = 0;$

Class II : Cases 2, 3, 4: $\mathcal{C}_1 = 0, \quad \mathcal{C}_2 \neq 0;$

Class III: Case 5 : $\mathcal{C}_1 \neq 0, \quad \mathcal{C}_2 \neq 0.$

Class I representations do not describe particles, because the momentum is identically zero.

Class II representations (as exemplified by the detailed consideration of Case 3) satisfy all requirements that one would expect from a representation corresponding to an exact symmetry. Since the labeling set of operators is $\Sigma = \{\mathcal{C}_2; W^2, W_3, P_\mu\}$, we see that the only continuous state label is the momentum vector p. We have the usual spin and spin component state labels[21] provided by W^2 and W_3. The value of the Casimir invariant $\mathcal{C}_2 = CP^2$ (which can take on any real value) is closely related to the squared mass. Finally, as will be seen in Sec. 4A, the Hilbert space which carries the representations corresponds to a meaningful realization of exact dilatation symmetry. It should be observed that spin (corresponding to W^2) is not a Casimir invariant, but is only a state label, which means that any fixed representation will contain all possible spin values, both integral and half-integral. In other words, each representation can be reduced into subrepresentations with arbitrary given mass and spin.

Class III representations involve, apart from p, the state-variable π which has no direct physical interpretation, and the same holds for its Casimir invariants. Furthermore, these representations cannot account for particle spin (even though it is not difficult to see that they yield total angular momentum labels[22]). Finally, as will be discussed in Sec. 4A, the unrestricted Hilbert space of a Class III representation corresponds to broken (rather than exact) dilatation symmetry. In spite of their unsuitability for accommodating a classification scheme for exact symmetry, the representations of Class III are far from being useless, and in fact they must be used when questions of broken symmetry are investigated. We shall come back to this point at the end of Sec. 4A.

After having familiarized ourselves with the major physical aspects comprised by our group, we now turn our attention to its possible applications to dilatation physics.[23]

3. DILATATIONS AND $\mathcal{K}_5$

A. Tensor-spinor calculus and scale dimension

One of the captivating features of our $\mathcal{K}_5$ group is *that dilatations occur as part of a semisimple subgroup*, viz., of $SU(1,1)_I$ generated by D, S, C. This fact permits us to establish a *classification scheme relative to dilatation behavior* which will resemble an isospin classification scheme. We point out that when dilatations are treated in the usual framework of the conformally extended Poincaré group (sometimes called the Liouville group), there is no analogous treatment available.

Following the well-known pattern, we must study the (nonunitary) finite-dimensional representations of $SU(1,1)_I$.

It will be convenient to introduce the notation

$$G^0 = D/2 = I_1, \quad G^+ = C = I_2 + I_3, \quad G^- = S = I_2 - I_3. \tag{3.1}$$

Here G^+ and G^- are raising (lowering) operators for the eigenstates of G^0. We also define a "spherical vector" G^1_α, $\alpha = 0, +1, -1$, by setting

$$\begin{aligned} G^1_0 &\equiv G^0 = D/2 = I_1, \\ G^1_{+1} &\equiv 2^{-1/2} G^+ = 2^{-1/2} C = 2^{-1/2}(I_2 + I_3), \\ G^1_{-1} &\equiv 2^{-1/2} G^- = 2^{-1/2} S = 2^{-1/2}(I_2 - I_3). \end{aligned} \tag{3.2}$$

The commutation relations are

$$[G^1_0, G^1_{\pm 1}] = \pm iG^1_{\pm 1}, \quad [G^1_{+1}, G^1_{-1}] = -iG^1_0, \tag{3.3}$$

and the Casimir invariant of $SU(1,1)_I$ can be written as

$$I^2 = (G^1_0)^2 - \tfrac{1}{2}(G^1_{+1} + G^1_{-1})^2 + \tfrac{1}{2}(G^1_{+1} - G^1_{-1})^2, \tag{3.4}$$

which is formally the same relation as one has for the Casimir invariant of $SU(2)$ in terms of spherical components. By definition and in consequence of (3.3), the transformation law of an arbitrary spherical vector V^1_α is given as

$$\begin{aligned} &[G^1_0, V^1_{\pm 1}] = \pm iV^1_{\pm 1}, \quad [G^1_{\pm 1}, V^1_{\mp 1}] = \mp iV^1_0, \\ &[G^1_{\pm 1}, V^1_0] = \mp iV_{\pm 1}, \quad [G^1_\alpha, V^1_\alpha] = 0. \end{aligned} \tag{3.5}$$

One advantage of using spherical vectors is that each component G^1_α spans a one-dimensional (nonunitary) representation for the T^D_1 dilatation subgroup generated by $I_1 \equiv G^0$. Because of its obvious physical importance, we study these representations.

The (nonunitary) representations of T^D_1 have the form

$$g(\lambda): \chi \to e^{il\lambda/2}\chi, \tag{3.6}$$

where λ is the parameter of dilatation and $l = 0, \pm 1, \pm 2, \cdots$. If X is a tensor (or spinor) operator belonging to the representation of T^D_1 characterized by l, then, clearly,[24]

$$e^{-i\lambda D} X e^{i\lambda D} = e^{l\lambda} X, \tag{3.7}$$

or, in infinitesimal form,

$$[D, X] = ilX. \tag{3.8}$$

This reveals the meaning of l: *the integer l is precisely the scale dimension of the dynamical observable X.* Furthermore, since $D = 2I_1 = G^1_0$, it is clear that *for any tensor (or spinor) operator X^n_α of $SU(1,1)_I$ the spherical component subscript α is exactly one-half of the scale dimension of X^n_α.* A further important remark is that, in general, *the (physical) scale dimension is different from the naive (geometrical) dimension.*[25]

We now study the tensor (spinor) transformation behavior of the generators of $\mathcal{K}_5$ under the $SU(1,1)_I$ subgroup. We find that

(a) $$G = \begin{pmatrix} 2^{-1/2}C \\ D/2 \\ 2^{-1/2}S \end{pmatrix} = \begin{pmatrix} G^1_{+1} \\ G^1_0 \\ G^1_{-1} \end{pmatrix}$$

is a vector, as already stated.

(b) $J_{\mu\nu}$ is a scalar.

(c) $$u = \begin{pmatrix} \Pi_\mu \\ P_\mu \end{pmatrix} = \begin{pmatrix} K^{1/2}_{+1/2} \\ K^{1/2}_{-1/2} \end{pmatrix}$$

is a fundamental spinor and

$$\bar{u} = (P_\mu, -Q_\mu) = (K^{1/2}_{1/2}, -K^{1/2}_{-1/2})$$

is its conjugate (contragredient) spinor. Observe that $\bar{u}u = 0$, as it should be. From the explicit form of u and G^1_α we get

$$\begin{aligned} &[G^1_0, K^{1/2}_{\pm 1/2}] = \pm \tfrac{1}{2} i K^{1/2}_{\pm 1/2}, \qquad [G^1_{\pm 1}, K^{1/2}_{\pm 1/2}] = 0, \\ &[G^1_{\pm 1}, K^{1/2}_{\mp 1/2}] = \mp (i/\sqrt{2}) K^{1/2}_{\pm 1/2}, \end{aligned} \tag{3.9}$$

which, then, determines the transformation law for an arbitrary fundamental spinor $V^{1/2}_\alpha$ $(\alpha = \pm 1/2)$.

We can now construct tensor (spinor) operators in the enveloping algebra. The simplest examples are collected in Table I.

TABLE I. Some simple covariants in the enveloping algebra. *Note*: $\underset{\smile}{AB}$ means the Hermitian product $\frac{1}{2}\{A, B\}$.

Symbol	Definition	$SU(1,1)_I$ transformation property	Scale dimension	Scale parity	Lorentz transformation property
V^0_0	$R_{\mu\nu} = P_\mu \Pi_\nu - P_\nu \Pi_\mu$	scalar	0	$-$	antisymm. tensor
V^1_{+1}	$\Pi_\mu \Pi_\nu$		+2		symmetric
V^1_0	$1/\sqrt{2}\,(P_\mu \Pi_\nu + P_\nu \Pi_\mu)$	vector	0	+	tensor
V^1_{-1}	$P_\mu P_\nu$		-2		
$V^{1/2}_{+1/2}$	$\underset{\smile}{CP^\mu} - \frac{1}{2}\underset{\smile}{D\Pi^\mu}$	$\frac{1}{2}$ spinor	+1	not applicable	vector
$V^{1/2}_{-1/2}$	$-\underset{\smile}{S\Pi^\mu} + \frac{1}{2}\underset{\smile}{DP^\mu}$		-1		
$V^{3/2}_{+3/2}$	$\underset{\smile}{C\Pi^\mu}$		+3		
$V^{3/2}_{+1/2}$	$\underset{\smile}{CP^\mu} + \underset{\smile}{D\Pi^\mu}$	$\frac{3}{2}$ spinor	+1	not applicable	vector
$V^{3/2}_{-1/2}$	$\underset{\smile}{S\Pi^\mu} + \frac{1}{2}\underset{\smile}{DP^\mu}$		-1		
$V^{3/2}_{-3/2}$	$\underset{\smile}{SP^\mu}$		-3		

One implication of this classification scheme is that when, on physical grounds, we know that a certain dynamical observable has a definite scale dimension, we can expect that it will have "partners" of different scale dimension, together with which it forms an $SU(1,1)_I$ tensor (spinor). Application of these ideas to currents are planned to be discussed in a later publication.

B. Scale conjugation

Inspection of the Eqs. (2.5a)–(2.5j) reveals that the Lie algebra of $\mathcal{K}_5$ admits the following, rather remarkable *involutive outer automorphism:*

$$\begin{aligned} &P_\mu \to \Pi_\mu, \quad \Pi_\mu \to P_\mu, \quad S \to -C, \quad C \to -S, \\ &D \to -D, \quad J_{\mu\nu} \to J_{\mu\nu}. \end{aligned} \tag{3.10}$$

We shall call this new symmetry transformation a *scale conjugation*, because of its strong resemblance to charge conjugation (or G-parity) for isospin multiplets, as will transpire below. Scale conjugation (3.10) can be represented by a unitary and self-adjoint operator[26] $\mathfrak{D}$:

$$\begin{aligned} &\mathfrak{D} P_\mu \mathfrak{D}^{-1} = \Pi_\mu, \quad \mathfrak{D} \Pi_\mu \mathfrak{D}^{-1} = P_\mu, \quad \mathfrak{D} S \mathfrak{D}^{-1} = -C, \\ &\mathfrak{D} C \mathfrak{D}^{-1} = -S, \quad \mathfrak{D} D \mathfrak{D}^{-1} = -D, \quad \mathfrak{D} J_{\mu\nu} \mathfrak{D}^{-1} = J_{\mu\nu}. \end{aligned} \tag{3.11}$$

Since $\mathfrak{D}^2 = 1$, the possible eigenvalues of $\mathfrak{D}$ are ± 1. We shall call the eigenvalue of $\mathfrak{D}$ *scale parity*. Equation (3.11) shows that D has negative scale parity.[27] Consequently, only those dynamical observables can have a sharp scale parity which have zero scale dimension. In general, $\mathfrak{D}$ causes transitions within $SU(1,1)_I$ multiplet members. For example, the neutral component G^1_0 of the vector G^1_α has negative scale parity, and, under $\mathfrak{D}$, we have the transitions $G^1_{+1} \to -G^1_{-1}$ and $G^1_{-1} \to G^1_{+1}$. Following the convention for the classification of integral isospin meson multiplets relative to G (or C) parity, we may say that G^1_α has negative scale parity. On the other hand, the vector V^1_α of Table I has positive scale parity, since under $\mathfrak{D}$, we have $V^1_0 \to V^1_0$ (and $V^1_{+1} \longleftrightarrow V^1_{-1}$). $SU(1,1)_I$ scalars have always sharp scale parity (because they have zero scale dimension). For example, the Casimir operator I^2 of $SU(1,1)_I$ has positive scale parity, as can be seen from Eq. (3.4). On the other hand, the scalar $V^0_0 = R_{\mu\nu}$ of Table I has negative scale parity. We observe that the Casimir invariant $\mathfrak{C}_1$ of $\mathcal{K}_5$ has positive scale parity (i.e., it is also an invariant for the extended group), but the second Casimir invariant $\mathfrak{C}_2$ has negative scale parity (so that it is not an invariant for the extended group).

It is interesting to note that in the usual treatment of dilatation symmetry within the framework of the conformal group, there is no room for a scale conjugation automorphism.[28] Further exploration of our $\mathfrak{D}$-symmetry is relegated to a later study.

4. SOME ASPECTS OF BROKEN SCALE INVARIANCE

A. Scale invariant subspaces

Whereas the group $\mathcal{K}_5$ incorporates dilatations, it by no means implies exact scale *invariance*. In fact, as we saw in Sec. 2A, S and C generate dynamical development and, since D does not commute with these operators, it cannot be a "constant of motion". Thus, we have a broken symmetry, and it becomes necessary to find conditions for a meaningful implementation of exact dilatation symmetry. This means that we must search for a subspace of the Hilbert space of states which is left invariant under D. Naturally, we must have this subspace invariant also under the whole kinematical subgroup, which is $SL(2,C)_J \otimes (T_4^P \times T_4^\Pi)$. A subspace is selected by a *subsidiary condition* $\Omega\psi = 0$ and, to have this subspace invariant under the transformations listed above, we must have

$$\Omega(D\psi) = 0, \quad \Omega(J_{\mu\nu}\psi) = 0, \quad \Omega(P_\mu\psi) = 0, \quad \Omega(\Pi_\mu\psi) = 0.$$

These conditions are met if we find an operator Ω such that

$$[D,\Omega] = c_1\Omega, \quad [J_{\mu\nu},\Omega] = c_2\Omega, \quad [P_\mu,\Omega] = c_3\Omega, \quad [\Pi_\mu,\Omega] = c_4\Omega, \tag{4.1}$$

where some of the constants c_k may be zero. Inspection of our Lie algebra (2.5) reveals that P^2, Π^2, or $P\Pi$ are three possible basic choices for Ω to satisfy Eq. (4.1). Correspondingly we have:

Choice (a): The subsidiary condition is $P^2\psi = 0$. This is not unexpected since it is well known that the dynamics of massless particles obeys scale invariance. The novelty is that, as easily seen, the subspace is also invariant under S.

Choice (b): The subsidiary condition $\Pi^2\psi = 0$ gives the interesting result that, in our framework, we can have scale invariance for massive particles too. It is also seen that this subspace has the additional invariance under C.

Choice (c): The subsidiary condition $P\Pi\psi = 0$ also permits scale invariance, but this possibility is not attractive because neither of the dynamical development operators C or S leaves this subspace invariant. In addition, the condition $P\Pi\psi = 0$ imposes an undesirable kinematical restriction among the components of P and Π: For example, if P is timelike, Π is *constrained* to be spacelike.

We remark here the following. The Class II representations of $\mathcal{K}_5$ are already so restricted that we have exact scale invariance. This is so because in this class $\mathcal{C}_1 = -\frac{1}{2}R^2 = 0$, so that $\Omega = R^2$ trivially satisfies the requirements of Eq. (4.1). In contrast, *the representations belonging to Class III are precisely those which permit the study of symmetry breaking in massive hadron systems.*

Here the exact scale invariance conditions are not automatically satisfied, and we select the exactly invariant reference system by *imposing* the condition $\Pi^2\psi = 0$. In the next subsection we shall show how broken symmetry arguments may lead to a mass spectrum.

B. Mass spectrum from broken symmetry

In the real world, the symmetry of $\mathcal{K}_5$ is certainly badly broken by the dynamics of interactions. Consequently, the squared mass operator $\mathfrak{M}^2$ cannot be an invariant of $\mathcal{K}_5$. But we expect that part of the $SU(1,1)_I$ symmetry, namely dilatation invariance, survives in first approximation. This means that we can write[29]

$$\mathfrak{M}^2 = A(1 + \gamma B), \tag{4.2}$$

where A, the unperturbed central squared mass, and the constant γ depend only on the Casimir invariants of $\mathcal{K}_5$, whereas B is a component of a tensor operator of $SU(1,1)_I$ which is invariant under the T_1^D dilatation symmetry subgroup. Moreover, B must be invariant under the entire kinematical subgroup $SL(2,C)_J \otimes (T_4^P \times T_4^\Pi)$. Thus, B should be constructed from $J_{\mu\nu}, P_\mu, \Pi_\mu$, it should have zero scale dimension, but it should not be invariant under S and C. The simplest choice for B is to take the neutral component of an $SU(1,1)_I$ *vector*. It is easily seen that the lowest order polynomial in the enveloping algebra which meets these requirements for our vector operator is given by[30]

$$\begin{aligned} Z^1_{+1} &= 2^{-3/2}J_{\alpha\beta}J^{\alpha\beta}\Pi^2 - 2^{-1/2}J_{\alpha\mu}J^{\beta\mu}\Pi^\alpha\Pi_\beta, \\ Z^1_0 &= \tfrac{1}{2}J_{\alpha\beta}J^{\alpha\beta}P\Pi - \tfrac{1}{2}J_{\alpha\mu}J^{\beta\mu}(P^\alpha\Pi_\beta + \Pi^\alpha P_\beta), \\ Z^1_{-1} &= 2^{-3/2}J_{\alpha\beta}J^{\alpha\beta}P^2 - 2^{-1/2}J_{\alpha\mu}J^{\beta\mu}P^\alpha P_\beta, \end{aligned} \tag{4.3}$$

and, in accord with the preceeding arguments, we take

$$B = Z^1_0 = \tfrac{1}{2}J_{\alpha\beta}J^{\alpha\beta}P\Pi - \tfrac{1}{2}J_{\alpha\mu}J^{\beta\mu}(P^\alpha\Pi_\beta + \Pi^\alpha P_\beta). \tag{4.4}$$

We now want to calculate the expectation value $\langle\mathfrak{M}^2\rangle$. In view of the discussion given at the end of subsection 4A, we select a representation of Class III and (so as to have, in our approximation, exact dilatation symmetry) take the subspace for which the $\Pi^2\psi = 0$ condition is met.

To do the calculations, it will be necessary to use a basis other than the canonical basis given by Eq. (2.23). The new basis can be obtained by considering the chain

$$\mathcal{K}_5 \supset (SL(2,C)_J \times T_1^D) \otimes (T_4^P \times T_4^\Pi) \supset SL(2,C)_J \times T_1^D \supset SU(2)_{\mathrm{T}} \supset SO(2)_{T_3}$$

and taking for labels the Casimir operators in each link of the chain.[31] We thus get[32] the new maximal commuting set

$$\Sigma' = \{\mathcal{C}_1, \mathcal{C}_3, WV, P^2\Pi^2, P\Pi, D, JJ, JJ^*, \mathbf{T}^2, T_3\},$$

where

$$JJ \equiv J_{\mu\nu}J^{\mu\nu}, \quad JJ^* \equiv \epsilon_{\mu\nu\rho\sigma}J^{\mu\nu}J^{\rho\sigma}, \quad \mathbf{T} = (J_{23}, J_{31}, J_{12}).$$

The basis corresponding to Σ' is unitarily equivalent to the original basis which corresponds to Σ.

We are now prepared to calculate $\langle B\rangle$ for states in our subspace. We take a *rest frame* in which $P = (P_0; 0,0,0)$ and observe that, since we have the subsidiary condition $\Pi^2\psi = 0$, Eq. (2.11) gives $\mathcal{C}_1 = (P\Pi)^2$, so that further, in our rest frame,

$$P\Pi = P_0\Pi_0 = P_0|\mathbf{\Pi}| = \sqrt{\mathcal{C}_1}.$$

With this in mind, an elementary calculation gives[33]

$$\langle B\rangle = \sqrt{\mathcal{C}_1}\,(\tfrac{1}{2}J_{\mu\nu}J^{\mu\nu} + \mathbf{N}^2) - P_0\mathbf{\Pi}\cdot(\mathbf{N}\times\mathbf{T}) + iP_0\mathbf{\Pi}\cdot\mathbf{N}. \quad (4.5)$$

At this point we recall[34] that

$$\tfrac{1}{2}J_{\mu\nu}J^{\mu\nu} \equiv \mathbf{T}^2 - \mathbf{N}^2 = k^2 + c^2 - 1, \quad (4.6a)$$

$$\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\mu\nu}J^{\rho\sigma} \equiv \mathbf{T}\cdot\mathbf{N} = 4ikc, \quad (4.6b)$$

where, for the principal series,

$$k = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots \quad \text{and} \quad -\infty < ic < +\infty,$$

but we must take $c = 0$ so as to have normalizable states. Then (4.6b) gives

$$\mathbf{T}\cdot\mathbf{N} = 0. \quad (4.7)$$

We further recall that in the rest system $\mathbf{T}$ becomes the spin, so that

$$\mathbf{T}^2 = s(s+1), \quad (4.8)$$

where $s = k, k+1, k+2, \cdots$. We need one more calculation to evaluate (4.5). With $V_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\nu\rho}\Pi^\sigma$ we have, by symmetry, $V_\mu\Pi^\mu = 0$, and since Π is lightlike, this implies the parallelism or antiparallelism $V_\mu = f\Pi_\mu$. In detail, this gives

$$\mathbf{T}\Pi_0 + \mathbf{N}\times\mathbf{\Pi} = f\mathbf{\Pi} \quad \text{and} \quad \mathbf{T}\cdot\mathbf{\Pi} = -f\Pi_0,$$

which, using (4.7), leads to

$$\mathbf{N}\cdot\mathbf{\Pi} = 0 \text{ and } \mathbf{T}\times\mathbf{\Pi} = -\mathbf{N}\Pi_0. \quad (4.9)$$

Substituting these results into (4.5) and using (4.6a) (with $c = 0$), we eventually get

$$\langle B\rangle = \sqrt{\mathcal{C}_1}\,[2s(s+1) - \tfrac{3}{2}k^2 + 2],$$

so that, finally, our broken symmetry mass formula (4.2) can be written in the form

$$\langle \mathfrak{M}^2\rangle = \alpha + \beta[s(s+1) - \tfrac{3}{4}k^2], \quad (4.10)$$

where α and β are (unknown) constants.

To test this formula, it must be remembered that approximate dilatation invariance is presumably a reasonable approximation only at higher energies which, when visualizing elementary particles as excited dynamical systems, implies that our formula may be reasonable for higher mass level sequences, and cannot shed light on the lower sequences (like the familiar octet or decuplet). It appears[35] that the first nucleonlike higher mass spin-tower starts with the $N''(1780)$ and $N(1860)$ states ($s = \frac{1}{2}$ and $\frac{3}{2}$, respectively). Assuming that $k = \frac{1}{2}$ for these, and, using their masses as inputs, we determine the parameters to be $\alpha = 3.1$ and $\beta = 0.1$ BeV^2, respectively. Table II shows the numerical results obtained with these input data, for a series of $k = \frac{1}{2}$ nucleon states, then for a $k = \frac{3}{2}$ sequence[36] of Δ-like states, and, finally, for a $k = 1$ meson tower.[37] Even though there are several reasons why one should not take a broken symmetry mass formula and its confrontation with the often hazy data too seriously, we find it rather impressive that, for this wide variety of quantum numbers, the agreement with experiment is extremely good. It is also worthwhile to note that, as expected, the accuracy of the formula seems to improve as we go to higher masses, say to above 2000 MeV.

TABLE II. Mass levels from broken symmetry.

Symbol (and resonance channel)	k	s	$\langle\mathfrak{M}^2\rangle$ calculated	$\mathfrak{M}^2$ experiment	Error: Stand. dev.	Error: %
$N''(1780)\ P'_{11}$	$\frac{1}{2}$	$\frac{1}{2}$	input	3.17 ± 0.51		
$N(1860)\ P_{13}$	$\frac{1}{2}$	$\frac{3}{2}$	input	3.46 ± 0.57		
See Footnote a	$\frac{1}{2}$	$\frac{5}{2}$	3.98	See Footnote a		
$N(1990)\ F_{17}$	$\frac{1}{2}$	$\frac{7}{2}$	4.67	3.96		+18
$N(2220)\ H_{19}$	$\frac{1}{2}$	$\frac{9}{2}$	5.57	4.93 ± 0.65	1	+13
$N(2650)$	$\frac{1}{2}$	$\frac{11}{2}$	6.67	7.02 ± 0.95	$\frac{1}{3}$	−5
⋮						
$\Delta(1690)\ P''_{33}$	$\frac{3}{2}$	$\frac{3}{2}$	3.31	2.86		+16
$\Delta(1890)\ F_{35}$	$\frac{3}{2}$	$\frac{5}{2}$	3.81	3.57 ± 0.49	$\frac{1}{2}$	+7
$\Delta(1950)\ F_{37}$	$\frac{3}{2}$	$\frac{7}{2}$	4.51	3.80 ± 0.39	<2	+18
See Footnote b	$\frac{3}{2}$	$\frac{9}{2}$	5.41	See Footnote b		
$\Delta(2420)$	$\frac{3}{2}$	$\frac{11}{2}$	6.51	5.86 ± 0.75	<1	+11
$\Delta(2850)$	$\frac{3}{2}$	$\frac{13}{2}$	7.81	8.12 ± 1.14	$\frac{1}{4}$	−4
⋮						
$R(1750)$ or $\Pi\Pi(1764)$	1	1	3.23	3.09		+5
$S(1930)$	1	2	3.68	3.72		−1
$X^-(2086)$?	1	3	4.23	4.37		−3
$T(2195)$?	1	4	5.03	4.84		+4
$X^-(2500)$ or $U(2375)$?	1	5	6.03	6.25 / 5.57 ± 0.07		−3 to +8

[a] There is no N state with $\frac{5}{2}^+$ in this region, but there is the $\Lambda(1815)F'_{05}$ with $\mathfrak{M}^2 = 3.30 \pm 0.15$ and the $\Sigma(1915)$ with $\mathfrak{M}^2 = 3.65 \pm 0.13$. The expected octet partner N would fit well the calculated value.

[b] There is no Δ known with $\frac{9}{2}^+$, but there is the spin $\frac{9}{2}^+$ $Y(2250)$ with $\mathfrak{M}^2 = 5.06 \pm 0.37$, which, if a partner of the missing Δ, fits well the calculated value.

ACKNOWLEDGMENT

We are greatly indebted to Professor R. M. Santilli for many interesting discussions and for advice on several points.

APPENDIX A: THE GALILEI GROUP IN PHASE SPACE

The standard defining transformations of the Galilei group $\mathcal{G}_4$ are given on the $E_3(\mathbf{q}) \times E_1(t)$ carrier space as follows:

$$\mathbf{J}: \begin{cases} q'_k = R_{kj}q_j, \\ t' = t, \end{cases}$$

$$\mathbf{P}: \begin{cases} q'_k = q_k + a_k, \\ t' = t, \end{cases}$$

$$\mathbf{G}: \begin{cases} q'_k = q_k + v_k t, \\ t' = t, \end{cases}$$

$$H: \begin{cases} q'_k = q_k, \\ t' = t + \tau. \end{cases}$$

Performing the transformations in the order $\mathbf{J}, \mathbf{G}, \mathbf{P}, H$ we have, in a condensed form,

$$\begin{aligned} \mathbf{q}' &= R\mathbf{q} + \mathbf{v}t + \mathbf{a}, \\ t' &= t + \tau. \end{aligned} \quad (A1)$$

Denoting a group element by $g = (\tau, \mathbf{a}, \mathbf{v}, R)$, we easily obtain the composition law

$$g'g = (\tau' + \tau, \mathbf{a}' + R'\mathbf{a} + \tau\mathbf{v}', \mathbf{v}' + R'\mathbf{v}, R'R). \quad (A2)$$

On the other hand, from the transformations (1.3a)–(1.3d), defined on the carrier space $E_3(\mathbf{x}) \times E_3(\boldsymbol{\xi})$, we find, using the same order of transformations as above,

$$\mathbf{x}' = R(\mathbf{x} - \tau\boldsymbol{\xi}) + \mathbf{a} - \tau\mathbf{v}, \quad \boldsymbol{\xi}' = R\boldsymbol{\xi} + \mathbf{v}. \tag{A3}$$

If we compute from here the composition law, we get precisely (A2). This proves that the two groups, defined on different carrier spaces, are actually isomorphic. Thus, the group of canonical transformations (1.3) in the phase space is simply another realization of the familiar Galilei group. QED

APPENDIX B: A NONLINEAR REALIZATION OF $\mathcal{H}_5$

Let $E_{3,1}(x)$ be the usual Minkowski space and let $E_1(u)$ be a one-dimensional space. Consider the following transformations on $E_{3,1}(x) \times E_1(u)$:

$$J_{\mu\rho}: \begin{cases} x'_\mu = \Lambda_{\mu\rho} x^\rho, \\ u' = u, \end{cases} \tag{B1a}$$

$$P_\mu: \begin{cases} x'_\mu = x_\mu + a_\mu \\ u' = u, \end{cases} \tag{B1b}$$

$$\Pi_\mu: \begin{cases} x'_\mu = x_\mu + b_\mu u, \\ u' = u, \end{cases} \tag{B1c}$$

$$S: \begin{cases} x'_\mu = x_\mu, \\ u' = u + \sigma, \end{cases} \tag{B1d}$$

$$C: \begin{cases} x'_\mu = x_\mu/(1 - \alpha u), \\ u' = u/(1 - \alpha u), \end{cases} \tag{B1e}$$

$$D: \begin{cases} x'_\mu = e^\lambda x_\mu, \\ u' = e^{2\lambda} u. \end{cases} \tag{B1f}$$

Performing the transformations in the order J, Π, P, S, C, D, we get, in condensed form,

$$x' = e^\lambda \left(\frac{\Lambda x + bu + a}{1 - \alpha(u + \sigma)}\right), \quad u' = e^{2\lambda}\left(\frac{u + \sigma}{1 - \alpha(u + \sigma)}\right). \tag{B2}$$

Denoting a group element by $g = (\lambda, \alpha, \sigma, a, b, \Lambda)$ and computing from (B2) the composition law, we get precisely the same result as given by Eq. (2.3) for the composition law of the group defined on $E_{3,1}(x) \times E_{3,1}(\xi)$ by the Eqs. (2.1a)–(2.1f). Thus, the group defined by the transformations (B1a)–(B1f) is isomorphic to our group $\mathcal{H}_5$. But the realization of $\mathcal{H}_5$ on $E_{3,1}(x) \times E_1(u)$ is nonlinear, as is evident from (B1e).

It is interesting to note that the nonrelativistic analog of the transformations (B1a)–(B1f) [when the carrier space is $E_3(\mathbf{q}) \times E_1(t)$] is precisely the set of defining transformations for Hagen's "conformal Galilei group" (cf. Ref. 2). Thus, our $\mathcal{H}_5$ is essentially the relativistic generalization of this group.

We also observe that if we omit the C and D transformations (i.e., set $\alpha = 0, \lambda = 0$), then Eqs. (B1a)–(B1d) become the original defining transformations of the group $\mathcal{G}_5$ whose central extension was the major topic of study in Refs. 5–8.

---

[1] In this paper we denote direct products of groups by × and semidirect products by ⊗.

[2] See C. R. Hagen, Phys. Rev. D 5, 377 (1972); and also P. Roman, J. J. Aghassi, R. M. Santilli, and P. L. Huddleston, "Nonrelativistic composite elementary particles and the conformal Galilei group," Nuovo Cimento (to be published); as well as U. Niederer (Zurich U.), Preprint, March 1972.

[3] Throughout this paper, $g_{00} = 1, g_{kk} = -1, g_{kl} = 0$ $(k \neq l)$.

[4] A further comment on this point will be given in Footnote 13.

[5] J. J. Aghassi, P. Roman, R. M. Santilli, Phys. Rev. D **1**, 2753 (1970).

[6] J. J. Aghassi, P. Roman, R. M. Santilli, J. Math. Phys. (N.Y.) **11**, 2297 (1970).

[7] J. J. Aghassi, P. Roman, R. M. Santilli, Nuovo Cimento A **5**, 551 (1971).

[8] R. M. Santilli, Particles Nuclei **1**, 81 (1970).

[9] In our earlier work, we actually concentrated on the central extension $\mathcal{G}_5$ of this group. These efforts met with limited success.

[10] In Appendix B we give another, nonlinear realization of the group, on a smaller carrier space.

[11] The unit element is (0,0,0,0,0,1).

[12] This is analogous to the situation in the nonrelativistic $\mathcal{H}_4$ (and $\tilde{\mathcal{H}}_4$) group, as was noticed originally by Hagen, Ref. 2.

[13] Forgetting the manner in which we arrive at $\mathcal{H}_5$ as a group of canonical transformations, one may wonder whether a doubling of the Lorentz part (one acting on x, one on ξ, independently) would serve a purpose. However, an elementary calculation with Jacobi identities reveals that then $SU(1,1)$ must commute with the rest of the group (which would consist of the two Lorentz and the two translation groups), so that we would have an entirely trivial structure, without any link to dynamics.

[14] For a simple account of this method, see, for example, G. W. Mackey, *Induced representations of groups and quantum mechanics* (Benjamin, New York, 1968), or the brief summary by G. W. Mackey, *Group representations in Hilbert space*, which is the Appendix in I. E. Segal, *Mathematical problems in relativistic physics* (Amer. Math. Soc., Providence, R. I., 1963).

[15] The procedure to be followed is a combination of the standard method for finding the little groups of the Poincaré group [E. P. Wigner, Ann. Math. **40**, 149 (1939)] and of the work of E. Inönü and E. P. Wigner, Nuovo Cimento **9**, 705 (1952), in which they obtained the little groups of the nonrelativistic Galilei group $\mathcal{G}_4$. Our results resemble closely those of the latter paper.

[16] We adopt the notation $\mathbf{T} = (J_{23}, J_{31}, J_{12})$ and $\mathbf{N} = (J_{01}, J_{02}, J_{03})$.

[17] Note that Eq. (2.7) tells us that I_2 generates the *compact* $SO(2)$ subgroup of our $SU(1,1)$. Of course, instead of the state label I_2 we could use any other one, like I_1.

[18] We shall adopt the same notational convention in all subsequent cases too.

[19] These are the covering groups of $SO(3)$, $SO(2,1)$, and the Euclidean group $E(2)$.

[20] The situation resembles closely that which is found in the classical Galilei group $\mathcal{G}_4$; cf. Ref. 15.

[21] Naturally, we consider the subcase when $p^2 > 0$, so that the Poincaré part of the little group is $SU(2)$.

[22] This can be seen by introducing a noncanonical basis and will be discussed in Sec. 4B.

[23] For a well readable survey of dilatation physics in the framework of quantum field theory, see, for example, P. Carruthers, Phys. Rep. **1**, 1 (1971), or S. Coleman, "Dilatations", in *The Proceedings of the 1971 International Summer School of Physics Ettore Majorana*, edited by A. Zichichi (in press).

[24] Recall that $D = 2I_1$.

[25] This can be seen from the defining transformations (2.1) which show that the parameters a_μ and b_μ have the same dimension as the coordinates whereas all other parameters are dimensionless.

[26] Obviously, $\mathfrak{D}$ is outside the enveloping algebra.

[27] This is analogous to the fact that electric charge Q has negative charge parity.

[28] The commutator between the generators K_μ of proper conformal transformations and the momenta P_μ is $[K_\mu, P_\nu] = -2i(g_{\mu\nu} D + J_{\mu\nu})$, and this prevents the implementation of a scale conjugation.

[29] The subsequent arguments are analogous to those used, for example, in the derivation of a mass formula for broken $SU(6)$ symmetry.

[30] It is interesting to note that (apart from a normalization factor) Z^1_{-1} is the standard W^2, Z^1_{+1} is its analog V^2, and Z^1_0 is $W_\mu V^\mu$.

[31] The procedure is the same as the one which is used in the Poincaré framework when an angular momentum basis is introduced in place of the canonical one by taking the chain $\mathcal{P} \supset SL(2,C) \supset SU(2) \supset SO(2)$.

[32] Observe that the Casimir operators of $(SL(2,C)_J \times T^P_1) \otimes (T^P_4 \times T^\Pi_4)$ are $WV, P^2\Pi^2, P\Pi$ and those of $SL(2,C)_J \times T^P_1$ are D, JJ, JJ^*.

[33] For the notation $\mathbf{T}$ and $\mathbf{N}$, see Footnote 16.

[34] See, for example, I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the rotation and Lorentz group* (Pergamon, New York, 1963).

[35] All particle data and symbols are taken from the tabulations of the Particle Data Group, Rev. Mod. Phys. **43** (2), Part II (1971).

[36] Here k cannot be $1/2$, because no $s = 1/2$ member occurs in the Δ-type family.

[37] All baryon states considered have positive parity. For the meson states, there are considerable uncertainties.

Converging bounds for the free energy in certain statistical mechanical problems

George A. Baker Jr.*

Baker Laboratory, Cornell University, Ithaca, New York 14850

Applied Mathematics Department, Brookhaven National Laboratory, Upton, L. I., New York 11973

(Received 9 June 1972)

We present sequences of converging upper and lower bounds to the partition function per spin specifically for a ferromagnetic Ising model which are valid in the entire finite, magnetic-field, inverse-temperature plane. They are based on the exact high-temperature expansions for a finite system, properties of generalized Padé approximants, and the Villani limit theorem. The results depend only on the general structure of the partition function and certain monotonicity with system size properties which hold fairly generally.

An important problem in theoretical physics[1] is the deduction of exact results for the general dimension, spin-$\frac{1}{2}$ Ising model. In particular, at the present time for dimension higher than 2, methods are lacking which can be *proved* to converge to compute the thermodynamic properties in zero magnetic field near the physically interesting critical point singularity, or even its exact location for that matter. In this work we show how to construct sequences of successive approximations to the free-energy per spin which bound it from above and below for all temperatures and magnetic fields and converge to the thermodynamic free-energy per spin in the limit. Our procedures can be generalized to apply to a variety of other statistical mechanical problems, but we have chosen the Ising problem for ease of presentation.

Gordon[2] showed that the generalized Padé approximant procedure could be used to give rigorous, converging upper and lower bounds to the free energy of a *finite* system from the coefficients of the high temperature expansion. At the time his results did not seem useful for thermodynamic systems because the necessary coefficients diverged with the system size. Recently, however, Villani[3] and Fogli *et al.*[4] have, in another connection, shown how to usefully employ particular types of series with all divergent coefficients. By generalizing their results, and by establishing a monotonicity property of the free energy, we have been able to obtain convergent upper and lower bounds to the free energy. By differentiating these successive approximations to the free energy, we, of course, obtain convergent approximations to the various thermodynamic properties.

We remark that, based on the theorem of Lee and Yang,[5] it has previously been shown[6] that converging upper and lower bounds on the magnetization can be given for $H \neq 0$. Also Gallavotti *et al.*[7] have shown for $H = 0$ and T large enough that the free energy is analytic, and hence exactly obtainable from the power series expansions. We extend our present results to cover the whole H-T plane, $T > 0$.

We introduce our Ising model Hamiltonian in the following form:

$$\mathcal{H} = \sum_{i,j} J_{ij}(1 - \sigma_i\sigma_j) - m \sum_i \sigma_i h_i. \tag{1}$$

The partition function is then

$$Z = \sum_{\{\sigma_i = \pm 1\}} \exp(-\beta\mathcal{H}). \tag{2}$$

We restrict the interactions to be ferromagnetic, $J_{ij} \geq 0$, and will normally take $h_i = H$, the magnetic field. Under these restrictions we can prove the following monotonicity property:

$$[1/\#(2A)]\ln Z_{2A} \leq [1/\#(A)] \ln Z_A, \tag{3}$$

where A is a set of the underlying lattice sites on which the spins are situated and $2A$ are two identical nonoverlapping such sets. The function $\#(X)$ is the number of sites in set X. To prove this result, suppose we consider $A \cap B = \phi$. Then

$$\begin{aligned} Z_{A\cup B} = \sum_{\sigma_i = \pm 1} & \exp\Big(-\beta \sum_{i,j\subset A} J_{ij}(1 - \sigma_i\sigma_j) + \beta m \sum_{i\subset A} \sigma_i h_i\Big) \\ & \times \exp\Big(-\beta \sum_{\substack{i\in A \\ j\in B}} J_{ij}(1 - \sigma_i\sigma_j)\Big) \\ & \times \exp\Big(-\beta \sum_{i,j\subset B} J_{ij}(1 - \sigma_i\sigma_j) + \beta m \sum_{i\in B} \sigma_i h_i\Big), \end{aligned} \tag{4}$$

or, as $\exp[-\beta J_{ij}(1 - \sigma_i\sigma_j)] \leq 1$ for any allowed state,

$$Z_{A\cup B} \leq Z_A Z_B. \tag{5}$$

We may rewrite Eq. (5) as

$$\begin{aligned} \frac{1}{\#(A\cup B)} \ln Z_{A\cup B} \\ \leq \frac{1}{\#(A)} \ln Z_A + \frac{B}{\#(A\cup B)} \left(\frac{1}{\#(B)} \ln Z_B - \frac{1}{\#(A)} \ln Z_A\right), \end{aligned} \tag{6}$$

which leads to (3) when we let the interactions and underlying configuration of set B be identical to set A. Thus, at least for doubling of the size of the set, we have shown that the $[\ln Z_A/\#(A)]$ is monotonically decreasing with $\#(A)$. For higher dimension d we can retain the same shape by increasing the system size by a factor of 2^d per step, which is to say doubling in each direction. We remark that the fundamental inequality (3) really depends only on having an interaction of fixed sign and is susceptible to wide generalization.

We next quote our modified version of the Villani limit theorem.[3]

Theorem: Let there be a sequence of functions $f_m(X)$ with the properties

(a) $f_n(1) = 2,\ \lim_{x\to\infty} f_n(x) = 2,\ f_n(x) < 2,\ 1 < x < \infty,$

(b) $\lim_{n\to\infty} f_n(x) = f(x),\ 1 \leq x < \infty,\ \lim_{x\to\infty} f(x) = \Lambda < 2,$

(c) $f(x_1) \leq f(x_2),\ x_1 \geq x_2,$

(d) $f_n(x) \geq f(x),\ 1 \leq x < \infty.$

Then there exists an infinite sequence of minima of $f_n(x), x_n$ such that

$$\lim_{n\to\infty} x_n = \infty, \tag{7}$$

$$\lim_{n\to\infty} f_n(x_n) = \Lambda, \tag{8}$$

where by properties (c) and (d) $f_n(x_n) \geq \Lambda$ for all n.

Proof: By property (a) there must exist at least one minimum of $f_n(x)$. By properties (b) and (c) the $f_n(x)$ can be made to agree arbitrarily closely to a monotonically decreasing function $f(x)$ over any pre-assigned range $1 \leq x \leq R < \infty$. Thus by choosing n large enough we must have a minimum at least as large as R. Thus, as R can be as large as we please, we have result (7). For any $\epsilon > 0$ we can by (a) and (b) find an $R(n, \epsilon)$ such that $|f(x) - f_n(x)| < \epsilon$ for $1 \leq x \leq R$. It must be, by (d) and (c), that

$$f(R) + \epsilon \geq f_n(R) \geq f_n(x_n) \geq f(x_n) \geq \Lambda \tag{9}$$

but, as $R(n, \epsilon) \to \infty$, as $n \to \infty$, Eq. (9) implies that the left-hand side of (8) differs from the right by at most ϵ. As ϵ is arbitrary, we therefore conclude (8).

The final step in establishing such a convergent bounding sequence is to exhibit such a sequence of $f_n(x)$ for $(Z_N)^{1/N}$. To this end we, with Gordon[2] consider the Stieltjes integral

$$Z_N = \int_0^{\Gamma N} e^{-\beta E}\, d\rho_N(E), \tag{10}$$

where ΓN is an upper bound on the total energy in any state implied by (1) and $d\rho_N \geq 0$. We approximate

$$Z_N \simeq {}_N B_{n,j}(\beta) = \sum_{k=0}^{j} b_k \beta^k + \sum_{m=1}^{n} a_m e^{-\beta\sigma_m}, \tag{11}$$

where the a_m, b_k, and σ_m are determined by equating the coefficients of β^s, $0 \leq s \leq 2n + j$, in $Z_n(\beta)$ and ${}_N B_{n,j}(\beta)$. The properties of these approximants are well known.[2,6,8] Briefly

$$\begin{aligned} &(-1)^{1+j}[{}_N B_{n+1,j}(\beta) - {}_N B_{n,j}(\beta)] \geq 0, \\ &(-1)^{1+j}[{}_N B_{n,j}(\beta) - {}_N B_{n-1,j+2}(\beta)] \geq 0, \\ &{}_N B_{n,0}(\beta) \geq Z_N(\beta) \geq {}_N B_{n,-1}(\beta), \end{aligned} \tag{12}$$

where $j \geq -1$, and it is certainly sufficient that the radius of convergence [by (10)] is finite to insure

$$\lim_{n\to\infty} {}_N B_{n,j}(\beta) = Z_N(\beta), \quad 0 \leq \beta < \infty,\ j \geq -1, \tag{13}$$

Finally, all $\sigma_m > 0$. From these results we can now verify that

$$\begin{aligned} &f_n(N) = [{}_N B_{n,0}(\beta)]^{1/N}, \quad f(N) = [Z_N(\beta)]^{1/N}, \\ &\Lambda(\beta) = \lim_{N\to\infty} [Z_N(\beta)]^{1/N} \end{aligned} \tag{14}$$

satisfy all the conditions (a)–(d) of the limit theorem. The allowed values of N are restricted to 2^{nd}, as those are the only ones for which we have proved (c). However, as by the linked cluster theorem, we expect $f(N) = f_0 + f_1/N + \cdots$, we expect (c) to also hold for all N large enough. Condition (a) follows from $Z_1(\beta) = {}_1B_{n,j}(\beta) = 2$, the result that all $\sigma_m > 0$, and an inequality obtained from the direct asymptotic solution for large N, $2^N C_n(\beta) \times N^{-n(n+1)} \leq {}_N B_{n,0}(\beta) \leq 2^N$. Thus by the limit theorem and the properties of the B's we conclude that

$$\begin{aligned} &\lim_{n\to\infty} [{}_{N(n)} B_{n,0}(\beta)]^{1/N(n)} = \Lambda(\beta), \\ &[{}_{N(n)} B_{n,0}(\beta)]^{1/N(n)} \geq \Lambda(\beta). \end{aligned} \tag{15}$$

We can also deduce convergent lower bounds to $\Lambda(\beta)$ in a similar way. Let us consider an Ising model with only nearest-neighbor interaction and coordination number q. There is no difficulty in greatly relaxing this condition. Then if instead of (1) we pick

$$\hat{\mathcal{K}} = -\sum_{\substack{\text{nearest}\\ \text{neighbors}}} J(1 + \sigma_i\sigma_j) - m \sum_i \sigma_i h_i, \tag{16}$$

it is easy to show

$$\lim_{N\to\infty} [\hat{Z}_N(\beta)]^{1/N} = e^{-q\beta J}\Lambda(\beta). \tag{17}$$

But for $\hat{Z}$ the fundamental inequality (3) is reversed as $\exp[\beta J(1 + \sigma_i\sigma_j)] \geq 1$. Clearly the conclusions of the limit theorem are equally valid if the inequalities are reversed for a monotonically increasing function instead of a decreasing one. A slightly more involved argument is required as $\lim_{x\to\infty} f_n(x) \neq 2$ here, but $\lim_{x\to\infty} f_n(x) \leq f(x)$ in this case turns out to be adequate. The results (12) change as the energies are now all negative instead of positive as before. The conclusion is that every $\hat{B}$ is a lower bound. Consequently, we conclude

$$\begin{aligned} &\lim_{n\to\infty} [{}_{N(n)} \hat{B}_{n,j}(\beta)]^{1/N(n)} = \Lambda(\beta) e^{q\beta J}, \\ &[{}_{N(n)} B_{n,j}(\beta)]^{1/N(n)} \leq \Lambda(\beta) e^{q\beta J}, \end{aligned} \tag{18}$$

where $j = 0, -1$ are best for a fixed number of coefficients and $N(n)$ here is determined by maximizing.

Hence we have constructed rigorous converging upper and lower bounds for the partition function per spin or equivalently the free energy per spin. By standard thermodynamics, we can by differentiation of these sequences of approximants generate convergent approximants for the various thermodynamic variables.

As a simple illustration, we will apply these results to the linear chain in zero field. Here, denoting βJ by K, we have by direct calculation

$$\begin{aligned} Z_N &= \sum_{\sigma_i = \pm 1} \exp\left(-K \sum_{i=1}^{N-1} (1 - \sigma_i\sigma_{i+1})\right) = 2(1 + e^{-2K})^{N-1} \\ &= 2^N[1 - (N-1)K + \tfrac{1}{2}N(N-1)K^2 \\ &\quad - \tfrac{1}{6}(N-1)^2(N+2)K^3 + \cdots \end{aligned} \tag{19}$$

and also

$$\begin{aligned} \hat{Z}_N &= 2(1 + e^{2K})^{N-1} \\ &= 2^N[1 + (N-1)K + \tfrac{1}{2}N(N-1)K^2 \\ &\quad + \tfrac{1}{6}(N-1)^2(N+2)K^3 + \cdots. \end{aligned} \tag{20}$$

Equations (15) and (18) become, for ${}_N B_{1,0}(\beta)$ and ${}_N B_{2-1}(\beta)$ which use the K^2 and K^3 terms respectively,

$$\begin{aligned} &2e^{-2K}\{\tfrac{1}{2}\exp[(N - 1 - \sqrt{N-1})K] \\ &\quad + \tfrac{1}{2}\exp[(N - 1 + \sqrt{N-1})K]\}^{1/N} \\ &\leq (1 + e^{-2K}) \leq 2[(1/N) + (1 - 1/N)e^{-NK}]^{1/N}. \end{aligned} \tag{21}$$

For $K = 0$, $N = 1$ extremizes to give the exact answer as expected. For $K = 1$, the left-hand side is a maximum for N near 16 and the right-hand side a minimum for N near 4. These bounds yield

$$0.84 \leq 1.135 \leq 1.43. \tag{22}$$

In the $K = \infty$ limit, the best result from (21) arise when N also goes to 16 and 4 and are $0 \leq 1 \leq 1.38$. Since our procedure is based on the exact high temperature expansions for finite sized systems, we expect, and find in this illustration, that the method converges to a given accuracy first at high temperatures and the error bound widens monotonically as the temperature decreases.

ACKNOWLEDGMENTS

The author is happy to acknowledge several helpful conversations with Professor J. L. Lebowitz and Professor M. E. Fisher. This work was supported in part by the U.S. A.E.C. and in part by the N.S.F. through the Cornell Materials Science Center.

*On sabbatical leave at Cornell from Brookhaven, 1971-72.

[1] S. G. Brush, Rev. Mod. Phys. **39**, 883 (1967).
[2] R. G. Gordon, J. Math. Phys. (N.Y.) **9**, 655 (1968).
[3] M. Villani, *Cargèse Lectures in Physics*, edited by D. Bessis, Vol. 5 (Gordon and Breach, Paris, in press).
[4] G. L. Fogli, M. F. Pellicoro, and M. Villani, Nuovo Cimento A **6**, 79 (1971).
[5] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[6] G. A. Baker Jr., Phys. Rev. **161**, 434 (1967).
[7] G. Gallavotti, S. Miracle-Sole, and D. W. Robinson, Commun. Math. Phys. **10**, 311 (1968).
[8] *The Padé Approximant in Theoretical Physics*, edited by G. A. Baker Jr. and J. L. Gammel (Academic, New York, 1970).

A spectral representation for Coulomb wavefunctions*

J. Townsend†, G. Feldman, and T. Fulton
Department of Physics, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 29 March 1972; revised manuscript received 25 July 1972)

We obtain a simple spectral representation for the momentum space wavefunctions of the Schrödinger equation with a Coulomb potential in the form of a contour integral. Both the bound state and the scattering solutions are evaluated as residues at the poles enclosed by this contour.

1. INTRODUCTION

A consideration of the relativistic or nonrelativistic two-body problem leads naturally to integral equations in momentum space for the wavefunction describing this system. Various methods are needed for dealing with problems of this type, and one that can be of help[1,2] is that of using a spectral representation for the wavefunction. In this paper we consider the well-known nonrelativistic Coulomb problem in momentum space. We demonstrate in detail how the use of a spectral representation easily leads to a complete set of solutions to this problem.

In general, two methods have been used to obtain the momentum-space solutions to the Schrödinger equation with a Coulomb potential. The first, used by Podolsky and Pauling,[3] is to Fourier transform the solutions to the equation in coordinate space. Fock,[4] on the other hand, was able to transform the equation to a four-dimensional hyperspace in which the bound-state solutions were given by the $O(4)$ spherical harmonics.

Our own approach in this paper is to employ a spectral representation. We express the momentum space wave function as a contour integral in this spectral function space. The spectral function is determined by solving a first order linear differential equation, and the wave function is then obtained by evaluation of the contour integral. Both the momentum space bound state and scattering solutions are given as residues at the poles enclosed by this contour.

2. BOUND-STATE SOLUTIONS

The Schrödinger equation in momentum space with a Coulomb potential can be written as

$$(\mathbf{p}^2 + k^2)\varphi(\mathbf{p}) = \frac{\lambda}{\pi^2}\int \frac{d^3p'\varphi(\mathbf{p}')}{(\mathbf{p}-\mathbf{p}')^2}, \tag{1}$$

where $\lambda = m\alpha$, α is the fine structure, and $k^2 = -2mE$. Here we are assuming k^2 is positive, so that initially we are solving the bound-state problem. This equation is supplemented by the boundary conditions that $\varphi(\mathbf{p})$ be finite at the origin and that for $p \to \infty$, $\varphi(\mathbf{p})$ vanishes fast enough for $\int d^3p\,\varphi(\mathbf{p})$ to be finite. This last condition is the counterpart of the requirement that the Fourier transform of $\varphi(\mathbf{p})$ be regular at the origin in coordinate space.

Since Eq. (1) is $O(3)$ symmetric, we can express the solution as

$$\varphi_{k\ell m}(\mathbf{p}) = \Psi_{k\ell}(p^2)\,\mathcal{Y}_{\ell m}(\mathbf{p}), \tag{2}$$

where $\mathcal{Y}_{\ell m}(\mathbf{p})$ is a solid harmonic of the three components of $\mathbf{p}$, that is $\mathcal{Y}_{\ell m}(\mathbf{p}) = p^\ell Y_{\ell m}(\theta,\varphi)$. We then try to find a solution through a spectral representation of the form

$$\varphi_{k\ell m}(\mathbf{p}) = \mathcal{Y}_{\ell m}(\mathbf{p})\int_C \frac{dx\, g_\ell(x)}{(p^2+x)^{\ell+2}}, \tag{3}$$

where $g_\ell(x)$ and the contour C are to be determined so that $\varphi_{k\ell m}(\mathbf{p})$ satisfies Eq. (1). The ansatz of Eq. (3) is suggested by the approach taken by Wick[1] and Cutkosky[2] to solve the Bethe–Salpeter equation for two scalar particles with a kernel involving exchange of a massless scalar meson.

If we substitute $\varphi_{k\ell m}(\mathbf{p})$ into Eq. (1), we obtain the equation

$$\frac{\mathcal{Y}_{\ell m}(\mathbf{p})}{\ell+1}\int_C \frac{dx}{(p^2+x)^{\ell+1}}\left((k^2-x)g_\ell' + \ell g_\ell - \frac{\lambda}{x^{1/2}}\,g_\ell\right) - \frac{\mathcal{Y}_{\ell m}(\mathbf{p})\,g_\ell(x)}{(p^2+x)^{\ell+1}}\Big|_a^b = 0. \tag{4}$$

In this equation a and b are the endpoints of the contour C. In reducing Eq. (1) to Eq. (4) the left-hand side of Eq. (1) has been rewritten as

$$\begin{aligned}(p^2+k^2)\varphi_{k\ell m}(\mathbf{p}) &= \mathcal{Y}_{\ell m}(\mathbf{p})\int_C dx\, g_\ell(x)\left(\frac{1}{(p^2+x)^{\ell+1}} - \frac{(k^2-x)}{\ell+1}\,\frac{d}{dx}\,\frac{1}{(p^2+x)^{\ell+1}}\right)\\ &= \mathcal{Y}_{\ell m}(\mathbf{p})\left[\int_C \frac{dx}{(p^2+x)^{\ell+1}}\left(g_\ell - \frac{1}{\ell+1}\,g_\ell + \frac{1}{\ell+1}(k^2-x)g_\ell'\right) - \frac{(k^2-x)g_\ell(x)}{(\ell+1)(p^2+x)^{\ell+1}}\Big|_a^b\right],\end{aligned} \tag{5}$$

where the last step follows after an integration by parts. The integration over $\mathbf{p}'$ in the right-hand side of Eq. (1) is done using the parametrization method of Feynman:

$$\begin{aligned}\frac{\lambda}{\pi^2}\int d^3p'\,\frac{\varphi_{k\ell m}(\mathbf{p}')}{(\mathbf{p}-\mathbf{p}')^2} &= \frac{\lambda}{\pi^2}\int_C dx\, g_\ell(x)\int d^3q\\ &\quad\times\int_0^1 \frac{du(\ell+2)(1-u)^{\ell+1}\mathcal{Y}_{\ell m}(\mathbf{q}+u\mathbf{p})}{[q^2+(1-u)(x+up^2)]^{\ell+3}}\\ &= \frac{\lambda}{2^{2\ell+1}}\,\frac{(2\ell+1)!}{(\ell+1)!}\,\mathcal{Y}_{\ell m}(\mathbf{p})\int_C dx\, g_\ell(x)\\ &\quad\times\int_0^1 \frac{du\, u^\ell}{(1-u)^{1/2}(x+up^2)^{3/2+\ell}}\\ &= \frac{\lambda}{\ell+1}\,\mathcal{Y}_{\ell m}(\mathbf{p})\int_C \frac{dx\, g_\ell(x)}{x^{1/2}(x+p^2)^{\ell+1}}.\end{aligned} \tag{6}$$

This result is valid for x restricted to the cut complex x plane with the cut running from 0 to $-\infty$ along the negative real axis. If the contour C is chosen in this region so that there is no contribution from the integrated term in Eq. (4), then $g\,(x)$ must satisfy the equation

$$(k^2-x)g_\ell' + \ell g_\ell - \lambda x^{-1/2}g_\ell = 0. \qquad (7)$$

The solution to this equation is given by

$$g_\ell(x) = A_{k\ell}(k^2-x)^\ell\left(\frac{k+x^{1/2}}{k-x^{1/2}}\right)^{\lambda/k}, \qquad (8)$$

where the constant $A_{k\ell}$ is determined by normalizing the solution.

In order to simplify the analysis, we make the change of variable $x^{1/2}=y$. Then

$$\varphi_{k\ell m}(\mathbf{p}) = A_{k\ell}\,\mathcal{Y}_{\ell m}(\mathbf{p})\int_C \frac{dy\, y(k^2-y^2)^\ell}{(p^2+y^2)^{\ell+2}}\left(\frac{k+y}{k-y}\right)^{\lambda/k}. \qquad (9)$$

The contour C is now restricted in the complex y plane to the region $\mathrm{Re}y>0$ or $\mathrm{Re}y<0$. The contour must not, of course, cross the $\mathrm{Re}y=0$ axis where the factor $(p^2+y^2)^{-(\ell+2)}$ is singular. In these two regions the singularities of the integrand in Eq. (9) are determined by the value of λ/k. For the case of λ/k not equal to an integer, the integrand has branch points at $y=k$ and $y=-k$. There are, therefore, two contours C_1 and C_2 with corresponding solutions $\varphi^{(1)}_{k\ell m}$ and $\varphi^{(2)}_{k\ell m}$ that can be chosen so that there is no contribution from the integrated term of Eq. (4). These two contours are shown in Fig. 1. For $p\to\infty$, these solutions do not approach zero fast enough for $\int d^3p\,\varphi_{k\ell m}(\mathbf{p})$ to be finite. In fact for large values of p

$$\varphi^{(i)}_{k00}(\mathbf{p}) \to \frac{iA_{k0}}{p^2}\,\mathcal{Y}_{00}\sin\left(\frac{\lambda\pi}{k}\right), \quad p\to\infty, \quad i=1,2, \qquad (10)$$

and for $\ell\neq 0$

$$\varphi^{(i)}_{k\ell m}(\mathbf{p}) > O(p^{\ell-4}), \quad p\to\infty, \quad i=1,2. \qquad (11)$$

A proof of these statements is given in the Appendix.

Since both $\varphi^{(1)}_{k00}$ and $\varphi^{(2)}_{k00}$ have precisely the same behavior for large p, it is natural to try $\varphi^{(1)}_{k00}-\varphi^{(2)}_{k00}$ as a possible solution which may behave better for $p\to\infty$. The integrand of Eq. (9) vanishes fast enough as $y\to\infty$ so that for any fixed p the contour C_1-C_2 may be closed at infinity and the function $\varphi^{(1)}_{k00}-\varphi^{(2)}_{k00}$ evaluated as $2\pi i$ times the sum of the residues at the poles at $y=\pm ip$ which are enclosed by the contour. In order to evaluate the residues we choose the phases of $y-k$ and $y+k$ so that $\arg(y-k)=\arg(y+k)=0$ just above the right hand cut. With this choice, at the location of the poles

$$\arg(y+k)=\pi-\arg(y-k), \quad y=\pm ip. \qquad (12)$$

Thus for the contour C_1-C_2, we obtain

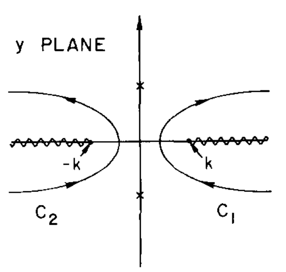


FIG. 1. Contours C_1 and C_2 for the bound-state solutions given by Eq. (9) for the case of λ/k not equal to an integer. The x's indicate the position of poles at $y=\pm ip$.

$$\varphi^{(1)}_{k00}-\varphi^{(2)}_{k00} = 2\pi i A_{k0}\mathcal{Y}_{00}\left(\left\{\frac{d}{dy}\left[\frac{y}{(y+ip)^2}\left(\frac{k+y}{k-y}\right)^{\lambda/k}\right]\right\}_{y=ip} + \left\{\frac{d}{dy}\left[\frac{y}{(y-ip)^2}\left(\frac{k+y}{k-y}\right)^{\lambda/k}\right]\right\}_{y=-ip}\right), \qquad (13)$$

where

$$\frac{(y+k)}{(y-k)} = e^{-i\pi}e^{2i\arg(y+k)} \quad \text{at} \quad y=\pm ip.$$

As $p\to\infty$, $\arg(y+k)|_{y=\pm ip}\to\pm\frac{1}{2}\pi$. Combining this result with Eq. (13), we obtain the result

$$\varphi^{(1)}_{k00}-\varphi^{(2)}_{k00} \xrightarrow[p\to\infty]{} \frac{\pi A_{k0}\mathcal{Y}_{00}}{p}\,\frac{e^{-i\pi\lambda/k}}{(p^2+k^2)}\,(1-e^{-i2\pi\lambda/k}). \qquad (14)$$

Thus, unless λ/k is an integer, $\varphi^{(1)}_{k00}-\varphi^{(2)}_{k00}$ behaves as p^{-3} for $p\to\infty$ and $\int d^3p\,\varphi(\mathbf{p})$ is infinite. These solutions therefore fail to satisfy the boundary conditions. This argument can be extended to the higher angular momentum states to show that for λ/k not an integer the phase of the residue at the pole $y=ip$ will be different from that at $y=-ip$ and, therefore, there can be no cancellation of the leading p behavior.

The only possibility remaining for a bound-state solution is $\lambda/k=n$, where n is an integer. Rather than continue to evaluate $\varphi^{(1)}_{n\ell m}-\varphi^{(2)}_{n\ell m}$ as the sum of residues at the poles $y=\pm ip$, it is easier to note that for $\lambda/k=n$, the integrand of Eq. (9) is analytic for $\mathrm{Re}y<0$ and, thus, $\varphi^{(2)}_{n\ell m}$ reduces to zero. Moreover, if $n\le\ell$, the integrand is also analytic for $\mathrm{Re}y>0$ and, thus, $\varphi^{(1)}_{n\ell m}$ is also zero, and there are no solutions. For $n=\ell+1, \ell+2, \ldots$, the integrand has a pole of order $n-\ell=1,2,\cdots$ at $y=k$. The contour C_1 can be deformed into one enclosing this pole. The bound state wavefunctions are given as the residues at this pole:

$$\varphi^{(1)}_{n\ell m}(\mathbf{p}) = \frac{2\pi i A_{n\ell}\,\mathcal{Y}_{\ell m}(\mathbf{p})}{(n-\ell-1)!}\left(\frac{d^{n-\ell-1}}{dy^{n-\ell-1}}\,\frac{y(k+y)^{n+\ell}}{(p^2+y^2)^{\ell+2}}\right)_{y=k} \qquad (15)$$

where the constant $A_{n\ell}$ normalizes $\varphi^{(1)}_{n\ell m}(\mathbf{p})$ to unity and $k=(-2mE)^{1/2}=\lambda/n$. Using the spectral representation, this normalization constant is determined to be

$$A_{n\ell} = \frac{-i\,2^{\ell+1}}{(2\pi)^{3/2}}\left(\frac{(n-\ell-1)!(2\ell+2)(\ell+1)!}{n(n+\ell)!(2k)^{2\ell-1}}\right)^{1/2}. \qquad (16)$$

3. SCATTERING SOLUTIONS

The spectral representation used to determine the bound state solutions also determines the scattering solutions if we let $k^2\to-\kappa^2$, where $\kappa^2=2mE$ and $E>0$. As an example we will discuss the $\ell=0$ partial wave solutions in some detail.

We begin by considering the outgoing wave solutions by assigning κ a small positive imaginary part $i\epsilon$. Letting $k\to i(\kappa+i\epsilon)$ in Eq. (9), we obtain the solution

$$\varphi_{\kappa 00}(\mathbf{p}) = \lim_{\epsilon\to 0} A_{\kappa 0}\,\mathcal{Y}_{00}(\mathbf{p})\int_C \frac{dy\,y}{(p^2+y^2)^2}\left(\frac{y+i(\kappa+i\epsilon)}{y-i(\kappa+i\epsilon)}\right)^{\lambda/i\kappa}, \qquad (17)$$

where there are two possible choices for the contour C. The integrand of Eq. (17) now has branch points at $y=-i\kappa+\epsilon$ and $y=i\kappa-\epsilon$. Thus for the contours C_3 and C_4 shown in Fig. 2(a) there are the corresponding solutions $\varphi^{(3)}_{\kappa 00}$ and $\varphi^{(4)}_{\kappa 00}$. Each of these solutions behaves as p^{-2} for $p\to\infty$. The proof is the same as for the bound-state solutions $\varphi^{(1)}_{k00}$ and $\varphi^{(2)}_{k00}$ and is presented in the Appendix. The sum of these solutions is also a solution which we denote by $\varphi^{(+)}_{\kappa 00}$. It can be evaluated as the sum of the residues of the poles at $y = \pm ip$ enclosed by the contour $C_3 + C_4$. In order to do this it is again necessary to specify which Riemann sheet we are on. We do this by choosing $\arg(y - i\kappa) \to 0$ and $\arg(y + i\kappa) \to 0$ as $y \to \infty$ along the positive real axis for all scattering states. With this choice we have the following results at the poles:

$$
\begin{aligned}
p > \kappa \quad &\begin{cases} y = ip & \arg(y - i\kappa) = \tfrac{1}{2}\pi \quad \arg(y + i\kappa) = \tfrac{1}{2}\pi, \\ y = -ip & \arg(y - i\kappa) = -\tfrac{1}{2}\pi \quad \arg(y + i\kappa) = \tfrac{3}{2}\pi, \end{cases} \\
p < \kappa \quad & \;\; y = \pm ip \quad \arg(y - i\kappa) = -\tfrac{1}{2}\pi \quad \arg(y + i\kappa) = \tfrac{1}{2}\pi.
\end{aligned}
\tag{18}
$$

With the aid of Eqs. (17) and (18) we can evaluate the outgoing solution $\varphi^{(+)}_{\kappa 00}$:

$$
\begin{aligned}
\varphi^{(+)}_{\kappa 00}(\mathbf{p}) &= \frac{A_{\kappa 0}\mathcal{Y}_{00}}{p}\,\frac{\lambda\pi}{(p^2 - \kappa^2)}\left[\left(\frac{p + \kappa}{p - \kappa}\right)^{\lambda/i\kappa} - e^{2\pi\lambda/\kappa}\left(\frac{p - \kappa}{p + \kappa}\right)^{\lambda/i\kappa}\right], \quad p > \kappa, \\
\varphi^{(+)}_{\kappa 00}(\mathbf{p}) &= \frac{A_{\kappa 0}\mathcal{Y}_{00}\lambda\pi}{p(p^2 - \kappa^2)}\, e^{\pi\lambda/\kappa}\left[\left(\frac{\kappa + p}{\kappa - p}\right)^{\lambda/i\kappa} - \left(\frac{\kappa - p}{\kappa + p}\right)^{\lambda/i\kappa}\right], \quad p < \kappa.
\end{aligned}
\tag{19}
$$

Thus for $p \to \infty$,

$$
\varphi^{(+)}_{\kappa 00}(\mathbf{p}) \to \frac{\lambda\pi A_{\kappa 0}\mathcal{Y}_{00}}{p^3}\,(1 - e^{2\pi\lambda/\kappa}), \tag{20}
$$

and $\varphi^{(+)}_{\kappa 00}$ is not regular at the origin in coordinate space.

For the incoming wave solution, when κ has a small negative imaginary part $- i\epsilon$, the analysis proceeds in the same manner as for the outgoing wave. The cut structure and the possible contours are shown in Fig. 2(b). We again consider the solution $\varphi^{(-)}_{\kappa 00} = \varphi^{(5)}_{\kappa 00} + \varphi^{(6)}_{\kappa 00}$, where the contour over which $\varphi^{(-)}_{\kappa 00}$ is evaluated is $C_5 + C_6$. For this case the cuts have been displaced relative to their positions for the outgoing solution, and thus the values of the phases at the poles $y = \pm ip$ enclosed by this contour are changed from the preceding results. From Fig. 2(b) we see that the phases are

$$
\begin{aligned}
p > \kappa \quad &\begin{cases} y = ip & \arg(y - i\kappa) = -\tfrac{3}{2}\pi \quad \arg(y + i\kappa) = \tfrac{1}{2}\pi, \\ y = -ip & \arg(y - i\kappa) = -\tfrac{1}{2}\pi \quad \arg(y + i\kappa) = -\tfrac{1}{2}\pi, \end{cases} \\
p < \kappa \quad & \;\; y = \pm ip \quad \arg(y - i\kappa) = -\tfrac{1}{2}\pi \quad \arg(y + i\kappa) = \tfrac{1}{2}\pi.
\end{aligned}
\tag{21}
$$

The incoming partial wave $\varphi^{(-)}_{\kappa 00}(\mathbf{p})$ is

$$
\begin{aligned}
\varphi^{(-)}_{\kappa 00}(\mathbf{p}) &= \frac{\lambda\pi A_{\kappa 0}\mathcal{Y}_{00}}{p}\,\frac{1}{(p^2 - \kappa^2)}\left[e^{2\pi\lambda/\kappa}\left(\frac{p + \kappa}{p - \kappa}\right)^{\lambda/i\kappa} - \left(\frac{p - \kappa}{p + \kappa}\right)^{\lambda/i\kappa}\right], \quad p > \kappa, \\
\varphi^{(-)}_{\kappa 00}(\mathbf{p}) &= \frac{\lambda\pi A_{\kappa 0}\mathcal{Y}_{00}}{p(p^2 - \kappa^2)}\, e^{\pi\lambda/\kappa}\left[\left(\frac{\kappa + p}{\kappa - p}\right)^{\lambda/i\kappa} - \left(\frac{\kappa - p}{\kappa + p}\right)^{\lambda/i\kappa}\right], \quad p < \kappa.
\end{aligned}
\tag{22}
$$

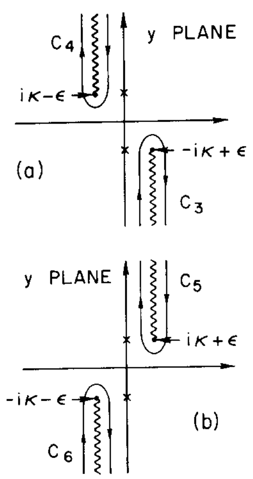


FIG. 2. (a) Contours C_3 and C_4 for the outgoing partial wave scattering solutions given by Eq. (17) when κ has a positive infinitesimal imaginary part $i\epsilon$. (b) Contours C_5 and C_6 for the incoming partial wave scattering solutions when κ has a negative infinitesimal imaginary part.

The asymptotic nature of this solution is given by

$$
\varphi^{(-)}_{\kappa 00}(\mathbf{p}) \to \frac{\lambda\pi A_{\kappa 0}\mathcal{Y}_{00}}{p^3}\,(e^{2\pi\lambda/\kappa} - 1), \qquad p \to \infty. \tag{23}
$$

Although neither $\varphi^{(+)}_{\kappa 00}$ nor $\varphi^{(-)}_{\kappa 00}$ vanishes fast enough as $p \to \infty$ for $\int d^3p\varphi$ to be finite, we see that it is possible to obtain such a solution by considering the sum $\varphi^{(+)}_{\kappa 00} + \varphi^{(-)}_{\kappa 00}$, provided that the normalization constants for both are chosen to be the same. This solution is the regular $\ell = 0$ Coulomb partial-wave solution.

Once we have obtained the regular partial wave solutions, the last point to be examined is the nature of the singularity at $p = \kappa$. Using the phase conventions already established, we can evaluate the various solutions at this point. For the outgoing wave solution, for example, we obtain the result

$$
\varphi^{(+)}_{\kappa 00}(\mathbf{p}) = \lim_{\epsilon \to 0} \frac{A_{\kappa 0}\mathcal{Y}_{00}\lambda\pi}{p[p^2 - (\kappa + i\epsilon)^2]}\left[\left(\frac{2\kappa e^{i\pi/2}}{[(p - \kappa)^2 + \epsilon^2]^{1/2}}\right)^{\lambda/i\kappa} - \left(\frac{[(p - \kappa)^2 + \epsilon^2]^{1/2} e^{3i\pi/2}}{2\kappa}\right)^{\lambda/i\kappa}\right], \quad p = \kappa. \tag{24}
$$

We may also observe that Eq. (17) with the contours C_3 and C_4 and the corresponding equation when κ has a negative infinitesimal part with contours C_5 and C_6 are, in fact, the Fourier transforms of the so-called irregular Coulomb partial wave solutions in coordinate space.[5]

The arguments we have given can, in principle, be extended to cover all the partial wave solutions. The general solution for arbitrary ℓ is given by

$$
\varphi_{\kappa\ell m}(\mathbf{p}) = \frac{2\pi i}{(\ell + 1)!}\, A_{\kappa\ell}\, \mathcal{Y}_{\ell m}(\mathbf{p})\left(\left\{\frac{d^{\ell+1}}{dy^{\ell+1}}\left[\frac{y(\kappa^2 + y^2)^{\ell}}{(y + ip)^{\ell+2}}\left(\frac{y + i\kappa}{y - i\kappa}\right)^{\lambda/i\kappa}\right]\right\}_{y = ip} + \left\{\frac{d^{\ell+1}}{dy^{\ell+1}}\left[\frac{y(\kappa^2 + y^2)^{\ell}}{(y - ip)^{\ell+2}}\left(\frac{y + i\kappa}{y - i\kappa}\right)^{\lambda/i\kappa}\right]\right\}_{y = -ip}\right), \tag{25}
$$

where in evaluating the residues, the phases of Eqs. (18) and (21) are to be used for outgoing and incoming partial wave solutions, respectively, and their sum is the regular standing wave solution.

4. CONCLUSION

We can look at our spectral representation approach to the solution of the Schrödinger equation from two points of view. On the one hand, the specific results we have obtained, because of their simplicity, may be useful in the evaluation of matrix elements of various momentum-space operators, or to study the general characteristics such as analyticity properties or high momentum limits of such matrix elements. On the other hand, the spectral representation technique may in itself be of some interest. Instead of having to solve an integral equation, as in momentum space, or a second-order partial differential equation as in coordinate space, we deal with only a first-order differential equation in the single spectral function variable. Thus, the solution becomes trivial and symmetries present in the problem are used from the start. Our treatment serves as another instance of the simplifications achieved in the spectral function approach first introduced by Wick and leads to the conjecture that it may have a wider applicability to other eigenvalue problems.

APPENDIX

We show here that the solutions given by Eq. (9) for C equal to C_1 or C_2 and λ/k not equal to an integer do not vanish fast enough as $p \to \infty$ for $\int d^3p\,\varphi_{k\ell m}(\mathbf{p})$ to be finite. First we consider the $\ell = 0$ state,

$$\varphi^{(j)}_{k00} = A_{k0}\,\mathcal{Y}_{00}\int_{C_j} \frac{dy\,y}{(p^2+y^2)^2}\left(\frac{k+y}{k-y}\right)^{\lambda/k}, \qquad j = 1, 2. \tag{A1}$$

Since

$$\int_{C_j} dy\,y\left(\frac{k+y}{k-y}\right)^{\lambda/k}$$

is not finite, $\varphi^{(j)}_{k00}(\mathbf{p})$ approaches zero more slowly than p^{-4} for large p. After an integration by parts

$$\varphi^{(j)}_{k00}(\mathbf{p}) = \lambda A_{k0}\,\mathcal{Y}_{00}\int_{C_j} dy\,(p^2+y^2)^{-1} \times (k^2-y^2)^{-1}\left(\frac{k+y}{k-y}\right)^{\lambda/k}. \tag{A2}$$

As $p \to \infty$

$$\varphi^{(j)}_{k00} \to \lambda A_{k0}\,\mathcal{Y}_{00}\,J^{(j)}/p^2, \tag{A3}$$

where

$$J^{(j)} = \int_{C_j} dy\,\frac{1}{k^2-y^2}\left(\frac{k+y}{k-y}\right)^{\lambda/k}, \tag{A4}$$

provided $J^{(j)} \neq 0$. The integral $J^{(j)}$ can be evaluated by deforming the contours C_1 and C_2 of Fig. 1 into ones running along the $\mathrm{Re}\,y = 0$ axis. The result is

$$J^{(j)} = i\lambda^{-1}\sin(\pi\lambda/k),$$

which is not zero unless λ/k is an integer.

For the $\ell \neq 0$ states, Eq. (9) becomes after ℓ integrations by parts,

$$\varphi^{(j)}_{k\ell m}(\mathbf{p}) = A_{k\ell}\,\mathcal{Y}_{\ell m}(\mathbf{p})\int_{C_j}\frac{dy\,y}{(p^2+y^2)^2}\,f_\ell(y), \tag{A5}$$

where $f_\ell(y)$ is finite everywhere along the contours C_1 and C_2 and

$$f_\ell(y) \sim O\left[\left(\frac{k-y}{k+y}\right)^{\lambda/k}\right], \qquad y \to \infty. \tag{A6}$$

Thus since $\mathcal{Y}_{\ell m}(\mathbf{p}) \xrightarrow[p\to\infty]{} O(p^\ell)$

$$\varphi^{(j)}_{k\ell m}(\mathbf{p}) > O(p^{\ell-4}), \qquad p \to \infty, \qquad \ell \neq 0. \tag{A7}$$

The above results are still valid if $k \to i(\kappa \pm i\epsilon)$ for the scattering states. Thus $\varphi^{(j)}_{\kappa\ell m}$, $j = 3, 4, 5, 6$, also fail to satisfy the boundary conditions. In particular,

$$|\varphi^{(j)}_{\kappa 00}(\mathbf{p})| \xrightarrow[p\to\infty]{} \frac{\mathcal{Y}_{00}}{p^2}\left|A_{\kappa 0}\sinh\left(\frac{\pi\lambda}{\kappa}\right)\right|, \tag{A8}$$

which is nonzero for all values of λ/κ.

---

*Supported in part by the National Science Foundation.
†N. S. F. Predoctoral Fellow.

[1] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[2] R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954).
[3] B. Podolsky and L. Pauling, Phys. Rev. **34**, 109 (1929).
[4] V. Fock, Z. Phys. **98**, 145 (1935); M. Levy, Proc. R. Soc. A **204**, 145 (1950).
[5] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1958), Vol. 1, p. 425, and Appendix B.

Erroneous bound state conditions from an algebraic misrepresentation of spin wave theory

S. L. Trubatch

California State University, Long Beach, Long Beach, California 90840

(Received 24 April 1972; Revised Manuscript Received 21 July 1972)

A recently derived condition [C. J. Liu and Yutze Chow, J. Math. Phys. 12, 2144 (1971)] for the existence of bound states in the Heisenberg ferromagnet is shown to be the erroneous result of an improper construction of the Hilbert space for an algebraic representation of the spin operators.

Recently,[1] the Heisenberg model of ferromagnetism has been reinvestigated in terms of a different algebraic representation for the spin operators. The resulting new condition for the existence of bound state contradicts a previous expression obtained in an earlier calculation[2] based on Green's function techniques. In this note it is demonstrated that the new bound state condition is the incorrect result of an improper construction of the Hilbert Space for the algebraic representation.

Heisenberg's model of ferromagnetism is described by the Hamiltonian

$$H = -\sum_{l,m} J_{lm} S_l \cdot S_m - B \sum_i S_l^z,$$

where S_l is the spin operator for the atom at the lth site, J_{lm} is the exchange integral between atoms at the lth and mth sites, and B is the externally applied magnetic field. The spin operators satisfy the usual commutation rules

$$[S_l^+, S_m^-] = 2\,\delta_{lm} S_l^z, \qquad [S_l^z, S_m^\pm] = \pm\delta_{lm} S_l^\pm,$$

where

$$S_l^\pm \equiv S_l^x \pm i S_l^y$$

and S_l^y, S_l^z are the cartesian components of the spin operator at site "l".

It is well known[3] that these spin operators can be realized by an algebraic representation in terms of two sets of commuting Bose operators. In the notation of Ref. 1, it is

$$S_l^+ = \beta_l^+ b_l, \qquad S_l^- = b_l^+ \beta_l,$$

$$S_l^z = \tfrac{1}{2}[\beta_l^+ \beta_l - b_l^+ b_l];$$

where the Bose operators satisfy the communtation rules

$$[\beta_l, \beta_m^+] = \delta_{lm}, \qquad [b_l, b_m^+] = \delta_{lm},$$

$$[\beta_l, \beta_m] = 0, \qquad [b_l, b_m] = 0,$$

$$[\beta_l, b_m^+] = 0, \qquad [b_l, \beta_m] = 0.$$

Although the algebraic representation reproduces all the spin commutation rules, it does not preserve the "kinematic constaint"[4]

$$(S_l^+)^{2S+1} = 0,$$

where S is the magnitude of the spin at the site "l".

Of course, this constraint is entirely equivalent to the "auxiliary condition" [Eq. (3.4) of Ref. 1]

$$\beta^+\beta + b^+ b = 2S.$$

Thus, in order to avoid the introduction of spurious "kinematic interactions,"[4] the Hilbert Space in which the operators act must be suitably limited. In other dynamical applications of this representation,[5] it has been shown by comparison with more usual techniques that the correct dynamics is preserved if and only if the Hilbert Space is restricted to vectors of the form

$$|P_1 \cdots P_N\rangle = \sum_{q_1 \cdots q_N} C(P_1 \cdots P_N | q_1 \cdots q_N) \prod_{i=1}^N (\beta_{q_i}^+)^{S+M_i} (b_{q_i}^+)^{S-M_i} |0\rangle.$$

In Ref. 1, Liu and Chow consider states of the form

$$|P_1 \cdots P_N\rangle = \sum_{n=0}^{2SN} \sum_{\{q\}} \frac{C^{(n)}(\{P\}|\{q\})}{(m!)^{1/2}(n!)^{1/2}} \frac{\sum e^{q \cdot l}}{N^{n/2} N^{m/2}} \times \Big(\sum_m \beta_m^+ - N\sqrt{2S}\Big)^n \prod_{i=1}^m b_{l_i}^+ |0\rangle,$$

where $q \cdot l = \sum_{i=1}^N q_i l_i$. These states are not clearly of the form discussed above. They do not satisfy the constaints imposed by employing a Bose representation for the spin operators.

As a result, these states introduce kinematic interactions. Any effective Hamiltonian derived in the space of these states will have a spectrum riddled with spurious eigenvalues. By extending the Hilbert space to include inadmissable vectors, all control over the dynamics of the problem, as described in this representation, have been lost.

The correct two deviate state for the discussion of bound states is given by

$$|2\rangle = \sum_p f_p \beta_p^{+2} b_p^{+2S-2} \prod_{q \neq p} b_q^{+2S} |0\rangle + \sum_{p \neq q}\sum f_{pq} \beta_p^+ \beta_q^+ b_p^{+2S-1} b_q^{+2S-1} \prod_{\substack{r \neq p \\ r \neq q}} b_r^{+2S} |0\rangle,$$

A straight forward solution of the eigenvalue problem using these states[6] leads to the usual bound-state conditions.[2]

Finally, the apparent reproduction by the authors of Ref. one of the correct bound state condition for the one-dimensional case cannot be taken seriously unless the correct binding energy M can be deduced by their method.

[1] C. J. Liu and Yutze Chow, J. Math. Phys. (N.Y.) **12**, 2144 (1971).
[2] M. Wortis, Phys. Rev. **132**, 85 (1963).
[3] J. Schwinger, *Quantum theory of angular momentum, perspectives in physics,* edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1956).
[4] F. J. Dyson, Phys. Rev. **102**, 1217, 1230 (1956).
[5] J. M. Robinson and S. L. Trubatch, Am. J. Phys. **39**, 1190 (1971).
[6] D. C. Mattis, *The theory of magnetism* (Harper and Row, New York, 1965).

The nonstandard $\lambda{:}\phi_2^4(x){:}$ model. I. The technique of nonstandard analysis in theoretical physics*

Peter J. Kelemen

Courant Institute of Mathematical Sciences and Department of Mathematics, University College, New York University, New York

Abraham Robinson

Department of Mathematics, Yale University, New Haven, Connecticut

(Received 22 May 1972; revised manuscript received 21 July 1972)

The methods of nonstandard analysis are demonstrated as a preliminary step for the construction of the nonstandard $\lambda{:}\phi_2^4{:}$ model. Elementary quantum mechanical problems are solved and the renormalization of the scalar field (Yukawa interaction) is investigated.

1. INTRODUCTION

In the applications of analysis, one often speaks of infinitesimal increments and of infinitesimal volume elements. Since Weierstrass, the above phrases were understood to be shorthand formulations of more complicated expressions involving limits. However, some time ago one of the authors[1,2] showed that expressions involving infinitesimals can be taken literally if one refers them to a suitable number system that contains infinitely large and infinitely small elements. The subject that arose out of this realization has come to be known as nonstandard analysis (n.s.a.). It is described informally in Sec. 2 below.

We feel that n.s.a. can be used advantageously in physics. Thus, many calculations can be simplified by its use through the avoidance of passages to the limit at certain stages. Also, using infinitely large numbers one can give a rigorous meaning to self-energies and renormalization. Finally, one may treat certain nonseparable Hilbert spaces with the same ease as separable ones.

Using n.s.a., we may retain results calculated by standard techniques whenever desirable, and, moreover, we may reinterpret them in the nonstandard system. On the other hand, the method is conservative; that is to say, any final result that has been obtained by nonstandard techniques but is itself formulated in standard terms might have been obtained by standard methods, though perhaps at the cost of a considerable effort.

Here we shall consider cases in which the basic assumptions of a problem were formulated originally in standard language and are then translated into the language of n.s.a. and solved by its methods. Several examples of this technique are given in Secs. 3 and 4 below. If we were to include an assumption that can be formulated only in nonstandard terms, then the result might not be amenable to a standard formulation and would have to be interpreted directly. We intend to develop this approach in a future paper.

2. NONSTANDARD ANALYSIS

Let R be the system of real numbers. An *ordered field* is an ordered number system which shares with R all the usual properties involving the operations of addition, subtraction, multiplication, and division. It has been known for a long time that there exist ordered fields which are extensions of R. Such fields are *non-archimedean*, i.e., they contain positive numbers which are greater than any natural number (in R), while their reciprocals are smaller than any positive element of R. However, generally speaking, one cannot extend most of the familiar functions of analysis [e.g., e^x, $\sin x$, $\ln x$, $J_n(x)$] to these new systems so as to preserve their usual properties. Nonstandard analysis is based on the existence of *particular* ordered fields which include R as a subfield and for which the extended functions in question are, in fact, available and, moreover, are provided automatically by certain model theoretic procedures. One of these is the so-called *ultrapower construction*[3] which, for the case of a countable index set, runs as follows. Let N be the set of natural numbers and let F be a free ultrafilter[4] on N. Then the field in question, $^*R$, is defined as the system of all sequences of real numbers R^N, where two sequences are regarded as equal if they coincide on a set of natural numbers which belongs to F. Functions on $^*R$ are defined termwise on representative sequences selected from the equivalence classes just defined, and a relation is said to hold between sequences if it holds termwise for an index which belongs to the ultrafilter. R can be embedded in $^*R$ by identifying any real $r \in R$ with the (equivalence class of the) sequence $(r, r, r, \cdots)$ in $^*R$.

While the ultrapower construction sketched above provides a relatively concrete realization of the type of structure required for nonstandard analysis, the class of these structures is quite large and contains fields which cannot be obtained in this way. They can be characterized in the following way.

A *nonstandard model of analysis* is a proper extension $^*R$ of the system of real numbers R, such that

Transfer Theorem: Any true assertion X about R is still valid in $^*R$, provided we reinterpret X in $^*R$ as follows:

For every class of objects in R (e.g., functions of one numerical variable, relations between numbers, relations between functions, functionals, i.e., mappings from functions to numbers), there exists a subclass, said to be the class of the corresponding *internal* objects. In particular, the class of internal entities of $^*R$, contains an element corresponding to each object in R. For example, the class $^*\Phi$ of internal functions of one numerical variable in $^*R$ contains, in particular, extensions of all functions on R; but these functions do not exhaust $^*\Phi$. However, all individuals (numbers) of $^*R$ are regarded as internal. Then the assertion X, supposed true in R, is still valid in $^*R$, provided we reinterpret each quantifier in X (e.g., "there exists a function," "for all relations") as referring only to the corresponding *internal* objects in $^*R$ ("there exists an *internal* function," "for all *internal* relations").

See Ref. 2. for a more rigorous formulation of these notions. An object which is not internal is said to be *external*. The following examples are instructive.

(1) Let N be the set of natural numbers. Then $N \subset R$. The corresponding *internal* subset $*N$ of $*R$ contains N. The elements of $*N$-N are called the *infinite* natural numbers. Any $a \in *N$-N is greater than all elements of N. If $*R$ has been obtained by the ultrapower construction sketched above, then an example of an infinite natural number is given by the sequence $(0,1,2,3,\cdots)$. The axiom of induction is satisfied in $*N$ if *set* (property) is interpreted as *internal set* (property). Accordingly, every nonempty internal subset of N possesses a smallest element. The set $*N$-N does not have a smallest element and, accordingly, is *external*.

(2) If $a \in *R$ is numerically smaller than any positive $r \in *R$, then a is said to be infinitely small or *infinitesimal*. The set of infinitely small numbers in $*R$ is said to be the *monad of zero*. More generally, for any $r \in R$ the set of all numbers $a \in *R$ which differ from r only by an infinitesimal amount is said to be the *monad* of r, $\mu(r)$. If $a \in \mu(r)$, then we write $r = {}^0a$ and we call r the *standard part* of a. All monads are external. The numbers of R ("standard numbers") are isolated points in the interval topology of $*R$.

(3) If a and b are any numbers in $*R$, finite or infinite, then the interval $a < x < a + b$ is internal. The interval of all $a \in *R$, a positive infinite, is external.

(4) The extensions[5] of x^2 and e^x to $*R$, $*(x^2)$ and $*(e^x)$, are internal and even standard. (We may omit the star on the extended functions, by convention.) The function

$$f(x) = \begin{cases} 0 & \text{for } x \text{ infinite} \\ 1 & \text{for } x \text{ finite} \end{cases}$$

is external. Among the functions which are internal but not standard are various representations of the Dirac delta function, e.g., $\delta(x) = (\pi\eta x)^{-1} \sin\eta^2 x$, where η is an infinite natural number

or

$$\delta(x) = \begin{cases} \eta & \text{for } -\frac{1}{2}\eta^{-1} \le x \le \frac{1}{2}\eta^{-1} \\ 0 & \text{otherwise} \end{cases}.$$

The first of these representations is analytic in $*R$; the second is not. In either case we have[6]

$${}^0\left[\int_{*R} \delta(x) f(x) dx\right] = f(0)$$

for any $f(x)$, which extends a bounded function which is defined and continuous in the neighborhood of 0 in R. The validity of this equation is, in fact, a condition which has to be imposed on any reasonable interpretation of the delta function. On the other hand, $\int_{*R} \delta^2(x) dx$ depends on our particular choice of the representations.

(5) Let $f(x)$ be a real function in R, defined for an interval $a < x < b$, and let x_0 be a point in that interval. $f(x)$ possesses an automatic extension to $*R$, as stated. It can then be shown that $f(x_0 + \eta)$ is infinitely close to $f(x_0)$, in symbols $f(x_0 + \eta) \simeq f(x_0)$, for all infinitesimal η if and only if $f(x)$ is continuous at x_0 in the classical (Weierstrass) sense, i.e., if and only if the following condition is satisfied:

(C) For every positive ϵ (in R), there exists a positive δ (in R) such that $|f(x_0 + h) - f(x_0)| < \epsilon$ provided $h < \delta$.

Now let $f(x)$ be an *internal* function in $*R$ Then condition (C) when interpreted in $*R$ according to the rule adopted earlier, refers to positive ϵ and δ in $*R$. If $f(x)$ satisfies the condition in this form, then we say that $f(x)$ is Q *continuous* (at x_0). On the other hand, if $f(x)$ satisfies (C) with ϵ and δ still assumed *in* R, then we say that $f(x)$ is S *continuous*. It turns out that the condition that $f(x_0 + \eta) \simeq f(x_0)$ for infinitesimal η is equivalent to S continuity. Thus, if $f(x)$ is standard then all these definitions coincide. For example, the two δ functions defined above are both Q continuous at the origin, but not S continuous. By contrast, the function

$$f(x) = \begin{cases} 0, & x \le 0 \\ \eta, & x > 0 \end{cases},$$

where η is infinitesimal, is S continuous at the origin, but not Q continuous.

In the sequel to this paper (Ref. 7) we shall require the notion of an *enlargement* which generalizes the kind of nonstandard model discussed here (see Ref. 2). However, the above indications should be sufficient for the study of the present paper.

3. EXAMPLES FROM QUANTUM MECHANICS

In this section we demonstrate the use of n.s.a. by finding the bound state solution of the one-dimensional Schrödinger equation

$$\left(\frac{d^2}{dx^2} + E - V(x)\right) \phi(x) = 0, \tag{1}$$

for the square well, infinite square well, δ-function, and singular square well potentials. The infinite square well and the δ-function potentials are limiting cases of "physical" potentials, but themselves are outside the framework of quantum mechanics. This only means that we have to prescribe conditions other than the continuity of the logarithmic derivative of the wavefunction or to treat these problems by taking limits.

In the n.s. translation we have

$$\left(\frac{d^2}{dx^2} + E - V(x)\right) \phi(x) = 0$$

where

$$V(x) = \begin{cases} V_0, & x < 0 \\ V_1, & 0 < x < L \\ V_0, & L < x \end{cases}, \tag{2}$$

and the conditions (i) $V_0 > E > V_1$, (ii) $\phi(x)$ is infinitesimal for $x \in *R$-R_0, where R_0 is the set of all finite reals; and (iii) the logarithmic derivative of $\phi(x)$ is Q continuous.

In our formulation all four potentials are "physical," since all four are "finite" square well potentials in the nonstandard universe. They differ only in the values of V_0, V_1, and L. For a finite square well potential the solution of (1) is well known. Since (2) contains only internal objects, by the transfer theorem its solution is given by

$$c_n\phi_n(x)=\begin{cases}\exp(k_nx), & x<0\\ \dfrac{p_n-ik_n}{2p_n}\exp(ip_nx)+\dfrac{p_n+ik_n}{2p_n}\exp(-ip_nx), & 0<x<L,\\ \dfrac{k_n^2+p_n^2}{2k_np_n}\sin p_n\exp(k_nL-k_nx), & L<x\end{cases}$$

where c_n is determined by normalization, p_n and k_n satisfy the equations $2k_np_n(k_n^2+p_n^2)^{-1}=\tan p_nL$, and $E_n=p_n^2+V_1=V_0-k_n^2$. Observe that for $n\in N$:

(α) For the finite square well V_0, V_1 and $L\in R$, thus $k_n\in R$ and ${}^0[\exp(k_nx)]=\exp(k_n{}^0x)$, ${}^0\phi_n(x)$ is the well-known solution of (1).

(β) For the infinite square well $V_1=0$, $V_0=+\eta$ (η positive infinite), and $L\in R$. Thus $p_n=(E_n)^{1/2}$ is finite, and ${}^0[2k_np_n(k_n^2+p_n^2)^{-1}]={}^0[2(E_n)^{1/2}(\eta+E_n)^{1/2}\eta^{-1}]=0={}^0\tan p_nL$, i.e., ${}^0p_nL=n\pi$ or ${}^0p_n=n\pi L^{-1}$. Again, ${}^0\phi_n(x)$ is the known solution of (1).

(γ) For $V_0=0$, $V_1=-AL^{-1}$, $0<A\in R$, and $0<L$ infinitesimal, i.e., the δ-function potential of strength $-A$, we have $2(-E_n)^{1/2}(AL^{-1}+E_n)^{1/2}A^{-1}=\tan[L(AL^{-1}+E_n)^{1/2}]$. Thus, ${}^0(-E_n)^{1/2}={}^0\{1/2AL^{-1}(AL^{-1}+E_n)^{-1/2}\tan[L(AL^{-1}+E_n)^{1/2}]\}$ or ${}^0(-E_n)^{1/2}=1/2A$, again leading to the known solution of (1).

The transfer theorem ensures that all representations of the δ function lead to the same ${}^0\phi(x)$, i.e., to the unique solution of (1).

(δ) For $V_0=0$, $V_1=-AL^{-2}$, $0<A\in R$, and $0<L$ infinitesimal, i.e., the singular square well, we have

$$2(-E_n)^{1/2}(AL^{-2}+E_n)^{1/2}A^{-1}L^2=\tan[(A+L^2E_n)^{1/2}]. \qquad (3)$$

Since $E_n<0$ and $A+L^2E_n>0$, there are only a finite number of bound state solutions. They have infinite energies, since if E_n were finite in (3) the right-hand side would not be infinitesimal while the left-hand side would be. Evaluating c_n from normalization we find that $c_n=b_nL^{-1/2}$, where b_n is finite. Therefore, $\phi_n(x)$ is infinite in the monad of zero and one cannot take the standard part of $\phi(0)$.

A more interesting case occurs when the potential in (1) is given by

$$V_a(x)=\begin{cases}\infty, & x<a, a>0\\ -Ax^{-2}, & x>a, A>\frac{1}{4}\end{cases}.$$

As $a\to 0$, $V_a(x)$ becomes a singular potential.

In the nonstandard formulation we have

$$\left(\frac{d}{dx^2}+E-V_a(x)\right)\phi(x)=0,$$

where

$$V_a=\begin{cases}\infty, & x<a,\ a>0\\ -Ax^{-2}, & x>a,\ A>\frac{1}{4},\end{cases} \qquad (4)$$

and the conditions: (i) $E<0$; (ii) $\phi(x)$ is infinitesimal for $x\in {}^*R\text{-}R_0$; (iii) $\phi(x)$ is Q continuous at $x=a$.

Note that $V_a(x)$ is not an n.s. physical potential.

Let $E=-k^2<0$, $z=kx$, and $\phi(x)=\psi(z)$. Then by the transfer theorem the two linearly independent solutions of (4) are

$$\psi_1(z)=\begin{cases}0, & z<ak\\ I_{(\frac{1}{4}-A)^{1/2}}(z) & z>ak\end{cases},$$

$$\psi_2(z)=\begin{cases}0, & z<ak\\ K_{(\frac{1}{4}-A)^{1/2}}(z), & z<ak\end{cases}.$$

Because of condition (ii) only $\psi_2(z)$ is acceptable. Denote the zeros of $K_{(\frac{1}{4}-A)^{1/2}}(z)$ by $z_0,z_1,\ldots,z_m$. By the transfer theorem m is finite. Condition (iii) requires that $k_i=z_ia^{-1}$, i.e., $E_i=-z_i^2a^{-2}$ for $i=0,1,\ldots,m$. Thus, we have a finite set of discrete eigenvalues. When a is infinitesimal; the E_i are infinite as expected on physical grounds.

Next we determine the normalization constants. Using the transfer theorem we find that[8]

$$\phi_i(x,a)=\begin{cases}0, & x<a\\ b_i^{1/2}a^{-1}x^{1/2}K_{(\frac{1}{4}-A)^{1/2}}(a^{-1}z_ix) & x\geq a\end{cases},$$

where

$$b_i^{-1}=2\mathrm{Re}\,[K_{(\frac{1}{4}-A)^{1/2}-1}(z_i)K_{(\frac{1}{4}-A)^{1/2}+1}(z_i)].$$

For standard $a\neq 0$, ${}^0\phi_i(x,a)$ is the well-known solution of (1). For infinitesimal a, $\phi_i(x,a)$ is infinite for some points in the monad of zero, and infinitesimal for x positive and not infinitesimal. We also see that ${}^0\langle\phi_i(x,a_1)|\phi_j(x,a_2)\rangle=0$ when a_1 is finite and a_2 is infinitesimal. That is $\phi_i(x,a)\to 0$ in the weak topology as $a\to 0$, or $\phi_i(x,a)$ rotates out of $H\subset {}^*H$ into ${}^*H\text{-}H$. The renormalized operator $\theta=a^2[(d^2/dx^2)+(A/x^2)]$ also rotates in such a manner that $\langle\phi_i(x,a)|\theta\phi_i(x,a)\rangle$ remains finite and independent of a.

4. THE SCALAR FIELD

In this section we give the n.s. version of the scalar field interacting with a (nonrecoiling) nucleon.[9] The form factor $f(\mathbf{k}^2)$ is taken to be the characteristic function

$$x_\eta(\mathbf{k}^2)\equiv\begin{cases}1 & \text{for } \mathbf{k}^2\leq\eta^2\\ 0 & \text{otherwise}\end{cases},$$

where η is some infinite integer.

The equivalent potential, under these conditions, differs infinitesimally from the Yukawa potential. We renormalize the resulting theory.

Following Ref. 9, we introduce the definitions:

(i) Let $\mathcal{H}=F=\sum_{n\in{}^*N}F_n$ be the n.s. Fock space. F will have ${}^*N$ mutually orthogonal axis, and the vectors in F have infinite, finite, or infinitesimal norms.

(ii) The Hamiltonian $H(\eta)\equiv H_0+H_I(\eta)$, where

$$H_0=m_0\int d\mathbf{p}\,\psi^*(\mathbf{p})\psi(\mathbf{p})d\mathbf{p}+\int d\mathbf{k}\omega(\mathbf{k})a^*(\mathbf{k})a(\mathbf{k}),$$

$$H_I(\eta)=\lambda(2\pi)^{-3/2}\int d\mathbf{p}\int d\mathbf{k}$$
$$\times\, x_\eta(\mathbf{k}^2)(2\omega(\mathbf{k}))^{-1/2}\psi^*(\mathbf{p}+\mathbf{k})\psi(\mathbf{p})[a(\mathbf{k})+a^*(-\mathbf{k})],$$

where $\psi(\mathbf{p})$, $\psi^*(\mathbf{p})$ and $a(\mathbf{k})$, $a^*(\mathbf{k})$ are the destruction and creation operators for the nucleons and mesons, respectively, $\omega(\mathbf{k}) = (\mathbf{k}^2 + \mu^2)^{1/2}$, μ is the mass of the meson, and m_0 is the bare mass of the nucleon, and λ is a coupling constant.

(iii) The following commutation rules are satisfied:

$$[\psi(\mathbf{p}), \psi(\mathbf{p}')]_+ = [\psi^*(\mathbf{p}), \psi^*(\mathbf{p}')]_+ = 0,$$
$$[a(\mathbf{k}), a(\mathbf{k}')] = [a^*(\mathbf{k}), a^*(\mathbf{k}')] = 0,$$
$$[\psi(\mathbf{p}), a(\mathbf{k})] = [\psi(\mathbf{p}), a^*(\mathbf{k})] = 0,$$
$$[\psi^*(\mathbf{p}), a(\mathbf{k})] = [\psi^*(\mathbf{p}), a^*(\mathbf{k})] = 0,$$
$$[\psi(\mathbf{p}), \psi^*(\mathbf{p}')]_+ = \delta^3(\mathbf{p} - \mathbf{p}') \text{ and } [a(\mathbf{k}), a^*(\mathbf{k}')] = \delta^3(\mathbf{k} - \mathbf{k}'),$$

where

$$\delta(x) \equiv \begin{cases} 0 & x \neq 0 \\ \text{undefined for } x = 0 \end{cases}$$

and

$$\int \delta(x) g(x) dx = g(0) \quad \forall g \in {}^*L_2.$$

(iv) The vector $|\eta, 0\rangle$ satisfying $\psi(\mathbf{p})|\eta, 0\rangle = a(\mathbf{k})|\eta, 0\rangle = 0$ $\forall \mathbf{p}$ and $\mathbf{k}$ is the physical vacuum.

(v) The physical nucleon state $|\eta, 1, \mathbf{p}\rangle$ is defined by the equation $H(\eta)|\eta, 1, \mathbf{p}\rangle = m(\eta)|\eta, 1, \mathbf{p}\rangle$ and $\| |\eta, 1, p\rangle \| = 1$, where $m(\eta)$ is the physical mass.

With the aid of the transfer theorem using the results of pp. 341–44 of Ref. 9, we get

(i) $H(\eta)|\eta, 0\rangle = 0$;

(ii) $H(\eta) a^*(\mathbf{k})|\eta, 0\rangle = \omega(\mathbf{k})|\eta, 0\rangle$;

(iii) $\psi(\mathbf{p})^*|\eta, 0\rangle$ is not an eigenstate of $H(\eta)$;

(iv) $|\eta, 1, p\rangle = \sum_{n \in {}^*N} \int d\mathbf{q} d\mathbf{k}_1 \cdots d\mathbf{k}_n$
$\times\, c_p^{(n)}(\eta, \mathbf{q}; \mathbf{k}_1, \ldots, \mathbf{k}_n)(1/n!) a^*(k_1) \cdots a^*(k_n)|\eta, 0\rangle$,

where

$$c_p^{(n)}(\eta, \mathbf{q}; \mathbf{k}_1, \ldots, \mathbf{k}_n) = z^{1/2} \delta^3 \left(\mathbf{q} + \sum_{i=1}^n \mathbf{k}_i - \mathbf{p}\right) \frac{(-\lambda)^n}{n!} \times \prod_{i=1}^n \frac{x_\eta(\mathbf{k}_i^2)}{[2(2\pi)^3 \omega^3(\mathbf{k}_i)]^{1/2}},$$

where

$$z = \frac{\lambda^2}{4\pi^2} \exp\left(\ln\mu - \ln[\eta + (\eta^2 + \mu^2)^{1/2}] + \frac{\eta}{\sqrt{\eta^2 + \mu^2}}\right)$$

{i.e., infinitesimal for finite λ and behaves as $\eta^{-\lambda^2/4\pi^2}$ if $m(\eta) = m_0 - (\lambda^2/4\pi^2)[\eta - \mu \tan(\eta/\mu)]$; ${}^0[\tan(\eta/\mu)] = \frac{1}{2}\pi$}.

For finite cutoff d the one-particle state $|d, 1, p\rangle$ is in $F \subset {}^*F$. As d becomes infinite the one-particle state rotates to ${}^*F - F$. As $|\mathbf{k}_i|$ increases the factor $[2(2\pi)^3\omega^3(\mathbf{k}_i)]^{-1/2}$ behaves as $|\mathbf{k}_i|^{-3/2}$. The volume grows as $|\mathbf{k}_i|^2$. Therefore, one is more likely to find mesons with large momentum than with small momentum. Hence as d increases the n meson state rotates out of $F_n \subset {}^*F_n$ into ${}^*F_n - F_n$. To $|d, 1, \mathbf{p}\rangle$ the contribution ratio from the $n+1$ and n meson states is proportional to $[\lambda d/(n+1)^{1/2}]$. Thus when $d = \eta$ (infinite), the main part of $(\eta, 1, \mathbf{p}\rangle$ will come from $\cup_{k \in I_k} {}^*F_{K+k}$ for some infinite integer K and $I_k = \{1, 2, \ldots, k\}$ for some finite k.

In accord with pp. 347–48 of Ref. 9 we define

$$S_\eta = i\lambda(2\pi)^{-3/2} \int d\mathbf{p} \int d\mathbf{k}\, \chi_\eta(\mathbf{k}^2)[2\omega^3(\mathbf{k})]^{-1/2} \times \psi(\mathbf{p} + \mathbf{k})[a^*(\mathbf{k}) - a\psi^*(\mathbf{p})]$$

and

$$\psi_r^*(\mathbf{p}) = e^{iS_\eta}\psi^*(\mathbf{p})e^{-iS_\eta} \text{ and } a_r^*(\mathbf{k}) = e^{iS_\eta}a^*(\mathbf{k})e^{-iS_\eta}.$$

Then $H_r(\eta) = H_{0_r} + H_{I_r}(\eta)$, where

$$H_{0_r} = m_0 \int d\mathbf{p} \psi_r(\mathbf{p})\psi_r(\mathbf{p}) + \int d\mathbf{k}\, \omega(\mathbf{k}) a_r^*(\mathbf{k}) a_r(\mathbf{k}),$$

$$H_{I_r}(\eta) = \lambda^2(2\pi)^{-3} \int d\mathbf{q} \int d\mathbf{p} \times \chi_\eta(\mathbf{k}^2)[2\omega^2(\mathbf{k})]^{-1}\psi_r^*(\mathbf{p} + \mathbf{k})\psi_r^*(\mathbf{q})\psi_r(\mathbf{p})\psi_r(\mathbf{q} + \mathbf{k}),$$

i.e., with this rotation of the n.s. Fock space the Hamiltonian no longer contains a self-interaction term, but contains interaction between "dressed" nucleons (nucleons with mesons clouds that contain most probably an infinite number of mesons with infinite momentum).

The equivalent static potential is

$$V(\mathbf{x} - \mathbf{x}') = \lambda^2(2\pi)^{-3} \int d\mathbf{k}[2\omega^2(\mathbf{k})]^{-1} \exp[i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')]$$
$$= \frac{\lambda^2}{8\pi} \frac{\exp\{-\mu|\mathbf{x} - \mathbf{x}'|\}}{|\mathbf{x} - \mathbf{x}'|} + \frac{i\lambda^2}{8\pi^2|\mathbf{x} - \mathbf{x}'|} \int_0^\pi \frac{\eta^2 \exp[i2\theta + i\eta(\cos\theta + i\sin\theta)]}{\eta^2 \exp(i^2\theta) + \mu^2}.$$

The second term is infinitesimal for infinite η and ${}^0|\mathbf{x} - \mathbf{x}'| \neq 0$. For θ in the monads of 0 and π the integrand is finite and, otherwise, the integrand is infinitesimal. Hence, if ${}^0|\mathbf{x} - \mathbf{x}'| \neq 0$, ${}^0V(\mathbf{x} - \mathbf{x}')$ is the Yukawa potential. Moreover, the second term is infinitesimal compared to the first term even when ${}^0|\mathbf{x} - \mathbf{x}'| = 0$.

5. CONCLUSION

We demonstrated the techniques of n.s.a. by working simple examples. In Sec. 3 we translated well-known one-dimensional bound state problems into n.s. language. We showed that potentials that are commonly called "idealized" or "limiting," are "physical" in the n.s. formulation. We showed how one recovers the standard results.

In Sec. 4 we showed the use of n.s.a. in investigating the properties of the wavefunction as a given parameter tends to some limit. We saw how the wavefunction "rotated" out of ordinary Hilbert space.

In Sec. 5 we found a field theory in which the equivalent potential differs infinitesimally from the Yukawa potential. We saw how the vacuum vector "rotated" out of Fock space into the n.s. Fock space as the form factor became one on an infinite set. One may renormalize the infinite cutoff Yukawa theory, by defining the mass renormalized Hamiltonian H_{ren} on a separable Hilbert space $\mathcal{H}_{\text{ren}}$, extracted from ${}^*F$. This construction is carried out for the $\lambda : \phi_2^4$: model in the sequel to this paper.[7]

---

*Research supported in part by the National Science Foundation, Grants Nos. GP-29218 and GP-32996X.

[1] A. Robinson, Proc. Roy. Acad. 64, 432 (1961).
[2] For a detailed treatment of nonstandard analysis see A. Robinson, *Non-Standard*

Analysis (North-Holland, Amsterdam, 1966).
[3]W. A. J. Luxemburg, *Non-Standard Analysis* (California Institute of Technology, Pasadena, Calif., 1962).
[4]F is a free-ultrafilter on N if: (i) F is a nonempty family of subsets of N; (ii) the empty set $\phi \notin F$; (iii) $N_1 \in F$ and $N_2 \in F$, then $N_1 \cap N_2 \in F$; (iv) $N_1 \in F$ and $N_1 \supset N_2$, then $N_2 \in F$; (v) $N_1 \subset N$, then either $N_1 \in F$ or $N - N_1 \in F$; (vi) the intersection of an infinite number of distinct elements of F is empty. Note that (i)–(iv) establish the equivalence classes of sequences, (v) ensures that every sequence will belong to an equivalence class, and (vi) ensures that the equivalence classes are not based on a single element of the sequence. For further detail see Ref. 3 or A. Vörös, *Introduction to Non-Standard Analysis* (to be published).
[5]The * in the upper left corner carries an object from the standard universe into the corresponding object of the nonstandard universe.
[6]$\int_{*R}$ is the extension of the linear functional $\int_R$.
[7]P. J. Kelemen and A. Robinson, J. Math. Phys. (N.Y) **13**, 1875 (1972).
[8]G. N. Watson, *Theory of Bessel Functions* (Cambridge, London, 1966), p.134.
[9]S. S. Schweber, *An Introduction of Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), pp. 339–348.

The nonstandard $\lambda{:}\phi_2^4(x)$: model. II. The standard model from a nonstandard point of view*

Peter J. Kelemen
Courant Institute of Mathematical Sciences and Department of Mathematics, University College, New York University, New York

Abraham Robinson
Department of Mathematics, Yale University, New Haven, Connecticut
(Received 22 May 1972)

As a second step in the construction of the nonstandard $\lambda{:}\phi_2^4{:}$ model we analyze Glimm and Jaffe's work from the nonstandard point of view.

1. INTRODUCTION

In this paper we analyze the $\lambda:\phi_2^4(x)$: model of quantum field theory using the tools of nonstandard analysis, n.s.a. The model was selected because it shows both the conceptual and the technical difficulties that one encounters in building a nontrivial model of quantum field theory. These problems will become more transparent through explicit constructions within the framework of n.s.a.

The $\lambda:\phi_2^4(x)$: model was investigated by Glimm and Jaffe[1] in three papers to which we will refer as I, II, and III. For comparison between the nonstandard and standard treatment of this model, we will restrict ourself to the subject matter covered and to the assumptions made in these papers. This permits us to concentrate on those aspects of this model where the nonstandard approach is advantageous.

In I and II it is found that with a space cutoff $g(x)$ imposed the theory is meaningful. There exists a self-adjoint operator on a Fock space, the Hamiltonian, that generates the time translations, provided the time interval is sufficiently short. The Hamiltonian possesses an isolated lowest eigenvalue, E_g of multiplicity one. The corresponding eigenvector Ω_g, the vacuum vector is an element of Fock space. However, as the support of $g(x)$ grows, the vacuum vector seems to move out of the Fock space, $E_g \to -\infty$ and the Hamiltonian ceases to be an operator.

In consequence, one is forced to change Hilbert spaces and to redefine the operators of the model. This renormalization is carried out in III using the GNS construction.

In the n.s. treatment both for finite and for infinite cutoffs the Hamiltonian is an n.s. self-adjoint operator on an n.s. Fock space with a unique vacuum vector. For finite cutoff the vacuum vector is an element of the standard Fock space which is imbedded in the n.s. Fock space. To renormalize the theory we map a certain subspace of the n.s. Fock space onto a standard Hilbert space and redefine the operators.

To carry out this program, we begin by describing n.s. objects such as $^*L_2(^*R)$, n.s. operators, etc. (Sec. 2). In Sec. 3 we outline the $\lambda:\phi_2^4(x)$: model. In Sec. 4 we build the n.s. model, and in Sec. 5 we renormalize it. In Sec. 6 we summarize our results and sketch lines of development that we intend to follow up in the future.

2. NONSTANDARD PRELIMINARIES

The method of n.s. extension that we are using in this paper is provided by model theory.[2] We consider R (reals), C (complex numbers), arithmetic, analysis, $L_2(R^n)$, D and D' (the spaces of test functions and distributions), F (Fock space), operators, and linear functionals on F as given within the framework of some structure M. A model of some nontrivial enlargement $^*M$ of M serves as our n.s. extension.[3]

For what follows, it is unnecessary to construct an explicit model for some specific enlargement $^*M$ of M.

However, it is convenient to picture the n.s. objects. The model we select is such that $^*R$, $^*C$, $^*S$, $^*F$ may be visualized through an ultrafilter construction.[4] This means that a vector of $^*F$ can be pictured as an infinite sequence of vectors of F (reduced with respect to a certain equivalence relation.)

For convenience we restate the main theorem of n.s.a.

Transfer theorem: All true assertions about analysis remain true in the nonstandard model provided we reinterpret them as referring to *internal objects* only.

See Ref. 2 for the notion of internal and external objects and Ref. 4 for an informal discussion of these concepts. We recall that any set, function, operator, operator algebra, etc. is either internal or external but not both. Among the internal objects are the nonstandard extensions of all standard objects. Thus, every function, set, operator, etc., in standard analysis possesses a canonical extension to the nonstandard model. The ultrapower method provides a relatively concrete construction of internal objects.

3. THE $\lambda:\phi_2^4$: MODEL

The model developed in I and II is a spin-zero boron field ϕ, with a nonlinear, $\lambda:\phi_2^4$:, self-interaction with a space cutoff in two dimensional space–time. The field $\phi(x,t)$ is a bilinear-form-valued solution of

$$\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial x^2}+m_o^2\right)\phi(x,t)=-4\lambda g(x)\phi^3(x,t), \tag{1}$$

where $g(x)$ is a smooth positive function that equals one on some bounded interval and vanishes outside some larger bounded interval containing the smaller one. In III the interval on which $g(x)=1$ is increased indefinitely in some prescribed manner, i.e., through a divergent sequence of intervals. An infinite sequence of standard intervals is replaced by a unique n.s. interval. Hence, "removing the cutoff" is equivalent to selecting a cutoff in the n.s. model.

The Hamiltonian corresponding to (1) is given by $H(g) = H_0 + \lambda\int_R :\phi^4(x): g(x)dx$, where H_0 is the free particle Hamiltonian. The corresponding vacuum vector is Ω_g. In I and II it is shown that for all finite cutoffs, Ω_g belongs to F. Hence, $\Omega_g \in {}^*F$ even when supp$g = (-\eta, \omega)$ where η and ω are infinite positive numbers. It is believed that as the length of the cutoff increases, Ω_g converges weakly to zero.[5] From the transfer theorem it would then follow that as the length of the n.s. cutoff increases, Ω_g would converge weakly to zero in $^*F$, i.e., there is no unique n.s. vacuum.

On the other hand in II it was shown that if $x_0 \in \text{supp} g_1(x)$ and $\text{supp} g_1(x)$ is large enough, then $\phi_{g_1}(x_0, t_0) = \phi_{g_2}(x_0, t_0)$ when $\text{supp} g_1(x) \subset \text{supp} g_2(x)$. Thus for finite (x_0, t_0), $\phi_g(x_0, t_0) = \phi(x_0, t_0)$ provided that $(-\eta, \eta) \subset \text{supp} g(x)$ for some infinite positive η. Note that $\phi(x, t)$ is an internal operator for each finite (x, t), but the collection of $\phi(x, t)$ for all finite (x, t) is an external set.

After the above intuitive remarks we could proceed by translating into n.s. language all the theorems and lemmas of I, II, and III. In some cases the standard proofs plus the Transfer theorem would provide the n.s. proofs. In others, especially where limits are taken to remove the momentum cutoffs, the n.s. proofs would be shorter.[6] But a different approach is more useful here. We apply the Transfer theorem only to the results of I and II in building an unrenormalized n.s. theory. (This exemplifies the fact that in building an n.s. theory one may incorporate as many of the standard results as desired.) We then extract from $^*F$ a standard Hilbert space $\mathcal{H}_{\text{ren}}$ which is identical with F_{ren}. Inthe sequel to this paper we will add external assumptions to the unrenormalized theory and find new interpretations.

4. THE n.s. MODEL

In this section we compile those n.s. definitions and theorems which define the n.s. model. The numbers in brackets after our numbering refer to the page on which the standard counterparts are found.

Definition 1 [II-364]: The Fock space $^*F$ is the Hilbert space completion of the symmetric tensor algebra over $L \equiv {}^*L_2({}^*R)$,

$$^*F = G(L) \equiv \oplus_{n \in {}^*N} {}^*F_n,$$

where $^*F_n \equiv L \otimes_s L \otimes_s \cdots \otimes_s L$ (n factors) is the space of n noninteracting particles. For $\psi \in {}^*F$, $\psi = \{\psi_0, \psi_1 \cdots\}$ we have $\|\psi\|^2 = \sum_{n \in {}^*N} \|\psi_n\|^2$.

Definition 2 [II-364]: The no particle space $^*F_0 = {}^*C$ (the complex numbers and $\Omega_0 = \{1, 0, 0, \cdots\} \in {}^*F$ is the bare vacuum or bare no-particle state vector.

Definition 3 [II-366]: The Hamiltonian $H(g)$ acts on $^*F$ and can be written as $H(g) = H_0 + \lambda \int_{*R} : \phi^4(x) : g(x) dx = H_0 + H_{I,g}$, where H_0 is the free particle Hamiltonian; $H_{I,g}$ is the interaction cutoff Hamiltonian; and $g(x)$ is an internal smooth positive function that equals one on an internal set B that contains in the interval $(-\eta, \eta)$, $\eta \in {}^*N$, and vanishes off an internal set that contains B.

Definition 4 [I-1946]: The domain $D_0 \equiv \bigcap_{n \in {}^*N} D(H_0^n)$.

Theorem 1 [I-1949]: (a) $H(g)$ is self-adjoint with domain $D(H(g)) = D(H_0 \cap D(H_{I,g})$; (b) $H(g)$ is essentially self-adjoint on D_0.

Definition 5 [II-364]: E_g is the lowest eigenvalue of the equation

$$H(g)\Omega_g = E_g \Omega_g, \qquad \|\Omega_g\| = 1.$$

Theorem 2 [II-368]: There exists a vacuum vector Ω_g for $H(g)$.

Theorem 3 [II-372]: The lower bound of $H(g)$ is a simple eigenvalue. (Note that we got the above result with the transfer theorem. Therefore, the gap between E_g and the rest of the spectrum may be infinitesimal but not zero.

Definition 6 [II-382]: Let $f(x, t)$ be the extension of a C^∞ function that vanishes off the rectangle $-n \le x, t \le n$ for some $n \in N \subset {}^*N$. Then $A_g(f) \equiv \int_{*R} \phi_g(x, t) f(x, t) dx$ and $\phi_g(f)$ is the closure of the operator defined by

$$(\psi, \phi_g(f)\psi) = \int_{*R} (\psi, A_g(f)\psi) dt, \qquad \psi \in D[(H(g) + b)^{1/2}],$$

where b is a suitably large constant.

Theorem 4 [II-388]: $A_g(f)$ and $\phi_g(f)$ are self-adjoint operators.

Theorem 5 [II-388]: $\phi_g(f) = \phi(f)$ provided $(-\eta, \eta) \subset \text{supp} g$ for some $\eta \in {}^*N - N$.

Theorem 6 [II-385]:

$$\pi(f) = \phi\left(-\frac{d}{dt} f\right) = i[H(g), \phi(f)].$$

Remark: $\phi(f)$ is an internal operator when $f(x, t)$ is the extension of a C^∞ function with support in the rectangle $-n < x, t < n$ for some $n \in N$; but the collection of all such operators is an external set. The importance of this fact cannot be over emphasized. Since $\phi(f)$ is internal its properties are determined by the standard operator that extends to $\phi(f)$. But there is no standard theorem that determines the properties of this external set of operators. This is why a renormalized nontrivial theory may exist.

5. THE RENORMALIZED MODEL

We reproduce the renormalized model of III. From $^*F$ we extract a standard Hilbert space $\mathcal{H}_{\text{ren}}$. Our method is equivalent to the Gel'fand–Naimark–Segal GNS construction. We redefine the operator of $^*F$ on $\mathcal{H}_{\text{ren}}$. Our construction illuminates the one employed in III.

To make the connection between the GNS construction and our extraction of $\mathcal{H}_{\text{ren}}$ more transparent first we discuss a case in which the linear functional used in the GNS construction is simpler than the one used in III. Let Ω_g be the vacuum vector for the cutoff $g(x)$, let $\hat{C}^\infty \equiv \{f(x) | f(x) = {}^*h(x), h(x) \in C^\infty$ and has compact support$\}$. Define $S \equiv \{z \in {}^*F | z = e^{i\phi(f_1)} \cdots e^{i\phi(f_k)} \Omega_g$, $f_j \in \hat{C}^\infty j = 1, 2, \ldots, k \in N\}$, and let $\mathcal{H}$ be the subspace of $^*F$ spanned by S. Note that each element of $\hat{C}^\infty$ and of $\mathcal{H}$ is internal, but that both $\hat{C}^\infty$ and $\mathcal{H}$ are external objects. We extract $\mathcal{H}_0$ from $\mathcal{H}$ by discarding all vectors which have infinite norms. To get $\mathcal{H}_{\text{ren}}$ from $\mathcal{H}_0$ we collapse into a single vector those vectors which differ from each other by a vector of infinitesimal norm; and redefine the innerproduct by passing from (z_1, z_2) to $^0(z_1, z_2)$. Equivalently, map $\mathcal{H}_0$ into $\mathcal{H}_{\text{ren}}$ a subspace of some standard Hilbert space by the rule that if $z_1, z_2 \in \mathcal{H}_0$ and $z_1 \to b_1, z_2 \to b_2$, then $^0(z_1, z_2)_{*F} = (b_1, b_2)_{\mathcal{H}_{\text{ren}}}$.

The elements of the C^*-algebra $\mathfrak{A}$, generated by $\{e^{i\phi(f)} | \cdot f \in \hat{C}^\infty\}$, are operators on $^*F$. From the Riesz representation theorem we infer that the linear functionals on $^*F$ are innerproducts. In particular, the positive linear functional $\phi(A)$ of the GNS construction in this concrete case is $^0(\Omega_g, A\Omega_g)$. Constructing the quotient space with the left ideal I, $I = \{\tau \in \mathfrak{A} \mid \phi(\tau^*\tau) = 0\}$ amounts to our "collapsing" of the vectors

The linear functional of the GNS construction employed in III is more complicated, because it uses an averaging process. As we will see the averaging serves two purposes. It ensures, first, that the energy per unit volume is finite space translation invariant.

To give the n.s. version of Sec. 2 of III we also take the cutoff function $g(x) \in \hat{C}^\infty$ to be nonnegative and equal to 1 on $[-3, 3]$, and define $g_n(x) \equiv g(x/n)$, $n \in {}^*N$. The corresponding vacuum vector is denoted by Ω_n. As in III we fix an $h(x) \in \hat{C}^\infty$ with support in $[-1,]$ that has the property $\int h(x)dx = 1$. We use the notation $E_j(\alpha) = e^{i\psi_\alpha(f_j)}$ where $\psi_\alpha(f_j) = \int_{*R}\psi(x+\alpha, t)f_j(x)dx$ and ψ stands for either ϕ or π. We define $S(\alpha, n) \equiv \{z(\alpha) \in {}^*F \mid z(\alpha) = E_1(\alpha)\cdots E_k(\alpha)\Omega_n,\ f_j \in \hat{C}^\infty,\ j = 1, 2, \cdots, k \in N\}$, and $\mathcal{K}_0(\alpha, n) \subset {}^*F$ the subspace that contains only finite normed vectors and spanned by $S(\alpha, n)$. To get $\mathcal{K}_{\text{ren}}(n)$ we average the $\mathcal{K}_0(\alpha, n) - s$. The vectors in $\mathcal{K}_{\text{ren}}(n)$ satisfy the conditions that if $z_1(0)$ and $z_2(0)$ map into b_1 and b_2, respectively, then

$$(b_1, b_2)_{\mathcal{K}_{\text{ren}}(n)} = {}^0[(1/n)\int h(\alpha/n)(z_1(\alpha), z_2(\alpha))_{*F}\, d\alpha].$$

For the proper choice of n, say η, $\mathcal{K}_{\text{ren}}(\eta)$ is identical to F_{ren} of III. The only difference between the construction of the two spaces is that in III $\mathfrak{A}$ is an abstract algebra while in the n.s. model it is an operator algebra. This is the case, because the space $\mathcal{K}_0(0, n)$ on which $\mathfrak{A}$ is defined and leaves invariant is nonstandard. The innerproduct of the GNS construction in III is defined by the mapping of $\mathcal{K}_0(0,\eta)$ onto $\mathcal{K}_{\text{ren}}(\eta)$, since for any standard bounded operator A, $\omega_n(A) = {}^0(\Omega_n, {}^*A\Omega_n)_{*F}$, and since a convergent subset of ω_n means selecting the corresponding Ω_η. Note that η is infinite, i.e., $\eta \in {}^*N - N$. Therefore $g_\eta(x)$ is equal to 1 on an infinite interval that contains $[-3\eta, 3\eta]$. But, by definition, the support of $g_\eta(x)$ is contained in $[-k, k]$ for some $K \in {}^*N$. Hence we have an infinite cutoff n.s. model, which means that the renormalized standard model without cutoff may not be unique. From the construction of $\mathcal{K}_{\text{ren}}(\eta)$ we see the effects of the averaging by $h(\alpha)$. The mapping of $\mathcal{K}_0(0,\eta)$ onto $\mathcal{K}_{\text{ren}}(\eta)$ leaves the norms invariant. $E_j^*(\alpha)E_j(\alpha) = I$, so that

$$\|b_j\|_{\mathcal{K}_{\text{ren}}(\eta)} = {}^0([1/\eta)\int h(\alpha/\eta)(E_1(\alpha)\cdots E_k(\alpha)\Omega_\eta,$$
$$E_1(\alpha)\cdots E_k(\alpha)\Omega_\eta)_{*F}d\alpha] = {}^0[(1/\eta)\int h(\alpha/n)(\Omega_\eta, \Omega_\eta)_{*F}d\alpha]$$
$$= {}^0[(\Omega_\eta, \Omega_\eta)_{*F}(1/\eta)h(\alpha/\eta)d\alpha] = (\Omega_\eta, \Omega_\eta)_{*F}$$
$$= (E_1(0)\cdots E_k(0)\Omega_\eta, E_1(0)\cdots E_k(0)\Omega_\eta)_{*F}.$$

But the map changes the angles between some of the vectors. They become larger, i.e., their innerproduct smaller.

It is easier to demonstrate some of the properties of F_{ren} on $\mathcal{K}_{\text{ren}}(\eta)$. For examples,

(i) H_{ren} is defined through the action of $H(g_\eta)$ on $\mathcal{K}_0(0,\eta)$. Thus, the spectrum of H_{ren} is nonnegative because the spectrum of $H(g_\eta)$ is nonnegative.

(ii) Finite time translation invariance follows from the finite propagation speed and from

$$e^{i\tau H(g_\eta)}\Omega_\eta = \Omega_\eta.$$

(iii) To see that the model is finite translation invariant it is sufficient to observe that

$$0 = {}^0[(1/\eta)\int_I h(\alpha/\eta)(\Omega_\eta, E_1(\alpha)\cdots E_k(\alpha)\Omega_\eta)_{*F}d\alpha]$$

for all finite intervals I; which is evident, since η is infinite and $|(\Omega_\eta, E_1(\alpha)\cdots E_k(\alpha)\Omega_\eta)_{*F}| \le 1$ so that

$$|(1/\eta)\int_I h(\alpha/\eta)(\Omega_\eta, E_1(\alpha)\cdots E_k(\alpha))_{*F}d\alpha| \le (1/\eta) \times \text{length}(I) = \text{infinitesimal}.$$

But there is yet another way to see this.

$$(1/\eta)\int_{*R}h(\alpha/\eta)(\Omega_\alpha, E_1(\alpha)\cdots E_k(\alpha)\Omega_\eta)_{*F}d\alpha$$
$$= (1/\eta)\int_{*R}h(\alpha/\eta)(\Omega_{\eta(\alpha)}E_1(0)\cdots E_k(0)\Omega_{\eta,\alpha})_{*F}d\alpha),$$

where $\Omega_{\eta,\alpha}$ is the vacuum vector of the $g[(x-\alpha)/\eta]$ cutoff. The last formula is interpreted the following way. For each $\alpha \in [-\eta, \eta]$ we construct a renormalized Hilbert space corresponding to the vacuum $\Omega_{\eta,\alpha}$ by the procedure given in the beginning of this section, and then we average over the renormalized Hilbert spaces. Clearly adding or deleting Hilbert space corresponding to a finite interval cannot affect the average.

6. CONCLUSIONS

In Sec. 4 we constructed an n.s. $\lambda:\phi_2^4$: model with an infinite cutoff. In Sec. 5 we recovered the renormalized Fock space.

Nonstandard analysis allowed us to work with operators on a Hilbert space, instead of an abstract operator algebra, and to employ intuitive ideas which are not available in the standard approach. It illuminated several interesting features of the renormalization. In III a renormalization by averaging is employed in addition to the energy renormalization by the subtraction of an infinite constant. This averaging "opens" the Hilbert space, i.e., it diminishes the inner product of two vectors. Vectors that are close together in $\mathcal{K}_{\text{ren}}(\eta)$ came from vectors that were even closer in $^*F$, and, hence, the finite space translation follows. This "opening" also decreases the energy density. On the other hand, this averaging procedure does not decrease the cardinality of the set of the basis vectors of $\mathcal{K}_{\text{ren}}(\eta)$. If the renormalized model constructed without averaging is only locally Fock, then the renormalized model constructed with averaging can be locally Fock only.

As it was pointed out, $\mathcal{K}_{\text{ren}}$ constructed without an averaging is already an external subspace of $^*F$. Hence, no standard theorem about F is transferable to F_{ren} directly by the use of the Transfer theorem. In particular, there is some hope that no analog of the Haag theorem will apply to F_{ren}. This statement demonstrates that it may be advantageous to investigate standard models with n.s. methods, bringing into play the distinction between external and internal objects.

Other approaches to the problem of finding a renormalized $\lambda:\phi_2^4$: model suggest themselves: (i) Retaining an infinite momentum cutoff may remove some of the difficulties. (ii) Using periodic boundary conditions, of infinite period, both in momentum and in position space would allow one to use n.s. Fourier series. (iii) Quantizing position space in an infinite box with rigid wall has its obvious advantages.

Field theory is probably best formulated on a nonseparable Hilbert space. The logical candidate is $^*F$. Having an established n.s. model (Sec. 4) one should check whether or not it satisfies a modified n.s. version of the Wightman axioms. We found in this paper that modification by external assumption is necessary. One can only require invariance for finite translations. Thus in the n.s. version of the Wightman axioms one should use the phrase "finite Lorentz transformation." What modifications, if any, are needed to assure that we do not need to average by $h(x)$ is not clear. Hopefully, one of the three approaches mentioned in the preceding paragraph will provide the answer.

ACKNOWLEDGMENTS

We would like to thank H. Georgi and J. Glimm for many helpful suggestions.

---

* Research supported in part by the National Science Foundation, Grant No. GP-29218, and in part by the Air Force Office of Scientific Research under Grant No. AFOSR-71-2013.

[1] J. Glimm and A. Jaffe Phys. Rev. **176**, 1945 (1968); Ann. Math. **91**, (1970); Acta Math. **125**, (1970).

[2] A. Robinson, *Non-Standard Analysis* (North-Holland, Amsterdam, 1966).

[3] The * in the upper left corner carries an object from the standard universe into the corresponding object of the nonstandard universe.

[4] P. J. Kelemen and A. Robinson, J. Math. Phys. (N.Y.) **13**, 1870 (1972).

[5] The vector moved out of F but remains in $^*F$. The occurrence of this phenomenon for the sharp cutoff case follows from F. Guerra, Phys. Rev. Letters 28, 1213 (1972).

[6] For this type of proof see p. 63 of Ref. 2.

Applications of the intertwining operators for representations of the restricted Lorentz group*

Robert W. Lind
Physics Department, Syracuse University, Syracuse, New York, 13210

James Messmer and Ezra T. Newman
Physics Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 23 June 1972)

Some properties of the intertwining operations, which carry over certain infinite-dimensional, reducible (but not completely reducible) representations of the restricted Lorentz group into finite-dimensional (irreducible) representations are studied.

1. INTRODUCTION

In the study of asymptotically flat solutions of the Einstein[1] or Einstein–Maxwell[2] equations, one naturally considers functions of three variables defined on null infinity. If null infinity has an $S^2 \times R^1$ topology (as is usually assumed) these functions can be viewed as time-dependent functions on the unit sphere.[3] Due to the fact that part of the asymptotic symmetry group[4] is the homogeneous Lorentz group L, these functions possess relatively simple transformation properties. In fact many of them transform as vectors in the representation space of a reducible (but not completely reducible) infinite-dimensional representation of the group L. From the theory of these representations, it is known that these spaces possess invariant subspaces and that frequently (depending on the type of representation) finite-dimensional representations can be constructed from the factor spaces.

Although it is common (in the study of asymptotically flat spaces) to assign to some of these tensors (associated with the finite-dimensional representations), a definite physical meaning such as energy–momentum, center-of-mass-angular momentum, 4-velocity, etc., it is not our purpose here to study the physical questions raised by such identifications. We will be concerned only with certain mathematical results having to do with the reduction of the infinite-dimensional representations to the finite-dimensional ones. These results, which deal with the reduction of products of infinite with finite-dimensional representations, though of interest in themselves, are to us of fundamental importance in analyzing asymptotically flat spaces. In the paper following this one we apply these ideas to the problem of equations of motion in general relativity.

It has been established through the use of the isomorphism between the conformal group of the sphere and the homogeneous (restricted) Lorentz group that the functions in question are the spin and conformally weighted functions on the sphere. Although the basic ideas come from the beautiful work of Gel'fand *et al.*,[5] we use the notation and techniques of Held, Newman, and Posadas (HNP),[6] which are reviewed in Sec. 2. In Sec. 3 we establish a link between tensors in Minkowski space and functions on the sphere with finite expansions in spin s spherical harmonics.[7] Section 4 contains the main results concerning the reduction of products of finite with infinite-dimensional representations.

2. THE LORENTZ GROUP AND THE SPHERE

By introducing the complex stereographic coordinate

$$\zeta = e^{i\varphi} \cot\tfrac{1}{2}\theta, \tag{2.1}$$

the standard line element of the unit sphere

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2$$

can be rewritten in the form

$$ds^2 = \frac{d\zeta d\bar{\zeta}}{P_0^2}, \tag{2.2}$$

where $P_0 = \frac{1}{2}(1 + \zeta\bar{\zeta})$.

Under the fractional linear transformation

$$\zeta' = (a\zeta + b)/(c\zeta + d), \quad ad - bc = 1, \tag{2.3}$$

the metric of the unit sphere transforms conformally, i.e.,

$$\frac{d\zeta' d\bar{\zeta}'}{P_0'^2} = K^2 \frac{d\zeta d\bar{\zeta}}{P_0^2} \tag{2.4}$$

with conformal factor K given by

$$\begin{aligned} K &= J^{-1/2} P_0/P_0' \\ &= (1 + \zeta\bar{\zeta})[(as + b)(\bar{a}\bar{\zeta} + \bar{b}) + (c\zeta + d)(\bar{c}\bar{\zeta} + \bar{d})]^{-1}, \end{aligned} \tag{2.5}$$

where

$$J^{-1} = \frac{d\zeta'}{d\zeta} \cdot \frac{d\bar{\zeta}'}{d\bar{\zeta}}.$$

From (2.3) we can also define the function $\lambda(\zeta, \bar{\zeta})$ by

$$e^{i\lambda} = \left(\frac{d\bar{\zeta}'/d\bar{\zeta}}{d\zeta'/d\zeta}\right)^{1/2} = \frac{c\zeta + d}{\bar{c}\bar{\zeta} + \bar{d}}, \tag{2.6}$$

where λ is interpreted geometrically as the local angle of rotation of the two coordinate grids, ζ = const and ζ' = const (after rotation).

Infinitely differentiable functions on the sphere, $\eta_{(s,w)}(\zeta,\bar{\zeta})$, which are expandable in spin s spherical harmonics

$$\eta_{(s,w)}(\zeta, \bar{\zeta}) = \sum_{l=|s|} a_{lm} \; {}_sY_{lm}(\zeta, \bar{\zeta}),$$

and which transform under (2.3) as

$$\eta_{(s,w)}(\zeta, \bar{\zeta}) \to \eta'_{(s,w)}(\zeta', \bar{\zeta}') = K^{w} e^{is\lambda} \eta_{(s,w)}(\zeta, \bar{\zeta}) \tag{2.7}$$

are said to have spin weight s and conformal weight w.

Because (2.3) is isomorphic to the restricted Lorentz group, it is possible to show that such functions, which transform under (2.3) with s an integer or half-integer and w any complex number, form the vector space of a representation of the restricted Lorentz group that can be labeled by (s, w). These representations are not necessarily irreducible, although a converse, namely that all irreducible representations are realizable on these spaces, is true.

For our purposes it will only be necessary to consider the representation labeled by (s,w) and $(s'=-s, w'=-w-2)\equiv(-s,-w-2)$, such that

$$s \text{ and } w \text{ are both either integer or half-integer}^{8} \text{ and } w \geq |s|. \quad (2.8)$$

The vector spaces associated with these representations, denoted by $D_{(s,w)}$ and $D_{(-s,-w-2)}$, respectively, turn out to be quite intimately related to one another. Both are reducible, but not completely reducible, possessing the respective invariant subspaces denoted by $E_{(s,w)}$ and $F_{(-s,-w-2)}$, such that the irreducible factor spaces $D_{(s,w)}/E_{(s,w)}$ and $D_{(-s,-w-2)}/F_{(-s,-w-2)}$ satisfy the isomorphisms

$$D_{(s,w)}/E_{(s,w)} \approx F_{(-s,-w-2)}$$

and

$$D_{(s,-w-2)}/F_{(-s,-w-2)} \approx E_{(s,w)}.$$

$E_{(s,w)}$ is finite-dimensional and spanned by the basis vectors ${}_sY_{lm}$, $|s|\leq l\leq w$, while $F_{(-s,-w-2)}$ is infinite-dimensional and spanned by the basis vectors ${}_{-s}Y_{lm}$, $l>w$.

If

$$\tilde{\eta}_{(-s,-w-2)} = \sum_{l=|s|}^{\infty} \tilde{a}_{lm}\ {}_{-s}Y_{lm} \in D_{(-s,-w-2)}$$

and

$$\varphi_{(s,w)} = \sum_{l=|s|}^{\infty} b_{lm}\ {}_sY_{lm} \in D_{(s,w)},$$

then the mappings from $D_{(s,w)} \to F_{(-s,-w-2)}$ and $D_{(-s,-w-2)} \to E_{(s,w)}$ can be given explicitly as follows:

$$\tilde{\varphi}_{(-s,-w-2)} = \sum_{l=w+1}^{\infty} \tilde{b}_{lm}\ {}_{-s}Y_{lm} = \bar{\eth}_0^{w+s+1}\eth_0^{w-s+1}\varphi_{(s,w)} \in F_{(-s,-w-2)},^{9} \quad (2.9)$$

$$\eta_{(s,w)} = \sum_{l=|s|}^{w} a_{lm}\ {}_sY_{lm} = \pi_{(s,w)}\tilde{\eta}_{(-s,-w-2)} \in E_{(s,w)}, \quad (2.10)$$

where

$$\pi_{(s,w)}\tilde{\eta}_{(-s,-w-2)} \equiv \int M_{(s,w)}(\zeta,\bar{\zeta};\zeta'\bar{\zeta}')\,\tilde{\eta}_{(-s,-w-2)}(\zeta',\bar{\zeta}')\,d\Omega'$$

with $d\Omega'$ the area element of the unit sphere and

$$M_{(S,W)}(\zeta,\bar{\zeta};\zeta'\bar{\zeta}') = \sum_{l=|s|}^{w}\sum_{m=-l}^{l}(-1)^{l+s}\frac{(w+|s|+1)!(w-|s|)!}{(w+l+1)!(w-l)!} \times {}_sY_{lm}(\zeta,\bar{\zeta})\ {}_{-s}\bar{Y}_{lm}(\zeta',\bar{\zeta}'). \quad (2.11)$$

The factor space $D_{(s,w)}/E_{(s,w)}$ is also isomorphic to the two equivalent representations $D_{(w+1,s-1)} \cong D_{(-w-1,-s-1)}$, so that if $\eta_{(s,w)} \in D_{(s,w)}$, then the mappings appropriate to these isomorphisms are given by

$$\eta'_{(w+1,s-1)} = \eth_0^{w-s+1}\eta_{(s,w)} \in D_{(w+1,s-1)}, \quad (2.12)$$

$$\eta''_{(-w-1,-s-1)} = \bar{\eth}^{w+s+1}\eta_{(s,w)} \in D_{(-w-1,-s-1)}. \quad (2.13)$$

The mappings (2.9), (2.10), (2.12), and (2.13) are all invariant under (i.e., commute with) the restricted Lorentz transformations (2.3).

3. $E(s,w)$ AND TENSORS IN MINKOWSKI SPACE

At a point in Minkowski space, we consider the family of all null directions (the null cone) parametrized by the points on a unit sphere. These (normalized) null vectors have the form

$$l^\mu(\zeta,\bar{\zeta}) = (1/2\sqrt{2}P_0)[1+\zeta\bar{\zeta}, \zeta+\bar{\zeta}, (\zeta-\bar{\zeta})/i, -1+\zeta\bar{\zeta}]$$
$$= \sqrt{2}\pi\,[{}_0Y_{00}, (-1/\sqrt{6})({}_0Y_{11} - {}_0Y_{1-1}), (i/\sqrt{6})({}_0Y_{11} + {}_0Y_{1-1}), (1/\sqrt{3})\,{}_0Y_{10}]. \quad (3.1)$$

For each null direction, one can define three additional vectors:

$$n^\mu = l^\mu + \eth_0\bar{\eth}_0 l^\mu, \quad (3.2a)$$

$$m^\mu = \eth_0 l^\mu, \quad \bar{m}^\mu = \bar{\eth}_0 l^\mu. \quad (3.2b)$$

Together with l^μ they satisfy the standard null tetrad orthogonality conditions

$$l^\mu n_\mu = -m^\mu\bar{m}_\mu = 1, \quad (3.3)$$

all other scalar products vanishing. In addition, for each null direction, we have the completeness relation[10]

$$\eta_{\mu\nu} = 2[l_{(\mu}n_{\nu)} - m_{(\mu}\bar{m}_{\nu)}], \quad (3.4)$$

where $\eta_{\mu\nu}$ is the usual Minkowski metric $\eta_{\mu\nu} = \text{diag}(1,-1,-1,-1)$.

Thus, at a point in Minkowski space we have constructed a family of null tetrad systems that are parametrized by the points on a sphere. That is, the tetrad $(l^\mu(\zeta,\bar{\zeta}), n^\mu(\zeta,\bar{\zeta}), m^\mu(\zeta,\bar{\zeta}), \bar{m}^\mu(\zeta,\bar{\zeta}))$ may be viewed as being a function on the sphere. If we contract a Minkowski tensor, at the point in question, with one or more of the tetrad vectors and allow the direction to vary over the sphere, the tensor will also be converted into a function on the sphere.

Under a restricted Lorentz transformation, i.e., under (2.3), the tetrad transforms as

$$l'^\mu = Kl^\mu, \quad (3.5)$$

$$m'^\mu = e^{i\lambda}(m^\mu + Hl^\mu), \quad (3.6)$$

$$n'^\mu = K^{-1}(n^\mu + \bar{H}m^\mu + H\bar{m}^\mu + H\bar{H}l^\mu), \quad (3.7)$$

where $H \equiv \eth\log K$ with K given by (2.5) and λ by (2.6).

We now wish to discuss the connection between irreducible Lorentz tensors and elements of $E_{(s,w)}$. In particular we will restrict[11] the discussion to trace-free tensors with the following symmetry properties:

$$A^{\mu_1\mu_2\cdots\mu_w} = A^{(\mu_1\mu_2\cdots\mu_w)} \quad (3.8)$$

and

$$B^{\mu_1\nu_1\mu_2\nu_2\cdots\mu_s\nu_s} = B^{([\mu_1\nu_1][\mu_2\nu_2]\cdots[\mu_s\nu_s])} \quad (3.9)$$

and show that

$$A(\zeta,\bar{\zeta}) \equiv A^{\mu_1\cdots\mu_w} l_{\mu_1}(\zeta,\bar{\zeta})\cdots l_{\mu_w}(\zeta,\bar{\zeta}) \quad (3.10)$$

and

$$B(\zeta,\bar{\zeta}) \equiv B^{\mu_1\nu_1\cdots\mu_s\nu_s} l_{\mu_1}m_{\nu_1}\cdots l_{\mu_s}m_{\nu_s} \quad (3.11)$$

are elements of $E_{(0,w)}$ and $E_{(s,s)}$, respectively.

From (3.5) and (3.6) it is clear that

$$A' = K^w A \quad (3.12)$$

and $B' = K^s e^{is\lambda} B$ (3.13)

under (2.3), so that A and B have the correct transformation properties under the restricted Lorentz group.

All that remains now is to show that A and B have the following expansions in spin s spherical harmonics:

$$A = \sum_{l=0}^{w} \sum_{m=-l}^{l} A_{lm}\ {}_0Y_{lm}, \tag{3.14}$$

$$B = \sum_{m=-s}^{s} B_{sm}\ {}_sY_{sm}. \tag{3.15}$$

By definition l_μ consists only of $l=0$ and $l=1$ spin-zero spherical harmonics ($\eth_0^2 l_\mu = 0$) and $l_{[\mu}m_{\nu]}$ is a purely $l=1$ spherical harmonic $[\eth_0(l_{[\mu}m_{\nu]} = \eth_0(l_{[\mu}\eth_0 l_{\nu]}) = 0]$, so that applying $\eth_0^{w+1}$ to A in (3.8) and $\eth_0$ to B in (3.9) leads to $\eth_0^{w+1}A = \eth_0 B = 0$, and (3.14) and (3.15) follow immediately.

As examples, the vector $v = (v^0, v^1, v^2, v^3)$ is related to the function

$$\begin{aligned} v &= v^\mu l_\mu \\ &= \sqrt{2\pi}\ v^0\ {}_0Y_{00} - \sqrt{\frac{2\pi}{3}}\left(\frac{(v' + iv^2)}{\sqrt{2}}\ {}_0Y_{1-1} + v^3\ {}_0Y_{10} \right.\\ &\left.\quad - \frac{(v^1 - iv^2)}{\sqrt{2}}\ {}_0Y_{11}\right) \in E_{(0,1)} \end{aligned} \tag{3.16}$$

and the antisymmetric tensor

$$S = \begin{bmatrix} O & S^{01} & S^{02} & S^{03} \\ -S^{01} & O & S^{12} & S^{13} \\ -S^{02} & -S^{12} & O & S^{23} \\ -S^{03} & -S^{13} & -S^{23} & O \end{bmatrix}$$

is related to the function

$$\begin{aligned} S = S^{\mu\nu}l_\mu m_\nu &= -(\tfrac{2}{3}\pi)^{1/2}(S^{03} + iS^{12})\ {}_1Y_{10} \\ &- (\tfrac{1}{3}\pi)^{1/2}(S^{13} + iS^{02})({}_1Y_{11} + {}_1Y_{1-1}) \\ &+ (\tfrac{1}{3}\pi)^{1/2}(S^{01} + iS^{23})({}_1Y_{11} - {}_1Y_{1-1}) \in E_{(1,1)}. \end{aligned} \tag{3.17}$$

We show, by example, how elementary tensor operations such as products and contractions can be performed on the equivalent functions on the sphere.

Suppose $a \equiv a^\mu l_\mu$ and $b \equiv b^\mu l_\mu$ are elements of $E_{(0,1)}$, $A \equiv A^{\mu_1\cdots\mu_w} l_{\mu_1}\cdots l_{\mu_w} \in E_{(0,w)}$ and $B \equiv B^{\mu\nu}l_\mu m_\nu \in E_{(1,1)}$, then if $C^{\mu\nu} \equiv a^{(\mu}b^{\nu)} - \frac{1}{4}a^\alpha b_\alpha \eta^{\mu\nu}$,

(1) $$C \equiv C^{\mu\nu}l_\mu l_\nu = ab - \tfrac{1}{4}a^\alpha b_\alpha \eta^{\mu\nu}l_\mu l_\nu = ab \in E_{(0,2)}, \tag{3.18}$$

(2) $$\begin{aligned} a^\alpha b_\alpha &= a^\alpha b^\beta \eta_{\alpha\beta} \\ &= a^\alpha b^\beta(l_\alpha n_\beta + n_\alpha l_\beta - m_\alpha \bar{m}_\beta - \bar{m}_\alpha m_\beta) \\ &= a^\alpha b^\beta(2l_\alpha l_\beta + l_\alpha \eth_0\bar{\eth}_0 l_\beta + l_\beta\eth_0\bar{\eth}_0 l_\alpha \\ &\quad - \eth_0 l_\alpha \bar{\eth}_0 l_\beta - \eth_0 l_\beta \bar{\eth}_0 l_\alpha) \\ &= 2ab + a\eth_0\bar{\eth}_0 b + b\eth_0\bar{\eth}_0 a - \eth_0 a\bar{\eth}_0 b - \eth_0 b\bar{\eth}_0 a \\ &\in E_{(0,0)}. \end{aligned} \tag{3.19}$$

Similarly,

(3) $$\begin{aligned} A^{\mu_1\cdots\mu_{w-1}\alpha} a_\alpha l_{\mu_1}\cdots l_{\mu_{w-1}} &= [(w+1)/w]Aa + A\eth_0\bar{\eth}_0 a + (1/w^2)a\eth_0\bar{\eth}_0 A \\ &- (1/w)\eth_0 A\bar{\eth}_0 a - (1/w)\bar{\eth}_0 A\eth_0 a \in E_{(0,w-1)}, \end{aligned} \tag{3.20}$$

where the vanishing of the trace of $A^{\mu_1\cdots\mu_w}$,

$$\begin{aligned} 0 &= A^{\mu_1\cdots\mu_{w-2}\alpha}{}_\alpha l_{\mu_1}\cdots l_{\mu_{w-2}} \\ &= 2A^{\mu_1\cdots\mu_{w-2}\alpha\beta}(l_\alpha l_\beta + l_\alpha\eth_0\bar{\eth}_0 l_\beta - \eth_0 l_\alpha \bar{\eth}_0 l_\beta) l_{\mu_1}\cdots l_{\mu_{w-2}}, \end{aligned} \tag{3.21}$$

has been used:

(4) $$B^{\mu\alpha}a_\alpha l_\mu = \mathrm{Re}[a^3\eth_0(B/a^2)] \in E_{(0,1)}. \tag{3.22}$$

From these examples it is clear that by simply using $\eta_{\mu\nu}$ in the form (3.4), any inner product can be expressed completely in terms of spin and conformally weighted functions on the sphere.

In the next section we will use these and similar results, together with the mapping (2.10), to obtain tensor expressions (with algebraic manipulations) directly from functions in $D_{(-s,-w-2)}$.

4. APPLICATIONS OF THE MAPPING $D_{(-s,-w-2)} \to E_{(s,w)}$

Given the functions $\tilde{T}_{(0,-w-2)} \in D_{(0,w-2)}$ and $\tilde{S}_{(-s,-s-2)} \in D_{(-s,-s-2)}$, by the application of (2.10), we immediately obtain

$$\begin{aligned} \pi_{(0,w)}\tilde{T}_{(0,-w-2)} &= T_{(0,w)} \\ &\equiv T^{\mu_1\cdots\mu_w} l_{\mu_1}\cdots l_{\mu_w} \in E_{(0,w)} \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} \pi_{(s,s)}\tilde{\tilde{S}}_{(-s,-s-2)} &= S_{(s,s)} \\ &\equiv S^{\mu_1\nu_1\cdots\mu_s\nu_s} l_{\mu_1}m_{\nu_1}\cdots l_{\mu_s}m_{\nu_s} \in E_{(s,s)}, \end{aligned} \tag{4.2}$$

where $T^{\mu_1\cdots\mu_w}$ and $S^{\mu_1\nu_1\cdots\mu_s\nu_s}$ are tensors in Minkowski space with the symmetries (3.8) and (3.9), respectively.

Suppose, for example, we have a function $V \in D_{(0,1)}$, from which we should like to obtain another function v, which is in the invariant subspace $E_{(0,1)} \subset D_{(0,1)}$. Because there is no invariant mapping of $D_{(0,1)} \to E_{(0,1)}$ this cannot be accomplished directly. However, we can form the function $V^{-3} \in D_{(0,-3)}$ and then use $\pi_{(0,1)}$ on it to obtain

$$v \equiv v^\mu l_\mu = \pi_{(0,1)}V^{-3}. \tag{4.3}$$

By writing $v = v_0 + v_1$ (where the subscripts refer to the l-values of the subscripted quantities) it is clear from (2.11) that V^{-3} can be expressed as

$$V^{-3} = v_0 - 3v_1 + O(l=2). \tag{4.4}$$

$[O(l=2)$ means the expression is expandable in harmonics with $l \geq 2.]$ Let us examine what happens when we form the function $V^{-3}v \in D_{(0,-2)}$ and apply $\pi_{(0,0)}$ to it:

$$\begin{aligned} \pi_{(0,0)}V^{-3}v &= \pi_{(0,0)}\{[v_0 - 3v_1 + O(l=2)][v_0 + v_1]\} \\ &= \pi_{(0,0)}[v_0^2 - (v_1^2 + \eth_0 v_1\bar{\eth}_0 v_1)_0 + O(l=1)] \\ &= v^2 + v\eth_0\bar{\eth}_0 v - \eth_0 v\bar{\eth}_0 v \\ &= \tfrac{1}{2}v^\alpha v_\alpha, \end{aligned}$$

where (3.19) and the identity

$$v_1^2 \equiv (\tfrac{1}{3}v_1^2 + \tfrac{1}{3}\eth_0 v_1\bar{\eth}_0 v_1)_0 + (\tfrac{2}{3}v_1^2 - \tfrac{1}{3}\eth_0 v_1\bar{\eth}_0 v_1)_2$$

have been used.

If the function V is such that v^μ is timelike and has the normalization

$$\tfrac{1}{2} v^\alpha v_\alpha = v^2 + v\eth_0\bar{\eth}_0 v - \eth_0 v \bar{\eth}_0 v = 1, \tag{4.5}$$

we will refer to v^μ or v as a unit 4-vector and obtain in this case

$$\pi_{(0,0)} V^{-3} v = \tfrac{1}{2} v^\alpha v_\alpha = 1. \tag{4.6}$$

For any unit vector we can show that

$$v^w \equiv \pi_{(0,w)} v^{-w-2} \tag{4.7}$$

is satisfied identically in the following way. Choose $K = v^{-1}$ and thus use the Lorentz transformation (2.3) to put $v = 1$. The proof of (4.7) is trivial in this frame and the Lorentz invariant nature of (2.10) guarantees the result in general.

By putting $w = 1$ in (4.7) we see immediately that the use of (4.3) to define a unit 4-vector $v \in E_{(0,1)}$ from a properly normalized function $V \in D_{(0,1)}$ was a good choice in that it not only still holds but becomes an identity for the case $V = v$.

Essentially the same method that led to the result (4.6) can be used to generalize it to the following: Given a tensor $T^{\mu_1\cdots\mu_w}$ defined from $\tilde{T} \in D_{(0,-w-2)}$ by (4.1) and any vector v, then

$$\pi_{(0,w-1)} \tilde{T} v = [w/(w+1)]\, T^{\mu_1\cdots\mu_{w-1}\alpha} v_\alpha l_{\mu_1} \cdots l_{\mu_{w-1}}. \tag{4.8}$$

The inner product between a vector and a tensor of the type (4.2) must be handled somewhat differently than the previous case, as the following example will illustrate.

Given the antisymmetric 2nd rank tensor $S^{\mu\nu}$ obtained from $\bar{\bar{S}} \in D_{(-1,-3)}$ by

$$\pi_{(1,1)} \bar{\bar{S}} = S = S^{\mu\nu} l_\mu m_\nu \tag{4.9}$$

and the vector $v = v^\mu l_\mu$, how do we go about finding a function from which we can obtain $S^{\mu\alpha} v_\alpha l_\mu$ by applying $\pi_{(0,1)}$ to it?

We know that the function must be linear in both $\bar{\bar{S}}$ and v and have conformal weight $w = -3$. The function must also have spin weight $s = 0$ so that it will be necessary to use the spin raising operator $\eth_0$. But (2.12) tells us that in order to obtain something with good spin and conformal weight, $\eth_0$ can only be applied to a function having $s = w$ in the first place.

The simplest function we can form satisfying all of these requirements is $v^{-1}\eth_0(\bar{\bar{S}}v^2)$. After working out the details we indeed discover that

$$\mathrm{Re}\,\pi_{(0,1)}[v^{-1}\eth_0(\bar{\bar{S}}v^2)] = \tfrac{1}{3} S^{\mu\alpha} v_\alpha l_\mu. \tag{4.10}$$

The proof of (4.10) proceeds as follows: Write $\bar{\bar{S}} = \bar{\bar{S}}_1 + \bar{\bar{S}}_2 + O(l=3)$ and $v = v_0 + v_1$. Then,

$$\begin{aligned}
&\mathrm{Re}\,\pi_{(0,1)}[v^{-1}\eth_0(\bar{\bar{S}}v^2)] \\
&\quad = \mathrm{Re}\,\pi_{(0,1)}[v\eth_0\bar{\bar{S}} + 2\bar{\bar{S}}\eth_0 v] \\
&\quad = \mathrm{Re}\,\pi_{(0,1)}[(v_0 + v_1)(\eth_0\bar{\bar{S}}_1 + \eth_0\bar{\bar{S}}_2) + 2(\bar{\bar{S}}_1 + \bar{\bar{S}}_2)\eth_0 v_1 \\
&\qquad + O(l=2)] \\
&\quad = \mathrm{Re}\,\pi_{(0,1)}[(-\tfrac{1}{3} v_1\eth_0\bar{\bar{S}}_1 + \tfrac{1}{3}\bar{\bar{S}}_1\eth_0 v_1 - \tfrac{1}{6}\bar{\eth}_0 v_1 \eth_0^2 \bar{\bar{S}}_1)_0 \\
&\qquad + (v_0\eth_0\bar{\bar{S}}_1 + \bar{\bar{S}}_1\eth_0 v_1 + \tfrac{1}{2}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_1 + O(l=2)] \\
&\quad = \mathrm{Re}(-\tfrac{1}{3} v\eth_0\bar{\bar{S}}_1 - \tfrac{1}{3}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1) \\
&\quad = \tfrac{1}{3}\mathrm{Re}[v^3\bar{\eth}_0(S/v^2)] \\
&\quad = \tfrac{1}{3} S^{\mu\alpha} v_\alpha l_\mu,
\end{aligned}$$

where we have used (3.22), the identities

$$\begin{aligned}
v_1\eth_0\bar{\bar{S}}_1 &\equiv (\tfrac{1}{3} v_1\eth_0\bar{\bar{S}}_1 - \tfrac{1}{3}\bar{\bar{S}}_1\eth_0 v_1 + \tfrac{1}{6}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_0 \\
&\quad + (\tfrac{2}{3} v_1\eth_0\bar{\bar{S}}_1 + \tfrac{1}{3}\bar{\bar{S}}_1\eth_0 v_1 - \tfrac{1}{6}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_2, \\
v_1\eth_0\bar{\bar{S}}_2 &\equiv (\tfrac{2}{5} v_1\eth_0\bar{\bar{S}}_2 - \tfrac{3}{5}\bar{\bar{S}}_2\eth_0 v_1 + \tfrac{1}{10}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_2)_1 \\
&\quad + (\tfrac{3}{5} v_1\eth_0\bar{\bar{S}}_2 + \tfrac{3}{5}\bar{\bar{S}}_2\eth_0 v_1 - \tfrac{1}{10}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_2)_3, \\
\bar{\bar{S}}_1\eth_0 v_1 &\equiv (\tfrac{1}{3}\bar{\bar{S}}_1\eth_0 v_1 - \tfrac{1}{3} v_1\eth_0\bar{\bar{S}}_1 - \tfrac{1}{6}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_0 \\
&\quad + (\tfrac{1}{2}\bar{\bar{S}}_1\eth_0 v_1 + \tfrac{1}{4}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_1 \\
&\quad + (\tfrac{1}{6}\bar{\bar{S}}_1\eth_0 v_1 + \tfrac{1}{3} v_1\eth_0\bar{\bar{S}}_1 - \tfrac{1}{12}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_1)_2, \\
\bar{\bar{S}}_2\eth_0 v_1 &\equiv (\tfrac{3}{10}\bar{\bar{S}}_2\eth_0 v_1 - \tfrac{1}{5} v_1\eth_0\bar{\bar{S}}_2 - \tfrac{1}{20}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_2)_1 \\
&\quad + (\tfrac{1}{2}\bar{\bar{S}}_2\eth_0 v_1 + \tfrac{1}{12}\bar{\eth}_0 v_1\eth_0^2\bar{\bar{S}}_2)_2 \\
&\quad + (\tfrac{1}{5}\bar{\bar{S}}_2\eth_0 v_1 + \tfrac{1}{5} v_1\eth_0\bar{\bar{S}}_2 - \tfrac{1}{30}\bar{\eth}_0 v_1\eth_0\bar{\bar{S}}_2)_3,
\end{aligned}$$

and the fact from (4.9) that

$$\bar{\eth}_0 S \equiv -\eth_0\bar{\bar{S}}_1.$$

Based on the preceding examples, we see that the general procedure for constructing tensors from the infinite-dimensional representation spaces is as follows: We start with functions that are expandable in spin-weighted spherical harmonics and which transform under (2.3) with spin-weight $s' = -s$ and conformal weight $w' = -w-2$, where s and w are integers satisfying $w \geqq |s|$. Then we use the mapping (2.10) to define tensors in Minkowski space from these functions. Finally, we derive relations between products of the spin-weighted functions and products of the tensors, similar to (4.6), (4.8), and (4.10), that may be useful to us.

We conclude by giving some further examples, which happen to be of particular interest in the study of gravitational radiation reaction is asymptotically flat spaces. Because the method of proof is basically the same in all of these cases, the results will simply be stated without additional proofs. Some results that have already been given will be included for convenience.

Suppose we are given the following functions on the sphere:

$$V = V(u, \zeta, \bar{\zeta}) \in D_{(0,1)},^{12} \tag{4.11a}$$

$$\tilde{p} = \tilde{p}(u, \zeta, \bar{\zeta}) \in D_{(0,-3)}, \tag{4.11b}$$

$$\tilde{T} = \tilde{T}(u, \zeta, \bar{\zeta}) \in D_{(0,-5)}, \tag{4.11c}$$

$$\bar{\bar{S}} = \bar{\bar{S}}(u, \zeta, \bar{\zeta}) \in D_{(-1,-3)} \tag{4.11d}$$

We can define the following tensor quantities from these functions:

$$\pi_{(0,1)} V^{-3} = v \equiv v^\mu l_\mu, \tag{4.12a}$$

$$\pi_{(0,1)}\tilde{p} = p \equiv p^\mu l_\mu, \tag{4.12b}$$

$$\pi_{(0,3)}\tilde{T} = T^{\mu\nu\rho} l_\mu l_\nu l_\rho, \tag{4.12c}$$

$$\pi_{(1,1)}\bar{\bar{S}} = S \equiv S^{\mu\nu} l_\mu m_\nu, \tag{4.12d}$$

$$\pi_{(1,1)}\overline{(-i\bar{S})} = iS = {}^*S^{\mu\nu} l_\mu m_\nu,$$

where $^*S^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} S_{\rho\sigma}$ is the dual of $S^{\mu\nu}$. Normalizing V such that

$$\pi_{(0,0)} V^{-3} v = \tfrac{1}{2} v^\alpha v_\alpha = 1. \tag{4.13}$$

We can then derive the following additional relations among these quantities in the manner of (4.10):

$$\pi_{(0,0)} \tilde{p} v = \tfrac{1}{2} p^\alpha v_\alpha \equiv m, \tag{4.14a}$$

$$\pi_{(0,1)}[v^{-1}\eth_0\bar{\eth}_0(\tilde{p}v^3)] = -2(p - mv) = -2(p^\mu - mv^\mu)l_\mu \tag{4.14b}$$

$$\pi_{(1,1)}[v^{-2}\bar{\eth}_0(\tilde{p}v^3)] = -3(p^\mu v^\nu - p^\nu v^\mu)l_\mu m_\nu, \tag{4.14c}$$

$$\mathrm{Re}\,\pi_{(0,1)}[v^{-1}\eth_0(\dot{\tilde{\bar{S}}}v^2)] = \tfrac{1}{3}\dot{S}^{\mu\alpha}v_\alpha l_\mu, \tag{4.14d}$$

$$\mathrm{Im}\,\pi_{(0,1)}[v^{-1}\eth_0(\dot{\tilde{\bar{S}}}v^2)] = -\tfrac{1}{3}{}^*\dot{S}^{\mu\alpha}v_\alpha l_\mu, \tag{4.14e}$$

$$\pi_{(0,2)}\tilde{T}v = \tfrac{3}{4} T^{\mu\nu\alpha}v_\alpha l_\mu l_\nu \equiv T^{\mu\nu}l_\mu l_\nu, \tag{4.14f}$$

$$\pi_{(0,1)}\tilde{T}v^2 = \tfrac{1}{2} T^{\mu\alpha\beta}v_\alpha v_\beta l_\mu = \tfrac{2}{3} T^{\mu\alpha}v_\alpha l_\mu \equiv T^\mu l_\mu, \tag{4.14g}$$

$$\pi_{(0,0)}\tilde{T}v^3 = \tfrac{1}{4} T^{\alpha\beta\gamma}v_\alpha v_\beta v_\gamma = \tfrac{1}{3} T^{\alpha\beta}v_\alpha v_\beta = \tfrac{1}{2} T^\alpha v_\alpha \equiv T, \tag{4.14h}$$

$$\pi_{(0,1)}\dot{\tilde{T}}v\dot{v} = \tfrac{1}{2}\dot{T}^{\mu\alpha\beta} v_\alpha \dot{v}_\beta l_\mu, \tag{4.14i}$$

$$\pi_{(0,1)}\tilde{T}\dot{v}^2 = \tfrac{1}{2} T^{\mu\alpha\beta}\dot{v}_\alpha \dot{v}_\beta l_\mu, \tag{4.14j}$$

$$\pi_{(0,1)}[v^{-1}\eth_0\bar{\eth}_0(\tilde{T}v^4\dot{v})] = (T^\alpha \dot{v}_\alpha v^\mu - \tfrac{4}{3} T^{\mu\alpha}\dot{v}_\alpha)l_\mu, \tag{4.14k}$$

$$\pi_{(1,1)}[v^{-2}\bar{\eth}_0(\tilde{T}v^4\dot{v})] = -2(T^{\mu\alpha}\dot{v}_\alpha v^\nu - T^{\nu\alpha}\dot{v}_\alpha v^\mu)l_\mu m_\nu, \tag{4.14l}$$

$$\pi_{(1,1)}[\tilde{T}v^3\bar{\eth}_0(\dot{v}/v)] = (T^{\mu\alpha}\dot{v}_\alpha v^\nu - T^{\nu\alpha}\dot{v}_\alpha v^\mu - T^{\mu\alpha}v_\alpha \dot{v}^\nu + T^{\nu\alpha}v_\alpha \dot{v}^\mu)l_\mu m_\nu. \tag{4.14m}$$

---

*This research was supported in part by the National Science Foundation under Grants GP-19378 and GP-22789.

[1]E. T. Newman and T. Unti, J. Math. Phys. (N.Y.) 3, 891 (1962).

[2]A. Exton, E. T. Newman, and R. Penrose, J. Math. Phys. (N.Y.) **10**, 1570 (1969).

[3]B. Aronson, R. Lind, J. Messmer, and E. T. Newman, J. Math. Phys. (N.Y.) **12**, 2462 (1971).

[4]H. Bondi, M. van der Berg, and A. Metzner, Proc. Roy. Soc., London **269**, 21 (1962); R. K. Sachs, Proc. Roy. Soc., London **270**, 103 (1962).

[5]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966). Vol. V.

[6]A. Held, E. T. Newman, and R. Posadas, J. Math. Phys. (N.Y.) **11**, 3145 (1970).

[7]E. T. Newman and R. Penrose, J. Math. Phys. (N.Y.) **7**, 863 (1966); J. N. Goldberg *et al.*, J. Math. Phys. (N.Y.) **8**, 2155 (1967).

[8]Any representation labeled by (s, w) that does not satisfy the condition (2.8) is, in fact, irreducible, infinite dimensional, and equivalent to the representation labeled by $(-s,-w,-2)$.

[9]$\eth\eta_s \equiv 2P_0^{1-s}\,\partial(P_0^s\eta_s)/\partial\zeta$, where P_0 is given by (2.2) and η_s is a function on the sphere expandable in spin s spherical harmonics. Further properties of $\eth_0$ may be found in Ref. 7.

[10]Parentheses enclosing indices denote symmetrization and brackets, antisymmetrization.

[11]The general connection between irreducible tensors and can be most easily obtained from spinor analysis (see HNP); but for our purposes the special cases considered here are sufficient. Actually, by multiplications and contractions of our cases, the general connection can be found.

[12]The timelike coordinate u, which we have introduced at this time for completeness, will have no effect on the techniques that have been developed in this paper since they depend only on the angular coordinates ζ and $\bar{\zeta}$. The derivative of a function with respect to u will be denoted by a dot, i.e., $df/du \equiv \dot{f}$.

Equations of motion for the sources of asymptotically flat spaces*

Robert W. Lind
Physics Department, Syracuse University, Syracuse, New York 13210

James Messmer and Ezra T. Newman
Physics Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 23 June 1972)

We present an exact calculation that leads to the equations of motion (which naturally contain gravitational radiation reaction terms) of a system subject to no external forces. The novelty of our approach lies in the fact that the system is to be considered as the source of an asymptotically flat space and that all the revelant physical quantities such as the velocity v^μ, 4-momentum p^μ, angular momentum-center of mass tensor $S^{\mu\nu}$ (as well as higher moments) are then defined in terms of surface integrals taken at infinity. A subset of the Einstein equations (equivalent to Bondi's supplemantary conditions) then yields the time-evolution equations for these variables.

1. INTRODUCTION

The study of equations of motion in general relativity is of great interest due (among other reasons) to the fact that, among physically interesting classical fields, the Einstein field equations are the only ones which determine the motion of their sources. This is basically a result of the general covariance and nonlinearity of the theory.[1] In other field theories, such as electromagnetic theory, the equations of motion must be postulated separately.

Another important area of study in general relativity has been the investigation of asymptotically flat spaces. Originally, these investigations were based on reasonable guesses concerning the behavior of the metric as spatial infinity was approached. The situation was greatly improved by the work of Bondi *et al.*[2] and then Sachs[3] who, utilizing the idea of approaching infinity along characteristic (or null) surfaces, determined from simple conditions the asymptotic behavior of the metric tensor. A further development was the spin coefficient formalism[4] (NP) and its applications[4,5] (NU) in which the emphasis was shifted from the metric to the Weyl Tensor and its behavior in the vicinity of null infinity.

It is the purpose of this paper to apply the ideas and techniques developed in the study of asymptotically flat spaces to the subject of equations of motion in general relativity. The notation we use is that of the spin coefficient formalism.

Although the essential ideas are very simple, the implementation of them is quite complicated, involving some very powerful results from the theory of infinite-dimensional representations of the Lorentz group. Basically we are considering a finite physical system and studying its properties at future null infinity. The basic physical variables of the system, such as the energy-momentum vector, the angular momentum-center of mass tensor, 4-velocity (and possibly higher moments) are *defined* by certain surface integrals over asymptotic values of the fields. The field equations (or, more specifically, the Bianchi identities or the supplementary conditions in Bondi's terminology) then yield the time evolution of these physical quantities and thereby constitute the equations of motion. We emphasize that we are concerned with the motion of a single composite system and not with the relative motion of its component parts. (We are thus, for example, not dealing with the two-body problem.) The final result (arrived at with no approximations) will resemble the equations of motion of a free particle with intrinsic angular momentum, but modified by radiation reaction terms arising from accelerations and changing quadrupole and higher moments.

The entire idea is almost perfectly analogous to *defining* electric charge in classical electrodynamics by a surface integral (Gauss' Theorem) at infinity and then using the vacuum Maxwell equations to prove that it is conserved.

In Sec. 2 we present a detailed review of the properties of asymptotically flat spaces. Although it is in essence a review of (NU), considerable modification and simplification of notation is achieved through the use of the operator $\eth$ and the notion of spin weighted functions. Here we shall find the time development equations for the tetrad components of the Weyl tensor which implicitly contain the equations of motion.

In Sec. 3 we investigate the asymptotic symmetry group, the so-called Bondi–Metzner–Sach (BMS) group[2,3,5–9] expressed in a very general null coordinate system. Utilizing the idea of the Winicour–Tamburino linkages[10] and the generators of the BMS group we find, in Sec. 4, integral expressions which are identified (by definition) with the energy-momentum 4-vector and the angular momentum-center of mass tensor of the source. They will be referred to as physical quantities or tensors. We point out that, although these quantities transform properly under the homogeneous Lorentz group (which is well defined in asymptotically flat spaces) and agree with expressions obtained from source properties in the linear theory, one cannot say with complete confidence that they are the unique expressions for the physical quantities. Any reasonable modification of the definitions would, however, not fundamentally change our final results.

In Sec. 5, there is first a review of some results from the preceding paper,[11] which relate (by integral operators) infinite-dimensional representations of the homogeneous Lorentz group to finite-dimensional ones. These integral operators are then related to the integral expressions (Winicour–Tamburino linkages), which have been identified with physical quantities. Finally, by applying the integral operators to the Bianchi Identities (after much manipulation), we obtain the evolution equations (or equations of motion) for the physical quantities.

Section 6 is devoted to a discussion of the coordinate freedom, or more precisely, to an attempt to eliminate the coordinate freedom. First, we use coordinate conditions[12] (different from the Bondi type), which are associated with "canonical" slicings of null infinity. These canonical slicings are generalizations of the slicings of future null infinity in Minkowski space, produced by the families of light cones emanating from arbitrary timelike world lines. From these slicings we conjecture that a unique one exists which can naturally be called the center of mass coordinate system.

2. THE ASYMPTOTICALLY FLAT SOLUTIONS

In this section we present a review of the asymptotically flat solutions. In a four-dimensional Riemannian space of signature $(+,-,-,-)$, one introduces a null tetrad $Z_{a\mu} = (l_\mu, n_\mu, m_\mu, \overline{m}_\mu)$ composed of two real null vectors l_μ and n_μ and two complex null vectors m_μ and $\overline{m}_\mu$, with the pseudo-orthogonality properties

$$l_\mu n^\mu = -m_\mu \overline{m}^\mu = 1, \tag{2.1}$$

all other inner products vanishing.

Equation (2.1) implies the completeness relation

$$g^{\mu\nu} = l^\mu n^\nu + n^\mu l^\nu - m^\mu \overline{m}^\nu - \overline{m}^\mu m^\nu, \tag{2.2}$$

or $g^{\mu\nu} = Z_m{}^\nu Z_n{}^\nu \eta^{mn}$, where η^{mn} is the null Minkowski metric

$$\eta^{mn} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} = \eta_{mn}, \tag{2.3}$$

which is used to raise and lower tetrad indices.[13]

A null coordinate system and its associated null tetrad system $Z_{a\mu}$ is introduced by first considering a family of null hypersurfaces labeled by a parameter $u = \text{const}$, i.e.,

$$g^{\mu\nu} u_{,\mu} u_{,\nu} = 0. \tag{2.4}$$

(A comma denotes ordinary partial differentiation and a semicolon, covariant differentiation.) We then choose the first tetrad vector l^μ to be orthogonal to these hypersurfaces, so that

$$l_\mu = u_{,\mu}. \tag{2.5}$$

Since these are null hypersurfaces, the vector l^μ will also be tangent to a family of null geodesics in the hypersurface and, therefore,

$$l^\mu{}_{;\nu} l^\nu \propto l^\mu. \tag{2.6}$$

It is convenient to choose for coordinates $x^0 = u$ and $x^1 = r$, where r is an affine parameter along the geodesics with tangent vector l^μ. Thus, $l_\mu = \delta^0_\mu$, $l^\mu = dx^\mu/dr = g^{\mu\nu} u_{,\nu} = g^{\mu 0} = \delta^\mu_1$, and $l^\mu{}_{;\nu} l^\nu = 0$ since r is affine.

The two remaining coordinates x^2 and x^3 will label directions in the null surface, i.e., they will label the geodesics on each hypersurface $u = \text{const}$.

In order to satisfy Eqs. (2.1) with the above conditions, the tetrad system must have the form

$$l_\mu = \delta^0_\mu, \quad l^\mu = \delta^\mu_1, \tag{2.7a}$$

$$n^\mu = \delta^\mu_0 + U\delta^\mu_1 + X^A \delta^\mu_A, \tag{2.7b}$$

$$m^\mu = \omega \delta^\mu_1 + \xi^A \delta^\mu_A, \tag{2.7c}$$

where ω, ξ^A, U, and X^A are arbitrary functions of the coordinates.

The completeness relation (2.2) then enables us to write the metric as

$$g = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & g^{11} & g^{12} & g^{13} \\ 0 & g^{21} & & \\ & & g^{AB} & \\ 0 & g^{31} & & \end{bmatrix}, \tag{2.8a}$$

where

$$g^{11} = 2(U - \omega\overline{\omega}), \tag{2.8b}$$

$$g^{1A} = X^A - (\xi^A \overline{\omega} + \overline{\xi}^A \omega), \tag{2.8c}$$

$$g^{AB} = -(\xi^A \overline{\xi}^B + \overline{\xi}^A \xi^B). \tag{2.8d}$$

Along with the above conditions on l^μ we impose the further condition that n^μ and m^μ be parallely propagated along the null geodesics.

From the tetrad one defines the Ricci rotation coefficients

$$\gamma^{mnp} = Z^m_{\mu;\nu} Z^{n\mu} Z^{p\nu} \tag{2.9}$$

and then the spin coefficients

$$\begin{aligned}
\kappa &= \gamma_{131} = l_{\mu;\nu} m^\mu l^\nu, & \nu &= -\gamma_{242} = -n_{\mu;\nu} \overline{m}^\mu n^\nu, \\
\rho &= \gamma_{134} = l_{\mu;\nu} m^\mu \overline{m}^\nu, & \mu &= -\gamma_{243} = -n_{\mu;\nu} \overline{m}^\mu m^\nu, \\
\sigma &= \gamma_{133} = l_{\mu;\nu} m^\mu m^\nu, & \lambda &= -\gamma_{244} = -n_{\mu;\nu} \overline{m}^\mu \overline{m}^\nu, \\
\tau &= \gamma_{132} = l_{\mu;\nu} m^\mu n^\nu, & \pi &= -\gamma_{241} = -n_{\mu;\nu} \overline{m}^\mu l^\nu,
\end{aligned} \tag{2.10}$$

$$\begin{aligned}
\alpha &= \tfrac{1}{2}(\gamma_{124} - \gamma_{344}) = \tfrac{1}{2}(l_{\mu;\nu} n^\mu \overline{m}^\nu - m_{\mu;\nu} \overline{m}^\mu \overline{m}^\nu), \\
\beta &= \tfrac{1}{2}(\gamma_{123} - \gamma_{343}) = \tfrac{1}{2}(l_{\mu;\nu} n^\mu m^\nu - m_{\mu;\nu} \overline{m}^\mu m^\nu), \\
\gamma &= \tfrac{1}{2}(\gamma_{122} - \gamma_{342}) = \tfrac{1}{2}(l_{\mu;\nu} n^\mu n^\nu - m_{\mu;\nu} \overline{m}^\mu n^\nu), \\
\epsilon &= \tfrac{1}{2}(\gamma_{121} - \gamma_{341}) = \tfrac{1}{2}(l_{\mu;\nu} l^\nu - m_{\mu;\nu} \overline{m}^\mu l^\nu).
\end{aligned}$$

where $\kappa = \pi = \epsilon = 0$, $\rho = \overline{\rho}$ and $\tau = \overline{\alpha} + \beta$ due to the conditions on l^μ and the parallel propagation of the tetrad.

Tetrad components of a tensor are defined by

$$A_{mn}\cdots = A^{\mu\nu}\cdots Z_{m\mu} Z_{n\nu}\cdots. \tag{2.11}$$

In particular, when applied to the Weyl tensor we have

$$\begin{aligned}
\psi_0 &= -C_{\mu\nu\rho\sigma} l^\mu m^\nu l^\rho m^\sigma, \\
\psi_1 &= -C_{\mu\nu\rho\sigma} l^\mu n^\nu l^\rho m^\sigma, \\
\psi_2 &= -C_{\mu\nu\rho\sigma} \overline{m}^\mu n^\nu l^\rho m^\sigma, \\
\psi_3 &= -C_{\mu\nu\rho\sigma} \overline{m}^\mu n^\nu l^\rho n^\sigma, \\
\psi_4 &= -C_{\mu\nu\rho\sigma} \overline{m}^\mu n^\nu \overline{m}^\rho n^\sigma.
\end{aligned} \tag{2.12}$$

Intrinsic or directional derivatives have the form

$$\begin{aligned}
D\varphi &= \varphi_{;\mu} l^\mu, & \Delta\varphi &= \varphi_{;\mu} n^\mu, \\
\delta\varphi &= \varphi_{;\mu} m^\mu, & \overline{\delta}\varphi &= \varphi_{;\mu} \overline{m}^\mu,
\end{aligned} \tag{2.13}$$

where

$$\begin{aligned}
D &= \frac{\partial}{\partial r}, & \Delta &= \frac{\partial}{\partial u} + U\frac{\partial}{\partial r} + X^A \frac{\partial}{\partial x^A}, \\
\delta &= \omega\frac{\partial}{\partial r} + \xi^A \frac{\partial}{\partial x^A}, & \overline{\delta} &= \overline{\omega}\frac{\partial}{\partial r} + \overline{\xi}^A \frac{\partial}{\partial x^A}.
\end{aligned} \tag{2.14}$$

The spin coefficient formalism (NP) consists of three sets of first order differential equations for the three sets of variables, the Weyl tensor components, the spin coefficients, and the tetrad components (also referred to as the metric variable), which are equivalent to the Einstein field equations.

Before summarizing the asymptotically flat solutions of these equations, we review[14] some properties of the differential operator edth $\eth$ and the idea of spin weighted functions. This will enable us to express the solution in a more compact form.

Consider an arbitrary two-dimensional surface with the metric in conformally flat form, i.e.,

$$ds^2 = \frac{1}{P^2}[(dx^2)^2 + (dx^3)^2] = \frac{d\zeta d\bar{\zeta}}{P^2}, \qquad (2.15)$$

$$\zeta = -(x^2 - ix^3).$$

(The coordinates x^2 and x^3 will eventually represent the coordinates introduced earlier to label a null geodesic. The peculiar form for the complex coordinate ζ has been chosen purely to make the conventions used here conform to those used in other papers.) Let

$$m^A = (1/\sqrt{2})(a^A + ib^A), \qquad (2.16)$$

where a^A and b^A are orthonormal tangent vectors to the surface. Under a rotation in the tangent plane we have

$$m^{A'} = e^{i\varphi} m^A. \qquad (2.17)$$

Any function η defined on the 2-surface which transforms under (2.17) as

$$\eta' = e^{is\varphi}\eta, \qquad (2.18)$$

is said to have spin weight s.

The operators $\eth$ and $\bar{\eth}$ are then defined as

$$\eth\eta = 2P^{1-s}\frac{\partial(P^s\eta)}{\partial\bar{\zeta}}, \qquad (2.19a)$$

$$\bar{\eth}\eta = 2P^{1+s}\frac{\partial(P^{-s}\eta)}{\partial\zeta}, \qquad (2.19b)$$

where η is any spin weight s function. $\eth(\bar{\eth})$ has the important property that it raises (lowers) the spin weight by unity. Their commutation relation is

$$(\bar{\eth}\eth - \eth\bar{\eth})\eta = 2s\eta\eth\bar{\eth}\ \ln P. \qquad (2.20)$$

For any suitably regular function η of integral spin weight $s > 0$ $(s < 0)$, there exists a function W of spin wieght zero such that

$$\eta = \eth^s W \quad (\eta = \bar{\eth}^{-s} W). \qquad (2.21)$$

This fact can be used to define the "electric" and "magnetic" parts of η, η_e, and η_m:

$s > 0$: $\qquad s < 0$:

$$\eta_e = \eth^s(\mathrm{Re}W), \quad \eta_e = \bar{\eth}^{-s}(\mathrm{Re}W), \qquad (2.22)$$

$$\eta_m = i\eth^s(\mathrm{Im}W), \quad \eta_m = i\bar{\eth}^{-s}(\mathrm{Im}W), \qquad (2.23)$$

where $\eta_e + \eta_m = \eta$.

If $P = P_0 = \frac{1}{2}(1 + \zeta\bar{\zeta})$, i.e., the 2-surface is the unit sphere, then the corresponding operator $\eth_0$ and the spherical harmonics can be used to define the spin s spherical harmonics ${}_sY_{lm}$. (See Ref. 14 for the definition.)

Any suitably regular function on the sphere η with spin weight s can be expanded in the form

$$\eta = \sum_{l=|s|}^{\infty}\sum_{m=-l}^{l} A_{lm}\ {}_sY_{lm}. \qquad (2.24)$$

The ${}_sY_{lm}$ are also eigenfunctions of $\bar{\eth}_0\eth_0$,

$$\bar{\eth}_0\eth_0\ {}_sY_{lm} = [-l(l+1) + s(s+1)]_s Y_{lm}. \qquad (2.25)$$

Finally, it can be shown that

$$\eth_0\ {}_sY_{sm} = 0 = \bar{\eth}\ {}_{-s}Y_{sm}, \qquad (2.26)$$

i.e., $\eth_0$ annihilates a function with spin weight $s = l$, and $\bar{\eth}_0$ annihilates a function with spin weight $s = -l$.

With the aid of $\eth$, defined above, and the complex coordinate $\zeta = -x^2 + ix^3$ we are able to present the following summary of the asymptotically flat solutions of the spin coefficient equations:

components of the Weyl tensor:

$$\psi_0 = \psi_0^0 r^{-5} + O(r^{-6}), \qquad (2.27a)$$

$$\psi_1 = \psi_1^0 r^{-4} + \bar{\eth}\psi_0^0 r^{-5} + O(r^{-6}), \qquad (2.27b)$$

$$\psi_2 = \psi_2^0 r^{-3} + \bar{\eth}\psi_1^0 r^{-4} + O(r^{-5}), \qquad (2.27c)$$

$$\psi_3 = \psi_3^0 r^{-2} + \bar{\eth}\psi_2^0 r^{-3} + O(r^{-4}), \qquad (2.27d)$$

$$\psi_4 = \psi_4^0 r^{-1} + \bar{\eth}\psi_3^0 r^{-2} + O(r^{-3}); \qquad (2.27e)$$

spin coefficients:

$$\rho = -r^{-1} - \sigma^0\bar{\sigma}^0 r^{-3} + O(r^{-5}), \qquad (2.28a)$$

$$\sigma = \sigma^0 r^{-2} + (\bar{\sigma}^0\sigma^{0^2} - \tfrac{1}{2}\psi_0^0)r^{-4} + O(r^{-5}), \qquad (2.28b)$$

$$\alpha = \alpha^0 r^{-1} + \bar{\sigma}^0\bar{\alpha}^0 r^{-2} + \sigma^0\bar{\sigma}^0\alpha^0 r^{-3} + O(r^{-4}), \qquad (2.28c)$$

$$\beta = -\bar{\alpha}^0 r^{-1} - \sigma^0\alpha^0 r^{-2} - (\sigma^0\bar{\sigma}^0\bar{\alpha}^0 + \tfrac{1}{2}\psi_1^0)r^{-3} + O(r^{-4}), \qquad (2.28d)$$

$$\tau = -\tfrac{1}{2}\psi_1^0 r^{-3} + \tfrac{1}{6}(\sigma^0\bar{\psi}_1^0 - 2\bar{\eth}\psi_0^0)r^{-4} + O(r^{-5}), \qquad (2.28e)$$

$$\lambda = \lambda^0 r^{-1} - \bar{\sigma}^0\mu^0 r^{-2} + (\sigma^0\bar{\sigma}^0\lambda^0 + \tfrac{1}{2}\bar{\sigma}^0\psi_2^0)r^{-3} + O(r^{-4}), \qquad (2.28f)$$

$$\mu = \mu^0 r^{-1} - (\sigma^0\lambda^0 + \psi_2^0)r^{-2} + (\sigma^0\bar{\sigma}^0\mu^0 - \tfrac{1}{2}\bar{\eth}\psi_1^0)r^{-3} + O(r^{-4}), \qquad (2.28g)$$

$$\gamma = \gamma^0 - \tfrac{1}{2}\psi_2^0 r^{-2} + (\tfrac{1}{6}\psi_1^0\alpha^0 - \tfrac{1}{6}\bar{\psi}_1^0\bar{\alpha}^0 - \tfrac{1}{3}\bar{\eth}\psi_1^0)r^{-3} + O(r^{-4}), \qquad (2.28h)$$

$$\nu = \nu^0 - \psi_3^0 r^{-1} - \tfrac{1}{2}\bar{\eth}\psi_2^0 + O(r^{-3}); \qquad (2.28i)$$

the metric variables:

$$U = (\dot{P}/P)r - \eth\bar{\eth}\ \ln P - \tfrac{1}{2}(\psi_2^0 + \bar{\psi}_2^0)r^{-1} - \tfrac{1}{6}(\bar{\eth}\psi_1^0 + \eth\bar{\psi}_1^0)r^{-2} + O(r^{-3}), \qquad (2.29a)$$

$$X^A = \tfrac{1}{6}(\psi_1^0\bar{\xi}^{0A} + \bar{\psi}_1^0\xi^{0A})r^{-3} + O(r^{-4}), \qquad (2.29b)$$

$$\xi^A = \xi^{0A}r^{-1} - \sigma^0\bar{\xi}^{0A}r^{-2} + \sigma^0\bar{\sigma}^0\xi^{0A}r^{-3} + O(r^{-4}), \qquad (2.29c)$$

where $\xi^{0A} = (P, iP)$,

$$\omega = \omega^0 r^{-1} - (\sigma^0\bar{\omega}^0 + \tfrac{1}{2}\psi_1^0)r^{-2} + O(r^{-3}). \qquad (2.29d)$$

(A dot denotes differentiation with respect to u.)

Further relations between the coefficients of the different powers of r are

$$\gamma^0 = -\tfrac{1}{2}(\dot{P}/P), \tag{2.30a}$$

$$\alpha^0 = -\partial P/\partial\bar{\zeta} = -\tfrac{1}{2}\bar{\eth}\ \ln P, \tag{2.30b}$$

$$\nu^0 = \bar{\eth}\dot{P}/P, \tag{2.30c}$$

$$\omega^0 = -\bar{\eth}\sigma^0, \tag{2.30d}$$

$$\lambda^0 = \dot{\bar{\sigma}}^0 - \bar{\sigma}^0\dot{P}/P, \tag{2.30e}$$

$$\mu^0 = U^0 = -\eth\bar{\eth}\ \ln P, \tag{2.30f}$$

$$(\psi_2^0 - \bar{\psi}_2^0) = \bar{\eth}^2\sigma^0 + \bar{\sigma}^0\bar{\lambda}^0 - \eth^2\bar{\sigma}^0 - \sigma^0\lambda^0, \tag{2.30g}$$

$$\psi_3^0 = \bar{\eth}\eth\bar{\eth}\ \ln P + \eth\lambda^0, \tag{2.30h}$$

$$\psi_4^0 = -\bar{\eth}^2(\dot{P}/P) - \dot{\lambda}^0 + 2(\dot{P}/P)\lambda^0. \tag{2.30i}$$

Finally, we have the dynamical relations

$$\dot{\psi}_0^0 - 3(\dot{P}/P)\psi_0^0 = -\eth\psi_1^0 + 3\sigma^0\psi_2^0, \tag{2.30j}$$

$$\dot{\psi}_1^0 - 3(\dot{P}/P)\psi_1^0 = -\eth\psi_2^0 + 2\sigma^0\psi_3^0, \tag{2.30k}$$

$$\dot{\psi}_2^0 - 3(\dot{P}/P)\psi_2^0 = -\eth\psi_3^0 + \sigma^0\psi_4^0. \tag{2.30l}$$

If the metric tensor is constructed from the tetrad components, using Eqs. (2.8), the 2-surface, u and r constant, in the limit as $r \to \infty$, has a metric of the form

$$2\lim_{r\to\infty}(r^{-2}ds^2) = \frac{d\zeta d\bar{\zeta}}{P^2}.$$

We assume that P can be written as

$$P(u,\zeta,\bar{\zeta}) = P_0 V(u,\zeta,\bar{\zeta}) = \tfrac{1}{2}(1+\zeta\bar{\zeta})V(u,\zeta,\bar{\zeta}), \tag{2.31}$$

where V is to be a regular function, with no zeros, on the sphere, i.e., expandable in spherical harmonics. $V-1$ can be interpreted as the deviation of this limiting 2-surface from sphericity.

A second interpretation, which we mention without proof, is the connection of V with the rate of change (at infinity) of our null coordinate system with respect to a Bondi type null coordinate u_B, namely

$$V^{-1} = \frac{\partial u}{\partial u_B}.$$

Closely associated with this interpretation is the fact that using the present type of coordinate system in flat space yields[15]

$$V = v \equiv \sum_{l=0}^{1} v_{lm}Y_{lm}(\zeta,\bar{\zeta}),$$

where the four v_{lm} are in one-to-one correspondence with the velocity vector of the world line defined by the apex of the null cones, $u = $ const. Later we shall show that even though the $u = $ cont null surfaces in asymptotically flat spaces are not exact cones and do not possess an apex, it will still be possible to interpret them as if they did and to extract by Lorentz invariant operations on V a 4-velocity v^μ. Under appropriate circumstances v^μ will be interpreted as the velocity of the center of the source.

Returning to the main discussion, if we use the operators $\eth_0$ and $\bar{\eth}_0$ defined by

$$\eth_0\eta = 2P_0^{1-s}\frac{\partial(P_0^{\ s}\eta)}{\partial\zeta}, \tag{2.32a}$$

$$\bar{\eth}_0\eta = 2P_0^{1+s}\frac{\partial(P_0^{\ -s}\eta)}{\partial\bar{\zeta}}, \tag{2.32b}$$

where η is any spin weight s function, and introduce the $s = -2$ function

$$R = \frac{\lambda^0}{V^2} + \frac{1}{V}\bar{\eth}_0^2 V, \tag{2.33}$$

we find that several of the equations in the set (2.30) may be greatly simplified. For instance, (2.30i) and (2.30h) become

$$\psi_4^0 = -V^2\dot{R} \tag{2.34}$$

and

$$\psi_3^0 = \eth(V^2R) = V^3\eth_0 R. \tag{2.35}$$

Furthermore, if we define the $s = 0$ function

$$\tilde{p} = -[\psi_2^0 - \bar{\eth}^2\sigma^0 + \sigma^0\dot{\bar{\sigma}}^0 - \sigma^0\bar{\sigma}^0(\dot{V}/V)]V^{-3}, \tag{2.36}$$

then after considerable manipulation (2.30l) and (2.30g) become

$$\dot{\tilde{p}} = -\eth_0^2\bar{\eth}_0^2 V + \eth_0^2(RV) + \bar{\eth}_0^2(\bar{R}V) - VR\bar{R} \tag{2.37}$$

and

$$\tilde{p} - \bar{\tilde{p}} = 0. \tag{2.38}$$

We note for future use that the first three terms on the right-hand side of (2.37) cannot possess any $l = 0$ and $l = 1$ spherical harmonic terms. This follows from the annihilation properties of $\eth_0^2$.

Finally, by introducing the $s = -1$ function

$$6\bar{\tilde{S}} = -[\bar{\psi}_1^0 - \tfrac{1}{2}\bar{\eth}(\sigma^0\bar{\sigma}^0) - \bar{\sigma}^0\bar{\eth}\sigma^0]V^{-3}, \tag{2.39}$$

the complex conjugate of (2.30k) becomes

$$\begin{aligned} 6\dot{\bar{\tilde{S}}} = &-V^{-3}\bar{\eth}(\tilde{p}V^3) + \sigma^0\bar{\sigma}^0V^{-3}\bar{\eth}(\dot{V}/V) \\ &-V^{-3}[2\bar{\sigma}^0\eth\bar{\eth}\eth\ \ln P + \tfrac{3}{2}\bar{\sigma}^0\bar{\eth}\dot{\sigma}^0 + \tfrac{1}{2}\dot{\sigma}^0\bar{\eth}\bar{\sigma}^0 \\ &-\bar{\eth}\eth^2\bar{\sigma}^0 - \tfrac{3}{2}\dot{\bar{\sigma}}^0\bar{\eth}\sigma^0 - \tfrac{1}{2}\sigma^0\bar{\eth}\dot{\bar{\sigma}}^0]. \end{aligned} \tag{2.40}$$

In order to keep the motivation clear, in the midst of this over-abundance of definitions, we anticipate results of the next several sections by pointing out that $\tilde{p}$ and $\tilde{S}$ will play a basic role. More precisely, we will argue that the four coefficients of the $l = 0$ and $l = 1$ spherical harmonics in the expansion of $\tilde{p}$ become the 4-momentum of the source [note the importance of the reality of $\tilde{p}$, (2.38)] and that the three complex (six real) coefficients of the $l = 1$ spherical harmonics in the expansion of $\tilde{S}$ become the angular momentum-center of mass tensor. It should, therefore, already be clear that Eqs. (2.37) and (2.40) implicitly contain the equations of motion. The analysis involved in obtaining them explicitly will be quite complicated.

As a last result of this section we write out the asymptotic form of the finite coordinate transformation (connected with the identity) which preserves all the relations developed up to this point [see (NU)]:

$$u' = G(u, \zeta, \bar{\zeta}) + O(r^{-1}),$$
$$r' = \dot{G}^{-1} r + O(1), \qquad (2.41)$$
$$\zeta' = (a\zeta + b)/(c\zeta + d) + O(r^{-1}),$$

where in principle the order symbols are determined in terms of $G(u, \zeta, \bar{\zeta})$ and the constants $a, b, c,$ and d. Notice the natural appearance of the homogeneous Lorentz group through the fractional linear transformation.

Although it is possible by the proper choice of G to make $V = 1$ and thereby arrive at Bondi coordinates, we will avoid doing this. Instead we leave the coordinate freedom open for the time being. In the final section we will argue for a different choice, dictated by a center of mass condition.

3. THE ASYMPTOTIC SYMMETRY GROUP

In this work we consider that the asymptotic coordinate group (discussed in the previous section) and the asymptotic symmetry group (BMS) are to be regarded as two distinct entities. This may be thought of as being analogous, for example, to the distinction made in three-dimensional Euclidean space between the transformations between arbitrary coordinate systems and the symmetry transformations generated by solutions to the Killing equation. In fact, the asymptotic symmetry group will not be thought of as a group of transformations at all, but rather as a set of descriptors (generators) from which we shall be able to define energy–momentum, angular momentum, etc. (We neglect the difficulties, which are not insurmountable, associated with the fact that the homogeneous Lorentz group is not an invariant subgroup of BMS.)

The infinitesimal BMS group is obtained from the asymptotic Killing equation

$$\underset{\xi}{£} g_{\mu\nu} \equiv \xi_{\mu;\nu} + \xi_{\nu;\mu} = O(r^{-n}), \qquad (3.1)$$

$$(\underset{\xi}{£} g_{\mu\nu}) l^\nu \equiv (\xi_{\mu;\nu} + \xi_{\nu;\mu}) l^\nu = 0, \qquad (3.2)$$

where n differs with the choice of components. (See Refs. 5 and 9.) Solutions to (3.1) may be found either by transforming the solutions, already known[5,9] in a Bondi coordinate system, to our coordinate system, or by direct integration. (See Appendix A.)

The results may be summarized by writing

$$\xi^\mu = A l^\mu + B n^\mu + C \bar{m}^\mu + \bar{C} m^\mu, \qquad (3.3)$$

where

$$A = A_1 r + A_0 + A_{-1} r^{-1} + O(r^{-2}), \qquad (3.4)$$
$$B = B_0, \qquad (3.5)$$
$$C = C_1 r + C_0 + C_{-1} r^{-1} + O(r^{-2}), \qquad (3.6)$$

and

$$A_1 = -(1/V)(B_0 V)^{\cdot},$$
$$A_0 = \eth\bar{\eth} B_0 + B_0 \eth\bar{\eth} \ln P, \qquad (3.7)$$
$$A_{-1} = \tfrac{1}{2}[B_0(\psi_2^0 + \bar{\psi}_2^0) + \bar{C}_1 \psi_1^0 + C_1 \bar{\psi}_1^0],$$
$$C_1 = c(\zeta, \bar{\zeta})/V, \quad \text{with } \eth_0 c = 0,$$
$$C_0 = \eth B_0 + \bar{C}_1 \sigma^0, \quad C_{-1} = 0, \qquad (3.8)$$
$$B_0 = b(\zeta, \bar{\zeta})/V - (1/2V)\int_0^u V^3[\eth_0(\bar{c} V^{-2}) + \bar{\eth}_0(c V^{-2})] du. \qquad (3.9)$$

Note that the only freedom in the solution is in $b(\zeta, \bar{\zeta})$, an arbitrary function on the sphere, which is the super translation freedom, and solutions to $\eth_0 c = 0$, which correspond to the homogeneous Lorentz transformation freedom. (See Ref. 16 for details). In fact, if we return to a Bondi system, i.e., $V = 1$, then (3.3) becomes the conventional form of the infinitesimal BMS transformation. In particular (3.9) becomes

$$B_0 = b(\zeta, \bar{\zeta}) + ku, \qquad (3.10)$$

where $k(\zeta, \bar{\zeta}) = -\tfrac{1}{2}(\eth_0 \bar{c} + \bar{\eth}_0 c)$ is the infinitesimal conformal conformal factor.[16]

In the next section the asymptotic Killing vectors, Eqs. (3.3)–(3.9), are used in the Winicour–Tamburino linkage expressions to obtain definitions of the energy-momentum 4-vector and the angular momentum-center of mass tensor.

4. THE WINICOUR-TAMBURINO LINKAGES

In this section we shall use the linkages defined in Ref. 10 to obtain definitions for the energy–momentum and angular momentum. When making use of the equations from this work, some caution must be exercised. Tamburino and Winicour use a metric with signature + 2, whereas the present metric has signature − 2. Also, only one of the two real null vectors k^μ and m^μ used in Winicour–Tamburino can be carried over directly into the present notation. Their k^μ can be equated to our l^μ. However, the m^μ of Winicour–Tamburino cannot be equated to n^μ for the following reason. In order to form the linkage integral correctly, the two complex null vectors must have no components in the x^0, x^1 directions. At the same time the tetrad must satisfy the pseudo-orthonormality conditions (2.1). In the present notation, the complex null vectors m^μ and $\bar{m}^\mu$, given by Eq. (2.7c) do contain components in the x^1 direction. Therefore, to make use of the linkage expressions we must find one real and two complex null vectors, denoted by $n^{\mu\prime}, m^{\mu\prime}$ and $\bar{m}^{\mu\prime}$, such that $m^{\mu\prime}$ and $\bar{m}^{\mu\prime}$ have no x^1 component, and $l^\mu, n^{\mu\prime}, m^{\mu\prime}$ and $\bar{m}^{\mu\prime}$ satisfy the pseudo-orthonormality conditions (2.1).

Since the set of vectors $l^\mu, n^{\mu\prime}, m^{\mu\prime}$, and $\bar{m}^{\mu\prime}$ must satisfy Eqs. (2.1), it is clear that they must be related to the tetrad l^μ, n^μ, m^μ, and $\bar{m}^\mu$ by a null rotation about l^μ given by

$$l^{\mu\prime} = l^\mu, \qquad (4.1a)$$

$$n^{\mu\prime} = n^\mu + \bar{H} m^\mu + H \bar{m}^\mu + H\bar{H} l^\mu, \qquad (4.1b)$$

$$m^{\mu\prime} = m^\mu + H l^\mu. \qquad (4.1c)$$

Thus, $m^{\mu\prime} = \omega \delta_1{}^\mu + \xi \delta_A{}^\mu + H \delta_1{}^\mu$ and by the choice $H = -\omega$ we can eliminate the component in the x^1 direction contained in $m^{\mu\prime}$ and $\bar{m}^{\mu\prime}$. This implies the following:

$$n^{\mu\prime} = n^\mu - \bar{\omega} m^\mu - \omega \bar{m}^\mu + \omega\bar{\omega} l^\mu. \qquad (4.2)$$

After correcting for the difference in the signatures of the metrics, it is possible to write the linkage integral in the present notation as

$$L_\xi(\mathscr{I}^+) = \lim_{r\to\infty} \int (\xi^{[\mu;\nu]} + \xi^\rho{}_{;\rho} l^{[\mu} n'^{\nu]}) l_\mu n'_\nu dS, \qquad (4.3)$$

where $L(\mathscr{I}^+)$ is the linkage evaluated at future null infinity $(\mathscr{I}^+)$.

By making use of the radial dependence of the spin coefficients, Eqs. (2.28), and the radial dependence of the tetrad components of the descriptors of the asymptotic symmetry group (previous section), Eq. (4.3) becomes, after a very tedious calculation,

$$L_\xi(\mathscr{I}^+) = \tfrac{1}{2}\int\{b[\psi_2^0 + \bar{\psi}_2^0 + \sigma^0\lambda^0 + \bar{\sigma}^0\bar{\lambda}^0 - \eth^2\bar{\sigma}^0 - \bar{\eth}^2\sigma^0]V^{-3} + c[2\bar{\psi}_1^0 - 2\bar{\sigma}^0\bar{\eth}\sigma^0 - \bar{\eth}(\sigma^0\bar{\sigma}^0)]V^{-3} + \bar{c}[2\psi_1^0 - 2\sigma^0\eth\bar{\sigma}^0 - \eth(\sigma^0\bar{\sigma}^0)]V^{-3}\}d\Omega, \tag{4.4}$$

where $d\Omega = d\zeta d\bar{\zeta}/P_0^2$.

When the angular dependence of b is restricted to each of the four $l = 0$ and 1 spherical harmonics, the four parameter translation subgroup of the BMS group will have been singled out. If, in addition, no homogeneous Lorentz transformation is allowed, i.e., $c = \bar{c} = k = 0$, then (4.4) yields, *by definition*, the four components of the energy-linear momentum vector.

If, on the other hand, $b = 0$ and $c \neq 0$, then, since c is a spin weight one quantity and satisfies $\eth_0 c = 0$, it has the form $c = a_{m1}Y_{1m}$, with a_m being three complex constants. The three complex (6 real) values of (4.4) are, once more by definition, the angular momentum-center of mass tensor.

We point out that the coefficients of b and c are, respectively, proportional to what we earlier called $\tilde{p}$ and $\bar{S}$.

Equation (4.4) is used only to justify the identification, up to a factor, of the $l = 0, 1$ parts of $\tilde{p}$ and the $l = 1$ parts of $\bar{S}$ with the 4-momentum and angular momentum-center of mass tensor, respectively. In the next section, using techniques from the theory of infinite-dimensional representations of the Lorentz group, we shall extract from $\tilde{p}$ and $\bar{S}$ in a Lorentz covariant fashion explicit expressions for p^μ and $S^{\mu\nu}$ and their dynamical laws.

5. EQUATIONS OF MOTION

Let us first review what has been accomplished thus far. Asymptotic symmetry considerations have led us to two functions, defined at future null infinity, $\tilde{p}(u, \zeta, \bar{\zeta})$ and $S(u, \zeta, \bar{\zeta})$, and to the identification of certain of their components with the physical quantities p^μ and $S^{\mu\nu}$. By examining asymptotically flat solutions to the field equations, we have learned that $\tilde{p}$ is real (2.38) and that the Bianchi Identities yield the time evolution of $\tilde{p}$ (2.37) and $\bar{S}$ (2.40). We have also introduced the function $V(u, \zeta, \bar{\zeta})$ and suggested that a part of it is related to still another physical quantity, namely, the 4-velocity v^μ. Furthermore, we have pointed out that, in asymptotically flat spaces, the homogeneous Lorentz group is well defined at future null infinity in terms of the fractional linear transformation

$$\zeta' = \frac{a\zeta + b}{c\zeta + d}, \qquad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1, \tag{5.1}$$

to which it is isomorphic. In this section we shall first explicitly define the physical quantities mentioned above and then obtain the equations of motion, both in a Lorentz covariant way.

In the preceding paper[11] (to which the reader may wish to refer at this time) we discuss spin and conformally weighted functions on the sphere, that is, functions which transform under (5.1) with spin weight s and conformal weight w, as defined by (2.7)′. (In this section reference to equations from the preceding paper will be indicated by a prime.) Such functions form a vector space, denoted by $D_{(s,w)}$, upon which infinite dimensional representations of the Lorentz group act. Of particular interest to us are spin and conformally weighted functions that form the vector spaces $D_{(s,w)}$ and $D_{(s'=-s,w'=-w-2)} \equiv D_{(-s,-w-2)}$ such that

$$s \text{ and } w \text{ are either integer or half-integer and } w \geq |s|. \tag{5.2}$$

Neither of these spaces is irreducible. $D_{(s,w)}$, for instance, contains a *finite*-dimensional invariant subspace $E_{(s,w)} \subset D_{(s,w)}$. Furthermore, there exists a mapping, $\Pi_{(s,w)}: D_{(-s,-w-2)} \to E_{(s,w)}$, defined by (2.10)′ and (2.11)′, which commutes with the Lorentz transformations. This means that, given any function $\tilde{\eta}_{(-s,-w-2)} \in D_{(-s,-w-2)}$, such that s and w satisfy (5.2), $\Pi_{(s,w)}$ can be used to define another function

$$\eta_{(s,w)} = \Pi_{(s,w)}\tilde{\eta}_{(-s,w-2)} \in E_{(s,w)} \tag{5.3}$$

in a Lorentz covariant way. As is shown in Sec. 3 of the preceding paper, the functions that form $E_{(s,w)}$ can be directly related to Minkowski tensors.

We shall now apply this procedure to our case. First, we point out that it can be shown that under (5.1)

$$V(u, \zeta, \bar{\zeta}) \in D_{(0,1)} \to V^{-3} \in D_{(0,-3)}, \tag{5.4a}$$

$$\tilde{p}(u, \zeta, \bar{\zeta}) \in D_{(0,-3)}, \tag{5.4b}$$

$$\bar{\tilde{S}}(u, \zeta, \bar{\zeta}) \in D_{(-1,-3)}. \tag{5.4c}$$

Based on the results of the preceding paper this means that we can define 4-vectors from V^{-3} and $\tilde{p}$ and a trace-free, antisymmetric second rank tensor from S as follows;

$$v^\mu l'_\mu = v = \Pi_{(0,1)}V^{-3} = Y_{00}(\Omega)\int\bar{Y}_{00}(\Omega')V(\Omega')^{-3}d\Omega' - \frac{1}{3}\sum_{m=-1}^{1} Y_{1m}(\Omega)\int\bar{Y}_{1m}(\Omega')V(\Omega')^{-3}d\Omega',$$

$$p^\mu l'_\mu = p = \Pi_{(0,1)}\tilde{p}, \tag{5.5}$$

$$S^{\mu\nu}l'_\mu m'_\nu = S = \Pi_{(1,1)}\bar{\tilde{S}} = \sum_{m=-1}^{1} {}_1Y_{1m}(\Omega)\int_{-1}Y_{1m}(\Omega')\bar{\tilde{S}}(\Omega')d\Omega',$$

where V has been normalized such that

$$\Pi_{(0,0)}V^{-3}v = \int V^{-3}v\,d\Omega = \tfrac{1}{2}v^\alpha v_\alpha = 1. \tag{5.6}$$

It is useful to define an additional function

$$\tilde{T} = \sigma^0\bar{\sigma}^0 v^{-5} \in D_{(0,-5)} \tag{5.7}$$

from which we can define the following trace-free, symmetric third-rank tensor

$$T^{\mu\nu\rho}l'_\mu l'_\nu l'_\rho = \Pi_{(0,3)}\tilde{T} = Y_{00}(\Omega)\int\bar{Y}_{00}(\Omega')\tilde{T}(\Omega')d\Omega' - \tfrac{3}{5}\sum_{m=-1}^{1} Y_{1m}(\Omega)\int\bar{Y}_{1m}(\Omega')\tilde{T}(\Omega')d\Omega' + \tfrac{1}{5}\sum_{m=-2}^{1} Y_{2m}(\Omega)\int\bar{Y}_{2m}(\Omega')\tilde{T}(\Omega')d\Omega' - \tfrac{1}{35}\sum_{m=-3}^{3} Y_{3m}(\Omega)\int\bar{Y}_{3m}(\Omega')\tilde{T}(\Omega')d\Omega'. \tag{5.8}$$

[The vectors l'_μ and m'_μ that appear in this section are given by (3.1) and (3.2b)′, respectively.]

Equations (2.37) and (2.40) can be rewritten as

$$\dot{\tilde{p}} = -\eth_0^2\bar{\eth}_0^2 V + \eth_0^2(RV) + \bar{\eth}_0^2(\bar{R}V) - \dot{\tilde{T}}v\dot{v} + 4\tilde{T}\dot{v}^2 + \tilde{F}, \tag{5.9}$$

$$6\dot{\bar{\tilde{S}}} = -v^{-2}\bar{\eth}_0(\tilde{p}v^3) + \tilde{T}v^3\bar{\eth}_0(\dot{v}/v) + 6\tilde{J}, \tag{5.10}$$

where

$$\tilde{F} = -[\dot{\sigma}^0\dot{\bar{\sigma}}^0 v^{-3}] - \left[R\bar{R}V - \left(\frac{\sigma^0}{v}\right)^{\cdot}\left(\frac{\bar{\sigma}^0}{v}\right)^{\cdot} v^{-1}\right] \tag{5.11}$$

and

$$\begin{aligned} 6\tilde{J} = {} & [\tfrac{3}{2}\dot{\bar{\sigma}}^0\bar{\eth}\sigma^0 + \tfrac{1}{2}\sigma^0\bar{\eth}\dot{\bar{\sigma}}^0 - \tfrac{3}{2}\bar{\sigma}^0\bar{\eth}\dot{\sigma}^0 - \tfrac{1}{2}\dot{\sigma}^0\bar{\eth}\bar{\sigma}^0]V^{-3} \\ & - [2V^{-3}\bar{\sigma}^0\eth\bar{\eth}\eth \ln P_0 V - V^{-3}\bar{\eth}\eth^2\sigma^0 + V^{-3}\eth(\tilde{p}V^3) \\ & - \sigma^0\bar{\sigma}^0V^{-3}\eth(\dot{V}/V) - v^{-2}\eth_0(\tilde{p}v^3) + \sigma^0\bar{\sigma}^0v^{-2}\bar{\eth}_0(\dot{v}/v)]. \end{aligned} \tag{5.12}$$

Each term in (5.9) has spin weight $s = 0$ and conformal weight $w = -3$, so that $\Pi_{(0,1)}$ can be applied to it directly to obtain

$$\dot{p}^\mu = F^\mu + \tfrac{1}{6}T\dot{v}^\alpha\dot{v}_\alpha v^\mu - \tfrac{1}{3}\dot{T}\dot{v}^\mu + \tfrac{1}{2}\dot{t}^{\mu\alpha\beta}v_\alpha\dot{v}_\beta + 2t^{\mu\alpha\beta}\dot{v}_\alpha\dot{v}_\beta, \tag{5.13}$$

Similarly, (5.10) consists entirely of $s = -1, w = -3$ terms and application of $\Pi_{(1,1)}$ yields

$$\dot{S}^{\mu\nu} + v^{[\mu}p^{\nu]} = J^{\mu\nu} - \tfrac{1}{3}Tv^{[\mu}\dot{v}^{\nu]} - \tfrac{1}{2}t^{[\mu}\dot{v}^{\nu]} + \tfrac{1}{3}t^{\alpha[\mu}\dot{v}^{\nu]}v_\alpha, \tag{5.14}$$

where extensive use has been made of the relations (4.14)′ from the preceding paper and where

$$F^\mu l'_\mu = \Pi_{(0,1)}\tilde{F}, \tag{5.15a}$$

$$J^{\mu\nu}l'_\mu m'_\nu = \Pi_{(1,1)}\tilde{J}, \tag{5.15b}$$

$$T^{\mu\nu} = \tfrac{3}{4}T^{\mu\nu\alpha}v_\alpha, \tag{5.16a}$$

$$T^\mu = \tfrac{2}{3}T^{\mu\alpha}v_\alpha = \tfrac{1}{2}T^{\mu\alpha\beta}v_\alpha v_\beta, \tag{5.16b}$$

$$T = \tfrac{1}{2}T^\alpha v_\alpha = \tfrac{1}{3}T^{\alpha\beta}v_\alpha v_\beta = \tfrac{1}{4}T^{\alpha\beta\gamma}v_\alpha v_\beta v_\gamma, \tag{5.16c}$$

$$t^{\mu\nu\rho} = T^{\mu\nu\rho} - T(v^\mu v^\nu v^\rho - \eta^{(\mu\nu}v^{\rho)}), \tag{5.17a}$$

$$t^{\mu\nu} = T^{\mu\nu} - T(v^\mu v^\nu - \tfrac{1}{2}\eta^{\mu\nu}), \tag{5.17b}$$

$$t^\mu = T^\mu - Tv^\mu, \quad t^\alpha v_\alpha = 0. \tag{5.17c}$$

In Eqs. (5.9) and (5.10) all of the explicit dependence on v and $\dot{v}$ is exhibited in the terms involving the function $\tilde{T}$. The expressions $\tilde{F}$ and $\tilde{J}$ depend only on σ^0 and the higher ($l \geq 2$) harmonics in V. (The second of the bracketed terms in both (5.11) and (5.12) vanishes if $V = v$.) Therefore, F^μ and $J^{\mu\nu}$ are independent of velocity and acceleration and are, respectively, the radiation reaction force and torque due to the mass (sometimes called "electric type") moments and the spin ("magnetic type") moments.

By contracting (5.13) with v^μ and defining the inertial mass by

$$m = \tfrac{1}{2}p^\alpha v_\alpha, \tag{5.18}$$

we obtain

$$p^\mu = mv^\mu + \dot{S}^{\mu\alpha}v_\alpha - J^{\mu\alpha}v_\alpha - \tfrac{1}{3}T\dot{v}^\mu + \tfrac{1}{4}t^\alpha\dot{v}_\alpha v^\mu - \tfrac{1}{3}t^{\mu\alpha}\dot{v}_\alpha. \tag{5.19}$$

Elimination of p^μ between (5.19) and (5.14) yields

$$\dot{S}^{\mu\nu} = J^{\mu\nu} - \dot{S}^{\alpha[\mu}v^{\nu]}v_\alpha + J^{\alpha[\mu}v^{\nu]}v_\alpha - \tfrac{1}{2}t^{[\mu}\dot{v}^{\nu]}. \tag{5.20}$$

Finally, substituting (5.19) into (5.13) yields

$$\begin{aligned} m\dot{v}^\mu = {} & F^\mu - \tfrac{1}{2}F^\alpha v_\alpha v^\mu - \ddot{S}^{\mu\alpha}v_\alpha + \dot{J}^{\mu\alpha}v_\alpha \\ & + \tfrac{1}{3}T(\ddot{v}^\mu + \tfrac{1}{2}\dot{v}^\alpha\dot{v}_\alpha v^\mu) + \tfrac{1}{4}t^\mu\dot{v}_\alpha\dot{v}^\alpha \\ & - \tfrac{1}{2}t^\alpha\dot{v}_\alpha\dot{v}^\mu + \tfrac{3}{4}(\dot{t}^{\mu\alpha\beta} - \tfrac{1}{2}\dot{t}^{\alpha\beta\gamma}v_\gamma v^\mu)v_\alpha\dot{v}_\beta \\ & + \tfrac{1}{4}(t^{\mu\alpha\beta} - \tfrac{1}{2}t^{\alpha\beta\gamma}v_\gamma v^\mu)(v_\alpha\ddot{v}_\beta + 9\dot{v}_\alpha\dot{v}_\beta) \end{aligned} \tag{5.21}$$

and $\dot{m} = \tfrac{1}{2}(F^\alpha v_\alpha - \dot{S}^{\alpha\beta}v_\alpha\dot{v}_\beta + J^{\alpha\beta}v_\alpha\dot{v}_\beta + \dot{t}^\alpha\dot{v}_\alpha + t^{\alpha\beta}\dot{v}_\alpha\dot{v}_\beta).$ (5.22)

It should be noted that although the fifth term in (5.21) is similar to the radiation reaction term in the Lorentz–Dirac force law with $\frac{2}{3}e^2$ replaced by $\frac{1}{3}T(T = \Pi_{(0,0)}\tilde{T}v^3 = \int\sigma^0\bar{\sigma}v^{-2}d\Omega \geq 0)$, the radiation reaction force is vastly more complicated in the gravitational case.

Although (5.22) yields the time development of the inertial mass, it is p^0 which is the Bondi mass, and the Bondi mass law is easily obtained from (5.13) in the following way. From (2.37) we see that

$$\dot{p}^\mu l'_\mu = -\Pi_{(0,1)}R\bar{R}V. \tag{5.23}$$

Taking the $l = 0$ part of (5.23) we obtain

$$\dot{p}^0 l'_0 = (1/\sqrt{2})\dot{p}^0 = -\int R\bar{R}V d\Omega \leq 0. \tag{5.24}$$

We have used units in which the gravitational constant is one. When conventional units are used and the limit of zero gravitational constant is taken, (5.13) and (5.14) reduce to

$$\dot{p}^\mu = 0, \tag{5.25}$$

$$\dot{S}^{\mu\nu} + v^{[\mu}p^{\nu]} = 0, \tag{5.26}$$

the usual Lorentz invariant equations of motion for a free particle with intrinsic angular momentum.[15]

6. DISCUSSION

In addition to the question of the reasonableness of the definitions of the physical quantities (in particular, the angular momentum and center of mass), there remains the difficult and important question of which family of 2-surface at infinity should be used in the evaluation of these quantities. (Due to the manner in which we have set up the problem, this question is equivalent to asking which null coordinate system should be used.)

To investigate this question let us first study an analogous situation in electrodynamics in Minkowski space. Consider a finite charge distribution and a null coordinate system based on an arbitrary timelike world line.[16] Relative to this coordinate system, one can describe (actually define) the multipole moments of the source distribution by looking at the angular behavior of the asymptotic field. (We emphasize that the moments defined in this manner do not agree with the usual definition of the moments, e.g., $Q_{lm} \propto \int Y_{lm}r^l\rho d^3x$, where the source properties are taken at a fixed time. In our case the integral is essentially taken on the light cone.)

It is clear that if a second world line is chosen as the base line for a different null coordinate system, then the associated moments will, in general, bear little relation to the first set of moments. (The transformation properties involve the history of the source.) It appears most likely (but to our knowledge not explicitly proven for these "null" type of moments) that a timelike world can be chosen such that the associated electric dipole moment is zero. Such a line could be called the center of charge line.

The natural question arises: Can the same idea be applied to the center of mass in either the linear of full theory of gravitation? In other words, can we find a coordinate system such that the associated center of mass is zero? In the linearized version the difficulties are exactly the same as in the electrodynamic case, while in the full theory the difficulties are vastly increased.

In the first place the null surfaces which we have been using do not, in general, have conelike behavior i.e., they do not possess an apex) and thus a family of them does not define a world line in the interior. However, in a recent paper[12] it was shown that there exist "canonical" families of null surfaces (surfaces of, in some sense, minimal asymptotic shear) which have many of the properties one associates with the families of null surfaces constructed from a world line in Minkowski space. In particular, the transformation freedom between two different "canonical" families is the same in both cases and depends essentially on three functions of u. This is (numerically) the correct amount of freedom to be able to set the center of mass (or charge) equal to zero. This observation, of course, does not constitute an existence proof. It, however, lends plausability to the conjecture that by an appropriate coordinate condition one can obtain supplementary conditions (to the equations of motion) of the form

$$S^{\mu\alpha} v_\alpha = 0 \tag{6.1}$$

or possibly

$$S^{\mu\alpha} p_\alpha = 0. \tag{6.2}$$

[From some preliminary calculations, (6.1) appears more likely than (6.2).]

APPENDIX A

Equations (3.1) and (3.2) are most easily solved by first rewriting them in spin-coefficient notation (Sec. 2). This is accomplished by substituting

$$\xi^\mu = Al^\mu + Bl^\mu + C\bar{m}^\mu + \bar{C}m^\mu \tag{A1}$$

and taking tetrad components of (3.1) and (3.2). The resulting equations are

$$DB = 0, \tag{A2}$$

$$DA + \Delta B = (\gamma + \bar{\gamma})B + \tau\bar{C} + \bar{\tau}C, \tag{A3}$$

$$\delta B - DC = \tau B + \sigma\bar{C} + \rho C, \tag{A4}$$

$$\Delta A = -(\gamma + \bar{\gamma})A - \delta\bar{C} - \nu C + O(1), \tag{A5}$$

$$\delta A - \Delta C = -2\tau A + \bar{\nu}B - \bar{\lambda}\bar{C} - (\mu + \gamma - \bar{\gamma})C + O(1), \tag{A6}$$

$$\delta C = \sigma A - \bar{\lambda}B + (\beta - \bar{\alpha})C + O(r^{-1}), \tag{A7}$$

$$\bar{\delta}C + \delta\bar{C} = 2\rho A - (\mu + \bar{\mu})B + (\bar{\alpha} - \beta)\bar{C} + (\alpha - \bar{\beta})C + O(r^{-2}). \tag{A8}$$

The powers of r^{-1} in the order symbols are chosen by realizing (not obviously) that if higher powers are chosen, the equations will in general have no solutions. (Other justifications are given in Refs. 5 and 6.)

Using the known spin-coefficients (Sec. 110), (A2), (A3), and (A4) can be integrated immediately for the r dependence of A, B, and C. The remaining equations then yield relations between the coefficients. These results are summarized in Eqs. (3.3)–(3.9).

*This research was supported in part by the National Science Foundation under Grants GP-19378 and GP-22789.

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

[2]H. Bondi, M. van der Burg, and A. Metzner, Proc. R. Soc. A **269**, 21 (1962).

[3]R. Sachs, Proc. R. Soc. A **270**, 103 (1962).

[4]E. T. Newman and R. Penrose, J. Math. Phys. (N.Y.) **3**, 566 (1962).

[5]E. T. Newman and T. Unti, J. Math. Phys. (N.Y.) **3**, 891 (1962).

[6]R. Sachs, Phys. Rev. **128**, 2851 (1962).

[7]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).

[8]R. Penrose, article in *Relativity, groups and topology,* edited by C. Dewitt and B. Dewitt (Gordon and Breach, New York, 1964).

[9]R. Sachs, article in *Relativity, groups and topology*, edited by C. Dewitt and B. Dewitt (Gordon and Breach, New York, 1964.)

[10]L. Tamburino and J. Winicour, Phys. Rev. **150**, 1039 (1966).

[11]R. Lind, J. Messmer, and E. T. Newman J. Math. Phys. (N.Y.) **15**, 1879 (1972).

[12]B. Aronson, R. Lind, J. Messmer, and E. T. Newman, J. Math. Phys. (N.Y.) **12**, 2462 (1971).

[13]Lower lase Latin indices are used for tetrad components and range over 1,2,3,4. Greek letters ranging over 0,1,2,3 and capital Latin letters ranging over 2,3 are used for tensor indices.

[14]For properties of the edth operator ð, see E. T. Newman and R. Penrose, J. Math. Phys. (N.Y.) **7**, 863 (1966), and J. N. Goldberg *et al.*, J. Math. Phys. (N.Y.) **8**, 2155 (1967).

[15]J. Frenkel, Z. Phys. **37**, 243 (1926); M. Mathisson, Acta Phys. **6**, 163 (1937); Acta Phys. **6**, 218 (1937).

[16]A. Held, E. T. Newman, and R. Posadas, J. Math. Phys. (N.Y.) **11**, 3145 (1970).

Irreducible tensor operators for finite groups

T. Buch

Laboratoire de Luminescence II, Université de Paris VI, Paris, France

(Received 23 May 1972; Revised Manuscript Received 25 July 1972)

Normalized tensor operators for a finite group $\mathcal{G}$ are defined by means of coefficients U which formalize the descent in symmetry from $\mathcal{R}_3$ to $\mathcal{G}$. The properties of these coefficients are demonstrated and tables given. Some examples of application show their use and utility.

Whereas in spherical symmetry irreducible tensor operators are defined by means of their commutation relations with angular momentum operators, for finite symmetry groups they are characterized by their transformation properties, i.e., by the irreducible representation and component to which they belong. Both definitions are of course equivalent; but they are not always handled in the same form. This introduces a discontinuity between the formalisms of atomic spectroscopy and crystal field theory. It is the purpose of the present paper to bridge this gap, by means of the formalization of the descent in symmetry from $\mathcal{R}_3$, to its finite subgroups, especially the cubic groups.

To show how this can be done, consider a basis, $|lm\rangle$ of the rep $\mathfrak{D}_l$ of $\mathcal{R}_3$. Upon descent in symmetry to the finite groups $\mathcal{G}$, $\mathfrak{D}_l$ breaks up in reps of $\mathcal{G}$, and if we choose bases $|l\Gamma\gamma\rangle$ for $\mathcal{G}$, we have

$$|lm\rangle = \sum_{\Gamma\gamma} (l\Gamma\gamma|lm)|l\Gamma\gamma\rangle. \tag{1}$$

Here the symbol for the basis $|l\Gamma\gamma\rangle$ contains l to indicate that Γ is contained in the decomposition of $\mathfrak{D}_l$. We can apply the development indicated in Eq. (1) to express an irreducible tensor operator $\{kq\} = C_q^{(k)}$, which obeys the usual commutation relations with the angular momentum operators $l_z, l_\pm$:

$$[\{kq\}, l_z] = q\{kq\},$$

$$[\{kq\}, l_\pm] = [k(k+1) - q(q\pm 1)]^{1/2}\{kq\pm 1\}. \tag{2}$$

Upon descent in symmetry from $\mathcal{R}_3$ to $\mathcal{G}$, we may write the equivalent to Eq. (1) for operators

$$\{kq\} = \sum_{\Gamma\gamma} (k\Gamma\gamma|kq)\{k\Gamma\gamma\}. \tag{3}$$

Equation (3) involves the same transformation coefficients as Eq. (1), and it may serve as a definition of the irreducible tensor operator component $\{k\Gamma\gamma\}$. We have

$$\{k\Gamma\gamma\} = \sum_q (kq|k\Gamma\gamma)\{kq\}. \tag{4}$$

We shall call the $\{k\Gamma\gamma\}$ a component of a $\mathcal{G}$-tensor of rank $k\Gamma$. The application of the Wigner–Eckhart theorem to the matrix elements of irreducible spherical or $\mathcal{G}$-tensors defines in each case a reduced matrix element

$$\langle lm|\{kq\}|l'm'\rangle = (-1)^{l-m}\begin{pmatrix} l\,k\,l' \\ -m\,q\,m'\end{pmatrix}\langle l\|\{k\}\|l'\rangle$$

$$= (-1)^{l-m}\overline{V}\begin{pmatrix} l\,l'\,k \\ -m\,m'\,q\end{pmatrix}\langle l\|\{k\}\|l'\rangle, \tag{5}$$

$$\langle l'\Gamma'\gamma'|\{k\Gamma\gamma\}|l''\Gamma''\gamma''\rangle = V\begin{pmatrix}\Gamma'\Gamma''\Gamma \\ \gamma'\gamma''\gamma\end{pmatrix}\langle l'\Gamma'\|\{k\Gamma\}\|l''\Gamma''\rangle \tag{6}$$

(when real components are chosen for the reps Γ).

The properties of the vector coupling coefficients V are well known[1,2]; we use here the $\overline{V}$ coefficient rather than the $3j$ coefficient even for $\mathcal{R}_3$ symmetry, to emphasize the similarity of the expressions.

Now, as the $\mathcal{G}$-tensor $\{k\Gamma\}$ is defined in terms of the spherical tensor components, we may develop the reduced matrix element in Eq. (6) and write according to a theorem of Racah[3]

$$\langle l'\Gamma'\|\{k\Gamma\}\|l''\Gamma''\rangle = U\begin{Bmatrix}\Gamma'\Gamma''\Gamma \\ l'\,l''\,k\end{Bmatrix}\langle l'\|\{k\}\|l''\rangle. \tag{7}$$

This defines the coefficient U which formalizes the symmetry reduction from $\mathcal{R}_3$ to $\mathcal{G}$. The reduced matrix element on the right-hand side of Eq. (7) is exactly the same as that in Eq. (5), and in it no reference to $\mathcal{G}$ is made.

The U coefficient is a sum of products of five factors, two coupling coefficients and three transformation coefficients, viz.

$$U\begin{Bmatrix}\Gamma_1\Gamma_2\Gamma_3 \\ l_1 l_2 l_3\end{Bmatrix} = \sum_{\substack{m_1m_2m_3 \\ \gamma_1\gamma_2\gamma_3}} (-1)^{l_1}\overline{V}\begin{pmatrix} l_1 l_2 l_3 \\ m_1 m_2 m_3\end{pmatrix} V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma_3 \\ \gamma_1\gamma_2\gamma_3\end{pmatrix}$$

$$\times (l_1m_1|l_1\Gamma_1\gamma_1)(l_2m_2|l_2\Gamma_2\gamma_2)(l_3m_3|l_3\Gamma_3\gamma_3). \tag{8}$$

It is clear from its definition that the U coefficient has the following properties:

(a) It is invariant, as is easily proved by a coordinate transformation.

(b) It is zero unless $(l_1 l_2 l_3)$ and $(\Gamma_1\Gamma_2\Gamma_3)$ separately satisfy a triangular condition, which for $\mathcal{G}$ means $\Gamma_1 \times \Gamma_2 \supset \Gamma_3$, and for $\mathcal{R}_3$ $|l_1 - l_2| \leq l_3 \leq l_1 + l_2$.

(c) It is invariant to even permutation of its columns; both elements of a column belong together: Under odd permutations, a factor $(-1)^{L+\Gamma}$ is introduced, where $L = l_1 + l_2 + l_3$, $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3$, and $(-1)^\Gamma$ are defined by Griffiths.[1]

(d) It is zero unless each Γ is contained in the branching of the corresponding l.

In the Appendix we give a short table of values of U for the cubic groups.

As a simple example of application of Eq. (7) and the U coefficient, we may generalize the theorem, due to Abragam and Pryce[4] of the proportionality of the matrix elements of the angular momentum operator $\mathbf{L}$ within the T_{2g} components of a d-state with those within a p-state. In fact we have, since in $\mathcal{G} = \mathcal{O}_h$, $\mathbf{L}$ transforms as T_{1g} and is proportional to $\{1T_{1g}\}$:

$$\langle 2T_{2g}\gamma_{2g}|\{1T_{1g}\gamma_{1g}\}|2T_{2g}\gamma'_{2g}\rangle = U\begin{Bmatrix}T_{2g}T_{2g}T_{1g} \\ 2\quad 2\quad 1\end{Bmatrix}$$

$$\times\langle 2\|\{1\}\|2\rangle V\begin{pmatrix}T_{2g}T_{2g}T_{1g} \\ \gamma_{2g}\gamma'_{2g}\gamma_{1g}\end{pmatrix},$$

$$\langle 1T_{1u}\gamma_{1u}|\{1T_{1g}\gamma_{1g}\}|1T_{1u}\gamma'_{1u}\rangle = U\begin{Bmatrix} T_{1u} T_{1u} T_{1g} \\ 1 \ 1 \ 1 \end{Bmatrix}\langle 1\|\{1\}\|1\rangle V\begin{pmatrix} T_{1u} T_{1u} T_{1g} \\ \gamma_{2g}\gamma'_{2g}\gamma_{1g}\end{pmatrix},$$

furthermore Griffiths[1] shows that $V\begin{pmatrix} T_1 T_1 T_1 \\ \gamma_1\gamma_2\gamma_3\end{pmatrix} = V\begin{pmatrix} T_2 T_2 T_1 \\ \gamma_1\gamma_2\gamma_3\end{pmatrix}$.

Thus the proportionality constant α between the matrices is given explicitly by

$$\alpha = \frac{U\begin{Bmatrix} T_{2g} T_{2g} T_{1g} \\ 2 \ 2 \ 1\end{Bmatrix}\langle 2\|\{1\}\|2\rangle}{U\begin{Bmatrix} T_{1u} T_{1u} T_{1g} \\ 1 \ 1 \ 1\end{Bmatrix}\langle 1\|\{1\}\|1\rangle},$$

a result which is immediately generalizable to any set of states $\{l\,T\}$, for any l and T being any triply degenerate rep of one of the cubic groups.

DESCENT IN SYMMETRY FOR IRREDUCIBLE PRODUCTS

An irreducible product of $\mathcal{G}$ tensors is a $\mathcal{G}$ tensor. We may define it by the equation

$$\{k_1\Gamma_1 \otimes k_2\Gamma_2 \to k\Gamma\gamma\} = \sum_{\gamma_1\gamma_2} [\Gamma]^{1/2} V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma \\ \gamma_1\gamma_2\gamma\end{pmatrix}\{k_1\Gamma_1\gamma_1\}\{k_2\Gamma_2\gamma_2\}, \quad (9)$$

which is analogous to the spherical case, where we have

$$\{k_1 \otimes k_2 \to kq\} = \sum_{q_1q_2} [k]^{1/2}(-1)^{k-q}\overline{V}\begin{pmatrix}k_1k_2 & k \\ q_1q_2 & -q\end{pmatrix}\{k_1q_1\}\{k_2q_2\}. \quad (10)$$

In each case $[k] = 2k+1$ is the dimension of $\mathfrak{D}_k$ in $\mathfrak{R}_3$ and $[\Gamma]$ is the dimension of Γ in $\mathcal{G}$.

If we reduce the irreducible product of Eq. (10) into a direct sum of $\mathcal{G}$-tensors using Eq. (3) and compare the result with Eq. (9), we find the following expression for the irreducible product of $\mathcal{G}$-tensors:

$$\{k_1\Gamma_1 \otimes k_2\Gamma_2 \to k\Gamma\gamma\} = \begin{Bmatrix}\frac{[k]}{[\Gamma]}\end{Bmatrix}^{1/2} U\begin{pmatrix}k_1k_2k \\ \Gamma_1\Gamma_2\Gamma\end{pmatrix}\{k_1 \otimes k_2 \to k\Gamma\gamma\}. \quad (11)$$

In the proof of Eq. (11) use is made of the orthogonality of the $\overline{V}$ coefficients and of the following property of the U coefficients:

$$\sum_{q_1q_2q_3} (-1)^{k_3-q_3}\overline{V}\begin{pmatrix}k_1k_2k_3 \\ q_1q_2 -q_3\end{pmatrix}(k_1q_1|k_1\Gamma_1\gamma_1) \times (k_2q_2|k_2\Gamma_2\gamma_2) \times (k_3\Gamma_3\gamma_3|k_3q_3) = V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma_3 \\ \gamma_1\gamma_2\gamma_3\end{pmatrix}U\begin{Bmatrix}k_1k_2k_3 \\ \Gamma_1\Gamma_2\Gamma_3\end{Bmatrix}. \quad (12)$$

The reciprocal of Eq. (11) gives the reduction of a spherical irreducible product in terms of irreducible products of $\mathcal{G}$-tensors:

$$\{k_1 \otimes k_2 \to k\Gamma\gamma\} = \sum_{\Gamma_1\Gamma_2}\begin{Bmatrix}\frac{[k]}{[\Gamma]}\end{Bmatrix}^{1/2} U\begin{Bmatrix}k_1k_2k \\ \Gamma_1\Gamma_2\Gamma\end{Bmatrix}\{k_1\Gamma_1 \otimes k_2\Gamma_2 \to k\Gamma\gamma\}. \quad (13)$$

Equation (13) is proved by a method similar to that of the proof of Eq. (11). The combination of both of these equations allows us to establish the following normalizations for the U-coefficients for any set of given $k_1k_2k_3$:

$$\sum_{\Gamma_1\Gamma_2} U\begin{Bmatrix}k_1k_2k_3 \\ \Gamma_1\Gamma_2\Gamma_3\end{Bmatrix}^2 = \frac{[\Gamma_3]}{[k_3]}, \quad (14a)$$

$$\sum_{\Gamma_1\Gamma_2\Gamma_3} U\begin{Bmatrix}k_1k_2k_3 \\ \Gamma_1\Gamma_2\Gamma_3\end{Bmatrix}^2 = 1. \quad (14b)$$

Equation (13) is particularly useful since it admits the immediate adaptation of mixed tensor operators for spherical symmetry to deal with lower symmetry. We thus consider a mixed tensor operator of rank $k\Gamma$, being of spherical rank k_1 with respect to part 1 of the system and k_2 with respect to part 2, $\{k_1 \otimes k_2 \to k\Gamma\}$. In a clumsy but fairly obvious notation we take the wavefunctions $|l_a a(1); l_b b(2){:}\ \Gamma'_{12}\rangle$ and $|l_c c(1); l_d d(2){:}\Gamma''_{12}\rangle$, and we calculate the matrix element

$$\langle l_a a(1); l_b b(2){:}\Gamma'_{12}\|\{k_1 \otimes k_2 \to k\Gamma\}\|l_c c(1); l_d d(2){:}\ \Gamma''_{12}\rangle$$
$$= \begin{Bmatrix}\frac{[k]}{[\Gamma]}\end{Bmatrix}^{1/2}\sum_{\Gamma_1\Gamma_2} U\begin{Bmatrix}k_1k_2k \\ \Gamma_1\Gamma_2\Gamma\end{Bmatrix}\langle l_a a(1); l_b b(2){:}\Gamma'_{12}\|\{k_1\Gamma_1 \otimes k_2\Gamma_2 \to k\Gamma\}\|l_c c(1); l_d d(2){:}\ \Gamma''_{12}\rangle \quad (15)$$
$$= \{[k][\Gamma'_{12}][\Gamma''_{12}]\}^{1/2}\langle l_a\|\{k_1\}\|l_c\rangle\langle l_b\|\{k_2\}\|l_d\rangle\sum_{\Gamma_1\Gamma_2} U\begin{Bmatrix}k_1k_2k \\ \Gamma_1\Gamma_2\Gamma\end{Bmatrix}U\begin{Bmatrix}l_al_ck_1 \\ a\ c\ \Gamma_1\end{Bmatrix}U\begin{Bmatrix}l_bl_dk_2 \\ b\ d\ \Gamma_2\end{Bmatrix}X\begin{Bmatrix}a & b & \Gamma'_{12} \\ c & d & \Gamma''_{12} \\ \Gamma_1 & \Gamma_2 & \Gamma\end{Bmatrix}.$$

Here $X\{\ \}$ is a $9j$ symbol for $\mathcal{G}$, as defined by Griffiths.[1] Equation (15) is useful for the description of two electron operators, or for certain vibronic interactions, when highly degenerate vibrational modes are described by means of angular momentum eigenfunctions[4-6] which couple to the electronic angular momentum.

The property of invariance of the U coefficients can be used to obtain expressions for wavefunctions of higher L values in unusual coordinate systems, when those of low L values are known or can be obtained by direct methods. The formula which will accomplish this expresses one of the transformation coefficients $(k\Gamma\gamma|kq)$ in terms of two other transformation coefficients, two V's and one U. It is, in fact, easy to prove the following equation:

$$(k_3\Gamma_3\gamma_3|k_3q_3) = [\Gamma_3]U\begin{Bmatrix}k_1k_2k_3 \\ \Gamma_1\Gamma_2\Gamma_3\end{Bmatrix}^{-1}$$

$$\times \sum_{\substack{q_1 q_2 \\ \gamma_1\gamma_2}} (-1)^{k_3 - q_3}\,\overline{V}\begin{pmatrix} k_1 k_2 k \\ q_1 q_2 - q\end{pmatrix}$$

$$\times V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma_3 \\ \gamma_1\gamma_2\gamma_3\end{pmatrix}(k_1q_1|k_1\Gamma_1\gamma_1)(k_2q_2|k_2\Gamma_2\gamma_2). \qquad (16)$$

It is to be noted that in case complex functions are desired or used, $V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma_3\\ \gamma_1\gamma_2\gamma_3\end{pmatrix}$ in Eq. (16) as well as in all other pertinent equations in this paper should be replaced by $[-1]^{\Gamma_3+\gamma_3} V\begin{pmatrix}\Gamma_1\Gamma_2\Gamma_3\\ \gamma_1\gamma_2-\gamma_3\end{pmatrix}$.[1]

Example: $\mathcal{G}$-tensorial expression of spin–spin coupling: Judd[2] gives the following expression for the spin-spin coupling Hamiltonian in spherical symmetry

$$V_{SS} = \sum_{kq}(-1)^k\left(\frac{(2k+5)!}{(2k)!}\right)^{1/2}$$

$$\times\left(\frac{r_j^k}{r_2^{k+3}}(\{k+2\}_1\otimes\{k\}_2\to\{2\}\mid\{S_1\otimes S_2\to 2\})\right.$$

$$\left.+\frac{r_2^k}{r_1^{k+3}}(\{k\}_1\otimes\{k+2\}_2\to\{2\}\mid\{S_1\otimes S_2\to 2\})\right), \qquad (17)$$

where the vertical bar indicates a scalar product of tensor operators. We have

$$(\{k\}\mid\{k\}) = \sum_q (-1)^q\{kq\}\{k-q\}.$$

This scalar product can be expressed in terms of our $\mathcal{G}$-tensors, and we have

$$(\{k\}\mid\{k\}\,) = \sum_{\Gamma\gamma}\{k\Gamma\gamma\}\{k\Gamma\gamma\} = \sum_{\Gamma}(\{k\Gamma\}\mid\{k\Gamma\}). \qquad (18)$$

The scalar product between the second rank tensors decomposes thus into a T_{2g} part and an E_g part, corresponding to O_h symmetry.

To make our example precise, suppose we wish to calculate the matrix element of V_{SS} between the ground state ${}^6A_{1g}$ and a charge transfer state ${}^6T_2$ of a $3d^5$-system in cubic symmetry. This limits the sum in Eq. (16) to just the $\Gamma = T_2$ term. Since in Eq. (16) $k = 2$ for V_{SS}, the matrix element reduces to bielectronic matrix elements between two-electron triplet functions only. This severely limits the extension of the fractional parentage expressions necessary to express the states. We have, for the spin part of $\langle {}^6A_1|V_{SS}|{}^6T_2\rangle$, in the notation of Eq. (13)

$$\left\langle\tfrac12\tfrac12 1\|\{1\otimes 1\to 2T_2\}\|\tfrac12\tfrac12 1\right\rangle$$

$$= \left\langle(\tfrac12\tfrac12)1T_1'\|\{1\otimes 1\to 2T_2\|(\tfrac12\tfrac12)1T_1'\right\rangle$$

$$= U\begin{Bmatrix}T_1' T_1' T_2\\ 1\ 1\ 2\end{Bmatrix}\left\langle(\tfrac12\tfrac12)1\|\{1\otimes 1\to 2\}\|(\tfrac12\tfrac12)1\right\rangle$$

$$= U\begin{Bmatrix}T_1' T_1' T_2\\ 1\ 1\ 2\end{Bmatrix}\sqrt{45}\begin{Bmatrix}\tfrac12\tfrac12 1\\ \tfrac12\tfrac12 1\\ 1 1 2\end{Bmatrix}\left\langle\tfrac12\|\{1\}\|\tfrac12\right\rangle^2 = \frac{\sqrt3}{2}, \qquad (19)$$

where use has been made of Eqs. (13) and (7) and the standard expansion of the 2-electron "spherical" matrix element[2] in terms of a $9j$ symbol. Here $T_1' = T_{1g}$ in O_h and T_2 in T_d symmetry.

For the orbital part, we first observe that, in general, if molecular orbitals are used, there is a "local" part of the orbital matrix element, in which all four atomic functions are centered at the same atom, and a nonlocal part[7] which is not *a priori* negligible, but which can be reduced to a sum over one-center functionals by means of the alpha-function development of Löwdin.[8,9] Once this reduction is made, we can apply directly Eq. (15) which leads us to the expressions found in atomic spectroscopy, and which can be expressed by means of a generalization of the radial integrals introduced by Horie.[10]

$$M^k(abcd) = \int_0^\infty \frac{r^2 dr}{r^{k+3}} f_a(r) f_c(r)\int_0^r r_1^2 r_1^k f_b(r_1) f_d(r_1) dr. \qquad (20)$$

APPENDIX A: VALUES OF $U\begin{Bmatrix}l_1 l_2 l_3\\ \Gamma_1\Gamma_2\Gamma_3\end{Bmatrix}$ FOR $l_1 l_2 l_3 \leq 4$ AND $l_1 = 6$, $l_2 = l_3 = 3$. GROUPS O OR T_d

The table gives U^2 in "prime factors" notation.[11] Successive figures are the exponents of 2, 3, 5 ··· in the decomposition of U^2 in prime factors, a negative exponent is underlined. An asterisk indicates that the negative square root should be taken (*). O means $U = 0$. $T_1' = T_1(O)$ or $T_2(T_d)$. When one $l = 0$, we have

$$U\begin{Bmatrix} l\ l'\ 0\\ \Gamma\Gamma' A_1\end{Bmatrix} = \left(\frac{[\Gamma]}{[l]}\right)^{1/2}\delta_{\Gamma\Gamma'}\delta_{ll'}\,.$$

$$U\begin{Bmatrix}1\ 1\ 1\\ T_1' T_1' T_1'\end{Bmatrix} = 1 \text{ evidently.}$$

APPENDIX B: PROOF OF EQ. (7)

In terms of the definition of the basis functions $|l\Gamma\gamma\rangle$ and of the $\mathcal{G}$-tensors $\{k\Gamma\gamma\}$ [Eqs. (1) and (3)] we may write the matrix element of a $\mathcal{G}$-tensor component:

$$\langle l'\Gamma'\gamma'|\{k\Gamma\gamma\}|l''\Gamma''\gamma''\rangle$$

$$= \sum_{m'm''q}(l'\Gamma'\gamma'|l'm')(k\Gamma\gamma|kq)(l''\Gamma''\gamma''|l''m'')$$

$$\times\langle l'm'|\{kq\}|l''m''\rangle$$

$$= \sum_{m'm''q}(l'\Gamma'\gamma'|l'm')(k\Gamma\gamma|kq)(l''\Gamma''\gamma''|l''m'')(-1)^{l'-m'}$$

$$\times V\begin{pmatrix} l'l''k\\ -m'm''q\end{pmatrix}\langle l'|\{k\}\|l''\rangle. \qquad (B1)$$

Via $(l'\Gamma'\gamma'|l'm') = (l'm'|l'\Gamma'\gamma')^* = (-1)^{m'}(l'-m|l'\Gamma'\gamma')$ (valid for the real component systems for the reps of $\mathcal{G}$), we have

$$\sum_{m'm''q}(l'-m'|l'\Gamma'\gamma')(kq|k\Gamma\gamma)(l''m''|l''\Gamma''\gamma'')\,\overline{V}\begin{pmatrix}l'l''k\\ -m'm''q\end{pmatrix}$$

$$\times(-1)^{l'}\langle l'\|\{k\}\|l''\rangle$$

$$= \sum_{m'm''q}(l'm'|l'\Gamma'\gamma')(kq|k\Gamma\gamma)(l''m''|l''\Gamma''\gamma'')\,\overline{V}\begin{pmatrix}l'l''k\\ m'm''q\end{pmatrix}$$

$$\times(-1)^{l'}\langle l'\|\{k\}\|l''\rangle.$$

We may compare this expression with Eq. (6) of the text.

TABLE.

$l_1 l_2 l_3$	$\Gamma_1\Gamma_2\Gamma_3$	U^2
2 1 1	$T_2 T_1' T_1'$	$01\underline{1}$
2 1 1	$E\ T_1' T_1'$	$10\underline{1}$
2 2 1	$T_2 T_2 T_1'$	$^*00\underline{1}$
2 2 1	$E\ T_2 T_1'$	$10\underline{1}$
2 2 2	$T_2 T_2 T_2$	$02\underline{1}\underline{1}$
2 2 2	$T_2 T_2 E$	$^*11\underline{1}\underline{1}$
2 2 2	$E\ E\ E$	$30\bar{1}\underline{1}$
3 2 1	$A_2 T_2 T_1'$	$^*00\bar{0}\underline{1}$
3 2 1	$T_1 E\ T_1'$	$^*02\bar{1}\underline{1}$
3 2 1	$T_2 E\ T_1'$	$00\bar{0}\underline{1}$
3 2 1	$T_1 T_2 T_1'$	$11\bar{1}\underline{1}$
3 2 1	$T_2 T_2 T_1'$	$10\bar{0}\underline{1}$
3 2 2	$A_2 E\ E$	$000\underline{1}$
3 2 2	$T_1 T_2 T_2$	$^*21\underline{1}\underline{1}$
3 2 2	$T_2 T_2 T_2$	0
3 2 2	$T_1 E\ T_2$	$^*\underline{1}1\underline{1}\underline{1}$
3 2 2	$T_2 E\ T_2$	$^*\underline{1}1\bar{0}\underline{1}$
3 3 1	$A_2 T_2 T_1'$	$\bar{0}00\underline{1}$
3 3 1	$T_1 T_1 T_1'$	$\underline{3}20\underline{1}$
3 3 1	$T_2 T_2 T_1'$	$\bar{3}00\underline{1}$
3 3 1	$T_1 T_2 T_1'$	$\underline{3}11\underline{1}$
3 3 2	$T_1 T_1 T_2$	$^*\underline{3}1\underline{1}\underline{1}$
3 3 2	$T_1 T_1 E$	$\underline{2}0\underline{1}\underline{1}$
3 3 2	$T_2 T_2 T_2$	$^*\underline{3}1\bar{1}\underline{1}$
3 3 2	$T_2 T_2 E$	0
3 3 2	$T_1 T_2 T_2$	$^*\underline{3}00\underline{1}$
3 3 2	$T_1 T_2 E$	$\bar{0}00\underline{1}$
3 3 3	$A_2 T_1 T_2$	$\underline{1}00\underline{1}$
3 3 3	$T_1 T_1 T_1$	$\bar{0}00\underline{1}$
3 3 3	$T_1 T_1 T_2$	0
3 3 3	$T_1 T_2 T_2$	$000\underline{1}$
3 3 3	$T_2 T_2 T_2$	0
4 2 2	$A_1 T_2 T_2$	$^*12\underline{1}$

$l_1 l_2 l_3$	$\Gamma_1\Gamma_2\Gamma_3$	U^2
4 2 2	$A_1 E\ E$	$0\underline{1}\underline{1}$
4 2 2	$E\ T_2 T_2$	$3\underline{2}\bar{0}\underline{1}$
4 2 2	$E\ E\ E$	$1\underline{1}0\underline{1}$
4 2 2	$T_2 T_2 T_2$	$^*2\bar{1}0\underline{1}$
4 2 2	$T_2 E\ T_2$	$^*\underline{1}\bar{0}0\underline{1}$
4 2 2	$T_1 T_2 T_2$	0
4 2 2	$T_1 T_2 E$	$\underline{1}\underline{1}$
4 3 1	$A_1 T_1 T_1'$	$^*\bar{0}\underline{2}$
4 3 1	$E\ T_2 T_1'$	$0\bar{0}0\underline{1}$
4 3 1	$E\ T_1 T_1'$	$^*0\underline{2}1\underline{1}$
4 3 1	$T_1 T_1 T_1'$	$^*\underline{3}\underline{1}1$
4 3 1	$T_1 T_2 T_1'$	$\underline{3}$
4 3 1	$T_2 A_2 T_1'$	$\bar{0}00\underline{1}$
4 3 1	$T_2 T_1 T_1'$	$^*\underline{3}\underline{1}1\underline{1}$
4 3 1	$T_2 T_2 T_1'$	$^*\underline{3}20\underline{1}$
4 3 2	$A_1 T_2 T_2$	$^*\bar{0}\underline{2}$
4 3 2	$E\ T_1 T_2$	$0\bar{0}0\underline{1}$
4 3 2	$E\ T_2 T_2$	$0\underline{2}\underline{1}\underline{1}$
4 3 2	$E\ A_2 E$	$^*3\underline{1}\underline{1}\underline{1}$
4 3 2	$T_1 T_1 T_2$	$\underline{3}$
4 3 2	$T_1 T_2 T_2$	$\underline{3}1\underline{1}$
4 3 2	$T_1 T_1 E$	$\bar{0}$
4 3 2	$T_1 T_2 E$	$0\underline{1}\underline{1}$
4 3 3	$A_1 T_1 T_1$	$^*\underline{1}\bar{0}00,\underline{1}$
4 3 3	$A_1 T_2 T_2$	$\underline{1}\underline{2}00,\underline{1}$
4 3 3	$A_1 A_2 A_2$	$^*\bar{1}\underline{1}00,\underline{1}$
4 3 3	$E\ T_1 T_1$	$\underline{1}01\underline{1},\underline{1}$
4 3 3	$E\ T_1 T_2$	$^*\underline{1}00\underline{1},\underline{1}$
4 3 3	$E\ T_2 T_2$	$^*\underline{1}\underline{2}1\bar{1},\underline{1}$
4 3 3	$T_1 T_1 T_1$	0
4 3 3	$T_1 T_1 T_2$	$0000,\underline{1}$
4 3 3	$T_1 T_2 T_2$	$^*0\underline{1}10,\underline{1}$
4 3 3	$T_2 T_1 T_1$	$^*0\bar{1}1\underline{1},\underline{1}$

$l_1 l_2 l_3$	$\Gamma_1\Gamma_2\Gamma_3$	U^2
4 3 3	$T_2 T_1 T_2$	$200\underline{1},\underline{1}$
4 3 3	$T_2 T_2 T_2$	$^*0\underline{1}1\underline{1},\underline{1}$
4 3 3	$T_1 T_2 A_2$	$\underline{1}\underline{1}1\bar{0},\underline{1}$
4 3 3	$T_2 T_1 A_2$	$^*\underline{1}\bar{0}0\underline{1},\bar{1}$
4 4 1	$A_1 T_1 T_1'$	$\bar{0}\underline{2}$
4 4 1	$E\ T_1 T_1'$	$^*0\underline{2}\underline{1}1$
4 4 1	$E\ T_2 T_1'$	$^*0\underline{1}\underline{1}$
4 4 1	$T_1 T_1 T_1'$	$^*\underline{3}\underline{1}\underline{1}$
4 4 1	$T_1 T_2 T_1'$	$\underline{3}\underline{1}\bar{1}1$
4 4 2	$A_1 E\ E$	$\bar{3}\underline{1}\bar{0}0,\underline{1}$
4 4 2	$A_1 T_2 T_2$	$0\underline{1}00,\underline{1}$
4 4 2	$E\ E\ E$	$^*6\underline{2}\underline{1}\underline{1},\underline{1}$
4 4 2	$E\ T_2 T_2$	$0\underline{1}\underline{1}\underline{1},\bar{1}$
4 4 2	$E\ T_1 T_2$	$0\bar{0}\underline{1}\bar{0},1$
4 4 2	$T_1 T_1 T_2$	$^*\underline{3}0\underline{1}1,\underline{1}$
4 4 2	$T_1 T_2 T_2$	$^*\underline{3}0\underline{1}2,\underline{1}$
4 4 2	$T_2 T_2 T_2$	$\underline{2}\underline{1}\underline{1}\underline{1},\bar{1}$
4 4 2	$T_1 T_1 E$	$^*\bar{2}\underline{1}\underline{1}1,\underline{1}$
4 4 2	$T_2 T_2 E$	$^*4\underline{1}\underline{1}\underline{1},\underline{1}$
4 4 2	$T_1 T_2 E$	$00\underline{1}\bar{0},\underline{1}$
4 4 3	$A_1 T_1 T_1$	$\underline{1}\underline{2}01,\underline{1}$
4 4 3	$A_1 T_2 T_2$	$\underline{1}\underline{1}10,\underline{1}$
4 4 3	$E\ T_1 T_1$	$^*\underline{1}\underline{2}12,\underline{1}$
4 4 3	$E\ T_1 T_2$	$\underline{1}\bar{0}00,\underline{1}$
4 4 3	$E\ T_2 T_1$	$^*\underline{1}\underline{1}\underline{1}\underline{1},\underline{1}2$
4 4 3	$E\ T_2 T_2$	$^*\underline{1}\underline{1}0\underline{1},\underline{1}$
4 4 3	$E\ E\ A_2$	$^*0\underline{2}0\underline{1},\underline{1}$
4 4 3	$T_1 T_1 T_1$	$0\underline{1}\underline{1}\bar{1},\underline{1}$
4 4 3	$T_1 T_1 T_2$	$\bar{0}$
4 4 3	$T_1 T_2 T_1$	$^*2\underline{1}\underline{1}0,\underline{1}$
4 4 3	$T_1 T_2 T_2$	$0\bar{0}00,\underline{1}$
4 4 3	$T_1 T_2 A_2$	$^*\underline{1}000,\underline{1}$
4 4 3	$T_2 T_2 T_2$	$^*\underline{4}001,\underline{1}$

$l_1 l_2 l_3$	$\Gamma_1\Gamma_2\Gamma_3$	U^2
4 4 3	$T_2 T_2 T_1$	$0\underline{1}1\underline{1},\underline{1}$
4 4 4	$A_1 A_1 A_1$	$1\underline{3}0\bar{2},\underline{1}\underline{1}$
4 4 4	$A_1 E\ E$	$2\underline{3}00,\underline{1}\underline{1}$
4 4 4	$A_1 T_1 T_1$	$^*\underline{1}\underline{2}02,\underline{1}\underline{1}$
4 4 4	$A_1 T_2 T_2$	$\underline{1}\underline{2}00,\underline{1}1$
4 4 4	$E\ E\ E$	$\underline{9}\underline{3}1\underline{1},\underline{1}1$
4 4 4	$E\ T_1 T_1$	$\underline{3}\underline{2}1\bar{3},\underline{1}\underline{1}$
4 4 4	$E\ T_2 T_2$	$\underline{3}\bar{2}1\underline{1},\underline{1}1$
4 4 4	$E\ T_1 T_2$	$\underline{1}\bar{1}1\bar{0},\underline{1}\bar{1}$
4 4 4	$T_1 T_1 T_1$	0
4 4 4	$T_1 T_1 T_2$	$\underline{4}111,\underline{1}\underline{1}$
4 4 4	$T_1 T_2 T_2$	$^*\underline{8}310,\underline{1}\underline{1}$
4 4 4	$T_2 T_2 T_2$	$\bar{0}\underline{1}1\underline{1},\underline{1}\underline{1}$
6 3 3	$A_1 A_2 A_2$	$^*3\bar{1}0\underline{1},\underline{1}\underline{1}$
6 3 3	$A_1 T_1 T_1$	$\underline{1}02\underline{1},\underline{1}\underline{1}$
6 3 3	$A_1 T_2 T_2$	$^*\underline{1}40\underline{1},\underline{1}\underline{1}$
6 3 3	$A_2 T_1 T_2$	$\underline{1}00\bar{0},0\underline{1}$
6 3 3	$E\ T_1 T_1$	$\underline{1}020,\underline{1}\underline{1}$
6 3 3	$E\ T_1 T_2$	$\underline{1}010,\underline{1}\underline{1}$
6 3 3	$E\ T_2 T_2$	$\underline{1}200,\underline{1}\underline{1}$
6 3 3	$T_1 T_1 T_1$	0
6 3 3	$T_1 T_1 T_2$	$\underline{1}110,\underline{1}\underline{1}$
6 3 3	$T_1 T_2 T_2$	0
6 3 3	$T_1 T_2 A_2$	$0200,\underline{1}\underline{1}$
6 3 3	$aT_2 T_1 T_1$	$\underline{6}110,\underline{1}\underline{1}$
6 3 3	$aT_2 T_1 T_2$	$^*\underline{6}000,\underline{1}\underline{1}$
6 3 3	$aT_2 T_2 T_2$	$^*\underline{6}112,\underline{1}\underline{1}$
6 3 3	$aT_2 T_1 A_2$	$\underline{3}000,\underline{1}\underline{1}$
6 3 3	$bT_2 T_1 T_1$	$\underline{6}120,0\underline{1}$
6 3 3	$bT_2 T_1 T_2$	$^*\underline{6}210,0\underline{1}$
6 3 3	$bT_2 T_2 T_2$	$\underline{6}300,0\underline{1}$
6 3 3	$bT_2 T_1 A_2$	0

* For example: $U\begin{Bmatrix} 3 & 3 & 1 \\ T_1 & T_2 & T_1 \end{Bmatrix} = +\left(\frac{3\times 5}{2^3\times 7}\right)^{1/2}$.

We equate the right-hand sides of Eqs. (B1) and (6), multiply both sides by $V\begin{pmatrix}\Gamma'\Gamma''\Gamma\\ \gamma'\gamma''\gamma\end{pmatrix}$ and sum over γ', γ'', and γ, using the orthonormality relations for the V coefficients.[1] This yields Eq. 7 when the U are defined as in Eq. 8.

[1] J. S. Griffith, *The irreducible tensor method for molecular symmetry groups* (Prentice-Hall, Englewood Cliffs, N.J., 1962).

[2] B. R. Judd, *Operator techniques in atomic spectroscopy* (McGraw-Hill, New York, 1963).

[3] G. Racah, Phys. Rev. **76**, 1352 (1949).

[4] A. Abragam and B. Bleaney, *Electron paramagnetic resonance of transition ions* (Clarendon Press, Oxford, England, 1970).

[5] W. Moffit and W. Thorson, Phys. Rev. **108**, 1251 (1957).

[6] M. C. M. O'Brien, J. Phys. C 4, 2524 (1971).

[7] R. R. Sharma, T. P. Das, and R. Orbach, Phys. Rev. **149**, 257 (1966).

[8] R. R. Sharma, J. Math. Phys. **9**, 505 (1968).

[9] P.-O. Löwdin, Adv. Phys. **5**, 1 (1956).

[10] H. Horie, Prog. Theor. Phys. **10**, 296 (1953).

[11] M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooteu, Jr., *The 3-j and 6-j symbols* (The Technology Press, M.I.T, Cambridge, Mass., 1959).

Solutions of a nonlinear integral equation for high energy scattering.* III. Analyticity of solutions in a parameter, explored numerically

G. R. Bart
Mayfair College, Chicago, Illinois 60630

R. L. Warnock†
Imperial College, Department of Physics, London SW7, England
(Received 30 May 1972)

Solutions of the Ball–Zachariasen equation, discussed in two previous papers, depend analytically on a parameter c which measures the strength of particle production. Numerical experiments, designed to elucidate the structure of the Riemann surface, are reported. The results are consistent with a very pretty hypothesis which describes the Riemann surface completely.

1. INTRODUCTION

The existence of an infinite class of solutions of the Ball–Zachariasen equation[1] was proved in Paper I of this investigation.[2] These solutions exist for sufficiently small values of a parameter c, which measures the strength of particle production. The solutions $\phi(b, c)$ (where ϕ is simply related to a Hankel transform of the two-particle scattering amplitude and b is the impact parameter), are analytic in c inside some circle $|c| = \gamma$ at each b. The radius γ is limited by the technical requirements of the existence proof, in such a way that the nearest singularity lies on a circle which is, undoubtedly, a good deal larger than $|c| = \gamma$.

Since the values of c for which the existence proof succeeds are too small to be interesting physically, a numerical continuation to larger values of c was attempted in Paper II.[2] Some representative small c solutions were computed, and were continued along the real c axis until a singularity of the Fréchet derivative of the nonlinear Ball–Zachariasen operator was encountered. This prevented further continuation along the real axis; but it was possible to circumvent the singularity by the detour into the complex c plane. Upon returning to the real axis it was found that the solutions were complex, which suggests that the singularity of the Fréchet derivative might be associated with a branch point of $\phi(b, c)$ regarded as a function of c.

In the present paper we report on numerical experiments which were designed to explore the Riemann surface of $\phi(b, c)$ as a function of c. The results were consistent with a remarkably simple hypothesis: namely, that for each b the function $\phi(b, c)$ is analytic on a two-sheeted Riemann surface, each sheet consisting of a plane cut along the real axis from some point $c(b)$ to infinity. The branch point $c(b)$ increases monotonically with b.

As in Papers I and II, we still find no evidence that the Ball–Zachariasen equation has solutions resembling experiment. We think, however, that this work is worth reporting for its mathematical interest, and as an unusual application of a computer for inductive determination of analyticity properties of a complicated equation. Our work is incomplete in that we cannot provide an analytic proof of our hypothesis. Nevertheless, our conjecture about the structure of the Riemann surface is quite definite. It might be pursued analytically, and it certainly can be subjected to more demanding numerical tests.

In Sec. 2 we recall some properties of the equation, and state carefully the analyticity properties which are suggested by the numerical data. Section 3 is a detailed description of the numerical work, and Sec. 4 contains an heuristic argument which is intended to make plausible our conjecture concerning analyticity.

2. THE RIEMANN SURFACE OF THE SOLUTION

After a Hankel transformation,[1] the Ball–Zachariasen equation has the form

$$\hat{f}(b) = \hat{f}^2(b) + C(b; \hat{f}, c), \tag{2.1}$$

where

$$C(b; \hat{f}, c) = \int_0^\infty dx J_0(bx) g(x)(e^{cg(x)} - 1), \tag{2.2}$$

$$g(x) = \int_0^\infty b db J_0(bx) \hat{f}^2(b). \tag{2.3}$$

The notation is the same as in Papers I and II:[2] The elastic scattering amplitude is $sf(x)$, where $x = (-t)^{1/2}$ is the magnitude of momentum transfer, and s is the squared energy. The Hankel transform of $f(x)$ by the zeroth-order Bessel function J_0 is $\hat{f}(b)$, where b is the impact parameter:

$$\hat{f}(b) = \int_0^\infty dx\, J_0(bx) f(x). \tag{2.4}$$

Hence $g(x)$ is a high energy approximation to the elastic unitarity integral. At $c = 0$ the integral equation (2.1) reduces to

$$\hat{f}(b) = \hat{f}^2(b), \tag{2.5}$$

which is solved by any function having values 0 or 1 only. Among all such step functions we consider only those having support in a finite region. An arbitrary member of this class is denoted by $h(b)$. As was shown in Paper I, it is convenient to write $\hat{f}(b)$ for $c \neq 0$ as $h(b)$ plus a remainder, as follows:

$$\hat{f}(b) = h(b) + [1 - 2h(b)]\phi(b). \tag{2.6}$$

The advantage of this change of variable is that $\phi(b)$ satisfies an integral equation which has a unique nontrivial solution in the subset K of a certain Banach space B for each h, provided that c is sufficiently small. The function ϕ is *continuous*, and it vanishes uniformly as c goes to zero, so that $h(b)$ is the limiting form of $\hat{f}(b)$. The continuity of $\phi(b)$ simplifies the proof of the existence theorem (Paper I) and eases numerical solution of the equation.

To find the equation for $\phi(b)$, we note the identities

$$\hat{f} - \hat{f}^2 = \phi - \phi^2, \quad (\phi - h)^2 = \hat{f}^2. \tag{2.7}$$

Thus

$$\phi(b) = \phi^2(b) + B(b; \phi, c), \tag{2.8}$$

where

$$B(b;\phi,c)=\int_0^\infty x\,dx\,J_0(bx)g(x)(e^{cg(x)}-1), \qquad (2.9)$$

$$g(x)=\int_0^\infty b\,db\,J_0(bx)\{h(b)[1-2\phi(b)]+\phi^2(b)\}. \qquad (2.10)$$

Our numerical calculations are based on Eq. (2.8) with a specific choice of $h(b)$; namely,

$$h(b)=\theta(r-b)=\begin{cases}1, & b<r,\\ 0, & b>r.\end{cases} \qquad (2.11)$$

Because of the scaling property of the equation explained in Paper I, the value of the radius r that one chooses is immaterial. As in Paper II, we put $r = 1.41$ (GeV)$^{-1}$, and measure x in units of GeV. The restriction to the simple step function (2.11) seemed unduly narrow in the work of Paper II. In the present report we shall find that infinitely many other solutions can be reached by continuation in c of $\phi(b;c)$, where one starts with the ϕ corresponding to (2.11) at small c. These solutions correspond to various step functions for $\hat{f}(b)$ in the limit $c = 0$. This comes about by $\phi(b;c)$ developing discontinuities in b as c passes through branch points, so that if $c = 0$ is reached by following an appropriate complex path, $\lim_{c\to 0}\phi(b,c)$ is itself a step function.

We note an alternative way of introducing the step functions explicity. If Eq. (2.1) is solved for $\hat{f}$ in terms of C, we can write

$$\hat{f}(b)=\tfrac{1}{2}\{1-S(b)[1-4C(b;\hat{f},c)]^{1/2}\}, \qquad (2.12)$$

where $S(b)$ is any function with values ± 1 almost everywhere (at points where S changes sign, a set of measure zero, it need not be defined). Since $C(b;\hat{f},0)=0$, a solution of (2.12) with an arbitrary S will be a solution of (2.1) having the arbitrary step function $\frac{1}{2}[1-S(b)]$ as its $c = 0$ limit. It will not be necessary to solve or analyze Eq. (2.12), since our discussion in terms of ϕ will also yield arbitrary step functions at $c = 0$.

We shall be interested, however, in a similar equation for ϕ, which is obtained by solving (2.8). In this case, we put the arbitrary step function in front of the square root equal to 1, and obtain

$$\phi(b)=\tfrac{1}{2}\{1-[1-4B(b;\phi,c)]^{1/2}\}. \qquad (2.13)$$

The unique solution of (2.8) in the subspace K at small c is continuous in b, real, positive, and less than $\frac{1}{2}$. Consequently, it is represented by (2.13) with the square root defined to be positive.

The numerical solution we begin with at small c has the representation (2.13). Furthermore, our numerical results are consistent with the following hypotheses about the solutions obtained by continuation in c of this solution [the continuation being performed by solving (2.8) at successive values of c]:

(i) $B(b;\phi(\cdot,c),c)$ is an entire function of c for each b.

(ii) For any b (except those in a set of measure zero) the solution $\phi(b,c)$ is given by formula (2.13) along any complex c path which starts at $c = 0$. At $c = 0$, the square root is defined to be 1, and at subsequent points on the path its value is determined by analytic continuation.

The analyticity of B in c implies that the only singularities of $\phi(b,c)$ in the finite plane, where ϕ is regarded as a function of c at fixed b, arise from zeros of $1-4B$ in (2.13). The numerical results suggest that there is just one such zero at $c = c(b)$ on the positive real axis, with $c(b)$ being a monotonically increasing function of b. In that case, $\phi(b,c)$ has only a simple square-root branch point, with the associated two-sheeted Riemann surface. Because of the dependence of the branch point on b, this simple structure can result in a very complicated b dependence of $\phi(b,c)$ after the path in the c plane has passed through branch cuts several times. This structure arises because as b is varied, with c fixed at the end of the continuation path, the value of the square root in Eq. (2.13) may come from first one and then the other of its two Riemann sheets. This gives discontinuities in b, and at such discontinuities $\phi(b,c)$ is undefined. The set of measure zero mentioned in hypothesis (ii) is just the set of points of discontinuity. At such points, $\phi(b,c)$ is easily defined by continuity from the right or left, and with that definition it is analytic in c.

Since the calculations were restricted to a small part of the Riemann surface, the evidence in favor of hypotheses (i) and (ii) is not overwhelming. One could easily weaken the hypotheses without contradicting the numerical data. For instance, in place of (i) we could assume analyticity in a sufficiently large finite region of the c plane. In Sec. 4 we shall give an argument for plausibility of (i) or a weakened form thereof. A real proof appears to be quite difficult.

3. NUMERICAL EVIDENCE

Equation (2.8), which we write as $F=\phi-\phi^2-B=0$, is solved by the Newton–Kantorovich (NK) method[3,2] in the modified form[3] in which the Fréchet derivative F_ϕ is computed just once, at the beginning of the sequence. In Paper II we used the result of a successful NK iteration with an altered value of c. We have now improved the program by making a linear extrapolation in c to obtain the starting point for the new sequence of approximations. That is, if $\phi(b,c)$ is our approximation for the limit of an NK sequence, the next sequence is begun with

$$\phi_0(b,c+\Delta c)=\phi(b,c)+\frac{\partial\phi}{\partial c}(b,c)\Delta c, \qquad (3.1)$$

where $\partial\phi/\partial c$ is obtained by solving the linear equation

$$F_\phi\frac{\partial\phi}{\partial c}+F_c=0, \qquad (3.2)$$

$$F=\phi-\phi^2-B. \qquad (3.3)$$

This procedure reduced the number of steps Δc required to move a given distance in c by a factor of five or so on the average. As in Paper II, automatic adjustment of Δc was employed: If the sequence failed to converge, Δc was reduced to $\alpha\Delta c$, $\alpha < 1$, and a new sequence was generated with the result of the last successful iteration as starting point. In the present calculations the limit on relative errors was $\epsilon = 0.01$ [see Eq. (2.6)ff, Paper II].

In Paper II we found the first singularity of the Fréchet derivative (in a continuation from $c = 0$ along the positive c axis) at $c = c_0 = 0.5535$. As was noted previously,[2] this is the first value of c for which $1-2\phi$ has a zero, and for which the Fréchet derivative F_ϕ becomes an integral operator of the third kind.[4] Equivalently, it is the first c where $1-4B$ has a zero. The zero of $1-2\phi$ and $1-4B$ occurs at $b = 0$, so according to hypotheses

(i) and (ii) of the previous section, $\phi(0, c)$ should have a square-root branch point at $c = c_0$.

The continuation by NK iteration came to a halt at $c = c_0$, in the sense that the allowed step Δc converged to zero as c_0 was approached. In order to bypass the difficult point, an excursion into the complex c plane was made. In Paper II, various points on the real c axis to the right of c_0 were reached via paths passing through the upper half plane. Since the solutions were complex for $c > c_0$, a branch point at $c = c_0$ was suggested.

To investigate the alleged branch point, we now follow paths which encircle c_0. The first such path is shown in Fig. 1. The point c_1 was reached in our previous calculations. The corresponding solution is plotted in Fig. 2. Note that we have left gaps in the curves near the breaks in slope. In the gaps we have no evaluations of the functions, because of the limited number of mesh points in the calculation. We continue into the lower half-plane to c_2, through the supposed cut. A discontinuity of Reϕ seems to develop as soon as one passes into the lower half-plane. This is illustrated in Fig. 3, where ϕ is plotted for $c = c_2$. The apparent discontinuity of Reϕ in Fig. 3 occurs at the value of b where the break of Reϕ is located in Fig. 1. This discontinuity persists when we return to the real axis at $c_3 < c_0$, as is seen in Fig. 4. The solution is now real to high accuracy.

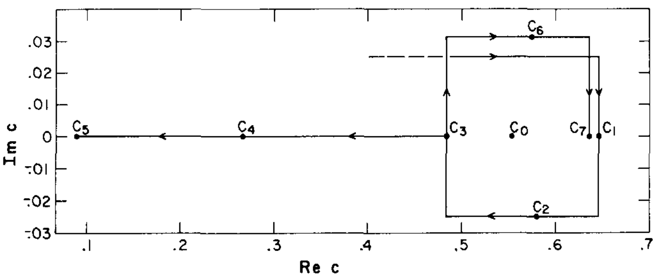


FIG. 1: Paths for analytic continuation of solution in complex c plane, beginning with solution computed in Paper I.

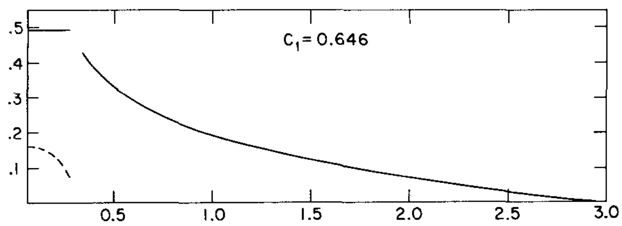


FIG. 2.

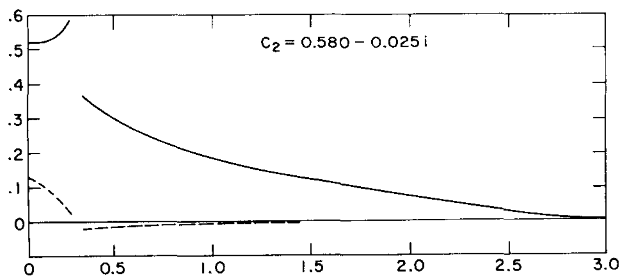


FIG. 3.

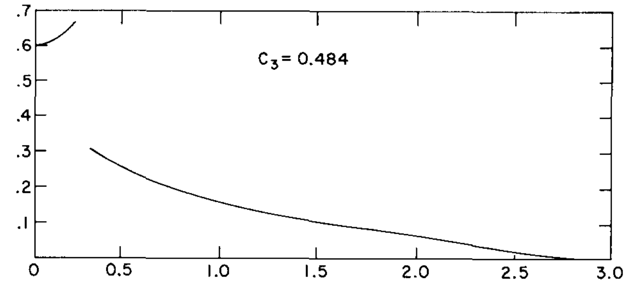


FIG. 4.

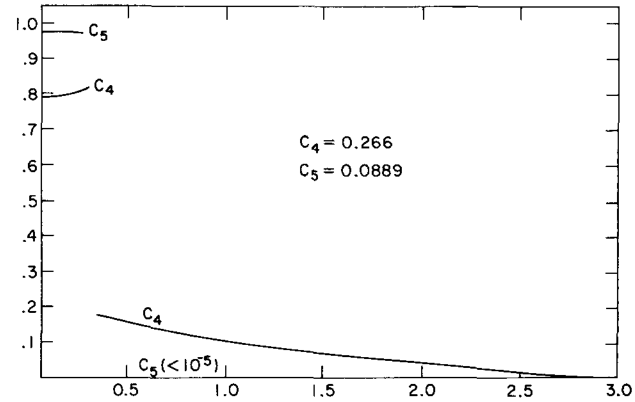


FIG. 5.

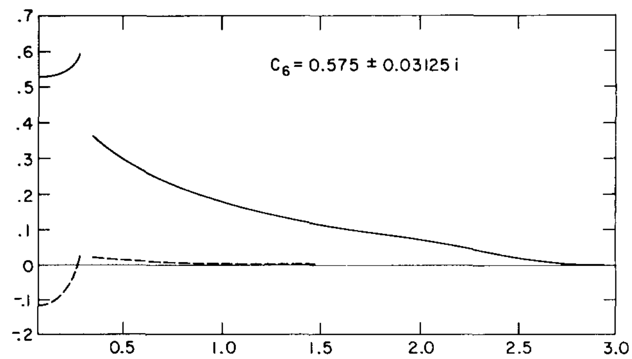


FIG. 6.

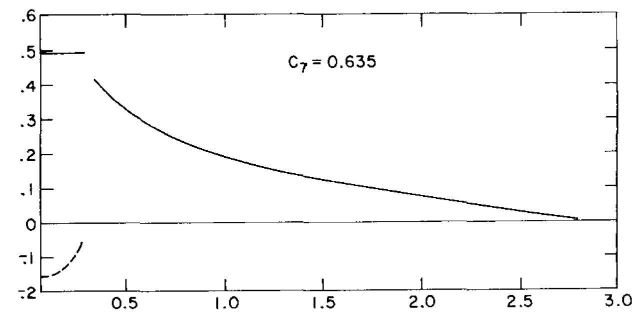


FIG. 7.

FIGS. 2–7: Real parts (solid line) and imaginary parts (dashed line) of solutions at the points c_1 through c_7 of Fig. 1.

We proceed toward the origin on the real axis and find that the solution tends toward a step function, which equals one for $b < b_1 \approx 0.3$ and zero for $b > b_1$ (see Fig. 5). By referring to Eq. (2.5), we see that the corresponding $\hat{f}(b)$, in the limit $c = 0$, is as follows:

$$\hat{f}(b) = \begin{cases} 0, & 0 < b < b_1, \\ 1, & b_1 < b < r, \\ 0, & r < b < \infty. \end{cases} \tag{3.4}$$

Since $\hat{f}(b)$, in the limit $c = 0$, must be a function with values 0 or 1 only, it was inevitable that we obtain some sort of step function in the limit. The result (3.4) is to be compared with $\hat{f}(b) = \theta(r - b)$, our original solution at $c = 0$ on the "first sheet." The value of b_1 in (3.4) depends on the point c_1 at which the real c axis was crossed.

Next we go from c_3 through c_6 to c_7, which is nearly equal to c_1. A comparison of Fig. 2 with Fig. 7 shows that the solutions for c_1 and c_7 are complex conjugates to an excellent approximation.

These results may be understood in terms of hypotheses (i) and (ii) of Sec. 2. To the largest values of c on the positive real axis that we have reached, B increases monotonically with c at each b. At each $c > 0$, b decreases monotonically with b. Let us assume that these monotonic behaviors persist to arbitrarily large values of c on the real axis. For $c > c_0$, we then have a unique zero of $1 - 4B$ as a function of b at a point $b(c) > 0$, since $B > \frac{1}{4}$ at small b. For $b \geqslant 0$ there is a unique real zero of $1 - 4B$ as a function of c, at a point $c(b) \geqslant c_0$. There could also be complex zeros of $1 - 4B$ as a function of c. There is no numerical evidence for complex zeros, however, so we shall assume that there are none. Now consider our initial solutions of Eq. (2.8), obtained by iteration for real $c < c_0$. These solve Eq. (2.13), where $(1 - 4B)^{1/2}$ is defined to be positive for all b. If $B(b; \phi(\cdot, c), c)$ is entire in c [Hypothesis (i)], then analytic continuation in c of the square root on its first Riemann sheet yields

$$\lim_{\gamma \to c \pm i0} [1 - 4B(b, \gamma)]^{1/2} = \begin{cases} |[1 - 4B(b, c \pm i0)]^{1/2}|, & c < c(b), \\ \mp i|[1 - 4B(b, c \pm i0)]^{1/2}|, & c > c(b), \end{cases} \tag{3.5}$$

where we use the abbreviated notation

$$B(b, c) = B(b; \phi(\cdot, c), c).$$

Since Eq. (3.5) holds for any b, and $c \lesseqgtr c(b)$ implies $b \lesseqgtr b(c)$, it follows that

$$\lim_{\gamma \to c \pm i0} [1 - 4B(b, \gamma)]^{1/2} = \begin{cases} \mp i|[1 - 4B(b, c \pm i0)]^{1/2}|, & b < b(c), \\ |[1 - 4B(b, c \pm i0)]^{1/2}|, & b > b(c). \end{cases} \tag{3.6}$$

Equations (3.5) and (3.6) hold when γ is on the first Riemann sheet. On the second sheet, of course, the signs of the right-hand sides of the equations are to be changed. In the following, it will be convenient to give the symbol $[1 - 4B(b, c)]^{1/2}$ an invariant meaning, so that its value depends only on the values of b and c, and not on the sheet in which c is located. Throughout its Riemann surface, the square root can then be written (in terms of its principal branch) as

$$S(b, \Gamma)[1 - 4B(b, c)]^{1/2}, \tag{3.7}$$

where $(1 - 4B)^{1/2}$ is defined so that it satisfies (3.5) and (3.6), and $S(b, \Gamma)$ is a step function with values ± 1 as a function of b. The argument Γ denotes the path in the Riemann surface by which the point c of interest is reached. Since the branch point $c(b)$ depends on b, S in (3.7) can be $+1$ for certain intervals of b, and -1 for other intervals.

The qualitative behavior of all our curves follows from (3.5), (3.6), and the formula

$$\phi(b, c) = \tfrac{1}{2}\{1 - S(b, \Gamma)[1 - 4B(b, c)]^{1/2}\}. \tag{3.8}$$

Consider first $\phi(b, c_1)$, graphed in Fig. 2, with Γ being the path shown in Fig. 1. Since $\mathrm{Im}\phi = 0$ for $b > b_1 \cong 0.3$, we conclude from (3.6) that $b(c_1) = b_1$, and $S(b, \Gamma) = \epsilon(b - b_1)$, where

$$\epsilon(x) = \begin{cases} -1, & x < 0 \\ 1, & x > 0 \end{cases}.$$

Then (3.8) and (3.6) imply $\mathrm{Im}\phi > 0$ and $\mathrm{Re}\phi = \frac{1}{2}$ for $b < b_1$ and $0 < \mathrm{Re}\phi < \frac{1}{2}$ for $b > b_1$, all of which agrees with the data except for a 2% discrepancy in $\mathrm{Re}\phi = \frac{1}{2}$. (The 2% discrepancy here, and slightly bigger disagreements elsewhere, may reasonably be ascribed to numerical error. The computer program was not designed to handle the Hankel transforms of discontinuous functions which arise. We are surprised that the program works as well as it does in such a difficult situation.) When the path of continuation passes downward through the real axis at $c = c_1$, we are passing onto the second sheet of $\phi(b, c)$ if $b < b(c_1)$, since for such b we have $c_1 > c(b)$. We remain on the first sheet of $\phi(b, c)$ for $b > b(c_1)$, since then $c_1 < c(b)$. This explains the apparent discontinuities in the curves at $c = c_2$, Fig. 3, which are now to be interpreted as genuine discontinuities. For $b < b(c_1)$ in Fig. 3, one is seeing second-sheet values, while for $b > b(c_1)$ one sees first-sheet values. In passing from c_1 to c_2, B has remained predominantly real, and ReB has decreased.

When the real axis is regained at $c = c_3$, Fig. 4, the solution is real to very high accuracy. We still have $S(b, \Gamma) = \epsilon(b - b_1)$. Since $c_3 < c_0$, this is expected if

$$\phi(b, c^*) = \phi(b, c)^*, \tag{3.9}$$

where c and c^* are both on a given cut plane, with cut $[c_0, \infty)$. The reality property (3.9) is in fact easily established by noting that $\phi(b, c^*)$ and $\phi(b, c)^*$ both satisfy the equation

$$\phi = \phi^2 + B(b; \phi, c^*). \tag{3.10}$$

For c inside a small circle about the origin, (3.10) has a unique solution in the space K described in Paper I. Hence (3.9) holds for c near the origin and, hence, everywhere on the cut plane. The reality of $\phi(b, c_3)$, $b < b(c_1)$, is achieved by $4B$ becoming less than 1, whereas it was greater than 1 in this region of b at $c = c_1$. As we pass from c_3 to c_4 and c_5 (Fig. 5), B continues to decrease. That is expected, since $B = 0$ at $c = 0$. At $c_5 = 0.0889$, we have a function which nicely resembles the step function which one should have at $c = 0$.

At c_6 and c_7, we have nearly the complex conjugates of the functions at c_2 and c_1, respectively. This agrees with (3.9), to the extent that $c_2 \cong c_6^*$, $c_1 \cong c_7^*$. (One would have liked $c_2 = c_6{}^*$, $c_1 = c_7{}^*$, but that was awkward to achieve, because of the automatic decrementing of Δc when sequences failed to converge).

Note that the definition

$$\phi(b(c_1), c) = \lim_{b' \to b(c_1) - 0} \phi(b', c), \tag{3.11}$$

mentioned in Sec. 2, makes $\phi(b(c_1), c)$ an analytic function of c, which at any particular c is on the same sheet (first or second) as $\phi(b, c)$, where $b(c_1) - \epsilon < b < b(c_1)$ for some $\epsilon > 0$. If instead the limit $b' \to b(c_1) + 0$ were adopted, the function at any c would be on the same sheet as $\phi(b, c)$, with $b(c_1) < b < b(c_1) + \epsilon$.

As a function of b, $\phi(b, c)$ retains a "memory" of the point at which the path of continuation Γ crosses the line $c_0 \leq c < \infty$; this is indicated by the factor $S(b, \Gamma)$ in (3.8). The phenomenon has a simple explanation in terms of hypotheses (i) and (ii), as we have seen; but it leads to the amusing possibility of multiple b discontinuities of $\phi(b, c)$ after several loops around c_0, provided the real axis is crossed at different points $c > c_0$ on different loops. We illustrate this by the additional continuations shown in Fig. 8. We pass from our previous point c_6 to a point $c_8 < c_1$ and also to $c_9 > c_1$. At c_8 (Fig. 9), the function is on the upper side of the second sheet cut for $0 < b < b(c_8) \cong 0.13$, to the left of the cut on the second sheet for $b(c_8) < b < b(c_1)$, and to the left of the cut on the first sheet for $b > b(c_1)$. Thus, in Eq. (3.8) one has $S(b, \Gamma) = \epsilon(b - b_1)\epsilon(b - b_8)$, where $b_i = b(c_i)$. This accounts for the discontinuities of Reϕ at $b = b(c_8)$ and $b = b(c_1)$, and for the vanishing of Imϕ for $b > b(c_8)$. At c_9 (Fig. 10), ϕ is evaluated on the upper side of the second sheet cut for $0 < b < b(c_1)$, on the upper side of the first sheet cut for $b(c_1) < b < b(c_9) \cong 0.53$, and to the left of the cut on the first sheet for $b > b(c_9)$. Thus, we get the change in sign of Imϕ at $b = b(c_1)$, etc. When the solution is continued from c_9 to c_{10} and c_{11} (Figs. 11 and 12), we see another step function developing. Passing from c_{11} to $c = 0$ would give a step function with support in the region $b(c_1) < b < b(c_9)$. According to Eq. (2.5), this means a $\hat{f}(b, 0)$ as follows:

$$\hat{f}(b,0) = \begin{cases} 1, & 0 < b < b(c_1), \\ 0, & b(c_1) < b < b(c_9), \\ 1, & b(c_9) < b < r, \\ 0, & r < b < \infty. \end{cases} \tag{3.12}$$

It is now clear that by appropriate loops around c_0, one can produce arbitrary step functions for $\hat{f}(b, 0)$.

4. REMARKS ON ANALYTICITY OF B

We have seen that the numerical results agree nicely with the hypothesis that $B(b; \phi(\cdot, c), c)$ is an entire func-

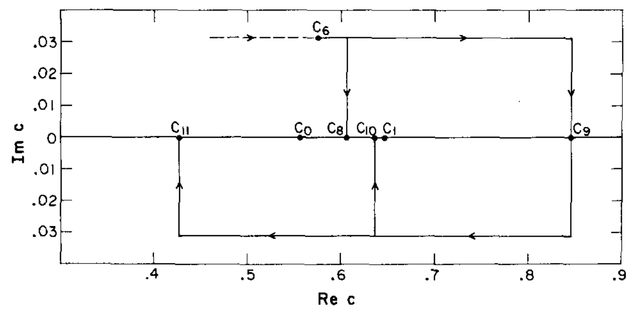


FIG. 8: Paths for analytic continuation of solution in complex c plane, beginning with solution at c_6 in Fig. 1.

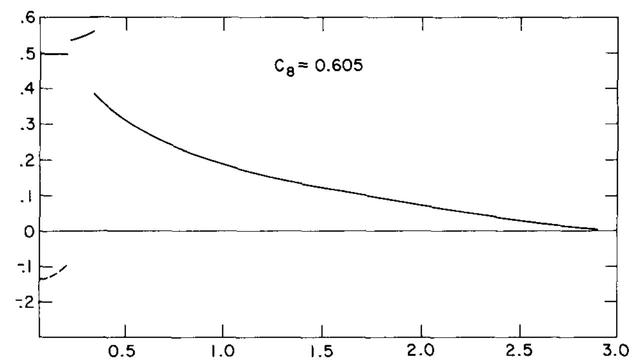


FIG. 9.

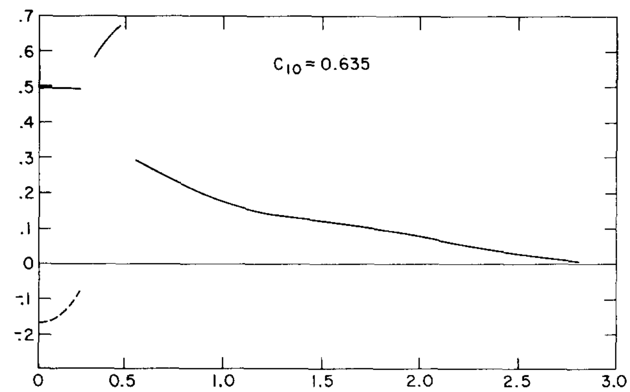


FIG. 11.

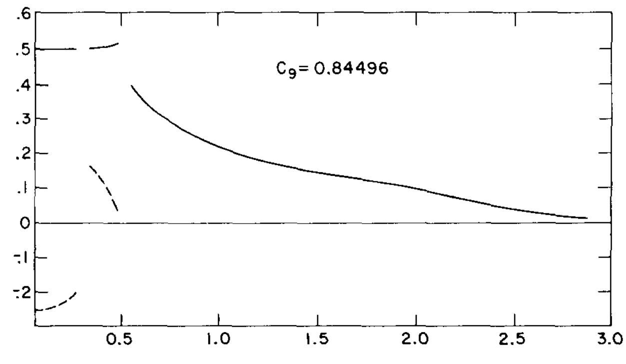


FIG. 10.

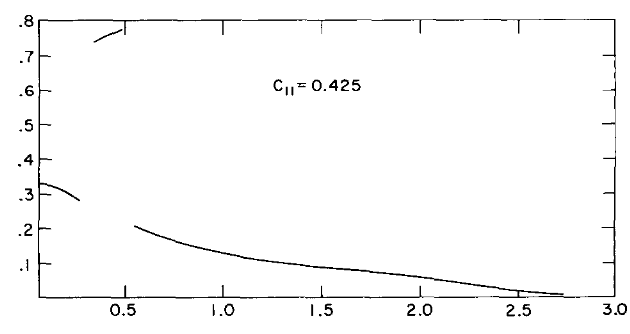


FIG. 12.

FIGS. 9–12: Real parts (solid line) and imaginary parts (dashed line) of solutions at the points c_8 through c_{11} of Fig. 8.

tion of c, or is at least analytic in a region including that we have explored. In order to see hueristically how this might come about, we can examine the Cauchy–Riemann equations for B. Let $c = x + iy$, and note that since $B = \phi - \phi^2$, we have

$$\frac{\partial B}{\partial x} = (1 - 2\phi)\frac{\partial \phi}{\partial x} = B_\phi \frac{\partial \phi}{\partial x} + B_x, \tag{4.1}$$

$$\frac{\partial B}{\partial y} = (1 - 2\phi)\frac{\partial \phi}{\partial y} = B_\phi \frac{\partial \phi}{\partial y} + iB_x, \tag{4.2}$$

where B_ϕ denotes the Fréchet derivative of B with respect to ϕ, and B_x is the derivative of B with respect to its explicit x dependence. In (4.2) we have used the obvious relation $B_y = iB_x$. The expression for the integral operator B_ϕ is as follows:

$$B_\phi \chi(b) = \int_0^\infty K(b, b'; \phi, c)\chi(b')b'\, db',$$

$$K(b, b'; \phi, c) = 2 \int_0^\infty x\, dx\, J_0(bx)J_0(b'x) \times \{1 - e^{cg(x)}[1 + cg(x)]\}[h(b') - \phi(b')]. \tag{4.3}$$

Equation (4.1) and (4.2) may be regarded as integral equations for $\partial\phi/\partial x$ and $\partial\phi/\partial y$, the two equations differing only by a factor of i in the inhomogeneous terms. When the operator $M = (1 - 2\phi)1 - B_\phi$, defined on some appropriate space S, has an inverse on S, we have solutions of the equations such that

$$\frac{\partial \phi}{\partial y} = i\frac{\partial \phi}{\partial x}. \tag{4.4}$$

If Eq. (4.4) holds at $c = \hat{c}$, and if $(1 - 2\phi)\partial\phi/\partial x$ and $(1 - 2\phi)\partial\phi/\partial y$ are continuous in x and y in a neighborhood of $\hat{c}$, then by (4.1) and (4.2) we see that B is analytic at $\hat{c}$. This follows from the Cauchy–Riemann condition

$$\frac{\partial B}{\partial y} = i\frac{\partial B}{\partial x}, \tag{4.5}$$

and the continuity of $\partial B/\partial x$ and $\partial B/\partial y$ in x and y near $\hat{c}$.

It is reasonable to suppose that M has an inverse, yielding solutions of (4.1) and (4.2) with the properties mentioned above, in the entire c plane minus a set of measure zero. The inverse can fail to exist at values of c for which $1 - 2\phi = 0$ at some b. Then M is an integral operator of the third kind.[4,2] In general, third-kind operators have no inverse on a space of continuous or piece-wise continuous functions. Suppose that $1 - 2\phi$ is initially free of zeros, but that c approaches a value c_* for which it acquires a zero. Then $\partial\phi/\partial x$ and $\partial\phi/\partial y$ might acquire singularities, but the Cauchy–Riemann conditions for B at c_* might be preserved by virtue of the fortunate factor $1 - 2\phi$ in (4.1) and (4.2). If $\partial\phi/\partial x$ and $\partial\phi/\partial y$ become infinite only at values of the pair (b, c) for which $1 - 2\phi = 0$, then it is possible for $\partial B/\partial x$ and $\partial B/\partial y$ to be well behaved at c_*.

If we assume the result we would like to prove, that B is analytic in c, we can see that this picture works out consistently. Since $1 - 2\phi = 0$ is equivalent to $B = \frac{1}{4}$, we have from Eq. (3.8) that

$$1 - 2\phi(b_0, c) = [1 - 4B(b_0, c)]^{1/2} \sim a(b_0)(c - c_*)^{1/2}, \tag{4.6}$$

$$\frac{\partial \phi}{\partial x}(b_0, c) \sim -\frac{a(b_0)}{4}(c - c_*)^{-1/2}, \tag{4.7}$$

$$\frac{\partial \phi}{\partial y}(b_0, c) \sim -i\frac{a(b_0)}{4}(c - c_*)^{-1/2}, \tag{4.8}$$

as c approaches c_*. Hence, the factor $1 - 2\phi(b_0, c)$ cancels the singularities of the derivatives as it should, and the Cauchy–Riemann conditions for B are maintained.

If the operator M is singular on a space of continuous functions, it still could be nonsingular on some bigger space, which might include functions with just the properties we need to satisfy the Cauchy–Riemann conditions on B. To see how this might happen, suppose we first try to find the behavior of $1 - 2\phi$ as a function of b near its zero. For this purpose it is reasonable to assume that $B(b; c_*)$ has a continuous first derivative in b near the point $b(c_*)$ at which $B(b; c_*) = \frac{1}{4}$. This assumption is in accord with the data; in fact, B seems to be a perfectly smooth, monotonic function of b at all c. Since our functions $\phi(b; c)$ seem to have the same asymptotic behavior in b as the small c solutions of Paper I, this can be attributed to the good convergence of the integral which defines B. In Paper I, this good convergence made $\partial B/\partial b$ continuous at all b. For b near $b_0 = b(c_*)$, we then find from Eq. (4.5) and Taylor's theorem for B that $1 - 2\phi$ behaves as $(b - b_0)^{1/2}$. Since $\partial B/\partial x$ is supposed to be bounded, we then expect from Eq. (4.1) that $\partial\phi/\partial x$ will behave as $(b - b_0)^{-1/2}$ near $b - b_0$. Now make the following changes of variable in the integral equation for $\partial\phi/\partial x$:

$$\chi(b) = [1 - 2\phi(b)]^{1/2}\frac{\partial \phi(b)}{\partial x}. \tag{4.9}$$

The equation for χ is

$$\chi(b) = \int_0^\infty \hat{K}(b, b'; \phi, c)\chi(b')b'\, db' + \hat{B}_x(b; \phi, c), \tag{4.10}$$

where

$$\hat{K}(b, b'; \phi, c) = \frac{K(b, b'; \phi, c)}{[1 - 2\phi(b)]^{1/2}[1 - 2\phi(b')]^{1/2}}, \tag{4.11}$$

$$\hat{B}_x(b; \phi, c) = \frac{B_x(b; \phi, c)}{[1 - 2\phi(b)]^{1/2}}. \tag{4.12}$$

Since $1 - 2\phi$ vanishes as $(b - b_0)^{1/2}$, the functions $\hat{K}$ and $\hat{B}_x$ will be square-integrable near $b, b' = b_0$. Let us suppose that they are also square-integrable at infinity (without trying to justify the assumption for now). Then we can apply L^2 Fredholm theory to (4.10). Barring unit eigenvalues of the kernel, we have a unique L^2 solution χ. This solution behaves as $(b - b_0)^{-1/4}$ near $b = b_0$, as one can see by applying Schwarz's inequality to the integral in Eq. (4.10). The expected behavior of $(b - b_0)^{-1/2}$ for $\partial\phi/\partial x$ then follows from (4.9). By a simple modification of the equation, we have succeeded in accommodating functions which lie outside the space of continuous (hence bounded) functions. Unfortunately, this discussion does not throw any light on the behavior of the solution as a function of c near $c = c_*$; that is, it does not help one establish Equations (4.6)–(4.8).

We found that it was possible to continue through the real c axis for $c > c_0$, but impossible to pass through the point $c = c_0$, regardless of the direction of approach to c_0. The reason for this appears to be that $1 - 2\phi(b, c_0)$ vanishes as $b - b_0$ when b tends to b_0, while $1 - 2\phi(b, c)$ vanishes only as $(b - b_0)^{1/2}$ when $c > c_0$. Thus, the equation is equivalent to a regular Fredholm equation in L^2 for $c > c_0$ [namely, Eq.(4. 10)], but is singular even in L^2 for $c = c_0$. The hypothesized linear zero of $1 - 2\phi(b, c_0)$ is fully consistent with the graph of $\phi(b, c_0)$ given in Fig. 1 of Paper II, and it means that $B(b, c_0)$ has vanishing slope at $b = b_0$. For $c > c_0$, $B(b, c)$ does not have zero slope at the point $b(c)$ where $1 - 2\phi$ vanishes.

In addition to the singularities of the operator M arising from zeros of $1 - 2\phi$, there could be ordinary Fredholm-type singularities of M. That is, $(1 - 2\phi)^{-1}B_\phi$ could have unit eigenvalues at isolated points in the c plane. Such points might give singularities of B as a function of c; but there is no hint of such in the small region of the c plane that we have investigated numerically.

ACKNOWLEDGMENTS

The computations were performed on the Univac 1108 computer at Illinois Institute of Technology, with funds provided by the Institute and the National Science Foundation. The paper was written while R. Warnock enjoyed the kind hospitality of Professor T. W. B. Kibble and the Department of Physics at Imperial College.

---

*Work partially supported by the National Science Foundation.

†Permanent address: Illinois Institute of Technology, Chicago, Illinois.

[1] J. S. Ball and F. Zachariasen, Phys. Lett. B 30, 558 (1969); Phys. Rev. 3, 1596 (1971).

[2] H. McDaniel, J. Uretsky, and R. L. Warnock, "Solutions of a nonlinear integral equation for high energy scattering. I. An existence theorem; II. Numerical solutions." Paper I is Argonne report ANL/HEP 7111 (April 1971), and paper II is ANL/HEP 7123 (June 1971). Both papers are to be published in Phys. Rev.

[3] L. V. Kantorovich and G. P. *Functional analysis in normed spaces* (Pergamon, Oxford, 1964).

[4] For a discussion of integral operators of the third kind, see G. R. Bart and R. L. Warnock, "Linear integral equations of the third kind," Argonne report ANL/HEP 7141 (November, 1971); to be published in SIAM J. Anal.

Scattering of a scalar wave from a random rough surface: A diagrammatic approach

G. G. Zipfel Jr. and John A. DeSanto
Naval Research Laboratory, Washington, D. C. 20390
(Received 6 March 1972; revised manuscript received 9 August 1972)

The solution to the problem of a scalar wave scattered from a rough surface is given under the conditions that the normal derivative of the field vanish at the surface and that the surface height be a single-valued function with Gaussian statistics. The solution is in terms of a series with a diagram representation. Partial summation of the series in terms of linear integral equations is briefly discussed.

I. INTRODUCTION

The problem of scalar waves scattering from a stationary randomly rough surface which averages to a plane is one of continuing interest from both theoretical and experimental points of view. It represents a realistic acoustical situation and, at least qualitatively, is relevant to the electromagnetic case. The usual and simplest approach is to calculate the scattered field in the Kirchhoff approximation and then calculate moments of fields using assumed surface statistics.[1,2] Clearly this procedure does not work for very rough surfaces. As a first step in a general attack on the problem of rough surface scattering, it is desirable to have a method which permits a systematic examination of the problem so that it is possible to say something about the errors committed when various approximations are made. In keeping with this goal we have constructed a solution in terms of a series expansion for moments of the Green's functions or alternatively for moments of the fields. A diagrammatic algorithm has been developed for the construction of an arbitrary term in the expansion analogous to those used in the study of propagation through a random medium.[3] The class of surfaces studied here is restricted to those with a vanishing normal derivative boundary condition where the height of the surface is a single-valued function with a mean value of zero. We also assume that spatial averages can be replaced by ensemble averages over a multivariate Gaussian distribution of surface heights.

In Sec. II, starting with the Helmholtz integral equation, a series expansion is developed for the Fourier transform of the Green's function associated with a representative sample of the ensemble of surfaces under consideration. An algorithm which is appropriate for subsequent ensemble averaging is then constructed for determining the individual terms in the series. A reduction formula is also derived which gives the scattered field in an asymptotic region far enough removed from the surface so that the cutoff surface modes can be neglected.

In Sec. III a cluster expansion is developed for the characteristic functions associated with the ensemble of surfaces. The properties of the moments which are needed in the series expansion of the moments of Green's functions are discussed using this cluster expansion. These results are then combined with the results of Sec. II to derive an algorithm for generating a series expansion for the n-point moments of Green's functions. The reduction formula for the mutual coherence function is derived and the lowest-order coherent and incoherent scattering terms are shown to correspond to the Kirchhoff approximation. The series are partially summed to generate an integral equation for the mean Green's function which is analogous to Dyson's equation and an integral equation for the mutual coherence function analogous to the Bethe–Salpeter equation. An approximation for coherent scattering that corresponds to using the lowest-order kernel in Dyson's equation is introduced and briefly discussed.

II. DETERMINISTIC SURFACES

A cross section through a representative surface is indicated in Fig. 1. In all future developments we will assume that the average surface is perpendicular to the z axis. The coordinates in planes perpendicular to the z axis will be x and y. We will henceforth use the abbreviation $\mathbf{x}_\perp = x\hat{i}_x + y\hat{i}_y$, where $\hat{i}_x$ and $\hat{i}_y$ are unit vectors in the x and y directions, respectively. The notation sub $\perp$ will always be used to indicate two-vectors in the transverse or x,y plane. Thus, $z = h(\mathbf{x}_\perp)$ defines the surface under consideration, and $z > h(\mathbf{x}_\perp)$ is the region of free propagation. The surface height $h(\mathbf{x}_\perp)$ is assumed to be a smooth single-valued bounded function of $\mathbf{x}_\perp$ with zero mean value. Strictly speaking the condition that $h(\mathbf{x}_\perp)$ is bounded is incompatible with a Gaussian height distribution. However, if the bound is made sufficiently large compared to the root mean squared height, then the probability of exceeding the bound can be made very small.

Since the surface is stationary, the wave equation reduces to the Helmholtz equation on separation of the time variable. The Green's function satisfies the inhomogeneous Helmholtz equation

$$(\nabla^2 + k_0^2)G_D(\mathbf{x}|\,\mathbf{x}') = \delta_3(\mathbf{x}-\mathbf{x}')$$

where k_0 is the frequency of a point source located at $\mathbf{x}'$ (in units where the phase velocity is unity). The source will always be assumed to be in the free propagation region, and thus $z' > h(\mathbf{x}'_\perp)$, the normal derivative of the Green's function vanishes, that is

$$\mathbf{n}\cdot\nabla_x G_D(\mathbf{x}|\mathbf{x}') = 0,$$

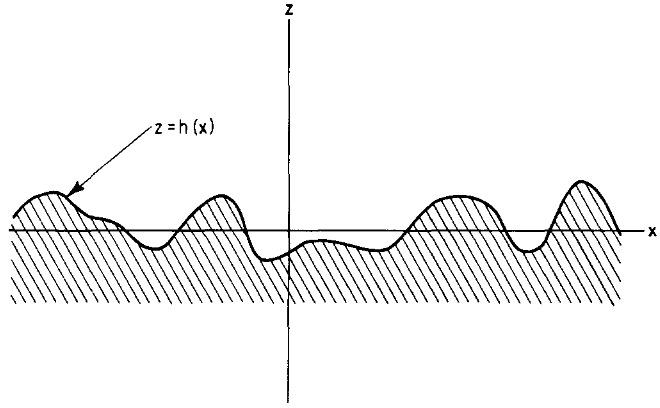


FIG. 1. A cross section through a representative surface.

where $\mathbf{n}$ is the unit normal to the surface. When $\mathbf{x}$ is below the surface, $G_D(\mathbf{x}|\mathbf{x}')$ is zero; the notation sub D indicates that the Green's function is discontinuous across the surface. In Appendix A the determination of G_D is reduced to the solution of an integral equation.[4,5]

The results are

$$G_D^{\pm}(\mathbf{x}'|\mathbf{x}'') = G_0^{\pm}(\mathbf{x}'-\mathbf{x}'') + 2(2\pi)^3 \int d^2x_\perp[\hat{i}_z - \nabla_\perp h(\mathbf{x}_\perp)] \times \{\nabla' G_0^{\pm}(\mathbf{x}' - \mathbf{x}_S(\mathbf{x}_\perp))\} G_S^{\pm}(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}''). \tag{1}$$

where we have introduced the notation $\mathbf{x}_S(\mathbf{x}_\perp)$ for the location of a point in the surface, that is $\mathbf{x}_S(\mathbf{x}_\perp) = \hat{i}_z h(\mathbf{x}_\perp) + \mathbf{x}_\perp$, where $\hat{i}_z$ is a unit vector in the z direction. The "surface Green's function," $G_S^{\pm}(\mathbf{x}'_S|\mathbf{x}'')$, satisfies the integral equation

$$G_S^{\pm}(\mathbf{x}_S(\mathbf{x}'_\perp)|\mathbf{x}'') = (2\pi)^{-3} G_0^{\pm}(\mathbf{x}_S(\mathbf{x}'_\perp) - \mathbf{x}'') + 2\int d^2x_\perp(\hat{i}_z - \nabla_\perp h(\mathbf{x}_\perp)) \cdot \mathbf{G}_0^{\pm}(\mathbf{x}_S(\mathbf{x}'_\perp) - \mathbf{x}_S(\mathbf{x}_\perp)) \times G_S^{\pm}(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}''), \tag{2}$$

where

$$G_0^{\pm}(\mathbf{x}) = (2\pi)^{-3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}} G_0^{\pm}(\mathbf{k}), \quad G_0^{\pm}(\mathbf{k}) = (k_0^2 - \mathbf{k}^2 \pm i\epsilon)^{-1}, \tag{3}$$

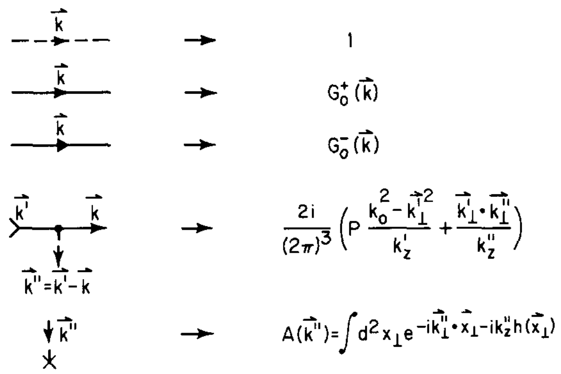


FIG. 2. Diagram rules for $D_S^{\pm}(\mathbf{k}'|\mathbf{k}'')$. Each diagram in the series has equal weight and is constructed by multiplying the indicated factors associated with the lines and vertices of the diagrams. Integration over the three momenta $\mathbf{k}$ associated with internal $G_0^{\pm}$ lines completes the construction of the term in the series associated with the diagram.

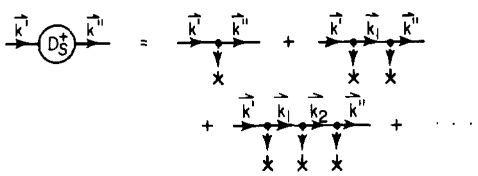


FIG. 3. Diagrammatic representation of the series for $D_S^{\pm}$.

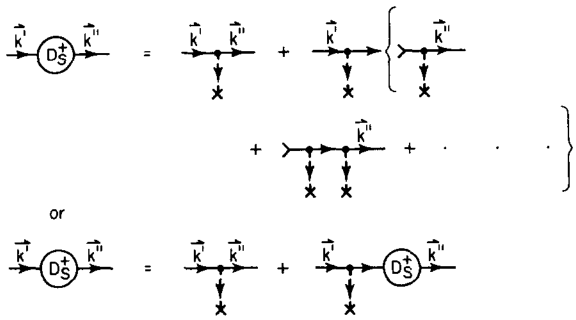


FIG. 4. Diagrammatic summation of the series to recover the integral equation for $D_S^{\pm}$.

$G_0^{\pm}$ being the Green's function for the situation where no surface is present, and where

$$\mathbf{G}_0^{\pm}(\mathbf{x}) = i(z\pi)^{-3}\int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k_0^2 - \mathbf{k}^2 \pm i\epsilon}\left(P\frac{k_0^2 - \mathbf{k}_\perp^2}{k_z}\hat{i}_z + \mathbf{k}_\perp\right) \tag{4}$$

The P indicates the Cauchy principal value distribution.

The integral equations (1) and (2) have a form more suitable for our purposes when $G_D^{\pm}(\mathbf{x}'|\mathbf{x}'')$ and $G_S^{\pm}(\mathbf{x}'|\mathbf{x}'')$ are expressed in terms of a Fourier transform. Let

$$G_{D,S}^{\pm}(\mathbf{x}'|\mathbf{x}'') = (2\pi)^{-6}\int d^3k'\, d^3k''\, e^{i(\mathbf{k}'\cdot\mathbf{x}' - \mathbf{k}''\cdot\mathbf{x}'')} G_{D,S}^{\pm}(\mathbf{k}'|\mathbf{k}''). \tag{5}$$

Substituting Eqs. (3) and (5) into Eq. (2) gives

$$\int d^3k'\, e^{i\mathbf{k}'\cdot\mathbf{x}_S(\mathbf{x}'_\perp)}\{G_S^{\pm}(\mathbf{k}'|\mathbf{k}'') - \delta_3(\mathbf{k}'-\mathbf{k}'')G_0^{\pm}(\mathbf{k}'') - [2i/(2\pi)^3]G_0^{\pm}(\mathbf{k}')\int d^3k\, F(\mathbf{k}',\mathbf{k})G_S^{\pm}(\mathbf{k}|\mathbf{k}'')\} = 0, \tag{6}$$

where

$$F(\mathbf{k}',\mathbf{k}) = \int d^2x_\perp \exp^{[-i(\mathbf{k}'_\perp - \mathbf{k}_\perp)\cdot\mathbf{x}_\perp]}\left(P\frac{k_0^2 - \mathbf{k}'^2_\perp}{k'_z} - \frac{i\mathbf{k}'_\perp\cdot\nabla_\perp}{k'_z - k_z}\right)\exp^{[-i(k'_z - k_z)h(\mathbf{x}_\perp)]}.$$

Integrating by parts gives

$$F(\mathbf{k}',\mathbf{k}) = \left(P\frac{k_0^2 - \mathbf{k}'^2_\perp}{k'_z} + \frac{\mathbf{k}'_\perp\cdot(\mathbf{k}'_\perp - \mathbf{k}_\perp)}{k'_z - k_z}\right)A(\mathbf{k}'-\mathbf{k}), \tag{7}$$

where

$$A(\mathbf{k}) = \int d^2x_\perp \exp^{[-i\mathbf{k}\cdot\mathbf{x}_S(\mathbf{x}_\perp)]}. \tag{8}$$

The surface term arising from the integration by parts has been neglected. This can be justified by noting that $h(\mathbf{x}_\perp)$ can be multiplied by a spatial cutoff, such as $\exp(-a^2\mathbf{x}_\perp^2)$, and the field, when defined with respect to an appropriate set of test functions (representing a set of finite detectors), will approach a limit as $a^2 \to 0$. This limit is to be taken with the source and detector in the near field region associated with the cutoff function.

The normal procedure to complete the transformation to $\mathbf{k}$ space would be to apply an inverse Fourier transform to (6). This cannot be done since (6) is only defined on the surface. However, if the quantity in the curly brackets in (6) is taken to be zero, then (6) is satisfied. Thus, we get the integral equation

$$G_S^{\pm}(\mathbf{k}'|\mathbf{k}'') = \delta_3(\mathbf{k}'-\mathbf{k}'')G_0^{\pm}(\mathbf{k}'') + \frac{2i}{(2\pi)^3}G_0^{\pm}(\mathbf{k}')\int d^3k\left(P\frac{k_0^2 - \mathbf{k}'^2_\perp}{k'_z} + \frac{\mathbf{k}'_\perp\cdot(\mathbf{k}'_\perp - \mathbf{k}_\perp)}{k'_z - k_z}\right) \times A(\mathbf{k}'-\mathbf{k})G_S^{\pm}(\mathbf{k}|\mathbf{k}''), \tag{9}$$

which is a sufficient condition that (6) be satisfied.

The Fourier transform of the full Green's function $G_D^{\pm}(\mathbf{k}'|\mathbf{k}'')$ can be related to $G_S^{\pm}(\mathbf{k}'|\mathbf{k}'')$ by Fourier transforming Eq. (1) and via (3) and (5). Following the same procedure used to derive Eq. (9) gives

$$G_D^{\pm}(\mathbf{k}'|\mathbf{k}'') = (2\pi)^3\delta_3(\mathbf{k}'-\mathbf{k}'')G_D^{\pm}(\mathbf{k}'') + (2\pi)^3 G_0^{\pm}(\mathbf{k}')\int d^3k \frac{2i}{(2\pi)^3}\frac{\mathbf{k}'\cdot(\mathbf{k}'-\mathbf{k})}{k'_z - k_z}A(\mathbf{k}'-\mathbf{k})G_S^{\pm}(\mathbf{k}|\mathbf{k}''). \tag{10}$$

Clearly (6) does not uniquely imply (9). In fact, an additional function can be added to the right-hand side of (9). This function is arbitrary except for the condition that it must be orthogonal to $e^{i\mathbf{k}\cdot\mathbf{x}_S(\mathbf{x}_\perp)}$ for all $\mathbf{x}_\perp$. We have chosen this function to be zero. If it is not assumed to be zero, then it can be shown that $G_S^\pm(\mathbf{k}'|\mathbf{k}'')$ is changed but $G_D^\pm(\mathbf{k}'|\mathbf{k}'')$ is not. This situation is analogous to the gauge transformations of electrodynamics. The basic formulation of the problem in momentum, or $\mathbf{k}$, space is now complete.

We will next consider an iterative series solution to Eq. (9). Let us define a new function $D_S^\pm$ by the equation

$$G_S^\pm(\mathbf{k}'|\mathbf{k}'') = \delta_3(\mathbf{k}'-\mathbf{k}'')G_0^\pm(\mathbf{k}'') + D_S^\pm(\mathbf{k}'|\mathbf{k}''). \tag{11}$$

An integral equation for $D_S^\pm$ is easily derived from (9) giving

$$\begin{aligned}D_S^\pm(\mathbf{k}'|\mathbf{k}'') &= G_0^\pm(\mathbf{k}')\frac{2i}{(2\pi)^3}\left(P\frac{k_0^2-\mathbf{k}_\perp'^2}{k_z'}+\frac{\mathbf{k}'_\perp\cdot(\mathbf{k}'_\perp-\mathbf{k}''_\perp)}{k_z'-k_z''}\right)\\ &\quad\times A(\mathbf{k}'-\mathbf{k}'')G_0^\pm(\mathbf{k}'')\\ &\quad+G_0^\pm(\mathbf{k}')\int d^3k\,\frac{2i}{(2\pi)^3}\left(P\frac{k_0^2-\mathbf{k}_\perp'^2}{k_z'}+\frac{\mathbf{k}'_\perp\cdot(\mathbf{k}'_\perp-\mathbf{k}''_\perp)}{k_z'-k_z}\right)\\ &\quad\times A(\mathbf{k}'-\mathbf{k})D_S^\pm(\mathbf{k}|\mathbf{k}'').\end{aligned} \tag{12}$$

This integral equation can be iterated to yield a series for $D_S^\pm$. Examination of this series yields a set of diagrammatic rules for the calculation of a general term in the series. These rules are shown in Fig. 2. The series for D_S^+ can be diagrammatically represented as shown in Fig. 3. A similar series of diagrams can be written for D_S^-. It is easy to see that this series can be summed to reproduce the integral equation for D_S^+ as is indicated in Fig. 4. The last equation is a diagrammatic representation of (12). It is also convenient to define the functions D_D^+ and D_D^- by the equation

$$G_D^\pm(\mathbf{k}'|\mathbf{k}'') = (2\pi)^3\delta_3(\mathbf{k}'-\mathbf{k}'')G_0^\pm(\mathbf{k}'') + (2\pi)^3D_D^\pm(\mathbf{k}'|\mathbf{k}'').$$

Via (10), $D_D^\pm$ can be expressed in terms of $G_S^\pm$ giving

$$D_D^\pm(\mathbf{k}'|\mathbf{k}'') = G_0^\pm(\mathbf{k}')\int d^3k\,\frac{2i}{(2\pi)^3}\,\frac{\mathbf{k}'\cdot(\mathbf{k}'-\mathbf{k})}{k_z'-k_z}\times A(\mathbf{k}'-\mathbf{k})G_S^\pm(\mathbf{k}|\mathbf{k}'').$$

The full Green's function $G_D^\pm(\mathbf{x}'|\mathbf{x}'')$ can be calculated from these two equations and (11) once $D_S^\pm$ is known.

Actually, the Green's function is a solution more general than one usually needs. Often it is the case that the sources and detectors are located in regions sufficiently far above the surface that the cutoff surface modes have decayed enough that they can be neglected. The connection between these solutions and the Green's function solutions is given by the reduction formula. The result is a scattering matrix for the surface which relates the plane wave decomposition of the scattered fields to the plane wave decomposition of the incident field $\psi_i(\mathbf{x})$ which satisfies the homogeneous Helmholtz equation.

In order to discuss the reduction formula, let us rewrite the Green's function $G_D^\pm$ in a form which explicitly displays the external free Green's functions evident in the expansion for $D_D^\pm$. This follows from the expansion for $D_S^\pm$ and the last equation above. Thus, we have

$$\begin{aligned}G_D^\pm(\mathbf{x}'|\mathbf{x}'') &= G_0^\pm(\mathbf{x}'-\mathbf{x}'')\\ &\quad+(2\pi)^3\int d^3x_1\,d^3x_2\,G_0^\pm(\mathbf{x}'-\mathbf{x}_1)\rho^\pm(\mathbf{x}_1|\mathbf{x}_2)G_0^\pm(\mathbf{x}_2-\mathbf{x}''),\end{aligned} \tag{13}$$

where

$$\begin{aligned}&\rho^\pm(\mathbf{x}_1|\mathbf{x}_2)\\ &=\int d^3k_1\,d^3k_2\,e^{i\mathbf{k}_1\cdot\mathbf{x}_1}(k_0^2-\mathbf{k}_1^2)D_D(\mathbf{k}_1|\mathbf{k}_2)(k_0^2-\mathbf{k}_2^2)e^{-i\mathbf{k}_2\cdot\mathbf{x}_2}.\end{aligned}$$

The advantage of this form is that ρ is localized on the surface.

The scattered or outgoing field $\psi_o(\mathbf{x})$, due to an incoming free field $\psi_i(\mathbf{x})$, can be found by observing that any incoming free field can be constructed by situating the appropriate source distributions at $z''=+\infty$. Following this line of reasoning it is easy to see that the outgoing field is given by

$$\psi_o(\mathbf{x}') = (2\pi)^3\int d^3x_1\,d^3x_2\,G_0^+(\mathbf{x}'|\mathbf{x}_1)\rho^+(\mathbf{x}_1|\mathbf{x}_2)\psi_i(\mathbf{x}_2). \tag{14}$$

The first term in (13) does not contribute to the scattered field since it gives back the incident field and can be dropped.

In order to reduce (14) further, let us rewrite G_0^+ in the form

$$G_0^+(\mathbf{x}) = \frac{1}{2i(2\pi)^2}\int d^2k_\perp\frac{e^{i\mathbf{k}_\perp\cdot\mathbf{x}_\perp+ik_z|z|}}{k_z},$$

where

$$k_z = +(k_0^2-\mathbf{k}_\perp^2)^{1/2}$$

Substituting this into (14) and taking the observation point inside the region where the cutoff surface waves have damped out gives

$$\psi_o(\mathbf{x}') = \int d^2k_\perp'\,e^{i\mathbf{k}'\cdot\mathbf{x}'}\varphi_o(\mathbf{k}_\perp'),\quad k_z' = +(k_0^2-\mathbf{k}_\perp'^2)^{1/2}, \tag{15}$$

where

$$\varphi_o(\mathbf{k}_\perp') = (\pi/ik_z')\int d^3x_1\,d^3x_2\,e^{-i\mathbf{k}'\cdot\mathbf{x}_1}\rho^+(\mathbf{x}_1|\mathbf{x}_2)\psi_i(\mathbf{x}_2).$$

Decomposing the incident field into a plane wave expansion

$$\psi_i(\mathbf{x}'') = \int d^2k_\perp''\,e^{i\mathbf{k}''\cdot\mathbf{x}''}\varphi_i(\mathbf{k}_\perp''), \tag{16}$$

where

$$k_z'' = -(k_0^2-\mathbf{k}_\perp''^2)^{1/2},$$

gives

$$\varphi_o(\mathbf{k}_\perp') = \int d^2k_\perp''\,T^+(\mathbf{k}_\perp'|\mathbf{k}_\perp'')\varphi_i(\mathbf{k}_\perp''), \tag{17}$$

where the scattering matrix $T^\pm$ is given by

$$T^\pm(\mathbf{k}_\perp'|\mathbf{k}_\perp'') = \lim_{\substack{k_z'\to\pm(k_0^2-\mathbf{k}_\perp'^2)^{1/2}\\ k_z''\to\mp(k_0^2-\mathbf{k}_\perp''^2)^{1/2}}}(\pi/ik_z')\times(k_0^2-\mathbf{k}'^2)D_D^\pm(\mathbf{k}'|\mathbf{k}'')(k_0^2-\mathbf{k}''^2). \tag{18}$$

This last equation can be put in a more convenient form by noticing that in the limits noted $D_D^\pm$ can be replaced by $D_S^\pm$. This gives

$$T^{\pm}(\mathbf{k}'_{\perp}|\mathbf{k}''_{\perp}) = \lim_{\substack{k'_z \to \pm(k_0^2 - \mathbf{k}'^2_{\perp})^{1/2} \\ k''_z \to \mp(k_0^2 - \mathbf{k}''^2_{\perp})^{1/2}}} (\pi/ik'_z) \times (k_0^2 - \mathbf{k}'^2) D_S^{\pm}(\mathbf{k}'|\mathbf{k}'')(k_0^2 - \mathbf{k}''^2). \quad (19)$$

Similarly for the complex conjugate field it follows that

$$\psi_o^*(\mathbf{x}') = \int d^2k'_{\perp}\, e^{-i\mathbf{k}'\cdot\mathbf{x}'} \varphi_o^*(\mathbf{k}'_{\perp}), \quad k'_z = +(k_0^2 - \mathbf{k}'^2_{\perp})^{1/2}, \quad (20)$$

$$\psi_i^*(\mathbf{x}'') = \int d^2k''_{\perp}\, e^{-i\mathbf{k}''\cdot\mathbf{x}''} \varphi_i^*(\mathbf{k}''_{\perp}), \quad k''_z = -(k_0^2 - \mathbf{k}''^2_{\perp})^{1/2}, \quad (21)$$

$$\varphi_o^*(\mathbf{k}'_{\perp}) = \int d^2k''_{\perp} T^{-}(-\mathbf{k}'_{\perp}|-\mathbf{k}''_{\perp}) \varphi_i^*(\mathbf{k}''_{\perp}). \quad (22)$$

Let us now turn to the application of these results to scattering from random rough surfaces.

III. RANDOM ROUGH SURFACES

Having considered scattering from a general surface in the preceding section, we will apply these results to the calculation of moments of Green's functions and moments of fields scattered from a random rough surface. The moments to be considered are

$$\langle \prod_{i,j} G_D^+(\mathbf{x}_i|\mathbf{x}_j) \prod_{m,n} G_D^-(\mathbf{x}_m|\mathbf{x}_n) \rangle$$

and

$$\langle \prod_i \psi_o(\mathbf{x}_i) \prod_j \psi_o^*(\mathbf{x}_j) \rangle.$$

We will assume that the moments can be calculated by averaging over an ensemble of surfaces characterized by a multivariate Gaussian height distribution.

Looking back at the results of Sec. II we see that the above moments can be determined from the moments of the $D_S^{\pm}(\mathbf{k}'|\mathbf{k}'')$ functions, so we are led to consider the moments

$$\langle \prod_{i,j} D_S^+(\mathbf{k}_i|\mathbf{k}_j) \prod_{m,n} D_S^-(\mathbf{k}_m|\mathbf{k}_n) \rangle.$$

These are the basic quantities of interest. As before we will determine a set of diagrammatic rules for constructing a general term in a series for these moments. From the rules derived in Sec. II we see that the surface height enters only through the function $A(\mathbf{k})$, where

$$A(\mathbf{k}) = \int d^2x_{\perp} \exp[-i(\mathbf{k}_{\perp}\cdot\mathbf{x}_{\perp} + k_z h(\mathbf{x}_{\perp}))]. \quad (23)$$

The series for $D_S^{\pm}$ and the series for products of $D_S^{\pm}$ contain products of A functions. Thus, the moments of the $D_S^{\pm}$ functions can be written in terms of a sum of terms each consisting of integrals over functions containing moments of the A function, $\langle \prod_i A(\mathbf{k}_i) \rangle$. In order to discuss the properties of these moments, it is useful to note that they are just Fourier transforms in $\mathbf{x}_{\perp}$ of the characteristic functions of the joint probability distribution of heights. The n-point characteristic function for the surface F_n is defined as

$$F_n(k_{1z}, \ldots, k_{nz}; \mathbf{x}_{1\perp}, \ldots, \mathbf{x}_{n\perp}) = \left\langle \exp\left(-i \sum_{j=1}^{n} k_{jz} h(\mathbf{x}_{j\perp})\right)\right\rangle. \quad (24)$$

For a multivariate Gaussian distribution in h with $\langle h \rangle = 0$, the n-point characteristic function is given by[6]

$$F_n(\{k_z\}_n; \{\mathbf{x}_{\perp}\}_n) = \exp\left(-\frac{1}{2}\sum_{j,l=1}^{n} k_{jz}\Gamma(\mathbf{x}_{j\perp} - \mathbf{x}_{l\perp})k_{lz}\right), \quad (25)$$

where $\{k_z\}_n$ is used to denote the set $k_{1z}, \ldots, k_{nz}$, and similarly $\{x_{\perp}\}_n$ the set $\mathbf{x}_{1\perp}, \ldots, \mathbf{x}_{n\perp}$. The correlation function $\Gamma(\mathbf{x}_{\perp})$ is defined as

$$\Gamma(\mathbf{x}_{1\perp} - \mathbf{x}_{2\perp}) = \langle h(\mathbf{x}_{1\perp}) h(\mathbf{x}_{2\perp}) \rangle. \quad (26)$$

It is a function of $\mathbf{x}_{1\perp} - \mathbf{x}_{2\perp}$ because we assume that the ensemble of surfaces is invariant under translation in the $\mathbf{x}_{\perp}$ plane.

An examination of the properties of the moments of A is now necessary. We will assume that $\Gamma(\mathbf{x}_{\perp})$ vanishes as $|\mathbf{x}_{\perp}| \to \infty$. This assumption has important implications for the singularity structure in $\{\mathbf{k}_{\perp}\}_n$ of $\langle \prod_i A(\mathbf{k}_i) \rangle$. These implications can be best exemplified by examining the two-point moment of A explicitly. The two-point moment of A is given by

$$\langle A(\mathbf{k}_1)A(\mathbf{k}_2) \rangle = \int d^2x_{1\perp} d^2x_{2\perp} \exp[-i(\mathbf{k}_{1\perp}\cdot\mathbf{x}_{1\perp} + \mathbf{k}_{2\perp}\cdot\mathbf{x}_{2\perp}) - \tfrac{1}{2}\Gamma(0)(k_{1z}^2 + k_{2z}^2) - k_{1z}k_{2z}\Gamma(\mathbf{x}_{1\perp} - \mathbf{x}_{2\perp})].$$

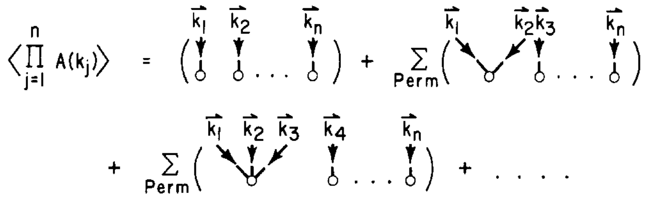


FIG. 5. Cluster expansion for moments of $A(\mathbf{k})$. The symbol $\sum_{\text{perm}}$ indicates a sum over all different labelings of the diagrams.

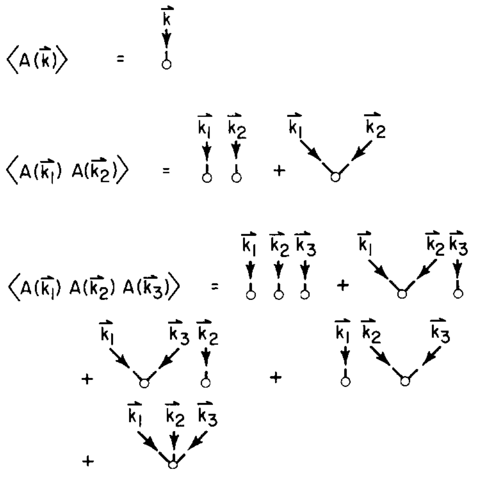


FIG. 6. Cluster expansion for the lowest three moments of $A(\mathbf{k})$.

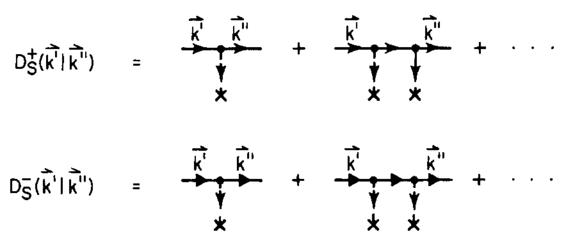


FIG. 7. Diagrammatic representation of the series for D_S^+ and D_S^-.

This can be rewritten as

$$\langle A(\mathbf{k}_1)A(\mathbf{k}_2)\rangle = (2\pi)^2 \exp[-\tfrac{1}{2}\Gamma(0)(k_{1z}^2 + k_{2z}^2)]\delta_2(\mathbf{k}_{1\perp} + \mathbf{k}_{2\perp}) \times \Big\{(2\pi)^2\delta_2(\mathbf{k}_{1\perp}) + \int d^2y_{1\perp}\, e^{(-i\mathbf{k}_{1\perp}\cdot\mathbf{y}_{1\perp})} \times [e^{-i\mathbf{k}_{1z}k_{2z}\Gamma(\mathbf{y}_\perp)} - 1]\Big\}.$$

The important point in the expression is that the second term in the curly brackets is free of δ-function singularities since

$$e^{-k_{1z}k_{2z}\Gamma(\mathbf{y}_\perp)} - 1 \xrightarrow[|\mathbf{y}_\perp|\to\infty]{} 0.$$

This separation of $\langle A(\mathbf{k}_1)A(\mathbf{k}_2)\rangle$ into terms with different singularity structure can be extended to the general function $\langle \Pi_{i=1}^n A(\mathbf{k}_i)\rangle$. The result is the cluster expansion discussed in Appendix B. There it is shown that this separation is accomplished by letting

$$\left\langle \prod_{j=1}^{n} A(\mathbf{k}_j)\right\rangle = \sum_{j\,\mathrm{perm}} \sum_{M=1}^{n} \sum_{\{m_i\}_M} \prod_{i=1}^{M} A_{m_i}(\{\mathbf{k}_j\}_{m_i}). \tag{27}$$

The notation $\sum_{\{m_i\}_M}$ denotes a sum over all unordered M element sets $\{m_i\}_m$ such that $\sum_{i=1}^M m_i = n$, and $\sum_{j\,\mathrm{perm}}$ denotes a sum over all different labelings, j, of the unordered m_i element sets $\{\mathbf{k}_j\}_{m_i}$ with $j = 1, \ldots, n$. The functions A_{m_i} are discussed in Appendix B. Equation (27) can be very conveniently represented by the series of diagrams shown in Fig. 5. The first few diagrams in the cluster expansion for $\langle \Pi A\rangle$ are explicitly shown in Fig. 6.

Since $D_S^\pm(\mathbf{k}'|\mathbf{k}'')$ can be expressed in terms of series involving integrals over products of A functions, we can use the cluster expansion for the expectation value of products of A functions to find a set of rules for generating a general term in the series for $\langle \pi D_S\rangle$. Let us recall from Sec. II the diagrammatic representation of D_S^+ as shown in Fig. 7, where each $\dot{x}^{\mathbf{k}}$ represents a factor of $A(\mathbf{k})$ and also recall the similar representation for D_S^-. Combining these rules for $D_S^\pm$ with the rules for the cluster expansion of $\langle \Pi_j A(\mathbf{k}_j)\rangle$ yields a set of rules for calculating a general term in the series for $\langle \Pi D\rangle$. These rules are shown in Fig. 8. A more complete discussion of the combinatorial problem leading to the rules can be found in Ref. 3 by Frisch.

An example of the application of these rules is shown in Fig. 9 where the diagram series for $\langle D_S^+\rangle$ is displayed.

The mutual coherence function $\langle \psi_o(\mathbf{x})\psi_o^*(\mathbf{x}')\rangle$ is a quantity of particular interest because of its close relationship to the scattered intensity, which is one of the most accessible pieces of data experimentally. It can be related to averages of D_S functions through the reduction formula derived in the previous section. From Eqs. (16)–(22) of Sec. II it follows that

$$\langle \psi_o(\mathbf{x})\psi_o^*(\mathbf{x}')\rangle = \int d^2k_\perp d^2k'_\perp e^{i(\mathbf{k}\cdot\mathbf{x}-\mathbf{k}'\cdot\mathbf{x}')}\langle \varphi_o(\mathbf{k}_\perp)\varphi_o^*(\mathbf{k}'_\perp)\rangle, \tag{28}$$

where

$$\langle \varphi_o(\mathbf{k}_\perp)\varphi_o^*(\mathbf{k}'_\perp)\rangle = \int d^2k_{i\perp}d^2k'_{i\perp}\langle T^+(\mathbf{k}_\perp|\mathbf{k}_{i\perp})T^-(-\mathbf{k}'_\perp|-\mathbf{k}'_{i\perp})\rangle \times \varphi_i(\mathbf{k}_{i\perp})\varphi_i^*(\mathbf{k}'_{i\perp}) \tag{29}$$

and where

$$\langle T^+(\mathbf{k}_\perp|\mathbf{k}_{i\perp})T^-(-\mathbf{k}'_\perp|-\mathbf{k}'_{i\perp})\rangle = \lim_{\substack{k_z\to +(k_0^2-\mathbf{k}_\perp^2)^{1/2}\\ k'_z\to +(k_0^2-\mathbf{k}'^2_\perp)^{1/2}}} \lim_{\substack{k_{iz}\to -(k_0^2-\mathbf{k}_{i\perp}^2)^{1/2}\\ k'_{iz}\to -(k_0^2-\mathbf{k}'^2_{i\perp})^{1/2}}} \times (\pi^2/k_zk'_z)(k_0^2-\mathbf{k}^2)(k_0^2-\mathbf{k}'^2)\langle D_S^+(\mathbf{k}|\mathbf{k}_i) \times D_S^-(-\mathbf{k}'|-\mathbf{k}'_i)\rangle(k_0^2-\mathbf{k}_i^2)(k_0^2-\mathbf{k}'^2_i). \tag{30}$$

Due to translational invariance in the $\mathbf{x}_\perp$ plane, a δ function can be extracted from this last equation, that is, we can write

$$\langle T^+(\mathbf{k}_\perp|\mathbf{k}_{i\perp})T^-(-\mathbf{k}'_\perp|-\mathbf{k}'_{i\perp}) = \delta_2[(\mathbf{k}_\perp-\mathbf{k}'_\perp)-(\mathbf{k}_{i\perp}-\mathbf{k}'_{i\perp})] \times F(\tfrac{1}{2}(\mathbf{k}_\perp+\mathbf{k}'_\perp), \mathbf{k}_\perp-\mathbf{k}'_\perp, \tfrac{1}{2}(\mathbf{k}_{i\perp}+\mathbf{k}_{i\perp})). \tag{31}$$

These general expressions are somewhat complex. They are considerably simpler if the incident wave is a plane wave, a case which is also of practical interest. For this case

$$\varphi_i(\mathbf{k}_\perp) = \delta_2(\mathbf{k}_\perp - \mathbf{k}_{i\perp})$$

and, thus,

$$\langle \varphi_o(\mathbf{k}_\perp)\varphi_0^*(\mathbf{k}'_\perp)\rangle = \langle T^+(\mathbf{k}_\perp|\mathbf{k}_{i\perp})T^-(-\mathbf{k}'_\perp|-\mathbf{k}_{i\perp})\rangle = \delta_2(\mathbf{k}_\perp-\mathbf{k}'_\perp)I(\mathbf{k}_\perp,\mathbf{k}_{i\perp}), \tag{32}$$

where I is the intensity scattered in the $\mathbf{k}_\perp$ direction due to a plane wave incident from the $\mathbf{k}_{i\perp}$ direction. Using the rules for averages of products of $D_S^\pm$ functions and Eqs. (30) and (32), we can write down the lowest-order coherent and incoherent scattered intensities.

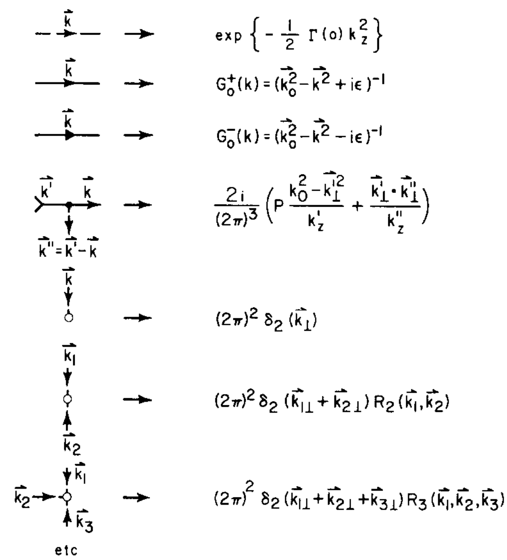


FIG. 8. Diagram rules for $\langle \Pi D_S\rangle$. Each diagram in the series has equal weight and is constructed by multiplying the indicated factors associated with the lines and vertices of the diagrams. Integration over the three momenta $\mathbf{k}$ associated with internal $G_0^\pm$ lines completes the construction of the term in the series associated with the diagram. The function R_n is discussed in Appendix B.

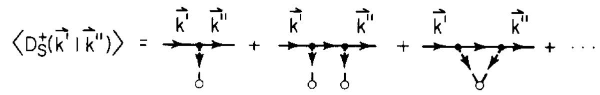


FIG. 9. Diagram series for $\langle D_S^+\rangle$.

The lowest-order coherent, or specularly-scattered, intensity is given by the diagram shown in Fig. 10, which on evaluation gives

$$I(\mathbf{k}_\perp,\mathbf{k}_{i\perp}) = \delta_2(\mathbf{k}_\perp - \mathbf{k}_{i\perp}) \exp[-4\Gamma(0)(k_0^2 - \mathbf{k}_\perp^2)].$$

The δ function indicates that the scattering is specular. The normalization of I follows from this expression since $\Gamma(0) = 0$ corresponds to mirror reflection.

Similarly, the lowest-order, incoherently-scattered intensity is given by the diagram shown in Fig. 11, which on evaluation gives

$$I(\mathbf{k}_\perp,\mathbf{k}_{i\perp}) = \left(\frac{\mathbf{k}\cdot(\mathbf{k}-\mathbf{k}_i)}{k_z(k_z - k_{iz})}\right)^2 \exp[-\Gamma(0)(k_z - k_{iz})^2]$$
$$\times (2\pi)^{-2} \int d^2y_\perp e^{i(\mathbf{k}_\perp - \mathbf{k}_{i\perp})\cdot \mathbf{y}_\perp}\{\exp[\Gamma(\mathbf{y}_\perp)(k_z - k_{iz})^2] - 1\},$$

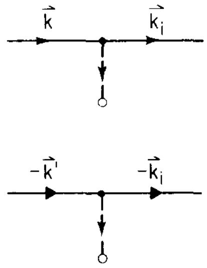


FIG. 10. Diagram for the lowest-order specular scattering.

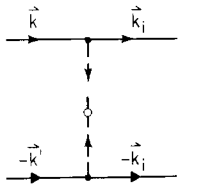


FIG. 11. Diagram for the lowest-order incoherent scattering.

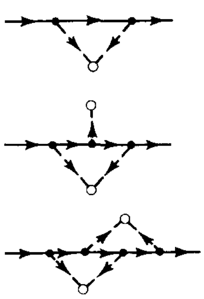


FIG. 12. Examples of connected diagrams contributing to $\langle D_S^+\rangle$.

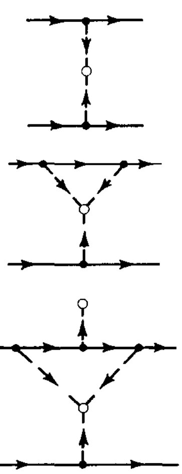


FIG. 13. Examples of connected diagram contributing to $\langle D_S^+ D_S^-\rangle$.

where

$$k_z = +(k_0^2 - \mathbf{k}_\perp^2)^{1/2}, \quad k_{iz} = -(k_0^2 - \mathbf{k}_{i\perp}^2)^{1/2}.$$

These results correspond to the usual results based on the Kirchhoff approximation.[1]

We will now rewrite the series for the first two moments of G in terms of linear integral equations by partially summing their series. There are basically two reasons for doing this. First, it considerably reduces the number of diagrams that must be individually considered when constructing approximate solutions and when performing formal manipulations, and second, it probably improves the convergence of the series. The reasons for this latter conjecture will be mentioned later. To facilitate this development, let us categorize the diagrams into classes. We will call a diagram disconnected if it can be broken into two parts, not connected to each other by any lines, by removing a G_0^+ and/or a G_0^- line. A diagram, which is not disconnected, is connected. Some examples of connected diagrams in the expansion of $\langle D\rangle$ are shown in Fig. 12. Similarly, Fig. 13 shows some connected diagrams for $\langle D^+D^-\rangle$.

Examination of the series for $\langle D_S^\pm(\mathbf{k}\mid \mathbf{k}_1)\rangle$ reveals that one can write

$$\langle D_S^\pm(\mathbf{k}|\mathbf{k}_1)\rangle = P_1^\pm(\mathbf{k}|\mathbf{k}_1)G_0^\pm(\mathbf{k}_1) + \int d^3k_2 P_1(\mathbf{k}|\mathbf{k}_2)\langle D_S^\pm(\mathbf{k}_2|\mathbf{k}_1)\rangle, \tag{33}$$

where $P_1^\pm(\mathbf{k}|\mathbf{k}_1)G_0^\pm(\mathbf{k}_1)$ is equal to the sum of the connected diagrams contained in the series for $\langle D_S^\pm(\mathbf{k}|\mathbf{k}_1)\rangle$. Equation (33) can be reduced to a one-dimensional integral equation. Because of translational invariance, a δ function can be factored out of P_1. Therefore, we can write

$$P_1^\pm(\mathbf{k}|\mathbf{k}') = G_0^\pm(k_z)\mathcal{P}_1^\pm(k_z|k_z')\delta_2(\mathbf{k}_\perp - \mathbf{k}_\perp') \tag{34}$$

and

$$\langle D_S^\pm(\mathbf{k}|\mathbf{k}')\rangle = G_0^\pm(k_z)\mathcal{D}^\pm(k_z|k_z')G_0^\pm(k_z')\delta_2(\mathbf{k}_\perp - \mathbf{k}_\perp'). \tag{35}$$

Combining (33), (34), and (35) gives

$$\mathcal{D}^\pm(k_z|k_z') = \mathcal{P}_1^\pm(k_z|k_z') + \int dk_z'' \mathcal{P}_1^\pm(k_z|k_z'')G_0^\pm(k_z'')\mathcal{D}^\pm(k_z''|k_z'). \tag{36}$$

For convenience the $\mathbf{k}_\perp$ dependence of the above quantities has not been explicitly noted. The integral equation for $\langle G_S^\pm\rangle$ is also easily found from (33). It follows from Eq. (11) that

$$\langle G_S^\pm(\mathbf{k}|\mathbf{k}_1)\rangle = \delta_3(\mathbf{k} - \mathbf{k}_1)G_0^\pm(\mathbf{k}_1) + \langle D_S^\pm(\mathbf{k}|\mathbf{k}_1)\rangle.$$

The resulting integral equation is

$$\langle G_S^\pm(\mathbf{k}|\mathbf{k}_1)\rangle = \delta_3(\mathbf{k} - \mathbf{k}_1)G_0^\pm(\mathbf{k}_1) + \int d^3k_2 P_1^\pm(\mathbf{k}|\mathbf{k}_2)\langle G_S^\pm(\mathbf{k}_2|\mathbf{k}_1)\rangle. \tag{37}$$

Similarly, examination of the series for $\langle D_S^+|\mathbf{k}|\mathbf{k}')$ $D_S^-(\mathbf{k}_1|\mathbf{k}_1')\rangle$ reveals that one can write

$$\begin{aligned}\langle D_S^+(\mathbf{k}|\mathbf{k}_1)D_S^-(\mathbf{k}'|\mathbf{k}_1')\rangle &= \langle D_S^+(\mathbf{k}|\mathbf{k}_1)\rangle\langle D_S^-(\mathbf{k}'|\mathbf{k}_1')\rangle \\ &+ \int d^3k_2 P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}_1')\langle D_S^+(\mathbf{k}_2|\mathbf{k}')\rangle G_0^-(\mathbf{k}_1') \\ &+ \int d^3k_2' P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}_2')G_0^+(\mathbf{k}_1)\langle D_S^-(\mathbf{k}_2'|\mathbf{k}_1')\rangle \\ &+ P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}_1')G_0^+(\mathbf{k}_1)G_0^-(\mathbf{k}_1') \\ &+ \int d^3k_2 d^3k_2' P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}_2')\langle D_S^+(\mathbf{k}|\mathbf{k}_1)D_S^-(\mathbf{k}'|\mathbf{k}_1')\rangle,\end{aligned} \tag{38}$$

where

$$P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}_1')G_0^+(\mathbf{k}_1)G_0^-(\mathbf{k}_1')$$

is equal to the sum of connected graphs contained in the series for

$$\langle D_S^+(\mathbf{k}|\mathbf{k}_1)D_S^-(\mathbf{k}'|\mathbf{k}_1')\rangle.$$

Equation (38) can be rewritten using (11) to give an integral equation for $\langle G_S^+(\mathbf{k}|\mathbf{k}')G_S^-(\mathbf{k}_1|\mathbf{k}_1')\rangle$. The result is

$$\langle G_S^+(\mathbf{k}|\mathbf{k}')G_S^-(\mathbf{k}_1|\mathbf{k}_1')\rangle = \langle G_S^+(\mathbf{k}|\mathbf{k}')\rangle\langle G_S^-(\mathbf{k}_1|\mathbf{k}_1')\rangle + \int d^3k_2 d^3k_2' P_2(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}_2')\langle G_S^+(\mathbf{k}_2|\mathbf{k}_1)G_S^-(\mathbf{k}_2'|\mathbf{k}_1')\rangle. \quad (39)$$

Equations (37) and (39) are the equivalents of Dyson's equation and the Bethe–Salpeter equation for a rough surface. Earlier we mentioned that the sequence of approximations to $\langle D_S^\pm\rangle$ and $\langle D_S^+D_S^-\rangle$ generated by a sequence of finite series approximations to P_1 and P_2 probably are more convergent than the sequence for $\langle D_S^\pm\rangle$ and $\langle D_S^+D_S^-\rangle$ formed directly. Our reasoning is: This partial summation is necessary for the case of a random medium with homogeneous statistics, in order to remove secular terms. The introduction of inhomogeneous statistics removes the secular terms and the need for partial summation; but a residual effect must still be present. It is probably in the form of poor convergence. The similarity of this problem to rough surface scattering led to the conjecture made earlier.

As an example of what can be done with these integral equations, let us consider the lowest-order approximation to P_1 shown in Fig. 14. On evaluation this diagram gives

$$\mathcal{P}^\pm(\mathbf{k}|\mathbf{k}') = \frac{i}{\pi} P \frac{k_0^2 - \mathbf{k}_\perp^2}{k_z} \exp[-\tfrac{1}{2}\Gamma(0)(k_z - k_z')^2]. \quad (40)$$

We will call this the average surface approximation since the correlation length does not appear. This is what one would expect if the surface, in some sense, were averaged over translations in the $\mathbf{x}_\perp$ plane before scattering. It is the logical extension of the lowest-order result for coherent scattering from a rough surface. Combining (40) with (36) and letting

$$\mathfrak{D}^\pm(k_z|k_z') = \frac{i}{\pi} P \frac{k_0^2 - \mathbf{k}_\perp^2}{k_z} \tau^\pm(k_z|k_z') \quad (41)$$

gives an integral equation for $\tau^\pm$:

$$\tau^\pm(k_z|k_z') = \exp[-\tfrac{1}{2}\Gamma(0)(k_z - k_z')^2] + \frac{i}{\pi}\int dk_z'' \exp[-\tfrac{1}{2}\Gamma(0)(k_z - k_z'')^2]\left(P\frac{1}{k''}\right) \times \frac{k_0^2 - \mathbf{k}_\perp^2}{k_0^2 - \mathbf{k}_\perp^2 - k_z''^2 \pm i\epsilon}\tau^\pm(k_z''|k_z'). \quad (42)$$

The corresponding scattered intensity follows from (30), (32), (35), and (41). We have

$$I(\mathbf{k}_\perp,\mathbf{k}_{i\perp}) = |\tau^+(k_z|-k_z)|^2\delta_2(\mathbf{k}_\perp - \mathbf{k}_{i\perp}), \quad (43)$$

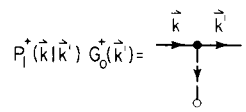


FIG. 14. Lowest-order approximation to P_1^+.

where

$$k_z = +(k_0^2 - \mathbf{k}_\perp^2)^{1/2}.$$

As the roughness of the surface vanishes or as $\Gamma(0) \to 0$ the solution of (42) should reduce to mirror reflection.

Observing that in this limit the right-hand side of (42) is independent of k_z and that the kernel of (42) is antisymmetric, one can immediately conclude that

$$\tau^\pm(k_z|k_z') \underset{\Gamma(0)\to 0}{\to} 1,$$

thus showing that mirror reflection is obtained.

IV. CONCLUDING REMARKS

A series solution to the moments of fields scattered from random rough surfaces has been developed. While explicit calculation of high-order terms in these series is probably impractical, the existence of a formally simple systematic procedure is very useful. For example, the series can be partially summed to construct integral equations for moments of Green's functions, two particularly interesting examples of which we have briefly discussed. These integral equations can be approximated in various ways depending on the particular problem under consideration. This latter area is one where good experimental data based on scattering from carefully controlled surfaces would be an aid in constructing useful approximations.

ACKNOWLEDGMENTS

We would like to thank Ken Flowers for his careful reading of our manuscript and suggestions. One of us, G. G. Z., would like to thank the National Research Council for its support.

APPENDIX A

In order to derive Eqs. (1)–(4), let us start with the continuous Green's function G_c which satisfies the inhomogeneous Helmholtz equation

$$(\nabla^2 + k_0^2)G_c(\mathbf{x}|\mathbf{x}') = \delta_3(\mathbf{x} - \mathbf{x}')$$

with the boundary condition that its normal derivative vanish on the surface given by $z = h(\mathbf{x}_\perp)$. The function G_c is regular everywhere except at $\mathbf{x} = \mathbf{x}'$. The free Green's function G_0 satisfies the same differential equation as G_c but it has pure outgoing or incoming boundary conditions as $|\mathbf{x} - \mathbf{x}'| \to \infty$. Applying Green's theorem gives

$$\nabla \cdot [G_0(\mathbf{x}|\mathbf{x}')\nabla G_c(\mathbf{x}|\mathbf{x}'') - G_c(\mathbf{x}|\mathbf{x}'')\nabla G_0(\mathbf{x}|\mathbf{x}')] = G_0(\mathbf{x}|\mathbf{x}')\delta_3(\mathbf{x} - \mathbf{x}'') - G_c(\mathbf{x}|\mathbf{x}'')\delta_3(\mathbf{x} - \mathbf{x}'). \quad (A1)$$

Multiplying (A1) by the unit step function $\theta[z - h(\mathbf{x}_\perp)]$

$$\theta(z) = \begin{cases} 1, & z > 0, \\ 0, & z < 0, \end{cases}$$

and integrating over all space gives

$$G_c(\mathbf{x}'|\mathbf{x}'')\theta[z' - h(\mathbf{x}_\perp')] = G_0(\mathbf{x}' - \mathbf{x}'')\theta(z'' - h(\mathbf{x}_\perp'')) + \int d^3x(\hat{i}_z - \nabla_\perp h(\mathbf{x}_\perp)) \cdot [G_0(\mathbf{x}' - \mathbf{x})\nabla G_c(\mathbf{x}|\mathbf{x}'') - G_c(\mathbf{x}|\mathbf{x}'')\nabla G_0(\mathbf{x}' - \mathbf{x})]\delta(z - h(\mathbf{x}_\perp)). \quad (A2)$$

The last term in (A2) has been integrated by parts. Applying the boundary condition

$$|\hat{\imath}_z - \nabla_\perp h(\mathbf{x}_\perp)] \cdot \nabla G_c(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}'') = 0,$$

where

$$\mathbf{x}_S(\mathbf{x}_\perp) = \mathbf{x}_\perp + \hat{\imath}_z h(\mathbf{x}_\perp)$$

gives

$$G_D(\mathbf{x}'|\mathbf{x}'') = G_0(\mathbf{x}' - \mathbf{x}'') + \int d^2x_\perp(\hat{\imath}_z - \nabla h(\mathbf{x}_\perp)) \cdot \{\nabla' G_0[\mathbf{x}' - \mathbf{x}_S(\mathbf{x}_\perp)]\} \times G_c(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}''), \tag{A3}$$

where

$$G_D(\mathbf{x}'|\mathbf{x}'') = G_c(\mathbf{x}'|\mathbf{x}'')\theta[z' - h(\mathbf{x}'_\perp)].$$

Also, we have assumed that $z'' > h(\mathbf{x}''_\perp)$, that is, the source is above the plane. In order to facilitate taking the surface limit of (A3) we will rewrite the kernel of (A3). Now

$$G_0(\mathbf{x}) = (2\pi)^{-3} \int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k_0^2 - \mathbf{k}^2 + i\epsilon} = G_0^+(\mathbf{x}) + G_0^{-*}(\mathbf{x})$$

carrying out the k_z integration gives ($\epsilon \to 0$)

$$G_0(\mathbf{x}) = (2\pi)^{-2} \int d^2k_\perp \frac{\exp\{i\mathbf{k}_\perp \cdot \mathbf{x}_\perp + i(k_0^2 - \mathbf{k}_\perp^2)^{1/2}|z|\}}{2i(k_0^2 - \mathbf{k}_\perp^2)^{1/2}}.$$

Thus,

$$\partial_z G_0(\mathbf{x}) = (2\pi)^{-2} \int d^2k_\perp \exp[i\mathbf{k}_\perp \cdot \mathbf{x}_\perp + i(k_0^2 - \mathbf{k}_\perp^2)^{1/2}|z|]\epsilon'(z), \tag{A4}$$

where

$$\epsilon'(z) = \begin{cases} +\frac{1}{2}, & z > 0, \\ -\frac{1}{2}, & z < 0. \end{cases}$$

Equation (A4) has a singularity at $z = 0$ which can be separated off in the following way;

$$\partial_z G_0(\mathbf{x}) = (2\pi)^{-2} \int d^2k_\perp \exp(i\mathbf{k}_\perp \cdot \mathbf{x}_\perp) \times (\exp[i(k_0^2 - \mathbf{k}_\perp^2)^{1/2}|z|] - 1)\epsilon'(z) + \delta_2(\mathbf{x}_\perp)\epsilon'(z).$$

Substituting this into (A3) gives

$$G_D(\mathbf{x}'|\mathbf{x}'') = G_0(\mathbf{x}' - \mathbf{x}'') + G_C(\mathbf{x}_S(\mathbf{x}'_\perp)|\mathbf{x}'')\epsilon[z' - h(\mathbf{x}'_\perp)] + \int d^2x_\perp(\hat{\imath}_z - \nabla h(\mathbf{x}_\perp)) \cdot \mathbf{G}_0(\mathbf{x}' - \mathbf{x}_S(\mathbf{x}_\perp))G_C(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}''), \tag{A5}$$

where

$$\mathbf{G}_0(\mathbf{x}) = i(2\pi)^{-3} \int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k_0^2 - \mathbf{k}^2 + i\epsilon}\left(P\frac{k_0^2 - \mathbf{k}_\perp^2}{k_z}\hat{\imath}_z + \mathbf{k}_\perp\right).$$

Since we are considering single-valued surfaces, we can let $\mathbf{x}'$ approach the surface from above and combine the second term on the right-hand side of (A5) with the left giving

$$G_S(\mathbf{x}_S(\mathbf{x}'_\perp)|\mathbf{x}'') = (2\pi)^{-3}G_0[\mathbf{x}_S(\mathbf{x}'_\perp) - \mathbf{x}''] + 2\int d^2x_\perp(\hat{\imath}_z - \nabla_\perp h(\mathbf{x}_\perp)) \cdot \mathbf{G}_0[\mathbf{x}_S(\mathbf{x}'_\perp) - \mathbf{x}_S(\mathbf{x}_\perp)] \times G_S(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}''), \tag{A6}$$

where

$$2(2\pi)^3 G_S(\mathbf{x}_S(\mathbf{x}'_\perp)|\mathbf{x}'') = \lim_{z' \to h(\mathbf{x}'_\perp)^+} G_D(\mathbf{x}'|\mathbf{x}'').$$

Rewriting (A3) in terms of G_S gives

$$G_D(\mathbf{x}'|\mathbf{x}'') = G_0(\mathbf{x}' - \mathbf{x}'') + 2(2\pi)^3 \int d^2x_\perp[\hat{\imath}_z - \nabla_\perp h(\mathbf{x}_\perp)] \cdot \{\nabla' G_0[\mathbf{x}' - \mathbf{x}_S(\mathbf{x}_\perp)]\} \times G_S(\mathbf{x}_S(\mathbf{x}_\perp)|\mathbf{x}'').$$

This equation and Eq. (A6) are the desired relations.

APPENDIX B

Let us define an irreducible characteristic function $K_n(\{k_z, \mathbf{x}_\perp\}_n)$ by [3,7]

$$F_n(\{k_z, \mathbf{x}_\perp\}_n) = \sum_{j\,\text{perm}} \sum_{M=1}^{n} \sum_{\{m_i\}_M} \prod_{i=1}^{M} K_{m_i}(\{k_{jz}, \mathbf{x}_{j\perp}\}_{m_i}), \tag{B1}$$

where $\sum_{\{m_i\}_M}$ denote a sum over all unordered M element sets $\{m_i\}_M$ such that $\sum_{i=1}^M m_i = n$ and $\sum_{j\,\text{perm}}$ denotes a sum over all different labelings, j, of the unordered sets $\{k_{jz}, \mathbf{x}_{j\perp}\}_{m_i}$ with $j = 1, \ldots, n$. Equation (B1) can be solved recursively for $K_n(\{k_z, \mathbf{x}_\perp\}_n)$. We will now show by induction that for $n > 1$ these irreducible characteristic functions have the property that they vanish as any $\mathbf{x}_\perp$ in the set is taken to be distant from any of the other members of the set.

Let us first rewrite (B1) in a more suitable form

$$K_n(\{k_z, \mathbf{x}_\perp\}_n) = F_n(\{k_z, \mathbf{x}_\perp\}_n) - \sum_{j\,\text{perm}} \sum_{M=2}^{n} \sum_{\{m_i\}_M} \prod_{i=1}^{M} K_{m_i}(\{k_{jz}, \mathbf{x}_{j\perp}\}_{m_i}). \tag{B2}$$

From Eq. (25) it follows that

$$F_n(\{k_z, \mathbf{x}_\perp\}_n) \xrightarrow[\{|\mathbf{x}'_\perp - \mathbf{x}''_\perp|\} \to \infty]{} F_m(\{k'_z, \mathbf{x}'_\perp\}_m)F_{n-m}), (\{k''_z, \mathbf{x}''_\perp\}_{n-m}) \tag{B3}$$

where $\{|\mathbf{x}'_\perp - \mathbf{x}''_\perp|\} \to \infty$ means that $|\mathbf{x}'_\perp - \mathbf{x}''_\perp| \to \infty$ for all $\mathbf{x}'_\perp$ in the set $\{\ \}_m$ and $\mathbf{x}''$ in the set $\{\ \}_{n-m}$. Assume that

$$K_m(\{k_z, \mathbf{x}_\perp\}_n) \xrightarrow[\{|\mathbf{x}'_\perp - \mathbf{x}''_\perp|\} \to \infty]{} 0, \quad 1 < m < n.$$

From this and Eqs. (B1) and (B2) it follows that

$$K_n(\{k_z, \mathbf{x}_\perp\}_n) \xrightarrow[\{|\mathbf{x}'_\perp - \mathbf{x}''_\perp|\} \to \infty]{} F_m(\{k'_z, \mathbf{x}'_\perp\}_m)F_{n-m}(\{k''_z, \mathbf{x}''_\perp\}_{n-m}) - \left(\sum_{j'\text{perm}} \sum_{M'=1}^{m} \sum_{\{m_i\}_{M'}} \prod_{i'=1}^{M'} K_{m_{i'}}(\{k_{j'z}, \mathbf{x}_{j'\perp}\}_{m_{i'}})\right) \times \left(\sum_{j''\text{perm}} \sum_{M''=1}^{n-m} \sum_{\{m_i\}_{M''}} \prod_{i''=1}^{M''} K_{m_{i''}}(\{k_{j''z}, \mathbf{x}_{j''\perp}\}_{m_{i''}})\right),$$

$(n > 1)$

and thus from (B1)

$$K_n(\{k_z, \mathbf{x}_\perp\}_n) \xrightarrow[\{|\mathbf{x}'_\perp - \mathbf{x}''_\perp|\} \to \infty]{} 0. \quad n > 1.$$

This concludes the proof.

The first three irreducible characteristic functions are

$$K_1(1) = F_1(1),$$
$$K_2(1,2) = F_2(1,2) - K_1(1)K_1(2),$$
$$K_3(1,2,3) = F_3(1,2,3) - K_2(1,2)K_1(3) - K_2(1,3)K_1(2) - K_2(2,3)K_1(1) - K_1(1)K_1(2)K_1(3),$$

where we have used the condensed notation

$$K_1(1) = K_1(k_{1z},\mathbf{x}_{1\perp}),$$
$$K_2(1,2) = K_2(k_{1z},\mathbf{x}_{1\perp};k_{2z},\mathbf{x}_{2\perp}),$$
etc.

Since the K_n functions are well behaved as the points are separated, we can now complete the program of separating the singularity structures of $\langle \Pi_j A(\mathbf{k}_j)\rangle$. We will use a new set of irreducible functions A_n, which are the Fourier transforms with respect to $\mathbf{x}_\perp$ of the irreducible characteristic functions, i.e.,

$$A_n(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \int d^2x_{1\perp}\cdots d^2x_{n\perp} \exp\left(-i\sum_{j=1}^{n}\mathbf{k}_{j\perp}\cdot\mathbf{x}_{j\perp}\right) \times K_n(k_{1z},\mathbf{x}_{1\perp};\cdots;k_{nz},\mathbf{x}_{n\perp}). \quad \text{(B4)}$$

From Eqs. (23), (24), (27), and (B4) we get

$$\langle\prod_{j=1}^{n} A(\mathbf{k}_j)\rangle = \sum_{j\,\mathrm{perm}}\sum_{M=1}^{n}\sum_{\{m_i\}_M}\prod_{i=1}^{M} A_{m_i}(\{\mathbf{k}_j\}_{m_i}). \quad \text{(B5)}$$

This is the desired cluster expansion. Writing out (B5) for $n = 1,2,3$ gives

$$\langle A(1)\rangle = A_1(1),$$
$$\langle A(1)A(2)\rangle = A_1(1)A_1(2) + A_2(1,2),$$
$$\langle A(1)A(2)A(3)\rangle = A_1(1)A_1(2)A_1(3) + A_2(1,2)A_1(3) + A_2(1,3)A_1(2) + A_2(2,3)A_1(1) + A_3(1,2,3),$$

where the condensed notation

$$A(\mathbf{k}_1) = A(1), \text{etc.}$$

has been used.

We will now explicitly display the δ-function singularity in A_n. Translational invariance in the $\mathbf{x}_\perp$ plane of the ensemble of surfaces implies

$$K_n(\{k_z,\mathbf{x}_\perp\}_n) = K_n(\{k_z,\mathbf{x}_\perp + \mathbf{Y}_\perp\}_n),$$

where $\mathbf{Y}_\perp$ is arbitrary. This invariance can be used to transform to a new set of coordinates $\{\mathbf{y}_\perp\}_n$ where one integration in (30) can be performed. Let

$$\mathbf{y}_{i\perp} = \mathbf{x}_{i\perp} - \mathbf{x}_{n\perp}, \quad i < n,$$
$$\mathbf{y}_{n\perp} = \mathbf{x}_{n\perp}.$$

The Jacobian of this transformation is unity. Using this transformation and taking $\mathbf{Y}_\perp = \mathbf{y}_{n\perp}$ gives

$$A_n(\{\mathbf{k}\}_n) = \int d^2y_{1\perp}\cdots d^2y_{n\perp} \exp\left[-i\sum_{j=1}^{n-1}\mathbf{k}_{j\perp}\cdot\mathbf{y}_{j\perp} - i\left(\sum_{j=1}^{n}\mathbf{k}_{j\perp}\right)\cdot\mathbf{y}_{n\perp}\right] K_n(k_{1z},\mathbf{x}_{1\perp};\cdots;k_{nz},0).$$

The $\mathbf{y}_{n\perp}$ integration can be performed giving

$$A_n(\{\mathbf{k}\}_n) = (2\pi)^2\delta_2\left(\sum_{j=1}^{n}\mathbf{k}_{j\perp}\right)\int d^2y_{1\perp}\cdots d^2y_{n\perp}\delta_2(\mathbf{y}_{n\perp}) \times \exp\left(-i\sum_{j=1}^{n}\mathbf{k}_{j\perp}\cdot\mathbf{y}_{j\perp}\right) K_n(\{k_z,\mathbf{y}_\perp\}_n).$$

From the characteristic function given in Eq. (25), we see that all of the $A_n(\{\mathbf{k}\}_n)$ have a factor of $\exp(-\frac{1}{2}\Gamma(0) \times \sum_{j=1}^{n} k_{jz}^2)$ that can be extracted. With this in mind we will define the functions R_n by

$$A_n(\{\mathbf{k}\}_n) = (2\pi)^2\delta_2\left(\sum_{j=1}^{n}\mathbf{k}_{j\perp}\right)\exp\left(-\frac{1}{2}\Gamma(0)\sum_{j=1}^{n}k_{jz}^2\right)R_n(\{\mathbf{k}\}_n).$$

It is now easy to write out the first three expressions for R_n explicitly

$$R_1(\mathbf{k}) = 1,$$
$$R_2(\mathbf{k}_1,\mathbf{k}_2) = \int d^2y_\perp e^{-i\mathbf{k}_{2\perp}\cdot\mathbf{y}_\perp}\{\exp[-\Gamma(\mathbf{y}_\perp)k_{1z}k_{2z}] - 1\},$$
$$R_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \int d^2y_{1\perp}d^2y_{2\perp}e^{-i(\mathbf{k}_{1\perp}\cdot\mathbf{y}_{1\perp}+\mathbf{k}_{2\perp}\cdot\mathbf{y}_{2\perp})} \times\{\exp[-\Gamma(\mathbf{y}_{1\perp}-\mathbf{y}_{2\perp})k_{1z}k_{2z} - \Gamma(\mathbf{y}_{1\perp})k_{1z}k_{3z} - \Gamma(\mathbf{y}_{2\perp})k_{2z}k_{3z}] - \exp[-\Gamma(\mathbf{y}_{1\perp}-\mathbf{y}_{2\perp})k_{1z}k_{2z}] - \exp[-\Gamma(\mathbf{y}_{1\perp})k_{1z}k_{3z}] - \exp[-\Gamma(\mathbf{y}_{2\perp})k_{2z}k_{3z}] + 2\}.$$

It is interesting to note that if we let $\Gamma(\mathbf{x}_\perp) = \Gamma(0)C(\mathbf{x}_\perp)$, where $C(0) = 1$, and expand R_n in a Taylor series in $\Gamma(0)$, then the lowest power of $\Gamma(0)$ appearing in the series is $\Gamma(0)^{n-1}$.

[1] P. Beckman and A. Spizzichino, *The scattering of electromagnetic waves from rough surfaces* (Pergamon, New York, 1963), and references therein.
[2] C. Eckart, J. Acoust. Soc. Am. 25, 566 (1953).
[3] U. Frisch, "Wave propagation in random media," in *Probabilistic methods in applied mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), and references therein.
[4] O. D. Kellogg, *Foundations of potential theory* (Dover, New York, 1953).
[5] P. M. Morse and H. Feshbach, *Methods of theoretical physics* (McGraw-Hill, New York, 1953).
[6] M. J. Beran, *Statistical continuum theories* (Interscience, New York, 1968).
[7] K. Huang, *Statistical mechanics* (Wiley, New York, 1963), pp. 305-06.

A kinetic theory for power transfer in stochastic systems

George C. Papanicolaou*
Courant Institute and Department of Mathematics, University Heights, New York University, New York, New York
(Received 12 June 1972; revised manuscript received 11 August 1972)

In the asymptotic limit of weak inhomogeneities and long times or distances, we obtain a system of kinetic equations governing the power transfer among the modes of oscillation of certain stochastic dynamical systems. We include applications to coupled oscillators, waveguides, beams, the quantized motion of a particle in a random potential, and the Klein-Gordon equation with random plasma frequency.

1. INTRODUCTION

The author and J. B. Keller employed previously[1] a method for computing moments of the solution of stochastic equations in a certain asymptotic limit. Subsequently, this method was justified for a large class of problems.[2,3]

We present here a kinetic theory for coupled systems of stochastic equations by employing the above method.[1] Specifically, we consider a general oscillatory system and find equations for the mean of the modulus square of the amplitudes. This problem is well known in statistical mechanics and the resulting kinetic equations are called master equations.[4-6] Our results, however, follow rigorously from previous results[1-3] and are obtained in a natural manner, without recourse to elaborate perturbation schemes, in the asymptotic limit of long times (or distances) and weak fluctuations.

In Sec. 2 we formulate the problem under consideration for a general class of coupled stochastic equations. In Sec. 3 we apply the above method[1-3] to obtain the kinetic equations.

In Sec. 4 we apply the results of Sec. 3 to wave propagation in a waveguide with random inhomogeneities and to Gaussian beams through random media. The same equations [(4.10), (4.12), and (4.13)] have obvious significance for quantum mechanical problems. They are master equations for the probability amplitudes first obtained by Pauli.[7] In connection with the waveguide problem, kinetic equations were first obtained in a heuristic manner by Marcuse.[8,9] The same problem has been treated by Young and Rowe[10] and Morrison and McKenna.[11] The latter employed the above method[1] as we do here.

In Sec. 5 we apply the results of Sec. 3 to a system of coupled harmonic oscillators. The single random harmonic oscillator has been treated by several authors beginning with Stratonovich.[12] Our results here generalize the results of Stratonovich and the author and Keller.[1]

In Sec. 6 we consider the Schrödinger equation for a particle in a random potential. We obtain kinetic or transport equations for the average probability density of the particle in momentum space in the usual asymptotic limit.

In Sec. 7 we consider the Klein–Gordon equation with random plasma frequency. We obtain a transport equation for the average field energy density in wavenumber space. The results of Secs. 6 and 7 are the continuous or infinite dimensional analogs of the results of Secs. 4 and 5, respectively. Naturally the application of the theory[2,3] is somewhat more involved here. We shall concentrate, however, on obtaining and interpreting the transport equations. In Secs. 6 and 7 we also omit details of some derivations since they are simple generalizations of those of previous sections.

2. FORMULATION OF THE PROBLEM

Let $v(t)$ be a complex n-vector function of t satisfying the system of equations

$$\frac{dv(t)}{dt} = (ik + \epsilon x(t))v(t), \qquad t \geq 0, \; v(0) = u_0, \; i = \sqrt{-1}, \tag{2.1}$$
$$k = \operatorname{diag}(k_1, k_2, \ldots, k_n), \qquad k_j \text{ real and distinct.}$$

Here $x(t) = (x_{pq}(t))$ denotes a real or complex matrix valued stochastic process and ϵ is a small parameter. We assume that the processes $x_{pq}(t)$ have mean zero

$$E\{x_{pq}(t)\} = 0, \qquad p, q = 1, \ldots, n, \tag{2.2}$$

and that they are wide sense stationary

$$E\{x_{pq}(t+s)x_{p'q'}(s)\} = R_{pq,p'q'}(t). \tag{2.3}$$

Equations (2.1) define the process $v(t)$. We wish to study the statistical properties of this process; in particular, we are interested in the quantities

$$E\{v_p(t)v_p^*(t)\}, \qquad p = 1, \ldots, n. \tag{2.4}$$

Here * stands for complex conjugate. In many problems, as we will see in Secs. 4 and 5, the quantities (2.4) represent power and are thus called power amplitudes. The main result of Sec. 3 is that in the limit $t \to \infty$, $\epsilon \to 0$, $\epsilon^2 t = \text{const}$, the quantities (2.4) satisfy a system of coupled equations which we call the kinetic equations because of their form.

Let us now transform (2.1) into a more convenient form. First we introduce the slowly varying amplitudes $u(t)$ by

$$v(t) = e^{ikt}u(t). \tag{2.5}$$

From (2.5) and (2.1) it follows that

$$\frac{du}{dt} = \epsilon(e^{-ikt}x(t)e^{ikt})u \equiv \epsilon\tilde{x}(t)u, \qquad u(0) = u_0. \tag{2.6}$$

Clearly the power amplitudes are not affected by the transformation (2.5). Next we introduce the tensor product[1] of u and u^*:

$$y(t) = u(t) \otimes u^*(t). \tag{2.7}$$

By differentiating y and using (2.6), we obtain the equations

$$\frac{dy(t)}{dt} = \epsilon(\tilde{x}(t) \otimes I + I \otimes \tilde{x}^*(t))y \equiv \epsilon V(t)y,$$
$$y(0) = u_0 \otimes u_0^*, \qquad I = \text{identity matrix.} \tag{2.8}$$

By using indices and the summation convention, (2.8) takes the form

$$\frac{dy_{pp'}}{dt} = \epsilon\{\tilde{x}_{pq}\delta_{p'q'} + \delta_{pq}\tilde{x}^*_{p'q'}\}y_{qq'}, \quad y(0) = u_{0p}u^*_{0p'},$$
$$y_{pp'} = u_p u^*_{p'}. \qquad (2.9)$$

In (2.9) δ_{pq} denotes the Kronecker delta function.

Equations (2.8) or (2.9) are set up appropriately now for the application of the method referred to in the Introduction.[1]

3. THE KINETIC EQUATIONS

Let τ and $y^{(\epsilon)}(\tau)$ be defined by

$$\tau = \epsilon^2 t, \quad y^{(\epsilon)}(\tau) = y(\tau/\epsilon^2), \quad y \text{ as in (2.7)}. \qquad (3.1)$$

According to previous results[1-3]

$$w(\tau) = \lim_{\epsilon\to 0} E\{y^{(\epsilon)}(\tau)\}, \quad 0 \le \tau \le \tau_0, \qquad (3.2)$$

exists and the vector $w(\tau)$ satisfies the system of equations

$$\frac{dw(\tau)}{d\tau} = \overline{V}w(\tau), \quad w(0) = u_0 \otimes u_0^*,$$
$$\overline{V} = \lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{V(s)V(\sigma)\}d\sigma ds. \qquad (3.3)$$

The precise hypotheses in the theory[2,3] can be satisfied easily here since $V(t)$ of (2.8) is finite dimensional. Of course, the process $x(t)$ must also satisfy certain hypotheses. For our purposes here it suffices to assume that

$$\int_0^\infty \sigma R_{pq,p'q'}(\sigma)d\sigma < \infty, \quad p,q,p',q' = 1,\ldots,n. \qquad (3.4)$$

We now proceed with the applications of this result.

From (2.8) we see that

$$V(s)V(\sigma) = (\tilde{x}(s)\otimes I + I\otimes\tilde{x}^*(s))(\tilde{x}(\sigma)\otimes I + I\otimes\tilde{x}^*(\sigma))$$
$$= \tilde{x}(s)\tilde{x}(\sigma)\otimes I + I\otimes\tilde{x}^*(s)\tilde{x}^*(\sigma) + \tilde{x}(s)\otimes\tilde{x}^*(\sigma)$$
$$+ \tilde{x}(\sigma)\otimes\tilde{x}^*(s). \qquad (3.5)$$

To find the form of $\overline{V}$, we insert (3.5) into (3.3) and compute the limit for each term in (3.5) separately. We have

$$\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{\tilde{x}(s)\tilde{x}(\sigma)\otimes I\}ds\right)_{pq,p'q'}$$
$$= \delta_{p'q'}\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{e^{i(k_r-k_p)s}x_{pr}(s) \times e^{i(k_q-k_r)\sigma}x_{rq}(\sigma)\}d\sigma ds$$
$$= \delta_{p'q'}\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s e^{i(k_r-k_p)s+i(k_q-k_r)(s-\sigma)} \times R_{pr,rq}(\sigma)d\sigma ds$$
$$= \delta_{p'q'}\delta_{pq}\int_0^\infty e^{-i(k_q-k_r)\sigma}R_{pr,rq}(\sigma)d\sigma. \qquad (3.6)$$

In (3.6) we have used the summation convention, (3.4), and the fact that $k_1, k_2, \ldots, k_n$ are distinct real numbers. A calculation identical to the above yields

$$\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{I\otimes\tilde{x}^*(s)\tilde{x}^*(\sigma)\}d\sigma ds\right)_{pq,p'q'}$$
$$= \delta_{pq}\delta_{p'q'}\int_0^\infty e^{i(k_{q'}-k_{r'})\sigma}R^{*,*}_{p'r',r'q'}(\sigma)d\sigma. \qquad (3.7)$$

Here we have introduced the notation

$$R^{*,*}_{pq,p'q'}(\sigma) = E\{x^*_{pq}(\sigma+t)x^*_{p'q'}(t)\}. \qquad (3.8)$$

We shall also need the notation

$$R^{*,}_{pq,p'q'}(\sigma) = E\{x^*_{pq}(\sigma+t)x_{p'q'}(t)\},$$
$$R^{,*}_{pq,p'q'}(\sigma) = E\{x_{pq}(\sigma+t)x^*_{p'q'}(t)\}. \qquad (3.9)$$

Before proceeding with the remaining computations for $\overline{V}$ we observe the following. Our principal interest lies not in the full tensor

$$w(\tau) = (w_{pp'}(\tau)), \qquad (3.10)$$

but rather in the "diagonal" part,

$$W_p(\tau) = w_{pp}(\tau). \qquad (3.11)$$

In general, the W_p, which are the limits of (2.4), do not satisfy a closed system of equations. For the particular system under consideration, however, they do satisfy a closed system. We now show this remarkable fact.

Let us continue with the computation of the limit $\overline{V}$. We have

$$\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{\tilde{x}(s)\otimes\tilde{x}^*(\sigma)\}d\sigma ds\right)_{pq,p'q'}$$
$$= \lim\frac{1}{t}\int_0^t\int_0^s E\{e^{i(k_q-k_p)s}x_{pq}(s) \times e^{-i(k_{q'}-k_{p'})\sigma}x^*_{p'q'}(\sigma)\}d\sigma ds$$
$$= \lim\frac{1}{t}\int_0^t\int_0^s e^{i(k_q-k_p)s}e^{-i(k_{q'}-k_{p'})(s-\sigma)} \times R^{,*}_{pq,p'q'}(\sigma)d\sigma ds. \qquad (3.12)$$

From the above remarks it follows that if the condition $p = p'$ implies that (3.12) is zero unless $q = q'$, then, in view of (3.6), (3.7), the claim of the preceding paragraph follows. Now it is easy to see that in fact when $p = p'$,

$$\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{\tilde{x}(s)\otimes\tilde{x}^*(\sigma)\}d\sigma ds\right)_{pq,pq'}$$
$$= \delta_{qq'}\int_0^\infty e^{i(k_{q'}-k_p)\sigma}R^{,*}_{pq,pq'}(\sigma)d\sigma. \qquad (3.13)$$

An identical calculation yields, when $p = p'$,

$$\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^s E\{\tilde{x}(\sigma)\otimes\tilde{x}^*(s)\}d\sigma ds\right)_{pq,pq'}$$
$$= \delta_{qq'}\int_0^\infty e^{-i(k_q-k_p)\sigma}R^{*,}_{pq',pq}(\sigma)d\sigma. \qquad (3.14)$$

On collecting our formulas above, we obtain the main result of this section, the kinetic equations for the asymptotic limit of the mean power amplitudes $W_p(\tau)$:

$$\frac{dW_p}{d\tau} = \sum_{q=1}^n Q_{pq}W_q, \quad W_p(0) = |u_{0p}|^2, \qquad (3.15)$$

$$Q_{pq} = \delta_{pq}\left(\int_0^\infty e^{-i(k_q-k_r)\sigma}R_{pr,rq}(\sigma)d\sigma + \int_0^\infty e^{i(k_q-k_r)\sigma}R^{*,*}_{pr,rq}(\sigma)d\sigma\right)$$
$$+ \int_0^\infty e^{i(k_q-k_p)\sigma}R^{,*}_{pq,pq}(\sigma)d\sigma$$
$$+ \int_0^\infty e^{-i(k_q-k_p)\sigma}R^{*,}_{pq,pq}(\sigma)d\sigma \qquad (3.16)$$

We may also obtain kinetic equations for the off diagonal terms of the tensor w by completing the calculation of $\overline{V}$. However, since they decouple from the diagonal, they will not concern us here.

4. APPLICATION TO WAVEGUIDES AND BEAMS

Our first application concerns mode coupling in a waveguide with random irregularities. The problem is as follows.

Let $u(x,y,z)$ denote the time harmonic complex field satisfying the boundary value problem

$$\begin{aligned} &u_{xx} + u_{yy} + u_{zz} + k^2 n^2(x,y,z)u = 0,\\ &\quad -\infty < x < \infty, \quad y,z \in \mathfrak{D} \subset R^2,\\ &\quad u(x,y,z) = 0 \quad \text{for } (y,z) \in \partial\mathfrak{D}. \end{aligned} \tag{4.1}$$

Here k is the free space wavenumber, $n(x,y,z)$ is the index of refraction, and $\partial\mathfrak{D}$ denotes the boundary of the region $\mathfrak{D}$. We shall assume that $n^2(x,y,z)$ is a stationary random field that deviates little from its mean value. Then we may expand the field in a series using the eigenfunctions of the cross section $\mathfrak{D}$:

$$\begin{aligned} &(\partial_{yy} + \partial_{zz})h_n = -\lambda_n^2 h_n, \quad n = 1,2,\ldots,\\ &\quad h_n(y,z) = 0, \quad (y,z) \in \partial\mathfrak{D},\\ &\quad (h_n, h_m) = \int h_n(y,z)h_m(y,z)\,dydz = \delta_{nm}. \end{aligned} \tag{4.2}$$

The ensuing system of ordinary differential equations for the amplitudes of the modes, i.e., the coefficients in the expansion as functions of x, cannot be treated by the method of Sec. 3 because problem (4.1) is not an initial value problem. The full analysis of this system of equations requires special considerations and is treated elsewhere.[13]

To study (4.1) by the methods of Sec. 3, we resort to the forward scattering or parabolic approximation. We omit details on the validity of this approximation and describe it as follows. We write u in the form

$$u(x,y,z) = e^{ikx}v(x,y,z), \tag{4.3}$$

insert (4.3) in (4.1), and neglect $\partial_{xx}v$ to obtain

$$(\partial_{yy} + \partial_{zz})v + 2ik\partial_x v + k^2(n^2 - 1)v = 0. \tag{4.4}$$

On scaling the transverse variables, y, z by the factor $\sqrt{k}$ and assuming that

$$\begin{aligned} &n^2 - 1 = (\epsilon/k)\mu(x,y,z),\\ &\qquad \mu \text{ a stationary, zero mean random field,} \end{aligned} \tag{4.5}$$

we arrive formally at an initial value problem for v:

$$\begin{aligned} &2i\partial_x v + (\partial_{yy} + \partial_{zz})v + \epsilon\mu v = 0, \quad x > 0,\\ &\quad (y,z) \in \mathfrak{D}, \quad v(x,y,z) = 0, \quad (y,z) \in \partial\mathfrak{D},\\ &\quad v(0,y,z) = v_0(y,z) \quad \text{given.} \end{aligned} \tag{4.6}$$

Now we expand the solution of (4.6) using the eigenfunctions (4.2). To conform with the notation of Sec. 3, we use t instead of x and set

$$v(t,y,z) = \sum_{n=1}^{\infty} v_n(t)h_n(y,z). \tag{4.7}$$

Upon inserting (4.7) into (4.6) and using (4.2), we obtain the following system of equations for $v_p(t)$:

$$\begin{aligned} &\frac{dv_p}{dt} = ik_p + \epsilon \sum_{q=1}^{\infty} i\mu_{pq}(t)v_q, \quad t \ge 0,\\ &\qquad v_p(0), \quad \text{given,} \quad p = 1,2,\cdots. \end{aligned} \tag{4.8}$$

Here we have the notation

$$k_p = -\tfrac{1}{2}\lambda_p^2, \quad \mu_{pq}(t) = \tfrac{1}{2}\int_{\mathfrak{D}} \mu(t,y,z)h_q(y,z)h_p(y,z)\,dydz. \tag{4.9}$$

Note that the matrix $\mu(t) = (\mu_{pq}(t))$ is real and symmetric. Since $\mu(t,y,z)$ is a random field, the matrix $\mu(t)$ is a process whose statistics follow from those of $\mu(t,y,z)$. It is convenient to also make the following approximation. We truncate the system (4.8) to the first n equations and write

$$\begin{aligned} &v = \begin{pmatrix} v_1\\ \vdots \\ v_n \end{pmatrix}, \quad k = \operatorname{diag}(k_1,\ldots,k_n), \quad x(t) = i(\mu_{pq}(t)),\\ &\qquad p,q = 1,2,\ldots,n,\\ &\frac{dv}{dt} = (ik + \epsilon x(t))v, \quad v(0) = u_0 \quad \text{given.} \end{aligned} \tag{4.10}$$

We have thus a problem of the form (2.1). Incidentally, the truncation above is not necessary (the theory[2,3] can handle infinite-dimensional systems). The results, however, are in better accord with (4.1), where for given k only finitely many modes propagate.

The kinetic equations (3.15) for the mean power amplitudes,

$$W_p(\tau) = \lim_{\epsilon\to 0} E\{|v_p(\tau/\epsilon^2)|^2\}, \tag{4.11}$$

become here

$$\frac{dW_p}{d\tau} = \sum_{q=1}^{n} Q_{pq}W_q, \tag{4.12}$$

$$\begin{aligned} &Q_{pq} = 2\int_0^{\infty} [\cos(k_p - k_q)\sigma]\rho_{pq,pq}(\sigma)\,d\sigma, \quad p \ne q,\\ &Q_{pp} = -\sum_{q\ne p} Q_{pq}, \quad \rho_{pq,pq}(\sigma) = E\{\mu_{pq}(\sigma + s)\mu_{pq}(s)\}. \end{aligned} \tag{4.13}$$

In (4.13) we have used the form of the matrix $x(t)$ as defined in (4.10). From (4.12), (4.13) it follows that

$$Q_{pq} = Q_{qp}, \tag{4.14}$$

and that

$$\sum_{p=1}^{n} W_p(\tau) = \text{const.} \tag{4.15}$$

The conservation equation (4.15) is in accord with

$$\sum_{p=1}^{n} |v_p(t)|^2 = \text{const}, \tag{4.16}$$

which follows from (4.10). Note also that $Q_{pq} \ge 0$, $p \ne q$, since Q_{pq} is a cosine transform of a correlation function.

Equations (4.12) are the kinetic equations for the mean power amplitudes in the asymptotic limit $\epsilon \to 0$, $t \to \infty$, $\epsilon^2 t = \tau$. They are the main result of this section.

If we take the constant in (4.15) equal to 1, then we may interpret the $W_p(\tau)$ as probabilities and (4.12) as a Kolmogorov equation[14] for a continuous time Markov chain. The Markov chain can be thought of as governing the mechanism of power transfer among the modes. The quantity

$$\alpha = \max_p\{-Q_{pp}^{-1}\}, \tag{4.17}$$

is the largest expected distance (t, τ are length parameters) between transitions of the chain. Thus when $\alpha \ll \tau$, then, $W_p(\tau)$ is approximated well by the equilibrium solution of (4.13),

$$W_p = 1/n, \quad p = 1, 2, \ldots, n. \tag{4.18}$$

We have arrived at an equipartition law for the mean power amplitudes under the approximations and assumptions made above. This result has been observed in numerical simulations of a simple model by Marcuse.[9]

The application of the above considerations to Gaussian beams[15] is immediate. We merely have to change the eigenfunctions (4.2). The transverse variables y, z, now vary over R^2, but the index of refraction is given by

$$n^2(t,y,z) = 1 - (y^2 + z^2)[1 + (\epsilon/k)\mu(t,y,z)]. \tag{4.19}$$

The eigenfunctions are now Hermite functions in two variables and the spectrum remains discrete. The orthonormality relation in (4.2) still holds if we define the inner product of eigenfunctions appropriately. For more details on the physical problem we refer to Arnaud[15] and the recent work of the author, McLaughlin, and Burridge.[16]

5. APPLICATION TO RANDOMLY COUPLED OSCILLATORS

In this section we shall apply the results of Section 3 to the following problem. Let $z(t)$ be a real n-vector function satisfying the equations

$$\frac{d^2z(t)}{dt^2} + (k^2 + \epsilon\tilde{\mu}(t))z(t) = 0,$$
$$z(0) = z_0, \quad \frac{dz(0)}{dt} = \dot{z}_0, \quad k^2 = \operatorname{diag}(k_1^2, \ldots, k_n^2). \tag{5.1}$$

In (5.1) we assume that $k_1, k_2, \ldots, k_n$ are real positive and distinct and $\tilde{\mu}(t)$ is a real symmetric $n \times n$ matrix valued process, stationary and with zero mean. We introduce new dependent variables $A(t)$ and $B(t)$ by the relations

$$z(t) = k^{-1/2}(e^{ikt}A(t) + e^{-ikt}B(t)),$$
$$\frac{dz(t)}{dt} = ik^{1/2}(e^{ikt}A(t) - e^{-ikt}B(t)). \tag{5.2}$$

Upon using these relations and (5.1), we obtain the following system of equations for the n-vector valued processes $A(t)$ and $B(t)$:

$$\frac{d}{dt}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} = \frac{i\epsilon}{2}\begin{pmatrix} e^{-ikt} & 0 \\ 0 & e^{ikt} \end{pmatrix}\begin{pmatrix} \mu(t) & \mu(t) \\ -\mu(t) & -\mu(t) \end{pmatrix}$$
$$\times \begin{pmatrix} e^{ikt} & 0 \\ 0 & e^{-ikt} \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix},$$
$$\mu(t) = k^{-1/2}\tilde{\mu}(t)k^{-1/2}. \tag{5.3}$$

Note that $\mu(t)$ is a real symmetric $n \times n$ matrix valued process. Also, from (5.3) we find that

$$\sum_{p=1}^{n} (|A_p(t)|^2 - |B_p(t)|^2) = \text{const.} \tag{5.4}$$

This conservation law is markedly different from (4.16). It can be used, however, to obtain a detailed analysis of the one-dimensional waveguide problem, *without* the parabolic approximation, where the equations (5.1) arise. This has been carried out elsewhere.[13] When the problem (5.1) concerns a mechanical or other system of oscillators then the quantities

$$|A_p(t)|^2 + |B_p(t)|^2, \quad p = 1, \ldots, n, \tag{5.5}$$

are of interest, because they represent the instantaneous energy of each oscillator. However, the sum of the energies (5.5) is not conserved and this will appear explicitly in the kinetic equation we now derive.

In order to use the results of Sec. 3 we introduce the following notation. Let J be the $n \times n$ matrix,

$$J = \begin{bmatrix} 0 \cdots 0 \; 1 \\ 0 \cdots 1 \; 0 \\ \cdots \\ 1 \; 0 \cdots 0 \end{bmatrix}, \tag{5.6}$$

and let

$$u(t) = \begin{pmatrix} JA \\ B \end{pmatrix} = \begin{bmatrix} u_{-n} \\ u_{-n+1} \\ \vdots \\ u_{-1} \\ u_{+1} \\ \vdots \\ u_n \end{bmatrix}. \tag{5.7}$$

Then we may rewrite (5.3) in the form

$$\frac{du(t)}{dt} = \epsilon e^{-iKt}x(t)e^{iKt}u(t), \tag{5.8}$$

$$e^{iKt} = \begin{pmatrix} Je^{ikt}J & 0 \\ 0 & e^{-ikt} \end{pmatrix}, \tag{5.9}$$

$$x(t) = \frac{i}{2}\begin{pmatrix} J\mu J & J\mu \\ -\mu J & -\mu \end{pmatrix}. \tag{5.10}$$

Note that the index in the vectors and matrices runs now as follows: $-n, -(n-1), \ldots, -1, 1, \ldots, n$. Equations (5.8) are indeed in the form (2.6), and so we can apply the results (3.15), (3.16) directly. Thus we obtain

$$\frac{dW_{-p}}{d\tau} = \sum_{q=1}^{n} (Q_{-p,-q}W_{-q} + Q_{-p,q}W_q), \quad p = 1, \ldots, n,$$

$$\frac{dW_p}{d\tau} = \sum_{q=1}^{n} (Q_{p,-q}W_{-q} + Q_{p,q}W_q), \quad p = 1, \ldots, n, \tag{5.11}$$

$$Q_{p,q} = \delta_{pq} \sum_{r=1}^{n} \tfrac{1}{2} \int_0^\infty [\cos(k_q + k_r)\sigma - \cos(k_q - k_r)\sigma]\rho_{pr,rp}(\sigma)d\sigma + \tfrac{1}{2}\int_0^\infty [\cos(k_q - k_p)\sigma]\rho_{pq,pq}(\sigma)d\sigma, \quad p,q > 0, \tag{5.12}$$

$$Q_{-p,q} = \tfrac{1}{2}\int_0^\infty [\cos(k_p + k_q)\sigma]\rho_{pq,qp}(\sigma)d\sigma, \quad p,q > 0, \tag{5.13}$$

$$Q_{p,-q} = Q_{-p,q}, \quad Q_{p,q} = Q_{-p,-q}. \tag{5.14}$$

Here the covariances $\rho_{pq,p'q'}(\sigma)$ are defined by (4.13). The results (5.11)–(5.14) follow by elementary considerations from (3.16) and the special form (5.10) of $x(t)$.

From (5.11)–(5.14) it follows immediately that

$$\frac{d}{d\tau}\sum_{p=1}^{n}(W_p(\tau) - W_{-p}(\tau)) = 0. \tag{5.15}$$

This, of course, was expected in view of (5.4) and the definition of $W_p(\tau)$:

$$W_p(\tau) = \lim_{\epsilon\to 0} E\{|B_p(\tau/\epsilon^2)|^2\},$$

$$W_{-p}(\tau) = \lim_{\epsilon\to 0} E\{|A_p(\tau/\epsilon^2)|^2\}, \quad p = 1,2,\ldots,n. \tag{5.16}$$

It also follows that the energies $W_p + W_{-p}$, $p = 1,\ldots,n$, satisfy the following system of kinetic equations:

$$\frac{d}{d\tau}(W_p + W_{-p}) = \sum_{q=1}^{n}\tilde{Q}_{pq}(W_q + W_{-q}) + \gamma_p(W_p + W_{-p}), \quad p = 1,2,\ldots,n, \tag{5.17}$$

$$\tilde{Q}_{pq} = \tfrac{1}{2}\int_0^\infty [\cos(k_p + k_q)\sigma + \cos(k_p - k_q)\sigma]\rho_{pq,qp}(\sigma)d\sigma, \quad p \neq q, \tag{5.18}$$

$$\tilde{Q}_{pp} = -\sum_{q\neq p}\tilde{Q}_{pq}, \tag{5.19}$$

$$\gamma_p = \sum_{q=1}^{n}\int_0^\infty \cos(k_p + k_q)\sigma\rho_{pq,pq}(\sigma)d\sigma \geq 0. \tag{5.20}$$

These kinetic equations constitute the main result of this section.

The interpretation of equations (5.17) is different from that of (4.12) because (5.17) are not conservative, that is, $\sum_{p=1}^{n}(W_p + W_{-p})$ is not a constant. Upon summing on p in (5.17) we obtain

$$\frac{d}{d\tau}\left(\sum_{p=1}^{n}(W_p + W_{-p})\right) = \sum_{p=1}^{n}\gamma_p(W_p + W_{-p}). \tag{5.21}$$

Note that the sum on the right side of (5.17) does not appear in (5.21). Let

$$\gamma = \max_{1\leq p\leq n}\{\gamma_p\}. \tag{5.22}$$

Since the $W_p + W_{-p}$, $p = 1,\ldots,n$, are nonnegative we have

$$\frac{d}{d\tau}\left(\sum_{p=1}^{n}(W_p + W_{-p})\right) \leq \gamma\sum_{p=1}^{n}(W_p + W_{-p}) \tag{5.23}$$

and hence

$$\sum_{p=1}^{n}(W_p + W_{-p}) \leq ce^{\gamma\tau}, \tag{5.24}$$

where c is a constant depending on initial conditions.

If we introduce dissipation in the system (5.1) by adding on the left side the term $2\epsilon^2\beta(dz_p/dt)$, $\beta > 0$, then, after a computation similar to that employed previously,[1] we obtain for $W_p + W_{-p}$ a system of equations identical with (5.17) except that γ_p is replaced by $\gamma_p - \beta$. Thus, in the case where dissipation is present, (5.24) leads to the estimate

$$\sum_{p=1}^{n}(W_p + W_{-p}) \leq ce^{(\gamma-\beta)\tau}. \tag{5.25}$$

From (5.25) we deduce a stability condition

$$\beta \geq \gamma, \tag{5.26}$$

which generalizes the result of Stratonovich[12] for the single random harmonic oscillator. By stability we mean here that the total average energy $\sum_{p=1}^{n}(W_p + W_{-p})$ is a bounded function of τ.

Of more interest is, however, the explicit way in which (5.17) expresses the mechanism of energy transfer among the oscillators. It is considerably more complicated than the corresponding result of Sec. 4.

6. SCHRÖDINGER EQUATION WITH RANDOM POTENTIAL

Let $\psi(t,\mathbf{x})$, $\mathbf{x} \in R^n$, denote the wavefunction satisfying the Schrödinger equation

$$i\psi_t = \tfrac{1}{2}[\Delta + \epsilon\mu(t,\mathbf{x})]\psi, \quad t \geq 0, \ \psi(0,\mathbf{x}) = \psi_0(\mathbf{x}), \quad i = \sqrt{-1}. \tag{6.1}$$

Here Δ denotes the Laplacian in R^n, and $\mu(t,\mathbf{x})$ is a stationary random process with mean zero. We denote the ensemble average by $E\{\ \}$. Thus

$$E\{\mu(t,\mathbf{x})\} = 0, \tag{6.2}$$

$$E\{\mu(t,\mathbf{x})\mu(t+s,\mathbf{x}+\boldsymbol{\xi})\} = \rho(s,\boldsymbol{\xi}). \tag{6.3}$$

We also assume that

$$|\mu(t,\mathbf{x})| \leq 1 \tag{6.4}$$

almost surely.

Let us transform (6.1) to momentum space. We define the Fourier transform of ψ by

$$\hat{\psi}(t,\mathbf{p}) = \frac{1}{(2\pi)^{n/2}}\int_{R^n} e^{i\mathbf{p}\cdot\mathbf{x}}\psi(t,\mathbf{x})d\mathbf{x}; \tag{6.5}$$

hence

$$\psi(t,\mathbf{x}) = \frac{1}{(2\pi)^{n/2}}\int_{R^n} e^{-i\mathbf{p}\cdot\mathbf{x}}\hat{\psi}(t,\mathbf{p})d\mathbf{p}. \tag{6.6}$$

On multiplying (6.1) by $e^{i\mathbf{p}\cdot\mathbf{x}}/(2\pi)^{n/2}$, integrating over R^n, and using (6.6), we find that

$$i\hat{\psi}_t = -\tfrac{1}{2}p^2\hat{\psi} + \tfrac{1}{2}\epsilon\int_{R^n}((2\pi)^{-n}\int_{R^n} e^{i(\mathbf{p}-\mathbf{p}')\cdot\mathbf{x}}\mu(t,\mathbf{x})d\mathbf{x}) \times \hat{\psi}(t,\mathbf{p}')d\mathbf{p}'. \tag{6.7}$$

The quantity in the braces in (6.7) represents a distribution-values stochastic process since for stationary processes the Fourier transform exists only in this generalized sense. p^2 denotes the squared modulus of the momentum $\mathbf{p}$. Define $v(t,\mathbf{p})$ by

$$v(t,\mathbf{p}) = e^{-ip^2t/2}\psi(t,\mathbf{p}). \tag{6.8}$$

Then from (6.7) we obtain

$$v_t(t,\mathbf{p}) = \epsilon \int_{R^n}\left(\frac{-ie^{-ip^2t/2}}{2(2\pi)^n}\int_{R^n} e^{i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\mu(t,\mathbf{x})d\mathbf{x}\, e^{itq^2/2}\right) \times v(t,\mathbf{q})d\mathbf{q}. \tag{6.9}$$

To simplify the notation, we shall denote the quantity in braces in (6.9) by $\tilde{\mu}(t,\mathbf{p},\mathbf{q})$ and omit R^n in the integrals throughout.

Let * stand for complex conjugate and define $y(t,\mathbf{p},\mathbf{p}')$ by

$$y(t,\mathbf{p},\mathbf{p}') = v(t,\mathbf{p})\,v^*(t,\mathbf{p}'). \tag{6.10}$$

Differentiating y with respect to t and using (6.9), we obtain the equation

$$\begin{aligned} y_t(t,\mathbf{p},\mathbf{p}') &= \epsilon \int \tilde{\mu}(t,\mathbf{p},\mathbf{q})y(t,\mathbf{q},\mathbf{p}')d\mathbf{q} \\ &+ \epsilon \int \tilde{\mu}^*(t,\mathbf{p}',\mathbf{q}')y(t,\mathbf{p},\mathbf{q}')d\mathbf{q}', \\ y(0,\mathbf{p},\mathbf{p}') &= \hat{\psi}_0(\mathbf{p})\hat{\psi}_0^*(\mathbf{p}') = y_0(\mathbf{p},\mathbf{p}'). \end{aligned} \tag{6.11}$$

We shall also write (6.11) in abstract form

$$y_t = \epsilon(\tilde{\mu}(t)\otimes I + I \otimes \tilde{\mu}^*(t))y \equiv \epsilon V(t)y, \quad y(0) = y_0, \quad y = v\otimes v^*. \tag{6.12}$$

We now proceed by applying to (6.12) the result stated at the beginning of sec. 3.

As before, the remarkable fact about this result, in connection with (6.12), is that it yields a closed equation for

$$W(\tau,\mathbf{p}) = w(\tau,\mathbf{p},\mathbf{p}) = \lim_{\epsilon\to 0} E\{|y(\tau/\epsilon^2,\mathbf{p},\mathbf{p})|^2\}, \tag{6.13}$$

the diagonal part of the tensor w. $W(\tau,\mathbf{p})$ is the expected value of the probability density in momentum space in the limit of small fluctuations and long times.

The details of the calculation of $\overline{V}$ are very similar to the previous ones and so we shall omit them. One very important difference here is the strong use we make of (6.3), the translation invariance in space of the covariance $\rho(s,\boldsymbol{\xi})$. In the discrete case this played no role, of course. We also assume that

$$\int_0^\infty \sigma\rho(\sigma,\boldsymbol{\xi})d\sigma < \infty, \quad \boldsymbol{\xi}\in R^n. \tag{6.14}$$

The result is that $W(\tau,\mathbf{p})$ satisfies the conservative transport equation

$$W_\tau(\tau,\mathbf{p}) = \int [\tilde{Q}(\mathbf{p},\mathbf{q})W(\tau,\mathbf{q}) - \tilde{Q}(\mathbf{p},\mathbf{q})W(\tau,\mathbf{p})]d\mathbf{q}, \quad \tau\ge 0,\ W(0,\mathbf{p}) = |\hat{\psi}_0(\mathbf{p})|^2, \tag{6.15}$$

$$\tilde{Q}(\mathbf{p},\mathbf{q}) = \tfrac{1}{2}\int_0^\infty\left[\frac{1}{(2\pi)^n}\int \cos\left(\frac{(p^2-q^2)\sigma}{2} + (\mathbf{q}-\mathbf{p})\cdot\boldsymbol{\xi}\right)\rho(\sigma,\boldsymbol{\xi})d\boldsymbol{\xi}\right]d\sigma \tag{6.16}$$

This is the main result of this section.

Let us make the following observations concerning (6.15) First the kernel $\tilde{Q}$ is nonnegative (because it is the space–time Fourier transform of a covariance); so (6.15) is indeed a transport equation, and

$$\int W(\tau,\mathbf{p})d\mathbf{p} \tag{6.17}$$

is manifestly independent of τ. If we normalize the integral (6.17) to equal 1, then $W(\tau,\mathbf{p})$ can be thought of as a probability density of a continuous time Markov process with values in R^n. These results are analogous to those of Sec. 4.

When $\rho(\sigma,\boldsymbol{\xi})$ is a product of a δ function in σ and a covariance function in $\boldsymbol{\xi}$, then $\tilde{Q}(\mathbf{p},\mathbf{q})$ is a function of $\mathbf{p}-\mathbf{q}$ only and (6.15) coincides with a result of Dolin[17] and Klyatskin and Tatarskii.[18]

Besieris and Tappert[19] have recently obtained further generalizations of (6.15), and they have also explored other aspects of the problem of this section.

7. THE KLEIN-GORDON EQUATION WITH RANDOM PLASMA FREQUENCY

We shall consider here the real-valued scalar field $u(t,\mathbf{x})$, $\mathbf{x}\in R^n$, satisfying the equation

$$u_{tt} - \Delta u + [m^2 + \epsilon\mu(t,\mathbf{x})]u = 0, \quad t\ge 0,\ u(0,\mathbf{x}) = u_0(\mathbf{x}), \quad u_t(0,\mathbf{x}) = \dot{u}_0(\mathbf{x}). \tag{7.1}$$

We adopt again (6.2)–(6.4). m^2 is the expected value of the plasma frequency, a positive constant. Equation (7.1) arises, of course, in contexts other than plasma physics but we shall adhere to this application for concreteness.

Let $\hat{u}(t,\mathbf{p})$ denote the Fourier transform (6.5) of $u(t,\mathbf{x})$. $\mathbf{p}\in R^n$ is the wavenumber vector. From (7.1) we obtain

$$\hat{u}_{tt} + l^2(p)\hat{u} + \epsilon\int\left(\frac{1}{(2\pi)^n}\int e^{i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\mu(t,\mathbf{x})d\mathbf{x}\right)\hat{u}(t,\mathbf{q})d\mathbf{q} = 0, \tag{7.2}$$

$$l(p) = +\sqrt{p^2+m^2}. \tag{7.3}$$

Next we introduce the normalized complex valued amplitudes $A(t,\mathbf{p}), B(t,\mathbf{p})$ by setting

$$\hat{u}(t,\mathbf{p}) = l^{-1/2}(p)[e^{il(p)t}A(t,\mathbf{p}) + e^{-il(p)t}B(t,\mathbf{p})], \tag{7.4}$$

$$\hat{u}_t(t,\mathbf{p}) = il^{1/2}(p)[e^{il(p)t}A(t,\mathbf{p}) - e^{-il(p)t}B(t,\mathbf{p})]. \tag{7.5}$$

Then, as in Sec. 5, we obtain the following system for A and B:

$$\begin{pmatrix} A_t(t,\mathbf{p}) \\ B_t(t,\mathbf{p}) \end{pmatrix} = \epsilon\int X(t,\mathbf{p},\mathbf{q})\begin{pmatrix} A(t,\mathbf{q}) \\ B(t,\mathbf{q}) \end{pmatrix}d\mathbf{q}. \tag{7.6}$$

Here $X(t,\mathbf{p},\mathbf{q})$ is the 2×2 matrix

$$\tilde{\mu}(t,\mathbf{p},\mathbf{q})\begin{pmatrix} e^{-i[l(p)-l(q)]t} & e^{-i[l(p)+l(q)]t} \\ -e^{i[l(p)+l(q)]t} & -e^{i[l(p)-l(q)]t} \end{pmatrix}, \tag{7.7}$$

$$\tilde{\mu}(t,\mathbf{p},\mathbf{q}) = l^{-1/2}(p)[i/2(2\pi)^n]\int e^{i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\mu(t,\mathbf{x})d\mathbf{x}\,l^{-1/2}(q). \tag{7.8}$$

The system (7.6) corresponds to (6.9). By taking tensor products we can obtain here also the analog of (6.11) or

(6.12). Then we may apply the result stated at the beginning of Sec. 3 and obtain closed equations for

$$W_A(\tau,\mathbf{p}) = \lim_{\epsilon\to 0} E\{|A(\tau/\epsilon^2,\mathbf{p})|^2\}, \tag{7.9}$$

$$W_B(\tau,\mathbf{p}) = \lim_{\epsilon\to 0} E\{|B(\tau/\epsilon^2,\mathbf{p})|^2\}, \tag{7.10}$$

If we assume that the covariance ρ satisfies

$$\rho(\sigma,\xi) = \rho(\sigma, -\xi), \tag{7.11}$$

then we obtain a transport equation for

$$W(\tau,\mathbf{p}) = W_A(\tau,\mathbf{p}) + W_B(\tau,\mathbf{p}) \tag{7.12}$$

as follows:

$$W_\tau(\tau,\mathbf{p}) = \int [\tilde{Q}(\mathbf{p},\mathbf{q})W(\tau,\mathbf{q}) - \tilde{Q}(\mathbf{p},\mathbf{q})W(\tau,\mathbf{p})]d\mathbf{q} + \gamma(\mathbf{p})W(\tau,\mathbf{p}), \tag{7.13}$$

$$Q(\mathbf{p},\mathbf{q}) = \tfrac{1}{2}\int_0^\infty \int \frac{[l(p)l(q)]^{-1}}{(2\pi)^n} (\cos\{[l(p)-l(q)]\sigma + (\mathbf{p}-\mathbf{q})\cdot\boldsymbol{\xi}\} + \cos\{[l(p)+l(q)]\sigma + (\mathbf{p}-\mathbf{q})\cdot\boldsymbol{\xi}\}) \times \rho(\sigma,\boldsymbol{\xi})d\boldsymbol{\xi}d\sigma \tag{7.14}$$

$$\gamma(\mathbf{p}) = \int \left(\int_0^\infty \int \frac{[l(p)l(q)]^{-1}}{(2\pi)^n} \cos\{[l(p)+l(q)]\sigma + (\mathbf{p}-\mathbf{q})\cdot\boldsymbol{\xi}\}\rho(\sigma,\boldsymbol{\xi})d\boldsymbol{\xi}d\sigma\right)d\mathbf{q}. \tag{7.15}$$

The nonconservative transport equation (7.13) in wave-number space is the main result of this section.

From the definition of $A(t,\mathbf{p})$ and $B(t,\mathbf{p})$ we find that

$$W(\tau,\mathbf{p}) = \lim_{\epsilon\to 0} E\left\{\frac{1}{2l(p)}\left[\hat{u}_t^2\left(\frac{\tau}{\epsilon^2},\mathbf{p}\right) + l^2(p)\hat{u}^2(\tau/\epsilon^2,\mathbf{p})\right]\right\}. \tag{7.16}$$

Thus, $W(\tau,\mathbf{p})$ is the normalized average field energy density.[20] We have remarked already that the transport equation (7.13) is nonconservative. This is due to the presence of the term involving $\gamma(\mathbf{p})$ which is nonnegative for all $\mathbf{p}\in R^n$. Equation (7.13) is in fact quite analogous to (5.17). In order for the total average field energy to remain bounded as a function of τ, we must introduce dissipative terms of order ϵ^2 in (7.1) as we did in (5.1). For example, if we introduce the term $2\epsilon^2\beta u_t(t,\mathbf{x})$ on the left side of (7.1), β constant, we obtain the stability condition

$$\beta > \gamma(\mathbf{p}) \quad \text{for all } \mathbf{p}\in R^n. \tag{7.17}$$

This condition implies that $\int W(\tau,\mathbf{p})d\mathbf{p}$ is a bounded function of τ.

ACKNOWLEDGMENTS

I wish to thank J. B. Keller for reading Secs. 2–5 and suggesting several improvements, J. McKenna for reading the manuscript and bringing to my attention an error in an earlier version of the remarks following (5.20), and I. M. Besieris and P. L. Chow for several stimulating discussions.

*Research supported by the Office of Naval Research under Contract No. N000 14-67-A-0467-0015.

[1] G. C. Papanicolaou and J. B. Keller, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. 21, 287 (1971).
[2] G. C. Papanicolaou and R. Hersh, Indiana Univ. Math. J. 21, 815 (1972).
[3] R. Cogburn and R. Hersh, Indiana Univ. Math. J., to appear.
[4] I. Prigogine, *Non-equilibrium statistical mechanics* (Interscience, New York, 1962).
[5] G. V. Chester, *The theory of irreversible processes in many-body problems*, edited by W. E. Parry (Benjamin, New York, 1969).
[6] L. Van Hove, "Master equations and approach to equilibrium for quantum systems," in *Fundamental problems in statistical mechanics*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1962).
[7] W. Pauli, *Festschrift zum 60. Geburtstage A. Sommerfelds* (Hirzel, Leipzig, 1928).
[8] D. Marcuse, Bell Syst. Tech. J. 51, 229 (1972).
[9] D. Marcuse, Bell Syst. Tech. J. 51, 429 (1972).
[10] H. E. Rowe and D. T. Young, IEEE Trans Microwave Theory Technology 20, 349 (1972).
[11] J. A. Morrison and J. McKenna, Bell Syst. Tech. J. 51, 209 (1972).
[12] R. L. Stratonovich, *Topics in the theory of random noise* (Gordon and Breach, New York, 1963), Vols. I, II.
[13] R. Burridge and G. Papanicolaou, Comm. Pure Appl. Math., to appear.
[14] W. Feller, *An introduction to probability theory* (Wiley, New York, 1966), Vol. II.
[15] J. Arnaud, Bell Syst. Tech. J. 49, 2311 (1970).
[16] G. Papanicolaou, D. McLaughlin, and R. Burridge, J. Math. Phys. to appear.
[17] L. C. Dolin, Radiophyzika 11, 840 (1968).
[18] V. N. Klyatskin, and V. N. Tatarskii, Radiophyzika 13, 1061 (1970).
[19] I. M. Besieris and F. D. Tappert, J. Math. Phys., submitted.
[20] The normalization is the factor $1/l(p)$. Thus, $W(\tau,\mathbf{p})$ is, in fact, the average field action density (remark due to F. D. Tappert).

Noumenon: Elementary entity of a new mechanics

E. Y. Rocher
International Business Machines Corporation, Mohansic Systems Laboratory, Yorktown Heights, New York 10598
(Received 15 June 1972; revised manuscript received 9 August 1972)

If the postulate of symmetry on which special relativity is built is rejected, a generalization of the relativistic notions of event and space-time can be proposed. The generalization leads to the notion of noumenon. The noumenon possesses a handedness; it is a seven-parameter entity obtained by associating with an event the angular momentum corresponding to a particular evolution of that event. The noumenon is defined in the complex extension C_4 of space-time R_4. The main purpose of the paper is to prove the acceptability of the new concepts when confronted with experimental results which are generally regarded as supporting the classical theory of relativity. The potential fruitfulness of the new concepts is shown by a short review of similar ideas developed independently by several authors in different fields of physics: A Maxwellian theory of gravitation is developed; interactions between gravitation and electromagnetism appear, which have common characteristics with weak interactions; and it is suggested that the extra degrees of freedom of the noumenon are related to the quantum numbers of elementary particles.

1. INTRODUCTION

Physics, mathematics, and mechanics are three complementary domains of science. Physics is the science treating of the material world and its happenings; the role of physics is to discover the laws describing the relations between happenings. Mathematics, starting from a minimum set of axioms, creates abstract entities and then studies all exact relations existing between these mathematical beings. Finally, mechanics is ultimately the mapping of physics into mathematics. From these short definitions, the role of mechanics appears to be twofold: choice of an elementary physical entity and choice of a mathematical being which the physical entity can be associated with.

According to this general scheme, relativity theory (special or general) is conventionally developed on the following two postulates: (1) The most fundamental physical entity is the time point or event and (2) space–time, the set of events, can be mapped one-to-one into the set of 4-vectors of a four-dimensional manifold. These basic hypotheses introduced in the early 1900's by Lorentz, Poincaré, Einstein, and Minkowski have proven so fruitful that they may seem impossible to challenge. In fact, however, further evolution of physics has shown that they apply only to the macroscopic world; they do not apply directly to microphysics. De Broglie (1923) suggested that particles behave like waves in space–time, Heisenberg's mechanics (1925) rejected the classical notion of position and trajectory, Dirac's theory (1928) showed that spinors play a more fundamental role in nature than vectors do.

The thesis sustained in this paper is the following: The success of relativity theory does not prove the ultimate validity of the relativistic mapping, as defined above. The numerous experimental "verifications" of the classical theory prove a more general property: the fundamental role played by the Lorentz-group in nature. In particular, corresponding to the more fundamental spinor representation of the group a new representation space will be mathematically defined. By inverse mapping onto physics, a new concept of universe will be proposed, as well as the definition of a new fundamental entity: the *noumenon*. The term is borrowed from Plato and Kant; it means thing-in-itself in contrast to the "phenomenon" or thing as it appears to us. These new concepts will be regarded as a generalization of the classical notions of space–time and event.

The Lorentz-group[1] is homomorphic to $SL(2, C)$, the group of unimodular 2×2 complex matrices. $SL(2,C)$ has two self-representations $D^{1/2\ 0}$ and $D^{0\ 1/2}$, the representation spaces of which are, respectively, a two-dimensional complex space and its complex conjugate. The elements of the representation space are two-component spinors belonging to the Minkowski-space R_4; they are each treated as a column matrix. The proper Lorentz-group corresponds to the $D^{1/2\ 1/2}$ representation of $SL(2, C)$; the elements of the representation space are spinors of order two of R_4. The classical relativistic formalism is based on the isomorphism between spinors of order 2 and 4-vectors of the Minkowski-space.

As early as 1911, another equivalent formalism had been discovered,[2] which has the advantage of displaying particularly well the relation between the representations $D^{1/2\ 1/2}$ and $D^{1/2\ 0}, D^{0\ 1/2}$: It is the *biquaternion formalism*. As will be discussed in more detail in the next section, the ring of biquaternions is isomorphic to the ring of 2×2 complex matrices. In the corresponding formalism, an event observed in two Galilean frames k′ and K is mapped into two biquaternions m' and M, which transform under a Lorentz-transformation according to

$$M \to m' = t{*}M{*}t^{\bullet}, \tag{1}$$

where t and its complex conjugate $t^{\bullet}$ are biquaternions of norm unity, i.e., are, respectively, elements of the representations $D^{1/2\ 0}$ and $D^{0\ 1/2}$ of $SL(2, C)$. It will be shown that t can be chosen such that $t = t^{\bullet}$. The entity mathematically defined by the biquaternion:

$$m = t{*}M \quad \text{or} \quad m^{\bullet} = M{*}t \tag{2}$$

will be mapped into a new physical entity possessing a handedness: the *noumenon*. Noumena are defined in the space of 2×2 complex matrices, which will appear to be the complex extension C_4 of space–time R_4.

Then the biquaternion mapping is generalized in two steps. First, a restricted definition of the physical concept of event is proposed: In a frame K an "event" is a *fixed* point at a given time. Second, the biquaternion mapping is redefined as follows: An "event" defined in a frame K is a noumenon in any other Galilean frame k, and not an "event" as assumed in classical relativity theory. The ultimate validity and the implications of the biquaternion mapping are discussed.

2. BIQUATERNION ALGEBRA-NOTATIONS

Biquaternions[3] are hypercomplex numbers of order 8. The biquaternion algebra (Pauli algebra) is the Clifford algebra[4] of order 3. The three unit-vectors, on which the algebra is built, can be represented by the three Pauli matrices $\boldsymbol{\sigma}$. Consequently a biquaternion q, noted

$$q = (q_0, \mathbf{q}), \tag{3}$$

can be represented by the 2×2 matrix q:

$$q = q_0 I + \mathbf{q}\cdot\boldsymbol{\sigma}, \tag{4}$$

where I is the unit matrix and where the four components $q_0, \mathbf{q}$ may be complex. Accordingly, the biquaternionic multiplication $*$ is defined by

$$\begin{aligned} q*Q &= (q_0, \mathbf{q})*(Q_0, \mathbf{Q}) \\ &= (q_0 Q_0 + \mathbf{q}\cdot\mathbf{Q},\ \mathbf{q}Q_0 + q_0\mathbf{Q} + i\mathbf{q}\wedge\mathbf{Q}). \end{aligned} \tag{5}$$

For a biquaternion q, one has the following definitions and the corresponding properties directly derived from Eq. (5):

The q conjugate of q is the biquaternion $\bar{q}$:

$$\bar{q} = (q_0, -\mathbf{q}). \tag{6}$$

The q conjugation is an anti-automorphism of the ring of biquaternions

$$\overline{q*Q} = \bar{Q}*\bar{q}. \tag{7}$$

The c conjugate of q is the biquaternion $q^{\cdot}$:

$$q^{\cdot} = (q_0^{\cdot}, \mathbf{q}^{\cdot}), \tag{8}$$

where the dot denotes the complex conjugate. The c conjugation is an anti-automorphism of the algebra

$$(q*Q)^{\cdot} = Q^{\cdot}*q^{\cdot}. \tag{9}$$

The qc conjugation is an automorphism[5] of the algebra

$$(\overline{q*Q})^{\cdot} = \bar{q}^{\cdot}*\bar{Q}^{\cdot}. \tag{10}$$

The norm of q is the generally complex scalar $N(q)$:

$$N(q) = q*\bar{q} = q_0^2 - \mathbf{q}^2 = \bar{q}*q. \tag{11}$$

Note that with the matrix notation of Eq. (4),

$$N(q) = \det(q_0 I + \mathbf{q}\cdot\boldsymbol{\sigma}). \tag{12}$$

It results from Eq. (7) or Eq. (12) that the biquaternion algebra is normed:

$$N(q*Q) = N(q)N(Q). \tag{13}$$

The inverse of a biquaternion q(nonzero norm) is the biquaternion q^{-1}:

$$q^{-1} = \bar{q}/N(q).$$

Any biquaternion q of norm unity can be written as the product of two biquaternions of norm unity:

$$q = t*r \tag{14}$$

with

$$t = (\cosh\alpha, \mathbf{t}\sinh\alpha), \quad \text{and} \quad r = (\cos\theta, i\mathbf{r}\sin\theta),$$

where α and θ as well as the unit-vectors $\mathbf{t}$ and $\mathbf{r}$ are real. Note that r is a quaternion[5] and that

$$t^{\cdot} = t, \quad r^{\cdot} = \bar{r}, \quad q^{\cdot} = \bar{r}*t. \tag{15}$$

3. BIQUATERNION MAPPING-LORENTZ TRANSFORMATION

Let M_0 be a point in a Galilean frame of reference K. We can measure the coordinates $\mathbf{X}$ of the point M_0 from the origin O of the frame K by using the "Gedanken-experiment" suggested by Einstein (1905), i.e., by observing the round trip time required by a light pulse on the path OM_0O. Then consider the physical entity $\mathfrak{M}$ represented by the same fixed point M_0 at a given time X_0 defined with the same techniques; this happening $\mathfrak{M}$ is considered in K as an *event* M and is represented by the biquaternion M with four real components:

$$\mathfrak{M} \xrightarrow{\mathrm{K}} M = (X_0, \mathbf{X}). \tag{16}$$

Note that according to Eq. (11) the norm of M is real and equal to the square of the length of the 4-vector $X_0, \mathbf{X}$ in space-time.

Now let k′ be another Galilean frame. Classical relativity theory requires $\mathfrak{M}$ to be considered as an *event* in k′, i.e., $\mathfrak{M}$ must be represented in k′ by a biquaternion m' having four real components:

$$\mathfrak{M} \xrightarrow{\mathrm{k}'} m' = (x_0', \mathbf{x}'). \tag{17}$$

To fulfill the condition of reality, taking Eq. (9) into account, the linear relation between the biquaternions M and m' must have the form

$$M \to m' = q*M*q^{\cdot} = m'^{\cdot}, \tag{18}$$

where the norm of q must be unity[6] for m' and M to have the same norm (Lorentz transformation). Via Eqs. (14) and (15), the transformation defined by Eq. (18) can be written

$$M \to m' = t*r*M*\bar{r}*t \tag{18'}$$

For $t = 1$ (i.e., $\alpha \to 0$), the transformation equation (18′) reduces to

$$M \to M' = r*M*\bar{r}, \tag{19}$$

which represents a spatial rotation -2θ of the frame K around the vector $\mathbf{r}$ (Hamilton—1853). Consequently when the coordinate axes of k′ and K are parallel, the transformation equation (18) can be written

$$M \to m' = t*M*t, \tag{20}$$

where

$$t = (\cosh\alpha, \mathbf{t}\sinh\alpha) = \gamma(1, \boldsymbol{\beta}) \tag{21}$$

with

$$\gamma = (1-\beta^2)^{-1/2}.$$

By developing Eq. (20) according to Eq. (5), one gets directly the vectorial expression[7] of the Lorentz transformation

$$\left.\begin{matrix} X_0 \\ \mathbf{X} \end{matrix}\right\} \to \left\{\begin{aligned} x_0' &= \gamma(X_0 + \boldsymbol{\beta}'\cdot\mathbf{X}) \\ \mathbf{x}' &= \mathbf{X} + \boldsymbol{\beta}'\left(\frac{\mathbf{X}\cdot\boldsymbol{\beta}'}{\beta'^2}(\gamma'-1) + \gamma' X_0\right), \end{aligned}\right. \tag{22}$$

where $\boldsymbol{\beta}'$ is the velocity of the frame K in k′:

$$\boldsymbol{\beta}' = 2\boldsymbol{\beta}/(1+\beta^2) \tag{23}$$

and

$$\gamma' = (1-\beta'^2)^{-1/2}.$$

In particular, we note that Eq. (23) shows that the velocity $\boldsymbol{\beta}'$ is relativistically twice the velocity $\boldsymbol{\beta}$.

4. THE NOUMENON

Equation (20) shows that the Lorentz transformation is performed in two steps, in two possible ways according to the schemes:

$$\begin{array}{ccccc} & & m = t*M & & \\ & \nearrow & & \searrow & \\ M & & & & m' = m*t = t*m^{\cdot}. \\ & \searrow & & \nearrow & \\ & & m^{\cdot} = M*t & & \end{array} \qquad (24)$$

We call the entity mathematically represented by the biquaternion m (or $m^{\cdot}$) a *noumenon*[8] and, more precisely, a left-noumenon (or a right-noumenon) to express the noncommutativity of the biquaternion product. The components $x_0, \mathbf{x}$ of the noumenon m are complex, and the norm of the biquaternion m is real and t invariant according to Eqs. (13) and (21); thus, the noumenon depends on seven parameters. The noumenon is a more fundamental entity than the event; it carries in itself the information pertaining not only to the position of the happening $\mathfrak{M}$, but also to the evolution (velocity) of that happening; furthermore, it permits attribution of a handedness characteristic of the evolution of the happening. This last property will be discussed in more detail in the last section.

By developing the relation $m = t*M$ we obtain the expressions of the four complex components of the noumenon m:

$$\left.\begin{array}{l} X_0 \\ \mathbf{X} \end{array}\right\} \to \left\{\begin{array}{l} x_0 = \gamma(X_0 + \boldsymbol{\beta}\cdot\mathbf{X}) \\ \mathbf{x} \;= \gamma(\mathbf{X} + \boldsymbol{\beta}X_0 + i\boldsymbol{\beta}\wedge\mathbf{X})^{\cdot} \end{array}\right. \qquad (25)$$

Comparison with the classical expression (22) of the Lorentz transformation shows a remarkable similarity: *The transformation formulas for the time component and the component in the direction of the motion have the same form.* The noumenon m (or $m^{\cdot}$) can be considered as defined in the complex extension of the frame k in which the velocity of K is $\boldsymbol{\beta}$ and in which the velocity of k′ is $-\boldsymbol{\beta}$, according to the concluding remark of Sec. 3. Since the direct "verifications"[9] of the classical theory of relativity were precisely designed to check the transformation formulas for the time component and the component in the direction of motion, we raise the question of the ultimate validity of the relativistic mapping, and we propose a more fundamental approach.

For that purpose we will restrict the classical notion of event as follows: In a frame K an "event"[10] is a *fixed point* at a given time. An "event" is represented by a biquaternion M with four real components. The generalized biquaternion mapping can then be redefined as follows:

> *An "event" defined in a Galilean frame* K *is a noumenon (left or right) in any other Galilean frame of references and not an event* as assumed in the classical theory of relativity.

This hypothesis can be interpreted as follows: When observing in the frame k the "event" M defined in K, the result of the measurement gives

(1) the four coordinates of M in k; they are taken as the real parts $x_0, \mathbf{x}_r$ of the components of the noumenon m (in classical relativity theory they are taken as the components of the "event" m' in k′) and

(2) the components $\boldsymbol{\beta}$ of the velocity[11] of M in k.

From Eq. (25) the noumenon m is the seven-dimensional happening

$$\mathfrak{M} \xrightarrow{\text{k}} m = (x_0, \mathbf{x}_r \pm i\boldsymbol{\beta}\wedge\mathbf{x}_r). \qquad (26)$$

In classical terms, the noumenon is obtained by associating with an event the angular momentum corresponding to the evolution of that particular "event." Reciprocally, given a noumenon and its handedness the immediate evolution of the corresponding "event" is well defined by Eq. (26). Thus, the new entity contains a dynamic property which characterizes the existence of the physical being it refers to; indeed, existence supposes some extension in space–time: The term noumenon is used here in a sense close to the sense of thing-in-itself given by the Greek philosophers and by Kant in his early works. Since physics is the science of existing beings it can be expected that it will be more fruitful to build a mechanics on the existential notion of noumenon instead of on the static notion of event.

The noumenon is defined in the complex extension C_4 of R_4: Under the biquaternion mapping, the universe has to be considered as a complex four-dimensional manifold. In particular, it appears that this concept, inspired by Dirac's formalism, leads us to replace the classical concepts of position and trajectory, rejected by Heisenberg's mechanics, with more generalized concepts, which will be shown in the next section to be compatible with experimental observations.

Further examination of the transformation t defined in Eqs. (25) shows that the formulas differ from the classical formulas for the components normal to the motion: First by the presence of the complex term and second by the dilatation factor γ. This last remark seems to indicate the possibility of a direct test of the theory, a test in which the lateral dilatation could be evidenced. Interpretation of the Michelson experiment (1881) will show the difficulty of the task.

In the classical interpretation, when considering the negative result of the Michelson experiment from a frame k attached to the Sun, it is supposed that the length of the north-south arm of the interferometer is not affected by the motion of the Earth, only the length of the east-west arm is contracted. In the present theory the classical interpretation applies also to the EW arm. For the NS arm, let $E = (0, \mathbf{0})$ represent the emission of a light pulse used in the Michelson experiment and let $R = (X_0, \mathbf{X})$ be the reflection of the light pulse at the extremity of the NS arm, as measured in the frame K attached to the interferometer. The norm of R is zero. In a frame k attached to the Sun, the corresponding noumena are $e = (0, \mathbf{0})$ and $r = (x_0, \mathbf{x})$. The norm of r is zero since $r = t*R$; furthermore, since x_0 transforms as in the classical theory, the length $|\mathbf{x}|$ of the trajectory of the photons used in the experiment is the same as in the classical case; the same conclusion applies for the return path. Consequently, as in the classical case, an observer in the frame attached to the Sun arrives at the same conclusions as an observer attached to the interferometer: The NS and EW light paths are equal.

The result was expected, because both theories are based on the same postulate: The measure of the velocity of light is independent of the frame in which the measurement is made. Translated into mathematical terms, Lorentz transformation and t transformation are representations of the same group $SL(2,C)$. For the same reason the law of composition of parallel velocities is the same in both theories; a second-order difference appears for the composition of nonparallel velocities. The main difference between the two theories is in the definition of the distance between two points (events or noumena) in space–time (real or complex), i.e., in the choice of the metric. In conclusion, a direct test of the theory should be a test of the definition of the metric. The test performed in a frame k would consist of observing a fast-moving object of known transverse dimensions at rest in $K(Y,Z)$, and in checking that in k the object *appears* dilated ($y_r = \gamma Y, z_r = \gamma Z$).

An indirect method to establish the validity of a theory is to prove its fruitfulness. The next section develops some elements of a unitary field theory in the complex space–time C_4, and, in particular, it introduces the connection between the biquaternion mapping and the work on quaternion-factorization of the metric of general relativity as initiated by Einstein (1929) and redeveloped more recently by Bergmann (1957) and Sachs (1968).

5. FIELD THEORY IN THE COMPLEX UNIVERSE C_4

Shortly after publication of Dirac's paper, several authors tried to introduce hypercomplex numbers in mechanics. In quantum field theory[12] the quaternion formalism appeared particularly well adapted to the two main relativistic field theories: Maxwell's theory and Dirac's theory. In general relativity,[13] a quaternion factorization of the metric offered the possibility of generalizing to electromagnetism the geometrization of gravitation. Ever after, these ideas have been further developed by many authors, but have generally looked too artificial to become fruitful. We shall show how the concept of noumenon, elementary entity of a complex universe, offers a simple and natural approach to unify the different branches of mechanics.

A. The quaternionic propagation operators

To describe the properties of the representation space C_4, it is necessary to introduce a differential operator. In the present context it is natural to define the linear biquaternion operator

$$P = (\partial_{X_0}, \partial_{\mathbf{X}}), \tag{27}$$

where

$$\partial_{X_j} = \frac{\partial}{\partial X_j} = \frac{\partial}{\partial X_{rj}} + i\,\frac{\partial}{\partial X_{cj}},$$

where X_{rj} and X_{cj} are the real and complex parts of the complex coordinate X_j in the frame K. The q conjugate of P is written $\bar{P}$ and is defined by

$$\bar{P} = (\partial_{X_0}, -\partial_{\mathbf{X}}) \tag{28}$$

By definition the operators P and $\bar{P}$ operate on biquaternionic functions placed on their right-hand side.[14] Successive applications of the operators P and $\bar{P}$ give the scalar complex operator $\boxdot$:

$$P*\bar{P} = \partial^2{}_{X_0^2} - \partial^2{}_{\mathbf{X}^2} = \boxdot = \bar{P}*P. \tag{29}$$

Consequently, the operators P and $\bar{P}$ appear as propagation operator in C_4 defined in the frame K.

Similarly, in a frame k the propagation operators are defined by

$$p = (\partial_{x_0}, \partial_{\mathbf{x}}), \quad \bar{p} = (\partial_{x_0}, -\partial_x). \tag{30}$$

Via Eq. (25) the relations between the operators p and P are

$$P \to p = P*\bar{t}, \quad \bar{P} \to \bar{p} = t*\bar{P} \tag{31}$$

Consequently,

$$p*\bar{p} = P*\bar{P}. \tag{32}$$

The scalar propagation operator $\boxdot$ is t invariant.

Let $\Phi = (\Phi_0, \mathbf{\Phi})$ be a biquaternion function defined at each point in C_4. Supposing the potential Φ differentiable in C_4, then the propagation equation of a wave propagating at speed $c = 1$ in C_4 can be written

$$\boxdot\,\Phi = R. \tag{33}$$

This second-order equation can be replaced by two coupled first-order equations

$$\bar{P}*\Phi = F, \tag{34}$$

$$P*F = R. \tag{35}$$

Under a t transformation the density R transforms according to[15]

$$R \to r = R*\bar{t}. \tag{36}$$

Consequently, since in Eq. (33) the operator $\boxdot$ is t invariant, Φ transforms according to

$$\Phi \to \phi = \Phi*\bar{t} \tag{37}$$

and

$$F \to f = t*F*\bar{t} \tag{38}$$

to assure the t covariance of Eqs. (34) and (35). Note that $\bar{F}$ transforms like F:

$$\bar{F} \to \bar{f} = t*\bar{F}*\bar{t} \tag{39}$$

Consequently, the scalar part of the biquaternion field F is t invariant.

If, now, the operators P and $\bar{P}$ are defined in R_4 instead of C_4, i.e., they are restricted to the real components, then the scalar operator $\Box$ is real, and since $\Box$ is invariant, it is also real in any other frame **k**. If the potential Φ is complex, the propagation equation (33) will split into two *uncoupled* equations

$$\Box\,\Phi_r = R_r, \quad \Box\,\Phi_c = R_c \tag{40}$$

between the real and complex part of Φ and R. A coupling will appear between the first-order equations (34) and (35).

B. Maxwell equations-Maxwellian gravitation

The preceding results can be applied to electromagnetism in R_4. First consider the equations of electrostatics in K. Let $\Phi = i\Phi_c = i(V, \mathbf{0})$ be the electrostatic potential, and let $R = iR_c = i(R, \mathbf{0})$ be the charge density. Then the second equation (40) is Poisson's equation in K, Eq. (34) defines the electrostatic field $\mathbf{E}$:

$$F = \bar{P}*\Phi = i(\partial_0 V, -\operatorname{grad} V) = (0, i\mathbf{E}), \tag{41}$$

and Eq. (35) is the expression of Gauss' theorem

$$P*F = (i\,\operatorname{div}\mathbf{E}, \mathbf{0}) = (iR, \mathbf{0}). \tag{42}$$

Finally, the density of electrostatic energy must be a quadratic invariant under charge conjugation, i.e., under c conjugation for reasons which will appear more clearly later:

$$W = F * F^{\cdot} = W^{\cdot} = (\mathbf{E}^2, 0) \tag{43}$$

The equations in a frame k are obtained by a t transformation according to Eq. (37):

$$\phi = i(v, -\mathbf{a}), \tag{44}$$

$$f = \bar{p} * \phi = i(\partial_0 v + \mathrm{div}\mathbf{a}, -\partial_0 \mathbf{a} - \mathrm{grad} v + i\, \mathrm{rot}\mathbf{a}) = (0, i\mathbf{e} - \mathbf{h}), \tag{41'}$$

where the scalar equation (Lorentz-condition) is deduced from Eqs. (38), (39), and (41).[16] In the frame k Eq. (42) gives

$$(i\, \mathrm{div}\mathbf{e} - \mathrm{div}\mathbf{h}, i(\partial_0 \mathbf{e} - \mathrm{rot}\mathbf{h}) - (\partial_0 \mathbf{h} + \mathrm{rot}\mathbf{e})) = (ir, -ij). \tag{42'}$$

When considering from k another source attached to a frame K′ similar equations are obtained; since they are linear they can be added without changing the form of Eq. (42′). Then Eq. (42′) splits into two scalar and two vector equations, which are the four Maxwell's equations (1864).

Finally, the density of electromagnetic energy is given by the C-invariant biquaternion

$$w = f * f^{\cdot} = (\mathbf{e}^2 + \mathbf{h}^2, -2i\mathbf{e} \wedge \mathbf{h}), \tag{43'}$$

where one recognizes the classical expression of electromagnetic energy density associated with the Poynting vector.

These results can be applied to calculate the equations of motion in C_4 of a charged particle (mass μ, charge ϵ) in an electromagnetic field. Taking Lorentz's approach as a first approximation, the equation of motion in a Galilean frame K, in which the particle is at rest at a given time, reads

$$\mu \frac{d^2 M}{ds^2} = -i\epsilon F. \tag{45}$$

In another Galilean frame k, we get after left multiplication by t and according to Eq. (38):

$$\mu \frac{d^2 m}{ds^2} = -i\epsilon f * t, \tag{46}$$

the real part of which gives the equations of the motion *observed in* R_4:

$$\begin{cases} \mu \dfrac{d\gamma}{dx_0} = \epsilon\, \mathbf{e}\cdot\boldsymbol{\beta}, \\ \mu \dfrac{d}{dx_0}\left(\gamma \dfrac{d\mathbf{x}_r}{dx_0}\right) = \epsilon(\mathbf{e} + \boldsymbol{\beta} \wedge \mathbf{h}). \end{cases} \tag{47}$$

These equations are the classical relativistic equations.[17] This important result has two consequences: (1) It assures the compatibility of the proposed theory with the most important class of experiments supporting the classical theory of relativity—the experiments involving the interaction of a charged particle with an electromagnetic field, and (2) it is a first example showing the acceptability of the new notion of complex trajectory.

Another particular case of the field equations (33)–(35) is $\Phi = \Phi_r = (V', 0)$ and $R = R_r = (R', 0)$. It is obtained by replacing the charge ϵ by $\mu = i\epsilon$. Consequently *Coulomb's repulsion transforms into an attraction, Newton's attraction as we will assume.* Under this hypothesis the real part of the complex wave Φ becomes the gravitational potential, the complex part being the electromagnetic potential. The quantities related to a gravitational origin will be represented with a prime. Equations (41)–(42′) apply up to the factor i. The density of gravitational energy is given by the C-invariant biquaternion

$$w' = -f' * f'^{\cdot} = (-\mathbf{e}'^2 - \mathbf{h}'^2, 2i\mathbf{e}' \wedge \mathbf{h}'). \tag{43''}$$

We will discuss in the last section some of the implications of the preceding hypothesis.

As in the electromagnetic case, the notion of complex trajectory of a massive body in a gravitational field can be shown to be acceptable. We are referring here to the interpretation of the three "tests" of general relativity. In fact Schiff[18] has shown that among the three implicated effects, at least two can be interpreted simply by introducing the principle of equivalence into special relativity. Adapting Schiff's method to the new theory gives the same interpretation of the gravitational red shift and with a somewhat simpler algebra, due to the introduction of the transverse dilatation, gives the conventional expression for the deviation of photons by a gravitational field. The third effect concerns the anomalous part of the perihelion precession of planetary orbits (only observable for Mercury); it is considered as characterizing the Riemannian structure of space–time. Recently, searching for the expression of this third effect under the assumption of a biquaternion-factorization of the Riemannian metric, Sachs[19] has found the same expression of the perihelion precession as found under the conventional hypothesis of a symmetric metric tensor. This formalism leads, as expected,[13] to a unitary theory of electromagnetism and gravitation, which has common characteristics with the proposed field theory in C_4.

At the flat space limit of Sachs' approach, the differential interval in R_4 is mapped, as in Eqs. (3), (4), into a biquaternion ds:

$$ds = (dx_0, d\mathbf{x}) = \sigma_\alpha \cdot dx^\alpha, \tag{48}$$

the norm of which is the classical invariant ds^2. In the presence of a field, the σ_α are replaced by another set of four field dependent biquaternions q_α. The correspondence with the metric tensor is then defined by

$$g^{\mu\nu} \Leftrightarrow -\tfrac{1}{2}(q^\mu * \bar{q}^\nu + q^\nu * \bar{q}^\mu) = -Sc(q^\mu * \bar{q}^\nu) \tag{49}$$

where $Sc(q)$ refers to the scalar part of biquaternion q. The q_α are determined by four metric field equations analogous to Einstein's field equations (1916):

$$\tfrac{1}{4}(\mathrm{K}_{\rho\lambda} q^\lambda + q^\lambda \mathrm{K}^\dagger_{\rho\lambda}) + \tfrac{1}{8} R q_\rho = \chi S_\rho, \tag{50}$$

where $\mathrm{K}_{\rho\lambda}$ is the "spin curvature" and R the scalar curvature of the Riemann space and where the biquaternions S_ρ characterize the source field. By proper combination of the field equation and its q conjugate, two tensor equations are built: The symmetric-tensor part (scalar part) leads to Einstein's original theory of gravitation, the antisymmetric-tensor part (vector part) leads to Maxwell equations.[20] With this approach, gravitation and electromagnetism receive a similar geometric interpretation, suggesting that both fields can be combined in a unique entity as we have done in Eqs. (40). However, our approach is very different in a basic and in a formal aspect: Equations (40) are defined in a flat space and we do not try to resymmetrize the formalism, as is done in Eq. (49); indeed, once the difficult concepts of noumenon and of complex space–time are accepted, the necessity of symmetrization does not appear. Nevertheless, some problems appear in the field theory when considering the interaction between electromagnetism and gravitation. They will be discussed in the last section.

C. Discussion-further developments

In the general case the potential ϕ is complex:

$$\phi = (v' + iv, - c(\mathbf{a}' + i\mathbf{a})), \tag{51}$$

where the speed of light c has been introduced to use the MKSA system. The biquaternion field defined by Eq.(34) is $f'' = (0, i\mathbf{e}'' - c\mathbf{h}'')$ with[21]

$$\mathbf{e}'' = \mathbf{e} + c\mathbf{h}', \tag{52}$$

$$c\mathbf{h}'' = c\mathbf{h} - \mathbf{e}', \tag{53}$$

where, following the electromagnetic model, a generalized principle of equivalence between radiating and gravitating masses is assumed, i.e., with

$$\epsilon_0' = 1/\mu_0' c^2 = 1/4\pi G \tag{54}$$

(G = gravitation constant). Contrary to the potential wave equation (33), the field equations (34) and (35) show a coupling between gravitational and electromagnetic fields. The coupling is very weak and can be evaluated. For example, Eq. (52) shows that a rotating mass creates a girogravitational field $\mathbf{h}'$ which is equivalent to an axial electric field. In the case of the Earth this field can be calculated adapting classical formulas of electromagnetism. One finds at the center of the Earth

$$c\mathbf{h}' = \frac{c\mu_0'}{2\pi} \frac{\mathfrak{M}'}{R^3}, \tag{55}$$

where $\mathfrak{M}' = 7.1 \times 10^{33}$ MKS is the kinetic momentum of the Earth and $R = 6.4 \times 10^6$ MKS its radius. Even in the case of as large a massive body as the Earth, the axial field is still very small:

$$ch' = 1.4 \times 10^{-5}\ \text{V/m}.$$

This field value compares favorably with the estimated 10^{-7} V/m electric field, the origin of which is still hypothetical, necessary to explain the geomagnetic field. This field creates in the core a small current (10^{-6} A/cm^2) which by dynamo effect is responsible for the Earth's magnetic field.

A basic difficulty appears in this simplified gravito-electromagnetic interpretation of the wave equation (33), when considering the inverse effect of an electromagnetic field on a mass. Solving the difficulty requires a detailed analysis of the symmetry properties of the biquaternion algebra; this analysis will be initiated here. The preceding results have then to be considered as fundamental solutions of the field equations, to be recombined to fulfill the proper symmetry condition.[22]

TABLE I. Internal Symmetries of the Complex Universe C_4.

Unit Vectors		e_0	$\mathbf{e}_l$	$\mathbf{e}_{lm} = \mathbf{e}_l \mathbf{e}_m$	$\mathbf{e}_{123}$
Matrix Repres.		I	σ_l	$i\sigma_l$	iI
Unit Square		$+1$	$+1$	-1	-1
Q Conj:	$(-T)$	+	−	−	+
C Conj:	(C)	+	+	−	−
QC Conj:	(P)	+	−	+	−
Noumenon		x_{0r}	$\mathbf{x}_r$	$\mathbf{x}_c$	x_{0c}
Potential		v'	$\mathbf{a}'$	$\mathbf{a}$	v
Source		r'	$\mathbf{j}'$	$\mathbf{j}$	r
Field		0	$c\mathbf{h} - \mathbf{e}'$	$\mathbf{e} + c\mathbf{h}'$	0

In Table I are gathered the elements of the discussion of the correspondence between symmetries of the biquaternion algebra and symmetry properties of the complex space–time C_4. The table is divided into three parts. The first three rows define which one of the four possible representations of the biquaternion algebra is chosen for the biquaternion mapping; the next three rows indicate which unit vectors are reversed under a given conjugation; the last part shows how the components of different physical entities are assigned to the unit vectors.

Q conjugation is a complex space inversion or a complex time reflection ($-T$). Under a c conjugation the electromagnetic potential and its source are inverted; c conjugation appears as a charge conjugation (C). Finally, qc conjugation is obtained by reversing the orientation of the real space (P). According to this interpretation, the product of the three fundamental symmetries C, P, and T of the complex universe C_4 is CPT $= -1$. Of particular interest here is the interpretation of the c conjugation. Classically, charge conjugation interchanges particles and anti-particles. Referring back to Eq. (24) defining the noumenon, right and left noumena are precisely c conjugate of each other. The close relationship between charge, handedness, and rotation appears particularly clearly in the proposed model of a complex universe.

Up to here we have only been concerned with the macroscopic aspect of the formal possibility of a concept of a complex universe. The macroscopic gravito–electromagnetic interaction, as defined by Eq. (55), has two characteristics in common with the microscopic weak interactions: order of magnitude (intermediate between electromagnetic and gravitational) and symmetry [in Table I the corresponding potential $v', \mathbf{a}$ or $v, \mathbf{a}'$ have the rotation symmetry of a four-dimensional real Euclidean space, i.e., $SU(2) \times SU(2)$[1,23]]. In fact, the requirement for a generalization of the basic concepts of event and space–time has always been even more pressing in microphysics than in macrophysics, probably due to the availability of more powerful experimental tools.

The most precise need for defining an entity possessing a handedness appeared with the experimental confirmation[24] of the Lee and Yang[25] hypothesis of "parity violation" in weak interactions. The electrons created in the decay of oriented radioactive cobalt nuclei emerged all in the direction of the magnetic field, thus showing that besides their momentum, moving electrons are characterized by an internal property: handedness. In the present context electrons are mapped into left noumena, anti-electrons into right noumena. *The biquaternion mapping (26) is uniquely defined by nature in this case.*

A broadening of the classical concept of space–time has also always been strongly suggested in particle physics.

The various internal characteristics of elementary particles (charge, baryon number, hypercharge, parity, ⋯) are often considered[23,26] as manifestations of extra-angular degrees of freedom in some abstract space (isospace, isobaric space, subquantized medium, ⋯) which is completely distinct from real space–time. In the present formalism those extra-angular variables are introduced in the general form of the transformation biquaternion q as defined by Eq. (14) and are given a precise geometric meaning.

6. CONCLUSION

The essential characteristic of this generalization of the concepts of space–time and event is its natural simplicity. The generalization is based on the renunciation of one of the postulates on which the classical theory of relativity was built, the postulate of symmetry. Instead, a more involved algebra of space–time symmetries has been introduced, the role of which appears to be essential in any further development of the theory. The main purpose of the paper is to prove the acceptability of the new concepts when confronted with the classical experimental results on which the classical theory is built. The noumenon possesses two main characteristics: (1) It has more degrees of freedom than the conventional point-like event, and (2) defined as a point in the complex universe it has a potential extension in real space–time. In the last section a rapid review of similar ideas developed recently and independently in different branches of physics shows the unifying properties of the new concepts. The geometric interpretation of the added degrees of freedom should lead to a better understanding of the different forms of interactions. The potential space–time extension built into the concept of noumenon confers on the noumenon an intrinsic existence, which should lead to a geometric approach to quantum field theory, in agreement with the wishes of de Broglie, Einstein, and Mie.

---

[1] See, for example, P. Roman, *Theory of elementary particles* (North-Holland, Amsterdam, 1961), 2nd ed., p. 60.

[2] F. Klein, Phys. Z. **12**, 17 (1911); C. Lanczos, thesis (1919) summarized in Z. Phys. **57**, 447 (1929).

[3] The term biquaternion algebra is chosen here instead of algebra of 2 × 2 complex matrices for two reasons: (a) complex numbers and quaternions can be used to represent rotations, respectively, in the plane R_2 and in space R_3. The next generalization in the family of hypercomplex numbers leads to biquaternions (complex quaternions after Hamilton): Biquaternions will be used to represent rotations in space-time. Note that the biquaternion algebra is the highest order associative algebra with a quadratic norm. (b) The three auto- or anti-automorphisms of the algebra will be associated in the last section with the three basic symmetries C, P, and T.

[4] D. Hestenes, *Space-time algebra* (Gordon and Breach, New York, 1966).

[5] With the notation of Eq. (3), a quaternion is represented by $q = (q_0, iq)$ where q_0 and q are real. If the basis vectors were the basis vectors (ie_j) of the quaternion algebra, the qc conjugation would be a c conjugation and, thus, an automorphism of the algebra of complex quaternions, as it is.

[6] Up to a phase factor $e^{i\Phi}$ which will be neglected.

[7] G. Hergoltz, Ann. Phys. (Leipz.) **36**, 497 (1911).

[8] Note that, under a Lorentz transformation the biquaternion m (or m·) transforms like a pair of two-component spinors. Similar entities have already been proposed in physics, but with a very different interpretation; for example, Penrose's twistors in J. Math. Phys. (N.Y.) **8**, 345 (1967). (See in particular Ref. 5.)

[9] H. E. Ives and G. R. Stilwell, J. Opt. Soc. Amer. **28**, 215 (1938); E. I. Williams and G. E. Roberts, Nature (Lond.) **145**, 102 (1940).

[10] When used in its restricted sense, the term event will be enclosed in quotation marks.

[11] Note that the information concerning the velocity β of the "event" is always available during the experimental determination of the position of the "event": Doppler effect, trajectory curvature ... Direct introduction of Planck's constant (1900) into the theory should then be made possible by noticing that the complex part of the noumenon, which corresponds in classical terms to an "uncertainty," is precisely equal to the angular momentum of the "event."

[12] See C. Lanczos, Ref. 2, and G. Rummer, Z. Phys. **65**, 224 (1930).

[13] A. Einstein, Math. Ann. **102**, 685 (1929).

[14] In the complex plane R_2 the operators P and $\bar{P}$ reduce to $(\partial_X + i\partial_Y)$ and $(\partial_x - i\partial_y)$. With the notations (30) Cauchy–Riemann's analyticity conditions are written $P \cdot F = 0$; when the analyticity conditions are not fulfilled in a region of the plane, the equation $P \cdot F = R$ characterizes the discontinuities of the field, i.e., its sources.

[15] An elementary volume dV can be defined in K by $dV = dM_1 * d\bar{M}_2 * dM_3$, in k by $t^* d\bar{V} = t^* dM_1 * d\bar{M}_2 * \bar{t}^* \bar{t}^* dM_3 = dm_1 * d\bar{m}_2 * dm_3 = dv$. Thus, a density transforms like $dv^{-1} = dV^{-1} * \bar{t}$.

[16] Note that the norm of the biquaternion field f is the invariant: $N(f) = f^* \bar{f} = \mathbf{e}^2 - \mathbf{h}^2 + 2i\mathbf{e} \cdot \mathbf{h}$ and in the particular case of a unique source, by comparing with Eq. (43): $\mathbf{e} \cdot \mathbf{h} = 0$.

[17] With the classical approach the same definitions (27), (28) apply in a frame K with the property (29) and with the same field equations (33)–(35). In k' the transformation (20) leads to $\bar{P}' \to \bar{p}' = t^* \bar{P}'^* t$ [Eq. (31′)], $\Phi' \to \phi' = \bar{t}^* \Phi'^* \bar{t}$ [Eq. (37′)] and, thus, $F' \to f' = \bar{p}'^* \phi' = t^* F'^* \bar{t} \equiv f$ [Eq. (38′)]. The same relation (45) applies in K, and in k' one gets after right and left multiplications by t: $\mu(d^2 m'/ds^2) = \mathrm{Re}(i\epsilon f'^* t^* t) = -\mathrm{Re}(i\epsilon f'^* t')$ [Eq. (46′)], where $\mathrm{Re}(q)$ means the real parts of q and with $t' = t^* t = \gamma'(1, \beta')$ according to Eq. (23). The equations of motion in k' (46′) and in k (46) have the same form.

[18] L. I. Schiff, Am. J. Phys. 28, 340 (1960).

[19] M. Sachs, Nuovo Cimento B **56**, 137 (1970).

[20] M. Sachs, Nuovo Cimento B **55**, 199 (1968).

[21] Quantities related to a gravitational (electromagnetic) origin are represented with (without) a prime.

[22] For example, the generalized Lorentz-condition can be considered as requiring the biquaternion field to reduce to its vector part, i.e., to be invariant under time reflection (T), as will be discussed.

[23] L. de Broglie, D. Bohm, P. Hillion, F. Halbwachs, T. Takabayasi, and J. P. Vigier, Phys. Rev. **129**, 438 (1963).

[24] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).

[25] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956).

[26] For example, F. Halbwachs, P. Hillion, and J. P. Vigier, Ann. Inst. Henri Poincaré **16**, 115 (1959); G. R. Allcock, Nucl. Phys. **27**, 204 (1961).

Diffraction by an infinite array of parallel strips

W. H. Kent
Bell Laboratories, Murray Hill, New Jersey

S. W. Lee
Department of Electrical Engineering, University of Illinois, Urbana, Illinois
(Received 3 August 1971; revised manuscript received 1 August 1972)

The problem of diffraction of a plane wave by an infinite array of parallel strips is attacked by the newly developed modified residue calculus method. The solution is found in terms of an infinite set of zeros of an analytic function. The asymptotic behavior of the set of zeros is specified by the edge condition, while the first several zeros are determined from a matrix equation. The rapid convergence of these zeros to their asymptotic values is demonstrated through numerical examples. For a given array of strips, it is shown that there exists a total reflection phenomenon at a critical frequency and incident angle. This fact suggests the possibility of constructing an open resonator with an extremely sparse resonance frequency.

I. INTRODUCTION

Diffraction of a plane wave by an infinite set of semi-infinite parallel plates was first solved by Carlson and Heins[1] by the Wiener–Hopf technique. Later, Berz[2] attacked the same problem by the residue calculus method.

Practical interest in this problem stems from the use of the structure as a microwave lens to focus beams in a desired direction or to produce multiple beams, and the use of the structure as an artificial dielectric medium.[3] In many such applications the plate length is in the order of the wavelength involved. Therefore, it is of importance to study the diffraction properties of a set of plates of finite length.

In the present paper, the problem is formulated in terms of an infinite set of linear equations, which is solved by the newly developed modified residue calculus method.

The same physical problem has been considered earlier by Meister.[4] He formulated the problem in terms of a modified Wiener–Hopf equation and solved it by a method developed by Jones.[5] Meister obtained only a formal solution in the form of an infinite set of simultaneous linear equations from which it is difficult to extract useful numerical results. The advantage of the modified residue calculus method used here over the method used by Jones has been detailed elsewhere[6–8] and will not be repeated here.

II. FORMULATION OF THE PROBLEM

The configuration of the infinite array of parallel strips, which have no y variations, is shown in Fig. 1.

A TM wave which is incident from the left at an angle θ_0 with respect to the x axis, is given by

$$H_y^{(i)} = e^{i\alpha_0 z} e^{-\beta_0 x}, \tag{1}$$

where the time convention $e^{-i\omega t}$ is understood and

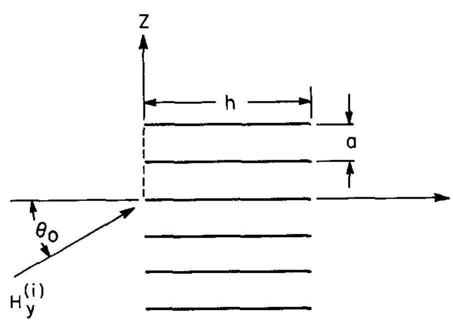


FIG. 1. Parallel strip configuration.

$$\alpha_p = \frac{2p\pi + ka \sin\theta_0}{a}, \tag{2}$$

$$\beta_p = \begin{cases} +\ (\alpha_p^2 - k^2)^{1/2}, & \text{if } \alpha_p^2 \geq k^2 \quad (3) \\ -i\ (k^2 - \alpha_p^2)^{1/2}, & \text{if } \alpha_p^2 < k^2 \quad (4) \end{cases}$$

Due to the periodic nature of the geometry, the Floquet theorem predicts that the scattered fields have a discrete spectrum. In the three regions $x \leq 0$, $0 \leq x \leq h$, and $x \geq h$ they can be represented by

$$H_y = \begin{cases} \sum_{p=-\infty}^{\infty} A_p e^{i\alpha_p z} e^{\beta_p x}, & \text{for } x \leq 0 \quad (5) \\ \sum_{n=0}^{\infty} \left(B_n e^{-\gamma_n x} + C_n e^{\gamma_n x}\right) \cos\left(\frac{n\pi z}{a}\right), & \text{for } 0 \leq x \leq h \quad (6) \\ \sum_{p=-\infty}^{\infty} D_p e^{i\alpha_p z} e^{-\beta_p (x-h)}, & \text{for } x \geq h \quad (7) \end{cases}$$

where

$$\gamma_n = \begin{cases} +\ [(n\pi/a)^2 - k^2]^{1/2}, & \text{if } (n\pi/a)^2 \geq k^2 \\ -i\ [k^2 - (n\pi/a)^2]^{1/2}, & \text{if } (n\pi/a)^2 < k^2 \end{cases} \tag{8}$$

Our main interest is to determine the reflected and transmitted fields, i.e., $\{A_p\}$ and $\{D_p\}$.

We begin the solution by deriving a set of linear equations for $\{A_p\}$ and $\{D_p\}$ obtained by enforcing the continuity conditions of the tangential field components at $x = 0$ and $x = h$. Fourier transformation of both sets of matching equations and straightforward algebraic manipulation results in

$$\sum_{p=-\infty}^{\infty} \left(\frac{1}{\beta_p - \gamma_n} \pm \frac{e^{-\gamma_n h}}{\beta_p + \gamma_n}\right) U_p^{\pm} - \left(\frac{1}{\beta_0 + \gamma_n} + \frac{e^{-\gamma_n h}}{\beta_0 - \gamma_n}\right) = 0, \quad n = 0, 1, 2, \ldots, \tag{9}$$

where the symmetrical and asymmetrical components U_p^+ and U_p^- are given by

$$(\alpha_p/\alpha_0)(A_p \pm D_p) = U_p^{\pm}. \tag{10}$$

III. SOLUTION BY THE MODIFIED RESIDUE CALCULUS METHOD

For the special case $h \to \infty$, corresponding to an array of semi-infinite plates, (9) can be solved by the conventional residue method, and such a solution has been carried out by Berz.[2] However, for finite h one has to modify (see Ref. 6) the conventional residue calculus method, as is to be detailed below.

The central step in the solution of (9) by the residue calculus method is the construction of two meromorphic functions of a complex variable w with specific pole-zero configurations. The poles and zeros are chosen such that contour integrals about a circle of infinite radius in the complex plane generate residue series which are identical to the left-hand side of (9).

Consider the integrals

$$\frac{1}{2\pi i}\oint_C\left(\frac{1}{w-\gamma_n}+\frac{e^{-\gamma_n h}}{w+\gamma_n}\right)f^+(w)dw, \tag{11}$$

$$\frac{1}{2\pi i}\oint_C\left(\frac{1}{w-\gamma_n}-\frac{e^{-\gamma_n h}}{w+\gamma_n}\right)f^-(w)dw. \tag{12}$$

Here $f^+(w)$ and $f^-(w)$ are the functions mentioned in the paragraph above, and C is a circle of infinite radius in the complex w plane. The functions $f^+(w)$ and $f^-(w)$ are required to satisfy the following conditions:

(1) $f^+(w)$ and $f^-(w)$ are analytic everywhere except for simple poles at $w=-\beta_0$ and $w=\{\beta_p\}$ for $p = 0, \pm 1, \pm 2, \cdots$.

(2) $f^\pm(w)$ has zeros at $\{\Gamma_n^\pm\}$, which are yet unknown and will be determined from the condition

$$f^\pm(\gamma_n) \pm e^{-\gamma_n h} f^\pm(-\gamma_n) = 0 \tag{13}$$

for $n = 0, 1, 2, \cdots$.

(3) $f^{+1}(w)$ and $f^{-1}(w)$ have algebraic behavior, explicitly $O(w^{-1/2})$, as $|w| \to \infty$. This is to satisfy the edge condition.[6]

(4) The residues of $f^+(w)$ and $f^-(w)$ at $w=-\beta_0$, denoted by $\mathrm{Res}f^+(-\beta_0)$ and $\mathrm{Res}f^-(-\beta_0)$, respectively, equal one.

The integrals (11) and (12) are identically zero by property (3). Evaluation of the integral (11) leads to

$$\sum_{p=-\infty}^{\infty}\left(\frac{1}{\beta_p\mp\gamma_n}\pm\frac{e^{-\gamma_n h}}{\beta_p+\gamma_n}\right)\mathrm{Res}f^\pm(\beta_p) + \left(\frac{-1}{\beta_0+\gamma_n}\pm\frac{e^{-\gamma_n h}}{-\beta_0+\gamma_n}\right)\mathrm{Res}f^\pm(-\beta_0) + [f^\pm(\gamma_n)\pm e^{-\gamma_n h}f^\pm(-\gamma_n)] = 0. \tag{14}$$

Clearly, if properties (2) and (4) are satisfied, then the residue series is identical to the left-hand side of (9). It follows immediately that

$$U_p^\pm = \mathrm{Res}f^\pm(\beta_p). \tag{15}$$

In the special case $h\to\infty$, (13) becomes

$$f^\pm(\gamma_n) = 0, \tag{16}$$

which means that $f^\pm(w)$ has zeros at $w=\gamma_n$ for $n = 0, 1, 2, \cdots$. For finite h, the zeros are shifted from $\{\gamma_n\}$ to $\{\Gamma_n^+\}$ and $\{\Gamma_n^-\}$ in order to satisfy (13). It is important to recall the fact that the edge condition remains the same whether h is finite or infinite. This requires $\{\Gamma_n^\pm\}$ to asymptotically coincide with $\{\gamma_n\}$; that is

$$\lim_{n\to\infty}\Gamma_n^\pm = \gamma_n. \tag{17}$$

It can be shown[6] that (17) is consistent with (13). The actual construction of $f^\pm(w)$ follows closely that in Ref. 4, 6–8 and only the result need be stated:

$$f^\pm(w) = \exp\left(\frac{-(w+\beta_0)a}{\pi}\ln 2\right)\frac{2}{(w+\beta_0)(1-w/\beta_0)} \times \frac{(1-w/\Gamma_0^\pm)}{(1+\beta_0/\Gamma_0^\pm)}\prod_{n=1}^{\infty}\frac{(1+\beta_0/\beta_n)(1+\beta_0/\beta_{-n})(1-w/\Gamma_n^\pm)}{(1-w/\beta_n)(1-w/\beta_{-n})(1+\beta_0/\Gamma_n^\pm)}. \tag{18}$$

It remains to evaluate the shifted zeros $\{\Gamma_n^\pm\}$ using (13).

It was concluded from the edge condition that for large n, the shifted zeros $\{\Gamma_n^\pm\}$ coalesce with the unshifted zeros. Then there exists an integer $M>0$ such that $|\Gamma_n^\pm-\gamma_n|$ is arbitrarily small for all $n>M$.

Thus, we need to solve only the first $(M+1)$ equations of (13) for $\{\Gamma_n^\pm\}$ while setting

$$\Gamma_n^\pm = \gamma_n, \quad \text{for } n>M. \tag{19}$$

Substituting (18) and (19) into (13) one has

$$\left.\frac{\phi^\pm(w)}{\phi^\pm(-w)}\right|_{w=\gamma_m} = t_m^\pm, \quad m=0,1,2,\ldots,M, \tag{20}$$

where $\phi^+(w)$ and $\phi^-(w)$ are polynomials in w with unknown coefficients

$$\phi^\pm(w) = \prod_{n=0}^{M}\left(1-\frac{w}{\Gamma_n^\pm}\right) = 1+\sum_{n=1}^{M}F_n^\pm w^n \tag{21}$$

and $\{t_m^+\}$ and $\{t_m^-\}$ are known constants and given explicitly by

$$t_m^+ = -t_m^- = -\exp\left(-\gamma_m\frac{(h-2a\ln 2)}{\pi}\right) \times \prod_{n=1}^{M}\frac{(1-\gamma_m/\beta_n)(1-\gamma_m/\beta_n)}{(1+\gamma_m/\beta_n)(1+\gamma_m/\beta_n)} \times \prod_{n=M+1}^{\infty}\frac{(1-\gamma_m/\beta_n)(1-\gamma_m/\beta_{-n})(1+\gamma_m/\gamma_n)}{(1+\gamma_m/\beta_n)(1+\gamma_m/\beta_{-n})(1-\gamma_m/\gamma_n)}. \tag{22}$$

Note that (20) is a set of $(M+1)$ linear equations for unknowns $\{F_n^+\}$; similarly for $\{F_n^-\}$. Once the $\{F_n^\pm\}$ are determined, the shifted zeros $\{\Gamma_n^\pm\}$ follow immediately from (21). (In actual computations, we need only $\phi^\pm(w)$; hence, it is not necessary to find $\{\Gamma_n^\pm\}$ explicitly.)

The final step in the solution is to calculate the residues in (15). The result is

$$U_0^\pm = -\exp\left(\frac{-2\beta_0 a\ln 2}{\pi}\right)\prod_{n=1}^{M}\frac{(1+\beta_0/\beta_n)(1+\beta_0/\beta_{-n})}{(1-\beta_0/\beta_n)(1-\beta_0/\beta_{-n})}\frac{\left(1+\sum_{n=1}^{M}F_n^\pm\beta_0^n\right)}{\left(1+\sum_{n=1}^{M}F_n^\pm(-\beta_0)^n\right)}\prod_{n=M+1}^{\infty}\frac{(1-\beta_0/\gamma_n)(1+\beta_0/\beta_n)(1+\beta_0/\beta_{-n})}{(1+\beta_0/\gamma_n)(1-\beta_0/\beta_n)(1-\beta_0/\beta_{-n})}, \tag{23}$$

$$U_p^{\pm} = \left[\exp\left(-\frac{(\beta_p+\beta_0)a\ln 2}{\pi}\right)\right]\frac{2}{(\beta_p+\beta_0)(1-\beta_p/\beta_0)}\,\frac{(1+\beta_0/\beta_p)(1+\beta_0/\beta_{-p})}{(-1/\beta_p)(1-\beta_p/\beta_{-p})}$$

$$\times\prod_{n=1}^{M}{}^{(|p|)}\frac{(1+\beta_0/\beta_n)(1+\beta_0/\beta_{-n})}{(1-\beta_p/\beta_n)(1-\beta_p/\beta_{-n})}\,\frac{\left(1+\sum_{n=1}^{M}F_n^{\pm}\beta_0^n\right)}{\left(1+\sum_{n=1}^{M}F_n^{\pm}(-\beta_0)^n\right)}\prod_{n=M+1}^{\infty}\frac{(1-\beta_p/\gamma_n)(1+\beta_0/\beta_n)(1+\beta_0/\beta_{-n})}{(1+\beta_0/\gamma_n)(1-\beta_p/\beta_n)(1-\beta_p/\beta_{-n})}$$

$$\text{for } p=\pm 1,\pm 2,\ldots,\quad |p|<M. \tag{24}$$

The notation $\Pi^{(|p|)}$ indicates that the pth term is deleted. The reflection coefficients $\{A_p\}$ and transmission coefficients $\{D_p\}$ are related to $\{U_p^{\pm}\}$ by (10). This completes the analytical solution to the problem.

IV. NUMERICAL RESULTS AND DISCUSSION

Machine computations for the coefficients of the propagating modes are carried out for parameters a and h with ranges $0 < a < \lambda$ and $0 < h \leq \lambda$. For these ranges there can exist, at most, two reflected and two transmitted beams exterior of the array. For computation, the infinite product terms were truncated at 200, which was found to be more than adequate. The question of how many zeros to shift was answered using two criteria. The first was convergence of the coefficients A_p and D_p as the number of zeros shifted was increased. Second was that of power check; namely,

$$p^{(i)} = p^{(r)} + p^{(t)}. \tag{25}$$

Calculating the values of $p^{(r)}$ and $p^{(t)}$, and normalizing to $p^{(i)}$ gives (25) to be

$$1 = \sum_p\left[(|A_p|^2 + |D_p|^2)\left(\frac{|\beta_p|}{|\beta_0|}\right)\right], \tag{26}$$

where the index p is for the propagating modes only.

Table I shows convergence of the coefficients A_{-1} and D_{-1} for $\theta_0 = 75$ degrees. The coefficients A_0 and D_0 were found to converge slightly faster in most cases, so that Table I represents a worse case. Generally speaking, a shift of eight or nine zeros is sufficient. Power checks were consistently good for six or more zeros shifted. The nine zeros shifted case in Table I represents a power check of better than 0.1%.

Figure 2 presents the dominant beam coefficient for a closely spaced array with $a = 0.25\lambda$, and $h = 1\lambda$. It is of interest to note that in the limiting case $h \to \infty$, the magnitude of the reflection coefficient is given by

$$|A_0| = (1 - \cos\theta_0)/(1 + \cos\theta_0) \tag{27}$$

provided $a < 0.50\lambda$, which is plotted in Fig. 2 by a dotted line. The difference between the cases with finite h and infinite h is very small when the incident angle θ_0 is close to broadside. This is due to the fact that the junctions at $x = 0$ and $x = h$ in Fig. 1 are nearly transparent for small $|\theta_0|$.

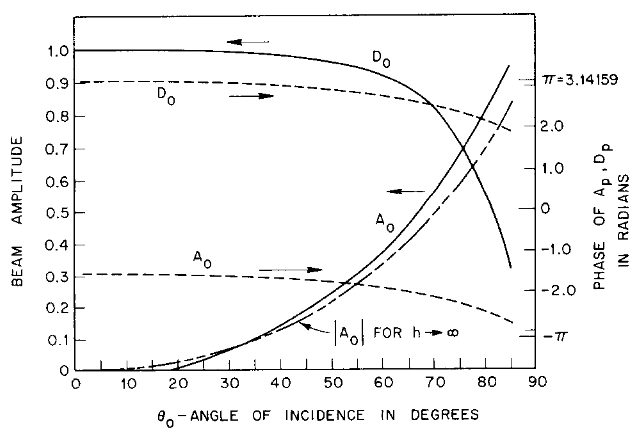


FIG. 2. Beam amplitude and phase vs incident angle for $h = 1\lambda$ and $a = 0.25\lambda$ (— — — — PHASE ——— MAGNITUDE).

TABLE I. Convergence of the coefficients A_{-1} and D_{-1}.[a]

Number of Zeros Shifted M	A_{-1}	D_{-1}
1	0.0691	0.1989
2	0.0867	0.2121
3	0.0983	0.2162
4	0.1039	0.2185
5	0.1071	0.2198
6	0.1097	0.2208
7	0.1101	0.2211
8	0.1103	0.2213
9	0.1103	0.2214

[a] The parameters used for this computation are $a = 0.75\lambda$, $h = 1\lambda$, and $\theta_0 = 75°$.

In Fig. 3, we use $h = 0.5\lambda$ and $a = 0.75\lambda$, which allows the (− 1) beam to propagate beginning at the angle $\theta_0 = 19.5°$. Note that at the grazing angle of the (− 1) beam, the scattered modal coefficients exhibit discontinuous derivatives, as expected. The transmitted power in the main beam drops from 100% at $\theta_0 = 0°$ to 73% at $\theta_0 = 60°$, while the transmitted power in the (− 1) beam is roughly 10% for $30° < \theta_0 < 60°$.

To demonstrate the dependence of h, we keep $a = 0.75\lambda$ and incerase h to 1λ. The result, presented in Fig. 4, is quite different from that in Fig. 3. A particularly interesting phenomenon occurs at $\theta_0 = 13.0°$. At this incident angle, $|A_0| = 1$ and $|D_0| = 0$ with the phase of D_0 undergoing a jump of π, which implies a *total reflection*! Thus, if the structure is used as a microwave lens with $a > 0.50\lambda$, a "blind spot" for the transmission is observed in the angular region where no propagating diffracted beams exist.

The total reflection phenomenon discussed above suggests the existence of a modal solution to the open structure formed by adding an identical set of strips at a distance $x = -(h + l)$ in Fig. 1. A direct computation of the modal solution to this structure is very difficult. However, by making use of the solution in the present paper, a simple and elegant solution can be described. Consider a plane wave traveling in the direction θ_c with field given by

$$H_y^{(1)} = \exp[ik(x\cos\theta_c + z\sin\theta_c)]. \tag{28}$$

The reflected field from the junction at $x = 0$ may be represented by

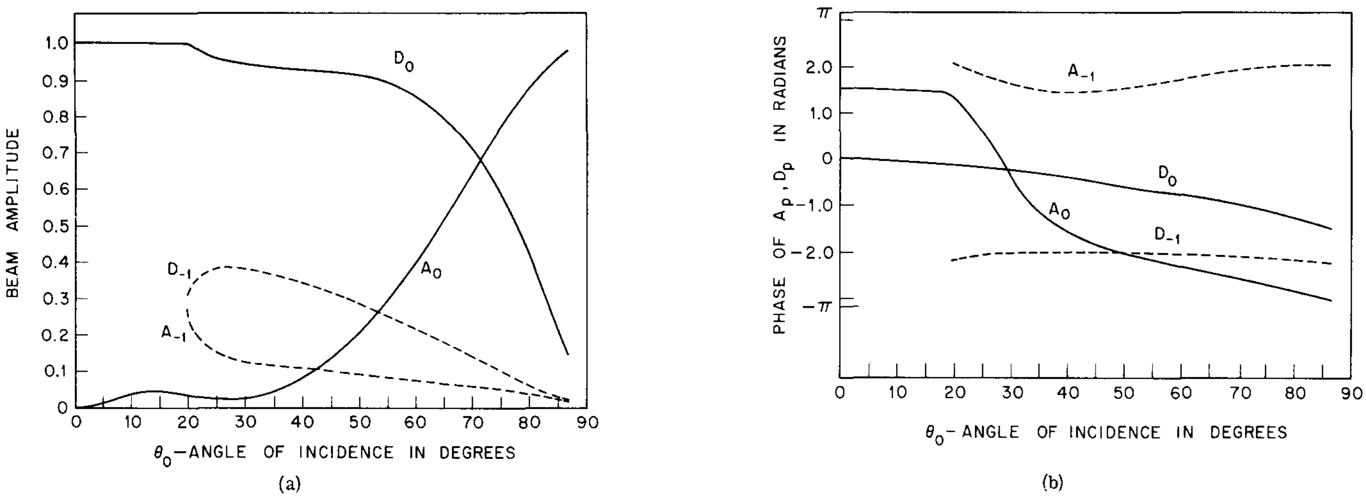


FIG. 3(a) Beam amplitude vs incident angle for $h = .5\lambda$ and $a = .75\lambda$; (b) Phase vs incident angle for $h = 0.5\lambda$ and $a = 0.75\lambda$.

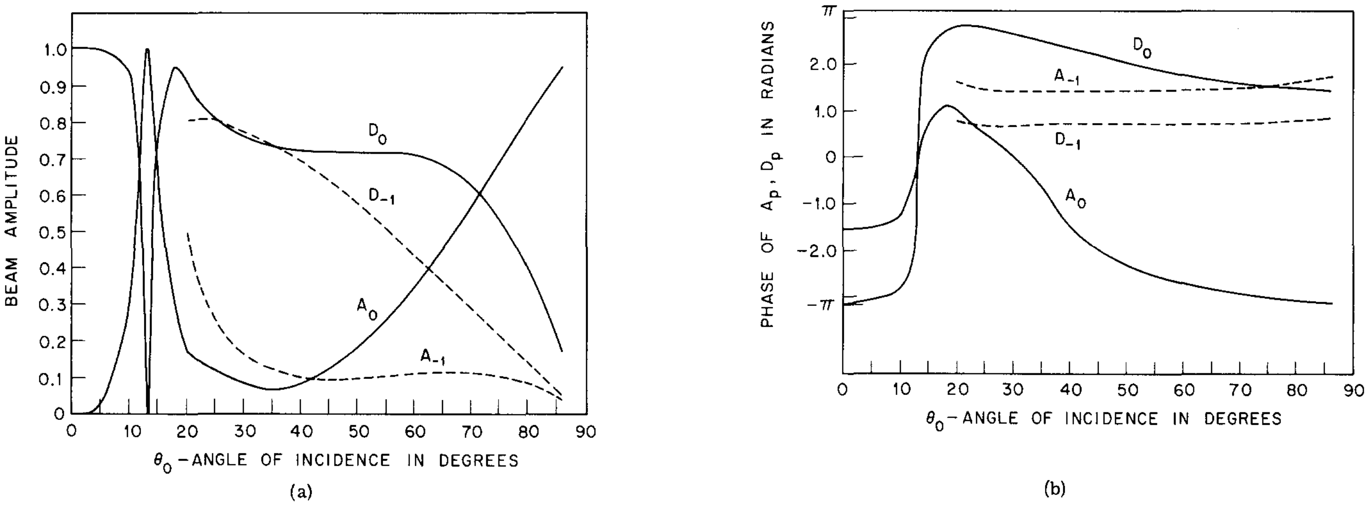


FIG. 4(a) Beam amplitude vs incident angle for $h = 1\lambda$ and $a = 0.75\lambda$; (b) Phase vs incident angle for $h = 1\lambda$ and $a = 0.75\lambda$

$$H_y^{(2)} = A_0 \exp[ik(-x\cos\theta_c + z\sin\theta_c)] + (\text{higher-order space harmonics}). \quad (29)$$

Since we are interested in the interaction between the junctions at $x = 0$ and $x = -l$, the attenuating (assuming θ_c is less than the grazing angle of next space harmonic) higher-order space harmonics may be neglected provided that kl is large. The reflected field in (29) is scattered again at $x = -l$ and yields a scattered field

$$H_y^{(3)} = A_0^2 \exp(i2kl\cos\theta_c)\exp[ik(x\cos\theta_c + z\sin\theta_c]. \quad (30)$$

Now, let us assume that θ_c is the critical angle for the occurrence of total reflection. Thus $|A_0| = 1$, or

$$A_0 = e^{i\xi}, \quad (31)$$

where ξ is the phase angle of A_0, and may be computed from the solution presented in this paper.

The condition for a self-consistent modal solution is that $H_y^{(1)} = H_y^{(3)}$ or by making use of (31) in (30) and (29):

$$\exp[i(2\xi + 2kl\cos\theta_c)] = \exp(\pm i2n), \quad n = 0, 1, 2, \cdots.$$

Solving for kl gives

$$kl = (\pm 2n\pi - \xi)/\cos\theta_c, \quad n = 0, 1, 2, \cdots. \quad (32)$$

Thus, for a given kh and ka, there may exist a total reflection angle θ_c. In such a case, one may use (32) to determine kl for the existence of a modal solution of the type described above. The modal field variation can be obtained by combining (28) and (29) and the result is

$$H_y \sim \cos kx \cos\theta_c - \tfrac{1}{2}\xi e^{ikz\sin\theta_c}, \quad \text{for } -l < x < 0. \quad (33)$$

If $\sin\theta_c < 1$ (or θ_c real), the modal field is a fast wave propagating along the z direction. If $\sin\theta_c > 1$ (or θ_c imaginary), the modal field is a slow wave, and its transverse variation is no longer oscillatory. The modal field outside the two sets of parallel strips, i.e., $x < -(h + l)$ or $x > h$, is of evanescent nature since the only propagating beam has an amplitude $D_0 = 0$. Thus, the energy carried in this mode is entirely confined between the two sets of parallel strips. This suggests the possibility of using this structure as a resonator. To form a resonator, we may close the gap between two sets of plates at $z = 0$, and $z = d$ such that[9]

$$kd\sin\theta_c = n\pi, \quad n = 1, 2, 3, \cdots. \quad (34)$$

For a ray impinging on the sides at $x = 0$ and $x = -l$ with an angle very close to θ_c, it is nearly totally reflected with very small diffraction loss. Suppose a source or an active medium is placed in the resonator; a steady oscillation will result when the diffraction loss is balanced by the gain from the source or medium. The apparent advantage of such a resonator over a conventional open resonator[10,11] formed by two plates (at $z = 0$ and $z = d$), lies in the fact that the two sets of strips at $x = 0$ and $x = -l$ perform a further "filtering" function for the field in the resonator, and, consequently, the resonance frequencies should be even more sparse. However, we emphasize that this resonator is only a preliminary idea, and its merit can be ascertained only after more quantitative analysis and experimental work.

[1] J. F. Carlson and A. E. Heins, Q. Appl. Math. **4**, 313 (1947).
[2] F. Berz, Proc. IEEE **98** (Pt. III), 47 (1951).
[3] See, e.g., R. E. Collins, *Field theory of guided waves* (McGraw-Hill, New York, 1961).
[4] E. Meister, Z. Angew. Math. Mech. **55**, T57 (1965).
[5] D. S. Jones, Proc. R. Soc. A **217**, 154 (1953).
[6] R. Mittra and S. W. Lee, *Analytical techniques in the theory of guided waves* (Macmillan, New York, 1971).
[7] R. Mittra and S. W. Lee, J. Math. Phys. (N.Y.) **11**, 775 (1970).
[8] R. Mittra, S. W. Lee, and G. F. Van Blaricum, Int. J. Eng. Sci. **6**, 395 (1968).
[9] More precisely, $kd \sin \theta_c = n\pi + p$, where $|p| \ll 1$. The real and imaginary parts of p are related to phase shift and loss factors of the resonator respectively. See Refs. 10 and 11.
[10] L. A. Vainshtein, Zh. Eksp. Teor. Fiz. **44**, 1050 (1963) [Sov. Phys. JETP **17**, 709 (1963)].
[11] L. A. Weinstein, *The theory of diffraction and factorization method* (Golem, Boulder, Colo., 1969).

The inverse scattering problem at fixed energy for L^2-dependent potentials*

M. A. Hooshyar†
Physics Department, Pahlavi University, Shiraz, Iran
(Received 19 April 1972)

The problem of finding central and L^2-dependent potentials, acting among spinless particles, from the knowledge of the S matrix as a function of angular momentum at a fixed energy is studied. The Newton method for central potentials is generalized to this case, and it is shown that phase shift information at fixed energy is not enough to give us both the central and the L^2-dependent potential.

1. INTRODUCTION

A systematic method for the construction of the scattering potential from the knowledge of the S matrix at one energy as a function of the angular momentum (inverse scattering problem at fixed energy) has not only direct physical significance; but it also gives us a better understanding of what kind of information on the potential we can obtain from scattering experiments. The first results for the inverse scattering problem at fixed energy for central potentials were reported by Newton,[1] and this work was then extended by Sabatier.[2] The inverse scattering problem for $\mathbf{L}\cdot\mathbf{S}$ potentials has been considered by Sabatier[3] and for the tensor force by Hooshyar,[4] but some work still remains for completely solving these two inverse problems. On observing the similarities between the methods used for attacking the mentioned inverse scattering problems, one may wonder if a similar method can be used for solving the inverse scattering problem at fixed energy for $\mathbf{L}^2$-dependent potentials. To make the problem simple, in this work we only consider the inverse problem for $\mathbf{L}^2$-dependent potentials for spinless particles. In other words, the radial Schrödinger equation which we would like to consider should have the following form:

$$r^2\left(\frac{d^2}{dr^2} + k^2 - V_c(r) - l(l+1)V_2(r)\right)\Phi_l(r) = l(l+1)\Phi_l(r) \quad (1.1)$$

where $V_c(r)$ is the central potential, $V_2(r)$ is associated with the radial-dependent part of the $\mathbf{L}^2$-dependent potential, and Φ_l is the regular solution to Eq. (1.1). The problem which we are interested in solving is to see what can be said about V_c and V_2 if the asymptotic behavior of Φ_l is given for all values of the angular momentum at a fixed energy. We would like to point out that we are not interested in this problem only because of our mathematical curiosity, but also because the Schrödinger equation for tensor force can be put into a matrix form related to the above equation, and one needs to know how to solve the above inverse scattering problem if one wishes to consider the inverse scattering problem at fixed energy for Nilsson potentials.[5]

Section 2 is devoted to the problem of connecting the phase shifts to the potentials V_c and V_2 through an auxiliary function which is the solution to the analog of the Regge–Newton equation.[1,6] For finding this analog of the Regge–Newton equation, we found it to be more convenient to rewrite Eq. (1.1) in the following form:

$$r^2\left(\frac{d^2}{dr^2} + k^2 - V_1(r) - \lambda^2 V_2(r)\right)\Psi_\lambda(r) = \left(\lambda^2 - \frac{1}{4}\right)\Psi_\lambda(r) \quad (1.2)$$

where $\lambda = l + \frac{1}{2}$, $V_1 = V_c - \frac{1}{4}V_2$, and $\Psi_\lambda = \Phi_l$. Also we noticed that the analog of the Regge–Newton equation that one may obtained for Eq. (1.2), is such that it can be used most conveniently if we assume that V_2 is an arbitrary but known potential. In this work we have not considered what one can find from our analog of the Regge–Newton equation if the above assumption is not made.

In Sec. 3 we give an example to demonstrate the method for the construction of V_1 in terms of V_2 and the phase shifts at a fixed energy.

2. THE PROCEDURE

In this section we shall develop the analog of Newton's method[1] for the Schrödinger equation where an $l(l+1)$-dependent potential is also present, that is, Eq. (1.1).

As it was stated in Sec. 1, the potential $V_1(r)$ is the potential which we are to find from the information on the phase shifts, and potential $V_2(r)$ is assumed to be arbitrary but given. In order that our procedure work, we also need the following condition to be satisfied:

$$1 + r^2V_2(r) \geq 0, \quad \text{for } r \geq 0,$$
$$\int_0^\infty dr\left|\frac{[1 + r^2V_2(r)]^{1/2} - 1}{r}\right| < \infty, \quad (2.1)$$

and $z^2V_2(z)$ should be an entire function of z.

With above conditions satisfied the following functions can be defined[7]:

$$a(r) = \int_0^r ds\,\frac{[1 + s^2V_2(s)]^{1/2} - 1}{s}, \quad b(r) = r\exp[a(r)],$$
$$F(r) = [\dot{b}(r)]^{-1/2} \quad \text{and} \quad c = \lim_{r\to\infty} b(r)/r. \quad (2.2)$$

Following the method developed in Ref. 1, we consider the spherical Riccati–Bessel functions $U_\lambda(k_1r) \equiv u_\lambda(r)$, which are solution of the following differential equation:

$$r^2\left(\frac{d^2}{dr^2} + k_1^2\right)u_\lambda(r) = (\lambda^2 - \tfrac{1}{4})u_\lambda(r), \quad \text{with } k_1 = \frac{k}{c}. \quad (2.3)$$

Next we define the input function $f(r, r')$ as

$$f(r, r') = \sum_{\lambda\in S} u_\lambda(r)d_\lambda u_\lambda(r'), \quad (2.4)$$

where the constants d_λ are arbitrary for now and are to be found later from information on phase shifts, and for our purpose, the set S is assumed to contain only the half-integers.

The analog of the Regge–Newton equation is then defined by

$$K(r, r') = F(r)f(b(r), r') - \int_0^{b(r)} ds\, s^{-2}K(r, s)f(s, r'). \quad (2.5)$$

Using the standard arguments[2] concerning the Fredholm determinants of Eq. (2.5), $\Delta(z)$, we notice that $\Delta(z)$ is analytic in the domain of analyticity of $b(z)$ and the zeros of $\Delta(z)$ are poles of $K(z, z')$. It then follows that in the domain of analyticity of $K(z, z')$, $\Delta(z) \neq 0$, (2.5) has a unique solution and the homogeneous version of Eq. (2.5) has only the trivial solution. This fact, together with the performance of some tedious differentiation and integration by parts, enables one to come to the conclusion that

$$r^2\left(\frac{d^2}{dr^2} + k^2 - V_1(r) - \tfrac{1}{4}V_2(r)\right) K(r,r') = [1 + r^2V_2(r)]r'^2\left(\frac{d^2}{dr'^2} + k_1^2\right)K(r,r') \quad (2.6)$$

if

$$V_1(r) = k^2 - \dot{b}k_1^2 - \tfrac{1}{4}V_2 + F^{-1}\left(\ddot{F} - \frac{d}{dr}[\dot{b}b^{-2}K(r,b(r))] - \dot{b}b^{-2}\frac{d}{dr}K(r,b(r))\right). \quad (2.7)$$

Choosing Eq. (2.7) as the definition of $V_1(r)$ in Eq. (1.2), using Eq. (2.6) and performing some differentiation and integration by parts, we can show that the regular solution $\Psi_\lambda(r)$ of Eq. (1.2) can be written as

$$\Psi_\lambda(r) = F(r)u_\lambda[b(r)] - \int_0^{b(r)} ds\, s^{-2}K(r,s)u_\lambda(s). \quad (2.8)$$

Since we have assumed that the asymptotic behavior of Ψ_λ is known (that is, the phase shifts), then it follows that $K(r, s)$ in Eq. (2.8) should have been so chosen so that it can give us the correct behavior for Ψ_λ. In other words, Eq. (2.8) should be used in such a way that it gives us the set d_λ which corresponds to the desired asymptotic behavior of Ψ_λ. In order to achieve this, let us substitute (2.4) in Eq. (2.5) and make use of (2.8) in order to get

$$K(r,r') = \sum_{\lambda\in S} \Psi_\lambda(r)d_\lambda u_\lambda(r'). \quad (2.9)$$

Substituting Eq. (2.9) in Eq. (2.8) gives us the desired relation

$$\Psi_\lambda(r) = F(r)u_\lambda[b(r)] - \sum_{\lambda\in S} \Psi_\gamma(r)d_\gamma L_{\lambda,\gamma}(r)$$

with $\quad (2.10)$

$$L_{\lambda,\gamma}(r) = \int_0^{b(r)} ds\, s^{-2}u_\lambda(s)u_\gamma(s).$$

We are interested in the asymptotic form of Eq. (2.10), and because of uniform bounds of u_λ and Ψ_λ, which are shown by Sabatier[8] to be of order of $\lambda^{1/3}$, we see that we can take the limit as $r \to \infty$ inside the summation in Eq. (2.10) if $|d_\lambda| < C\lambda^{1/3-\epsilon}$, where C is some constant. With this assumption realized we let r tend to infinity in (2.10). Using the following asymptotic forms

$$\begin{aligned} &\lim_{r\to\infty} \Psi_\lambda(r) = A_\lambda \sin[kr - \tfrac{1}{2}\pi(\lambda - \tfrac{1}{2}) + \delta_\lambda], \\ &\lim_{r\to\infty} u_\lambda(r) = \sin[k_1r - \tfrac{1}{2}\pi(\lambda - \tfrac{1}{2})], \quad (2.11) \\ &\lim_{r\to\infty} b(r) = \lim_{r\to\infty} cr, \lim_{r\to\infty} F(r) = c^{-1/2}, \end{aligned}$$

writing the sine function in terms of exponentials, separating the coefficients of e^{ikr} and e^{-ikr}, and equating the coefficients separately, we find that

$$A'_\lambda e^{i\delta_\lambda} = 1 - \sum_{\gamma\in S} L'_{\lambda,\gamma}(\infty)d'_\gamma A'_\gamma e^{i\delta_\gamma}e^{i(\pi/2)(\lambda-\gamma)}, \quad (2.12)$$

where

$$d'_\lambda = k_1 d_\lambda, \quad A'_\lambda = c^{1/2}A_\lambda$$

and

$$L'_{\lambda,\gamma}(\infty) = \sin[(\gamma-\lambda)(\pi/2)]/(\gamma^2 - \lambda^2).$$

Equation (2.12) is identical with Eq. (17) of Ref. 1 and the question of finding the d_λ from the phase shifts, via Eq. (2.12), becomes identical with the similar problem which arises when one tries to do the inverse problem for only the central potentials, with no $l(l+1)$-dependent potentials. This problem has been considered in detail by Newton[1] and Sabatier,[9] and they have shown that, in general, one can find sets of d_λ corresponding to a set of phase shifts, if the phase shifts tend to zero sufficiently rapidly with increasing value of the angular momentum.

Having found a set of d_λ corresponding to a given set of phase shifts, we can define the input function for (2.5) from Eq. (2.4). From Eq. (2.5) we can then find $K(r,r')$, from which, using Eq. (2.7), we can find the potential $V_1(r)$. Clearly the construction of $V_1(r)$ then assures us that this potential, $V_1(r)$, together with $V_2(r)$, is such that the wavefunction $\Psi_\lambda(r)$ will have the desired asymptotic behavior.

3. EXAMPLE

For the purpose of illustrating the method let us assume we are given a set of phase shifts such that the corresponding d_λ are found to be

$$d_\lambda = 0 \quad \text{for all} \quad \lambda \neq \gamma, \qquad d_\gamma \neq 0, \quad (3.1)$$

Since V_2 is arbitrary in this case, let us choose it to be

$$V_2(r) = V^2(r) + 2V(r)/r, \quad (3.2)$$

where

$$V(r) = we^{-r}$$

and w is some constant. Clearly V_2 will satisfy conditions of Eq. (2.1), and the defined functions a, b, c and F are then given as

$$a(r) = w - V(r), \quad b(r) = cre^{-V(r)}, \quad c = e^w, \quad \text{and } F(r) = b^{-1/2}[V(r) + 1/r]^{-1/2}. \quad (3.3)$$

From Eq. (2.10) it follows that

$$\Psi_\gamma(r) = F(r)u_\gamma[b(r)]/[1 + d_\gamma L_{\gamma,\gamma}(r)]. \quad (3.4)$$

It then follows that

$$\Psi_\lambda(r) = F(r)\left[u_\lambda[b(r)] - \frac{u_\gamma[b(r)]d_\gamma L_{\lambda,\gamma}(r)}{1 + d_\gamma L_{\gamma,\gamma}(r)}\right]. \quad (3.5)$$

Substitution of Eq. (3.4) in Eq. (2.9) implies that

$$K(r,r') = F(r)u_\gamma(r)d_\gamma u_\gamma(r')/[1 + d_\gamma L_{\gamma,\gamma}(r)]. \quad (3.6)$$

Knowing $K(r,r')$, one can find potential $V_1(r)$ from Eq. (2.7):

$$\begin{aligned} V_1(r) = {} & [1 - e^{-2V}(1 + rV)^2]k^2 - \tfrac{1}{4}V_2 + \tfrac{1}{4}[r^2V_2^2 - r^2V_2 \\ & + 2V^2(r^2 - 1) + 6V]/(1 + r^2V_2) \\ & - 2e^V(1 + rV)[-\dot{G}(r) \\ & + (V + 1/r)G(r)]/cr^2, \end{aligned} \quad (3.7)$$

where

$$G(r) = u_\gamma^2[b(r)]d_\gamma/[1 + d_\gamma L_{\gamma,\gamma}(r)].$$

From our construction procedure it follows that Ψ_λ, given by Eq. (3.5), is the regular solution to Eq. (1.2) if the potential V_1 and V_2 in that equation are given by Eqs (3.7) and (3.2). If we let r tend to infinity in Eq. (3.5), we find that the phase shifts associated with the potentials V_1 and V_2 have the following form:

$$\tan(\delta_\lambda) = 0 \quad \text{if } |\lambda - \gamma| \text{ is even,}$$
$$\tan(\delta_\lambda) = \frac{k_1 d_\gamma}{(1 + \pi k_1 d_\gamma/4\gamma)(\gamma^2 - \lambda^2)} \text{ if } |\lambda - \gamma| \text{ is odd.} \tag{3.8}$$

One should notice that when $V_2 = 0$, $k_1 = k$, and, if we choose units such that $k = 1$, then the example which we have considered is the same as the one in Ref. 1. As expected, the results become the same.

From the construction procedure and the results of our example, we observe that in a scattering experiment in which Eq. (1.1) is the governing equation, we cannot deduce both the central potential V_1 and the $\mathbf{L}^2$-dependent potential V_2 from the information on phase shifts of all angular momenta. In other words, the information on the phase shifts gives us only a relation between the central and $\mathbf{L}^2$-dependent potentials. In this work we have been able to give a method for finding the central potential V_1 if the $\mathbf{L}^2$-dependent potential V_2 and the phase shifts at a fixed energy are given.

We also find it interesting to point out that for the $\mathbf{L}^2$-dependent potential in our example, Eq. (3.2), the phase shifts associated with this potential and V_1 tend to zero for large values of the angular momentum. This observation is not *a priori* obvious and suggests that the method could also be used to study the asymptotic behavior of phase shifts associated with $\mathbf{L}^2$-dependent potentials.

ACKNOWLEDGMENT

The author would like to thank Professor R. G. Newton for interesting discussions and for reading the manuscript.

---

*Part of this work was done while the author was at Indiana University, Bloomington, Indiana and was supported by the United States Army Research Office, Durham, North Carolina.

† Presently at the Physics Department, Arya-Mehr University of Technology, Tehran, Iran.

[1] R. G. Newton, J. Math. Phys. (N.Y.) **3**, 75 (1962).
[2] P. C. Sabatier, J. Math. Phys. (N.Y.) **8**, 905 (1967).
[3] P. C. Sabatier, J. Math. Phys. (N.Y.) **9**, 1241 (1968).
[4] M. A. Hooshyar, J. Math. Phys. (N.Y.) **12**, 2243 (1971).
[5] S. G. Nilsson, Kgl. Dan. Vidensk. Selsk. Mat.-Fys. Medd. **29**, 16 (1955).
[6] T. Regge, Nuovo Cimento **14**, 951 (1959).
[7] A dot on top of a function of a single variable indicates the derivative of that function.
[8] P. C. Sabatier, J. Math. Phys. (N.Y.) **7**, 2079 (1966).
[9] P. C. Sabatier, J. Math. Phys. (N.Y.) **7**, 1515 (1966).

Elastic general relativistic systems

Edward N. Glass*
Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221

Jeffrey Winicourt
Aerospace Research Laboratories, Wright-Patterson A.F.B., Ohio 45433
(Received 28 June 1972)

A theory of elastic deformations of general relativistic systems is presented. The theory is derived from a generalized Hooke's law. An important feature of this theory is that its classical limit corresponds to the classical elasticity theory of prestressed materials. A perturbation description of small deformations is developed and applied to the test body case. For the first time, the strain-curvature equation for an elastic test body interacting with a gravitational wave has been derived from a complete theory. The semi-classical work of Dyson, showing the interaction of a gravitational wave with the inhomogeneities of the shear modulus, is rederived and placed within the framework of general relativity. The theory presented is quite comprehensive in scope and applicable to fully relativistic situations such as the elastic behavior of neutron stars.

1. INTRODUCTION

Elastic phenomena in the relativistic domain of influence have recently come into prominence. The most notable of these is Weber's observation[1,2] of the elastic response of an aluminum cylinder to gravitational radiation. Along the same lines are investigations[3,4] of the excitation of the earth's and moon's vibrational modes by gravitational waves. These gravitational radiation experiments do not directly involve the relativistic properties of elastic bodies, but rather their interactions with relativistic fields. However, it has been argued[5] that the crusts of neutron stars are in elastic states. If so, they would manifest full relativistic elasticity in which the velocity of sound waves is comparable to the velocity of light.

Relativistic theories of elasticity split into two categories characterized by their formulation of Hooke's law:

(i) Rate of change of stress is proportional to rate of change of strain.[6,7]

(ii) Stress is proportional to strain.[8,9] In the first case, the rate of strain is formulated in terms of the derivative of the space–time metric along the four-dimensional hydrodynamical streamlines, but the concept of strain itself is not introduced. Modern terminology would classify such treatments as theories of *hypoelasticity*. In the second category, an auxiliary spatial metric is introduced. This metric describes the equilibrium separation of the streamlines. The concept of strain is formulated in terms of the difference between the equilibrium separation and the actual separation determined by the space–time metric. This description corresponds to classical elasticity theory.

We adopt formulation (ii) of Hooke's law as the basis of our approach. In particular, we base our treatment very closely upon the work of Rayner,[8] in which the auxiliary metric describes a state of rigid motion of the elastic body. The auxiliary metric is to be regarded as part of the thermodynamic specification of the equilibrium state of the body. An alternative approach[9] which has been taken introduces the auxiliary metric by imagining small portions of the body removed to a distant stress-free region where their natural state can be examined. The auxiliary metric is then defined in terms of the equilibrium separations in the natural state. The conceptual awkwardness of this approach is that the natural state may not be in the solid phase, but instead a liquid or even an expanding gas with no finite equilibrium configuration. Such is likely to be the case with the elastic material in a neutron star.

Rayner's statement of Hooke's law does not include the possibility of initial stresses in the equilibrium state of the body. Such stresses are likely to be important to the elastic properties of bodies of astronomical size. In Sec. 2, we extend Rayner's theory to conform with a generalized Hooke's law of the form:

(iii) Stress minus equilibrium stress is proportional to strain.

Our formulation of a general theory of elasticity is primarily for the purpose of a starting point to describe small elastic deformations. It is in this case that relativistic elasticity theory can be expected to be most meaningful and applicable. We develop a general perturbation theory in Sec. 3 based upon a rigid equilibrium space–time and an associated one-parameter family of space–times representing elastic motion. We will not apply the results of Sec. 3 to fully relativistic systems, such as neutron stars, here. Instead, as a check on these results, we proceed in Sec. 4 to derive the test body limit for elastic perturbations. In this limit, our results are in essential agreement with other works describing the interaction of elastic test bodies with gravitational waves. Furthermore, in the nonrelativistic limit we find agreement with the classical theory of elasticity of prestressed materials.[10-12] To simplify the discussion we will treat adiabatic motion only. Effects of damping may be included using the techniques of relativistic viscosity theory.

2. GENERAL THEORY

In order to formulate an energy–momentum tensor $\bar{T}_{\alpha\beta}$ appropriate to the description of an elastic body, we begin with the standard hydrodynamical description. (In this section, we use a "bar" over symbols representing physical quantities to facilitate later notation). The trajectories of the material particles trace out world lines with unit 4-velocity $\bar{u}^\alpha$

$$\bar{u}^\alpha \bar{u}_\alpha = 1, \tag{2.1}$$

so that the metric has the natural decomposition[13]

$$\bar{g}_{\alpha\beta} = \bar{u}_\alpha \bar{u}_\beta + \bar{\gamma}_{\alpha\beta}, \tag{2.2}$$

where the spatial part $\bar{\gamma}_{\alpha\beta}$ satisfies

$$\bar{\gamma}_{\alpha\beta}\bar{u}^\beta = 0. \tag{2.3}$$

The tensor $\bar{\gamma}_\alpha{}^\beta$,

$$\bar{\gamma}_\alpha{}^\beta \equiv \bar{g}^{\beta\sigma}\bar{\gamma}_{\alpha\sigma} = \delta_\alpha{}^\beta - \bar{u}_\alpha \bar{u}^\beta,$$

satisfies the idempotent relation

$$\bar{\gamma}_{\alpha}{}^{\beta}\bar{\gamma}_{\beta}{}^{\sigma} = \bar{\gamma}_{\alpha}{}^{\sigma}$$

and acts as a spatial projection operator on tensor fields. The energy-momentum tensor takes the usual form

$$\bar{T}_{\alpha\beta} = \bar{\rho}\,\bar{u}_{\alpha}\bar{u}_{\beta} - \bar{P}_{\alpha\beta}, \tag{2.4}$$

where the energy density $\bar{\rho}$ and the stress tensor $\bar{P}_{\alpha\beta}$ are measured in the local rest frames,

$$\bar{P}_{\alpha\beta}\bar{u}^{\beta} = 0. \tag{2.5}$$

Following Rayner,[8] we now introduce an auxiliary spatial metric $\underset{0}{\bar{\gamma}}_{\alpha\beta}$ which imposes a rigid structure on the three-dimensional manifold of trajectories. It satisfies the orthogonality condition

$$\underset{0}{\bar{\gamma}}_{\alpha\beta}\bar{u}^{\beta} = 0 \tag{2.6}$$

and has vanishing Lie derivative along the trajectories

$$\underset{\bar{u}}{\mathcal{L}}\,\underset{0}{\bar{\gamma}}_{\alpha\beta} = 0. \tag{2.7}$$

This auxiliary metric describes the equilibrium distances between neighboring streamlines. For physical purposes, we may adopt the point of view that in the infinite past the body was in an equilibrium state satisfying the conditions of Born rigidity with initial conditions

$$\bar{\gamma}_{\alpha\beta} \rightarrow \underset{0}{\bar{\gamma}}_{\alpha\beta}. \tag{2.8}$$

Note that the nondegenerate metric

$$\bar{g}^{*}_{\alpha\beta} = \bar{u}_{\alpha}\bar{u}_{\beta} + \underset{0}{\bar{\gamma}}_{\alpha\beta} \tag{2.9}$$

satisfies the conditions of rigid motion for all times. However, this latter metric, while of mathematical interest, cannot be interpreted as an equilibrium space-time metric (see Sec. 3).

The strain tensor $\bar{S}_{\alpha\beta}$ is defined by

$$\bar{S}_{\alpha\beta} := \tfrac{1}{2}(\bar{\gamma}_{\alpha\beta} - \underset{0}{\bar{\gamma}}_{\alpha\beta}). \tag{2.10}$$

As a generalized statement of Hooke's law, we now postulate that the stress tensor is given by[14]

$$\bar{P}_{\alpha\beta} = \underset{0}{\bar{P}}_{\alpha\beta} + 4\bar{S}^{\sigma}{}_{(\alpha}\underset{0}{\bar{P}}_{\beta)\sigma} + \bar{\Lambda}_{\alpha\beta}{}^{\mu\nu}\bar{S}_{\mu\nu}, \tag{2.11}$$

where $\underset{0}{\bar{P}}_{\alpha\beta}$ are the stresses present in the unstrained equilibrium state and the second-order adiabatic elastic coefficients $\bar{\Lambda}_{\alpha\beta}{}^{\mu\nu}$ have the Voigt symmetry

$$\bar{\Lambda}^{\alpha\beta\mu\nu} = \bar{\Lambda}^{\mu\nu\alpha\beta} = \bar{\Lambda}^{\mu\nu\beta\alpha}. \tag{2.12}$$

These quantities refer to the local rest frame, i.e.,

$$\underset{0}{\bar{P}}_{\alpha\beta}\bar{u}^{\beta} = 0 \tag{2.13}$$

and

$$\bar{\Lambda}_{\alpha\beta}{}^{\mu\nu}\bar{u}^{\beta} = 0. \tag{2.14}$$

In this formulation of Hooke's law, both the background stress and the elastic tensor give rise to stresses which are linear in the strains. The justification for this is given by the correspondence with the classical theory (see Sec. 3). Note that the effective elastic tensor

$$4\bar{g}^{\nu(\alpha}\underset{0}{\bar{P}}{}^{\beta)\mu} + \bar{\Lambda}^{\alpha\beta\mu\nu}$$

does not possess the Voigt symmetry.

The complete thermodynamic description of an elastic body would include an equation of state for the equilibrium energy density and stress, the details of which vary from system to system. Here we implicitly regard the equilibrium properties as included in the specification of the system. Equation (2.11) then gives the additional stresses in nonequilibrium configurations. The change in energy density due to elastic deformations follows in the usual way from the conservation equation

$$\bar{u}_{\alpha}\bar{\nabla}_{\beta}\bar{T}^{\alpha\beta} = 0$$

(where $\bar{\nabla}_{\beta}$ denotes covariant differentiation with respect to $\bar{g}_{\alpha\beta}$). The full set of equations of motion is given by

$$\bar{\nabla}_{\beta}T^{\alpha\beta} = 0. \tag{2.15}$$

The Einstein equations are necessary to complete the description. Outside the body, the Einstein tensor vanishes, and inside we have

$$\bar{G}_{\alpha\beta} = -(8\pi Gc^{-4})\bar{T}_{\alpha\beta}. \tag{2.16}$$

On the surface, we have the boundary condition that the normal component of stress vanishes:

$$\bar{P}_{\mu\nu}\bar{n}^{\nu} = 0. \tag{2.17}$$

Equation (2.11) is a realistic description of the elastic stresses only when the higher-order elastic coefficients are not important as in the case of small strains. This is the situation for which relativistic elasticity theory is most relevant and which we now treat by perturbation methods.

3. ELASTIC PERTURBATIONS

Elastic motion will now be developed as a perturbation of an equilibrium system undergoing Born rigid motion.[15,16] In practice, static or stationary equilibrium states are more often of interest; but we will defer such specialization until Sec. 4. Born rigidity is the necessary and sufficient condition for strain-free elastic motion and is, therefore, the most natural choice for an elastic equilibrium state. In general relativity, Born rigidity imposes much weaker restrictions on the motion of a body than in special relativity, so the equilibrium system has considerable dynamical freedom.

The hydrodynamical description of the equilibrium state follows from specializing the conditions of Sec. 2 to the strain-free case. We use the same symbols for physical properties of the equilibrium state as in Sec. 2, but without a "bar". Thus,

$$u^{\alpha}u_{\alpha} = 1, \tag{3.1}$$

$$g_{\alpha\beta} = u_{\alpha}u_{\beta} + \gamma_{\alpha\beta}, \tag{3.2}$$

$$\gamma_{\alpha\beta}u^{\beta} = 0, \tag{3.3}$$

$$T_{\alpha\beta} = \rho u_{\alpha}u_{\beta} - P_{\alpha\beta}, \tag{3.4}$$

$$P_{\alpha\beta}u^{\beta} = 0, \tag{3.5}$$

$$G_{\alpha\beta} = -(8\pi Gc^{-4})T_{\alpha\beta}, \tag{3.6}$$

and

$$P_{\mu\nu} n^\nu = 0. \tag{3.7}$$

The necessary and sufficient condition for rigid motion is

$$\underset{u}{\mathcal{L}} \gamma_{\alpha\beta} = 0. \tag{3.8}$$

The kinematic properties of the rigid body are determined by the acceleration vector[17]

$$a_\alpha := u_{\alpha;\beta} u^\beta \tag{3.9}$$

and the rotation tensor

$$\omega_{\alpha\beta} := u_{[\alpha;\beta]} - a_{[\alpha} u_{\beta]} \tag{3.10}$$

with

$$u_{\alpha;\beta} = a_\alpha u_\beta + \omega_{\alpha\beta}. \tag{3.11}$$

In addition, the various dynamical corollaries of rigid motion apply, for instance

$$\underset{u}{\mathcal{L}} \rho = 0. \tag{3.12}$$

Now consider a one-parameter family of elastic systems $\bar{S}(\epsilon)$. For each value of ϵ, the system consists of a four-dimensional manifold $\bar{M}(\epsilon)$ with space-time metric $\bar{g}_{\mu\nu}(\epsilon)$ and physical properties as described in Sec. 2. We denote the relevant tensor fields on $\bar{M}(\epsilon)$ by $\bar{u}^\alpha(\epsilon)$, $\bar{T}_{\alpha\beta}(\epsilon)$, etc. For $\epsilon = 0$, we choose a rigid system as described above. In that case, we simply write $\bar{S}(0) = S$, $\bar{u}^\alpha(0) = u^\alpha$, $\bar{g}_{\mu\nu}(0) = g_{\mu\nu}$, etc.

In order to compare the systems $\bar{S}(\epsilon)$ with S we must map them all onto a common manifold, which we choose to be $M = \bar{M}(0)$. There are as many ways to do this as there are one-parameter families of diffeomorphisms of M. This is the gauge freedom of general relativistic perturbation theory. However, there is one natural class of gauges for this problem, namely the comoving gauge in which corresponding streamlines are mapped into each other. We can picture this in the following way. Asymptotically, in the infinite past, all systems $\bar{S}(\epsilon)$ are chosen to satisfy the same initial conditions as the rigid system S. This induces a natural identification of material points and, consequently, an identification of streamlines in the manifolds $\bar{M}(\epsilon)$. The remaining gauge freedom corresponds to the ways in which identified streamlines can be mapped into each other. We restrict this by requiring that unit proper time intervals along the streamlines in $\bar{M}(\epsilon)$ be mapped into unit intervals along the streamlines of M. The resulting gauge transformation group has infinitesimal descriptors

$$\eta^\alpha = f u^\alpha,$$

where f is constant along each streamline,

$$f_{;\alpha} u^\alpha = 0.$$

The following conditions are immediate consequences of the comoving gauge:

$$\underset{0}{\bar{\gamma}}_{\alpha\beta}(\epsilon) = \gamma_{\alpha\beta}, \tag{3.13}$$

$$\bar{u}^\alpha(\epsilon) = u^\alpha, \tag{3.14}$$

and

$$\bar{n}^\alpha(\epsilon) = n^\alpha. \tag{3.15}$$

In addition, because we are considering purely elastic perturbations, we take

$$\underset{0}{\bar{P}}_{\alpha\beta}(\epsilon) = P_{\alpha\beta} \tag{3.16}$$

Since the parameter ϵ can be arbitrarily rescaled, without loss of generality, we choose $\epsilon = 1$ for the perturbed system $\bar{S}$ we wish to treat, so that $\bar{S} = S(1)$. Then, to lowest order in the perturbation, quantities in $\bar{S}$ are related to their counterparts in S by the rule

$$\Delta A := \left.\frac{d\bar{A}(\epsilon)}{d\epsilon}\right|_{\epsilon=0} = \bar{A} - A + \text{higher-order terms}, \tag{3.17}$$

where A symbolizes a generic tensor field. The kinematic properties of the perturbation are given by the velocity

$$v_\alpha := \Delta u_\alpha \tag{3.18}$$

and the strain

$$S_{\alpha\beta} := \tfrac{1}{2}\Delta\gamma_{\alpha\beta}. \tag{3.19}$$

They satisfy

$$v_\alpha = \perp v_\alpha \tag{3.20}$$

and

$$S_{\alpha\beta} = \perp S_{\alpha\beta}, \tag{3.21}$$

where the projection operator $\perp$ projects every free index with

$$\gamma_\alpha{}^\beta = \delta_\alpha{}^\beta - u_\alpha u^\beta. \tag{3.22}$$

The perturbation of the metric is given by

$$h_{\alpha\beta} := \Delta g_{\alpha\beta} = 2u_{(\alpha} v_{\beta)} + 2S_{\alpha\beta}. \tag{3.23}$$

In the comoving gauge, the streamlines of an oscillating elastic system coincide with the streamlines of the unperturbed rigid system. As a result, there is no kinematic concept of displacement from equilibrium as in classical elasticity theory. The acceleration due to the perturbation satisfies

$$\Delta a_\alpha = \underset{u}{\mathcal{L}} v_\alpha. \tag{3.24}$$

Consequently, for cases of practical interest Δu_α will normally be nonzero, even though Δu^α vanishes because of the gauge conditions. This is why the rigid metric $\bar{g}^*_{\alpha\beta}$, introduced in Eq. (2.9) of Sec. 2, is not suitable for describing the unperturbed system.

Hooke's law for small perturbations takes the form

$$\Delta P_{\alpha\beta} = 4S^\nu{}_{(\alpha} P_{\beta)\nu} + \Lambda_{\alpha\beta}{}^{\mu\nu} S_{\mu\nu}, \tag{3.25}$$

where

$$\Lambda_{\alpha\beta}{}^{\mu\nu} = \lim_{\epsilon\to 0} \bar{\Lambda}_{\alpha\beta}{}^{\mu\nu}(\epsilon). \tag{3.26}$$

The complete set of dynamical equations governing the perturbation are obtained by applying the Δ operator to Eqs. (2.15)–(2.17). A collection of useful formulas and results are presented in Appendix A.

The perturbation theory developed above is applicable to fully relativistic systems for which the equilibrium space–time is curved. We do not treat such applications here, but proceed to compare the test body limit of the theory with other descriptions of elastic test bodies interacting with gravitational waves.

4. TEST BODY LIMIT

We now restrict our considerations to the test body limit[18] in which the elastic system does not give rise to space–time curvature. In addition, we restrict the streamlines of the material points in the body to be non-rotating geodesics.[19] Thus we have

$$g_{\mu\nu} \to \eta_{\mu\nu}, \tag{4.1}$$

with $\eta_{\mu\nu}$ the Minkowski metric, and the trajectories satisfy

$$u_{\mu;\nu} = \omega_{\mu\nu} = a_\mu = 0. \tag{4.2}$$

The test body equation of motion (A14) becomes

$$\rho(\dot{v}^\mu + \dot{S}u^\mu) + (\Delta\rho u^\mu + P^{\mu\alpha}v_\alpha)^{\cdot} - (\Lambda^{\mu\sigma\alpha\beta}S_{\alpha\beta})_{;\sigma} - P^{\alpha\beta}(2S^\mu{}_{\alpha;\beta} - S_{\alpha\beta}{}^{;\mu} + S_{;\alpha}\delta^\mu{}_\beta) = 0. \tag{4.3}$$

The dynamical equations of the perturbation theory developed in Sec. 3 now allow us to relate the physical observables: the strain of the elastic test body and the curvature of the gravitational wave. However, the traditional elastic equation, the equation of motion of the deformation itself, requires the introduction of the concept of displacement from equilibrium. This must be done in a gauge different than the comoving one.

A. The comoving gauge

Proceeding in the comoving gauge of Sec. 3, the u^μ component of Eq. (4.3) is

$$(\Delta\rho)^{\cdot} = -\rho\dot{S} - P^{\alpha\beta}\dot{S}_{\alpha\beta}, \tag{4.4}$$

which gives the rate of change of density, where the first term on the right is due to the expansion of the material and the second term is due to the work done by the stresses.

In order to relate the change in strain of the test body with the curvature of the gravitational wave, we use the spatial projection of the equation of motion (4.3) and the kinematical relation given by the perturbed Ricci identity (A10). Restricted to the test body, Eq. (A10) reduces to

$$\perp\{\ddot{S}^{\mu\nu} - \dot{v}^{(\mu;\nu)} + \Delta R^\mu{}_\alpha{}^\nu{}_\sigma u^\alpha u^\sigma\} = 0, \tag{4.5}$$

where we have contracted the equation with u^σ and projected the free indices. Restricting the test body equation of motion (4.3) to the case of negligible initial stress, we find by projection

$$\perp\dot{v}^\mu = \rho^{-1}(\Lambda^{\mu\sigma\alpha\beta}S_{\alpha\beta})_{;\sigma}. \tag{4.6}$$

In the comoving gauge v^μ has no direct physical interpretation. Geometrically, it represents the difference between rest frames of a material point in the equilibrium system and its vibrating counterpart, both having the same streamline. In order to obtain a relationship between physical observables, v^μ is eliminated by differentiating Eq. (4.6) and substituting it into Eq. (4.5).

We find

$$\perp\{\ddot{S}^{\mu\nu} - \rho^{-1}\nabla^{(\mu}\nabla_\sigma[\Lambda^{\nu)\sigma\alpha\beta}S_{\alpha\beta}] + \rho^{-2}\rho^{;(\mu}\nabla_\sigma[\Lambda^{\nu)\sigma\alpha\beta}S_{\alpha\beta}] + \Delta R^\mu{}_\alpha{}^\nu{}_\sigma u^\alpha u^\sigma\} = 0. \tag{4.7}$$

This describes the response of an elastic test body, such as Weber's aluminum bar, to incoming gravitational waves. The second term of (4.7) arises from the usual elastic restoring force and the third term from inhomogeneities in the density of the body. The last term represents the gravitational tidal force.

B. The displacement gauge

We now explicitly introduce the deformation in order to obtain its equation of motion. Once again we compare the systems $\bar{S}(\epsilon)$ with S, where S is now the flat equilibrium system of the test body. However, we no longer choose the comoving gauge in mapping $\bar{S}(\epsilon)$ and S onto the common manifold $M = \bar{M}(0)$. Instead, we introduce a vector field ξ^μ in M to represent the elastic displacement and map the streamlines of $\bar{S}$ onto the displaced streamlines of S. This gives

$$\bar{u}'^\mu = u^\mu + \epsilon\dot{\xi}^\mu, \tag{4.8}$$

where a prime denotes quantities in the displacement gauge. As in Sec. 3, we choose $\epsilon = 1$ for the perturbed system. In this gauge the first order perturbation quantities in $\bar{S}$ are related to their counterparts in S by

$$\Delta' A := \left.\frac{d\bar{A}'(\epsilon)}{d\epsilon}\right|_{\epsilon=0} = \bar{A}' - A + \text{higher-order terms}, \tag{4.9}$$

for a generic tensor field A. Quantities in the displacement gauge are related to quantities in the comoving gauge by

$$\bar{A}' = \bar{A} - \underset{\xi}{\pounds}A, \tag{4.10}$$

as exemplified by the gauge transformation of the metric tensor. For further discussion of this point see Ref. 20. Equation (4.10) then relates the operators Δ' and Δ by

$$\Delta' A = \Delta A - \underset{\xi}{\pounds}A \tag{4.11}$$

For the perturbed metric

$$h'_{\mu\nu} := \Delta' g_{\mu\nu} = \Delta g_{\mu\nu} - \underset{\xi}{\pounds}\eta_{\mu\nu},$$

we use Eq. (3.23) to obtain

$$h'_{\mu\nu} = 2S_{\mu\nu} + 2u_{(\mu}v_{\nu)} - 2\xi_{(\mu;\nu)}. \tag{4.12}$$

From the u^ν component and from the projection of (4.12), we find, respectively,

$$v_\mu = h'_{\mu\nu}u^\nu + 2\xi_{(\mu;\nu)}u^\nu \tag{4.13}$$

and

$$S_{\mu\nu} = \perp[\tfrac{1}{2}h'_{\mu\nu} + \xi_{(\mu;\nu)}]. \tag{4.14}$$

It is evident at this stage that the quantities v_μ and $S_{\mu\nu}$, which describe the kinematic properties of the perturbation in the comoving gauge, can be replaced by combinations of ξ_μ and $h'_{\mu\nu}$ in the displacement gauge. Rather than maintain complete generality in the equations of motion, we now restrict the gauge freedom in order to obtain a clear interpretation of ξ^μ.

We demand the radiation gauge condition

$$h'_{\mu\nu}u^\nu = 0. \tag{4.15}$$

Contraction of Eq. (4.12) with $u^\mu u^\nu$ results in

$$h'_{\mu\nu}u^\mu u^\nu + 2(\xi_\mu u^\mu)^{\cdot} = 0. \tag{4.16}$$

In the radiation gauge, $\xi_\mu u^\mu$ is constant, according to Eq. (4.16), so that without loss of generality we choose

$$\xi_\mu u^\mu = 0. \tag{4.17}$$

Equations (4.13) and (4.14) become

$$v_\mu = \dot{\xi}_\mu \tag{4.18}$$

and

$$S_{\mu\nu} = \tfrac{1}{2}h'_{\mu\nu} + \xi_{(\mu;\nu)} - \dot{\xi}_{(\mu}u_{\nu)}. \tag{4.19}$$

Here v_μ can be interpreted as the velocity of the displacement ξ_μ orthogonal to the streamlines. We use Eqs. (4.18) and (4.19) to rewrite the equation of motion (4.3) of the displacement. The u^μ component gives

$$(\Delta\rho)^{\cdot} = -\rho(\xi^\mu{}_{;\mu} + \tfrac{1}{2}\eta^{\mu\nu}h'_{\mu\nu})^{\cdot} - P^{\alpha\beta}(\xi_{\alpha;\beta} + \tfrac{1}{2}h'_{\alpha\beta}). \tag{4.20}$$

for the rate of change of density, and the spatial projection gives

$$\begin{aligned}&(\rho\dot{\xi}^\mu + P^{\mu\alpha}\dot{\xi}_\alpha)^{\cdot} + \perp(\tfrac{1}{2}P^{\alpha\beta}h'_{\alpha\beta}{}^{;\mu})\\ &-[\Lambda^{\mu\sigma\alpha\beta}(\xi_{\alpha;\beta} + \tfrac{1}{2}h'_{\alpha\beta}) + P^{\alpha\sigma}h'^{\mu}{}_{\alpha} + \tfrac{1}{2}P^{\mu\sigma}\eta^{\alpha\beta}h'_{\alpha\beta}\\ &-\underset{\xi}{\mathcal{L}}P^{\mu\sigma}]_{;\sigma} = 0,\end{aligned} \tag{4.21}$$

where we have used the equilibrium condition $P^{\alpha\beta}{}_{;\beta} = 0$.

Let us now consider the case treated by Dyson of an isotropic body with no initial stresses which interacts with a $p-p$ wave. There are only two independent elastic moduli, the Lamé parameters,

$$\Lambda^{\mu\sigma\alpha\beta} = \lambda\gamma^{\mu\sigma}\gamma^{\alpha\beta} + 2\mu\gamma^{\mu(\alpha}\gamma^{\beta)\sigma}, \tag{4.22}$$

with $\gamma^{\mu\nu}$ now the negative Euclidean metric of the spatial hypersurfaces. The details of the $p-p$ wave are given in Appendix B. Equation (4.21) now reduces to

$$\rho\ddot{\xi}^\mu - (\Lambda^{\mu\sigma\alpha\beta}\xi_{\alpha;\beta})_{;\sigma} - \mu_{;\beta}h'^{\mu\beta} = 0. \tag{4.23}$$

This shows that the gravitational interaction is between the wave and the inhomogeneity of the shear modulus, in agreement with the result obtained previously by Dyson.[4] In order to recover the equation of motion for the strain tensor, we rewrite Eq. (4.23) as

$$\ddot{\xi}^\mu - \rho^{-1}(\Lambda^{\mu\sigma\alpha\beta}S_{\alpha\beta})_{;\sigma} = 0. \tag{4.24}$$

With the help of Eqs. (4.19) and (B9), differentiation and projection leads to Eq. (4.7) for this special case.

C. The classical limit

In the displacement gauge, Hooke's law, Eq. (3.25), for small perturbations takes the form

$$\Delta' P_{\alpha\beta} = 4S^\nu{}_{(\alpha}P_{\beta)\nu} + \Lambda_{\alpha\beta}{}^{\mu\nu}S_{\mu\nu} - \underset{\xi}{\mathcal{L}}P_{\alpha\beta}. \tag{4.25}$$

For the contravariant version, this gives

$$\bar{P}^{\alpha\beta} = P^{\alpha\beta} - \underset{\xi}{\mathcal{L}}P^{\alpha\beta} + \Lambda^{\alpha\beta\mu\nu}S_{\mu\nu}, \tag{4.26}$$

in agreement with Hooke's law for prestressed materials in classical elasticity theory.[10–12]

The special relativistic limit of Eq. (4.21) is

$$(\rho\dot{\xi}^\mu + P^{\mu\alpha}\dot{\xi}_\alpha)^{\cdot} = (\Lambda^{\mu\sigma\alpha\beta}\xi_{\alpha;\beta} - \underset{\xi}{\mathcal{L}}P^{\mu\sigma})_{;\sigma}. \tag{4.27}$$

The term $P^{\mu\alpha}\dot{\xi}_\alpha$ is a relativistic correction to the energy density due to initial stresses. In the nonrelativistic limit, we obtain the classical equations of motion for elastic deformations of prestressed materials.

5. SUMMARY

We have presented a general relativistic description of elastic deformations which extends the initial work of Rayner. This is based upon a generalized Hooke's law for prestressed materials.

We have treated in detail the perturbations of an equilibrium system undergoing Born rigid motion. A most important feature of this perturbation treatment is its classical correspondence with the elasticity theory of prestressed materials.

For the first time, the strain-curvature equation for an elastic test body interacting with a gravitational wave has been derived from a complete theory. Previous treatments have introduced assumptions concerning test body motion in an *ad hoc* manner. We have corroborated the semiclassical work of Dyson showing the interaction of a gravitational wave with the inhomogeneities of the shear modulus. This places Dyson's results within the framework of general relativity.

The theory developed here is quite comprehensive in scope, although we have only applied it to the test body case. The agreement of our test body results with other physically reasonable descriptions leads us to believe that our work can be properly applied to fully relativistic systems.

APPENDIX A

Application of the Δ operator, defined in Eq. (3.17), to the indicated quantities leads to the following perturbation equations:

Metric:

$$\Delta g_{\mu\nu} = h_{\mu\nu}, \tag{A1}$$

$$\Delta g^{\mu\nu} = -h^{\mu\nu}. \tag{A2}$$

Christoffel symbols:

$$\Delta\left\{{\alpha \atop \beta\gamma}\right\} = \tfrac{1}{2}(h^\alpha{}_{\beta;\gamma} + h^\alpha{}_{\gamma;\beta} - h_{\beta\gamma}{}^{;\alpha}), \tag{A3}$$

$$\begin{aligned}&= 2S^\alpha{}_{(\beta;\gamma)} - S_{\beta\gamma}{}^{;\alpha}\\ &+ u^\alpha v_{(\beta;\gamma)} + v^\alpha u_{(\beta;\gamma)} + v^\alpha{}_{;(\beta}u_{\gamma)}\\ &+ u^\alpha{}_{;(\beta}v_{\gamma)} - [v_{(\beta}u_{\gamma)}]^{;\alpha}.\end{aligned} \tag{A4}$$

Riemann tensor:

$$\Delta R^\alpha{}_{\mu\nu\beta} = \nabla_\nu\Delta\left\{{\alpha \atop \mu\beta}\right\} - \nabla_\beta\Delta\left\{{\alpha \atop \mu\nu}\right\}, \tag{A5}$$

$$\Delta R_{\sigma\mu\nu\beta} = g_{\alpha\sigma}\Delta R^\alpha{}_{\mu\nu\beta} + h_{\alpha\sigma}R^\alpha{}_{\mu\nu\beta}, \tag{A6}$$

where we use the convention corresponding to

$$2A_{\mu;[\rho\sigma]} \equiv A_\alpha R^\alpha{}_{\mu\rho\sigma}.$$

Ricci tensor: With the definition $R_{\mu\nu} := R^\alpha{}_{\mu\nu\alpha}$, we have

$$\Delta R_{\mu\nu} = \nabla_\nu\Delta\left\{{\alpha \atop \mu\alpha}\right\} - \nabla_\alpha\Delta\left\{{\alpha \atop \mu\nu}\right\}, \tag{A7}$$

and using (A3) and (A4) we find

$$\Delta R_{\mu\nu} = S_{;\mu\nu} - S^{\alpha}{}_{\mu;\nu\alpha} - S^{\alpha}{}_{\nu;\mu\alpha} + S_{\mu\nu}{}^{;\alpha}{}_{\alpha} + [u_{(\mu}v_{\nu)}]^{;\alpha}{}_{\alpha} - [v_{\alpha}u_{(\mu}]_{;\nu)}{}^{\alpha} - [u_{\alpha}v_{(\mu}]_{;\nu)}{}^{\alpha}, \tag{A8}$$

where $S: = g^{\alpha\beta}S_{\alpha\beta}$.

Curvature scalar:

$$\Delta R = g^{\mu\nu}\Delta R_{\mu\nu} - h^{\mu\nu}R_{\mu\nu}. \tag{A9}$$

Ricci identity: For the identity $2u_{\nu;[\rho\sigma]} \equiv u_{\alpha}R^{\alpha}{}_{\nu\rho\sigma}$, we have

$$\nabla_{\sigma}(v_{\nu;\rho} - u_{\alpha}\Delta\{^{\alpha}_{\nu\rho}\}) - \nabla_{\rho}(v_{\nu;\sigma} - u_{\alpha}\Delta\{^{\alpha}_{\nu\sigma}\}) + (\omega_{\alpha\sigma} + a_{\alpha}u_{\sigma})\Delta\{^{\alpha}_{\nu\rho}\} - (\omega_{\alpha\rho} + a_{\alpha}u_{\rho})\Delta\{^{\alpha}_{\nu\sigma}\} + v_{\alpha}R_{\nu}{}^{\alpha}{}_{\rho\sigma} + u_{\alpha}\Delta R_{\nu}{}^{\alpha}{}_{\rho\sigma} = 0. \tag{A10}$$

Einstein equations:

$$\Delta G_{\mu\nu} = -(8\pi G c^{-4})\Delta T_{\mu\nu}, \tag{A11}$$

where the Einstein tensor is defined as

$$G_{\mu\nu}: = R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R.$$

Matter tensor:

$$\Delta T_{\mu\nu} = \Delta\rho u_{\mu}u_{\nu} + 2\rho u_{(\mu}v_{\nu)} - \Delta P_{\mu\nu} \tag{A12}$$

with $\Delta P_{\mu\nu}$ given by Hooke's law Equation (3.24).

Equations of motion:

$$\Delta(T^{\mu\nu}{}_{;\nu}) = 0 \tag{A13}$$

which can be written in detail as

$$\rho(\dot{v}^{\mu} + v_{\alpha}\omega^{\alpha\mu} - 2S^{\mu}{}_{\alpha}a^{\alpha} + \dot{S}u^{\mu} + v^{\alpha}a_{\alpha}u^{\mu}) + (\Delta\rho u^{\mu} + P^{\mu\alpha}v_{\alpha})^{\cdot} - (\Lambda^{\mu\nu\alpha\beta}S_{\alpha\beta})_{;\nu} - P^{\alpha\beta}(2S^{\mu}{}_{\alpha;\beta} - S_{\alpha\beta}{}^{;\mu} + S_{;\alpha}\delta^{\mu}{}_{\beta} + \omega^{\mu}{}_{\beta}v_{\alpha} - v_{\alpha}a_{\beta}u^{\mu}) = 0, \tag{A14}$$

where the dot is defined as the covariant derivative along u^{α}:

$$\dot{A}^{\beta}: = u^{\alpha}\nabla_{\alpha}A^{\beta}.$$

Boundary conditions:

$$n^{\alpha}\Delta P_{\alpha\beta} = 0. \tag{A15}$$

APPENDIX B

We list the properties of $h'_{\mu\nu}$ for a linearized p–p wave: Let the complex vector m^{μ} have the decomposition

$$m^{\mu}: = p^{\mu} + iq^{\mu}$$

in terms of two real orthonormal spatial vectors P^{μ} and q^{μ},

$$p^{\mu}p_{\mu} = q^{\mu}q_{\mu} = -1$$

and

$$p^{\mu}q_{\mu} = m^{\mu}m_{\mu} = 0.$$

We choose m^{μ} such that

$$u^{\alpha}m_{\alpha} = k^{\alpha}m_{\alpha} = 0, \tag{B1}$$

where k^{α} is a real null vector corresponding to the propagation direction of a p–p wave

$$k^{\alpha}k_{\alpha} = 0. \tag{B2}$$

To specify the p–p wave, we take m^{μ} and k^{μ} to be covariant constant

$$k_{\mu;\nu} = m_{\mu;\nu} = 0. \tag{B3}$$

The wave is then given by

$$h'_{\mu\nu} = f\ \mathrm{Re}\{m_{\mu}m_{\nu}\}. \tag{B4}$$

The derivative of the scalar f satisfies

$$f_{;\rho} = f'k_{\rho}.$$

(The prime of f symbolizes a derivative and should not be confused with the prime of $h'_{\mu\nu}$.) It follows that

$$\eta^{\mu\nu}h'_{\mu\nu} = h'^{\mu\nu}{}_{;\nu} = 0, \tag{B5}$$

and

$$h'_{\mu\nu;\rho} = (\log f)'k_{\rho}h'_{\mu\nu}. \tag{B6}$$

The Riemann tensor is given by

$$\Delta R_{\mu\alpha\nu\sigma} = 2\nabla_{[\nu}h'_{\sigma][\mu;\alpha]}. \tag{B7}$$

It satisfies the equation

$$\Delta R_{\mu\alpha\nu\sigma}k^{\sigma} = 0. \tag{B8}$$

The contraction with $u^{\alpha}u^{\sigma}$ is useful:

$$\Delta R_{\mu\alpha\nu\sigma}u^{\alpha}u^{\sigma} = -\tfrac{1}{2}\ddot{h}'_{\mu\nu}. \tag{B9}$$

*Present Address: Department of Physics, Trenton State College, Trenton, N. J. 08625

†Present address: Department of Physics, University of Pittsburgh, Pittsburgh, Pa. 15213

[1] J. Weber, *General relativity and gravitational waves* (Interscience, New York, 1961).
[2] J. Weber, Phys. Rev. Lett. **22**, 1320 (1969).
[3] J. Weber, Phys. Rev. Lett. **18**, 498 (1967).
[4] F. J. Dyson, Astrophys. J. **156**, 529 (1969).
[5] D. Pines, "Inside Neutron Stars" in *Proceedings of the twelfth international conference on low temperature physics, Kyoto, Japan*, edited by E. Kanda (Academic Press of Japan, 1971).
[6] J. L. Synge, Math. Z. **72**, 82 (1969).
[7] A. Papapetrou, Ann. Inst. Henri Poincaré A 16, 63 (1972).
[8] C. B. Rayner, Proc. R. Soc. A **272**, 44 (1963).
[9] W. C. Hernandez, Phys. Rev. **1**, 1013 (1970).
[10] C. Truesdell and W. Noll, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1965), Vol. III/3.
[11] D. C. Wallace, in *Solid state physics*, edited by F. Seitz *et al.* (Academic, New York, 1970), Vol. 25.
[12] E. Glass and J. Winicour, "A geometric generalization of Hooke's law" (to be published).
[13] Our metric has signature −2, and Greek indices range over 0,1,2,3.
[14] Parentheses about indices denote symmetrization: $A_{(\alpha\beta)}: = 1/2(A_{\alpha\beta} + A_{\beta\alpha})$; and brackets similarly denote antisymmetrization.
[15] For a summary of the properties of rigid motion, see F. Pirani, "Foundations

and current problems of general relativity," in *Lectures on general relativity.* 1964 Brandeis Summer Institute (Prentice-Hall, Englewood Cliffs, N. J., 1965), Vol. 1, p. 201.

[16]An attractive alternative description of rigid systems can be developed, analogous to the treatment of stationary systems by R. Geroch, J. Math. Phys. (N.Y.) **12**, 918 (1971). We thank Dr. Geroch for pointing out this generalization.

[17]Covariant differentiation is denoted by a semicolon or ∇_a.

[18]The test body limit is achieved by a family of equilibrium systems depending upon a parameter α, with the matter tensor vanishing for $\alpha = 0$. We again assume a parameter calibration such that $\alpha = 1$ describes the real system which the test body approximates.

[19]By using the same techniques, the effects of accelerating and rotating streamlines can readily be incorporated.

[20]R. K. Sachs, "Gravitation radiation," in *Relativity, groups, and topology,* 1963 Les Houches Summer School (Gordon and Breach, New York, 1964).

On the possibility of observing first-order corrections to geometrical optics in a curved space–time

I. Dwivedi and R. Kantowski
The University of Oklahoma, Norman, Oklahoma
(Received 5 June 1972)

Gravity's effect on the polarization of test electromagnetic fields is presented. It is shown that under ordinary circumstances the effect is not measurable.

INTRODUCTION

Almost all calculations which involve electromagnetic waves moving in a curved space–time resort to the geometrical optics (g.o.) approximation for test fields. g.o. gives information about the intensity of point sources, the bending of light rays, and the distortion of wavefronts as the light moves through a curved space–time. However, it neglects wavelength dependent properties such as polarization, e.g., a plane or circularly polarized wave at the source will be a plane or circularly polarized wave at the observer. To estimate what effect the gravitational field has on the polarization, one must go to the first-order correction in wavelength for g.o. In the next section the first-order correction is presented, and in the following sections it is applied to a point source in a Schwarzschild field.

FIRST-ORDER CORRECTIONS TO g.o.[1]

Following Ehlers the E & M field in vacuum can be written as a self-dual bivector G^{ab} and the amplitude of the wave expanded in a power series to first order in the wavelength:

$$G^{ab}(x^c, \epsilon) = K_+^{ab}(x^c, \epsilon)e^{iS(x^d)/\epsilon} + K_-^{ab}(x^c, \epsilon)e^{-iS(x^d)/\epsilon},$$
$$\approx [K_+^{ab}(0) + K_+^{ab}(1)\epsilon]e^{iS/\epsilon} + [K_-^{ab}(0) + K_-^{ab}(1)\epsilon]e^{-iS/\epsilon}, \quad (1)$$

where the positive phase terms represent the right circularly polarized part of the wave and negative the left. $S =$ con. are the null surfaces of constant phase and ϵ is related to the wavelength by

$$\lambda = -2\pi\epsilon/(S_{,a}u^a), \quad (2)$$

u^a being the observers 4-velocity ($u^a u_a = -1$).

There are only three independent self-dual bivectors, and one is given by $K_+(0)$ and $K_-(0)$, i.e., by the g.o. limit, $\epsilon \to 0$. These basis bivectors are constructed in terms of a null tetrad ($k_a \equiv S_{,a}, m_a, t_a$) which is parallely transported along the characteristics of the null surfaces $S =$ con,[2]

$$k^a k_a = m^a m_a = t^a t_a = t^a m_a = t^a k_a = 0,$$
$$k^a m_a = t^a \bar{t}_a = +1,$$
$$\dot{k}_a = \dot{m}_a = \dot{t}^a = 0, \quad \text{where } \dot{(\cdot)} \equiv k^a \nabla_a. \quad (3)$$

The basis bivectors are then simply written as

$$V^{ab} = 2k^{[a}\bar{t}^{b]}, \quad U^{ab} = 2m^{[a}t^{b]},$$
$$M^{ab} = 2k^{[a}m^{b]} + 2\bar{t}^{[a}t^{b]}. \quad (4)$$

The g.o. limit as given by Maxwell's equation is

$$G^{ab} = A_+(0)V^{ab}e^{iS/\epsilon} + A_-(0)V^{ab}e^{-iS/\epsilon}, \quad (5)$$

where

$$\dot{A}_\pm(0) + \theta A_\pm(0) = 0. \quad (6)$$

θ in the above is the usual expansion parameter ($\theta \equiv t^a k_{a;b}\bar{t}^b = \frac{1}{2}k^a{}_{;a}$), and Eq. (6) says that the intensity of a g.o. wave falls off as (area)$^{-1}$. The polarization is unaffected because only V^{ab} appears and it is parallely transported.

To do the first-order corrections to g.o., K_+^{ab} and K_-^{ab} are expanded in terms of V^{ab}, U^{ab}, M^{ab}, and Maxwell's equations are imposed to first order in ϵ:

$$K_\pm^{ab} = A_\pm V^{ab} + B_\pm U^{ab} + C_\pm M^{ab}, \quad A_\pm = A_\pm(0) + A_\pm(1)\epsilon,$$
$$B_\pm = B_\pm(1)\epsilon, \quad C_\pm = C_\pm(1)\epsilon. \quad (7)$$

Maxwell's equations give

$$B_\pm(1) = \mp iA_\pm(0)\sigma,$$
$$C_\pm(1) = \pm iA_\pm(0)_{,\bar{t}} - A_\pm(0)\zeta, \quad (8)$$
$$A_\pm(1) = \pm iA_\pm(0)f_\pm,$$

where ${}_{,t} \equiv t^a\nabla_a$.

$f_\pm(x^c)$ are functions satisfying

$$\dot{f}_\pm = \left(\frac{A_\pm(0)_{,\bar{t},t}}{A_\pm(0)} - \frac{A_\pm(0)_{,t}}{A_\pm(0)}\bar{\zeta} - \bar{\zeta}_{,t} - \sigma\bar{\sigma}'\right), \quad (9)$$

and σ, ζ, σ' are scalars defined in terms of the tetrad by

$$\sigma = \bar{t}^a k_{a;b}\bar{t}^b, \quad \text{usual shear},$$
$$\zeta = t^a \bar{t}_{a;b}t^b, \quad \sigma' = \bar{t}^a m_{a;b}\bar{t}^b. \quad (10)$$

The procedure for doing first-order optics is to first construct S and its tetrad field; second, to evaluate the scalars θ, σ, ζ, and σ' and integrate Eq. (6) for the geometrical optics terms $A_\pm(0)$; and finally, turn to Eqs. (8) and (9) to find the first-order correction. Equation (9) is clearly the hardest to solve in most applications, however, in the next section it is shown that $B_\pm(1)$ are the only terms needed to estimate polarization effects.

FIRST-ORDER EFFECT ON POLARIZATION

The electric and magnetic vectors seen by an observer u^a are given by

$$\mathcal{E}^a = E^a - iB^a = G^{ab}u_b, \quad (11)$$

and the maximum and minimum values of E^2 are

$$E^2_{\max,\min} = \tfrac{1}{2}\{\mathcal{E}\cdot\mathcal{E} \pm |\mathcal{E}\cdot\mathcal{E}|\}. \quad (12)$$

Gravity's effect on polarization can be estimated by considering right circularly polarized light at the source. If the light stays right circularly polarized, then $\mathcal{E}\cdot\mathcal{E}$ remains zero as it does when $K_+^{ab} = A_+ V^{ab}$. The following can be taken as alteration of the polarization,

$$P \equiv |\mathcal{E}\cdot\mathcal{E}|/(\mathcal{E}\cdot\bar{\mathcal{E}}) = 2|A_+B_\mp - C_\mp^2|/(\boldsymbol{\mathcal{E}}\cdot\bar{\boldsymbol{\mathcal{E}}}), \tag{13}$$

where $\boldsymbol{\mathcal{E}}\cdot\bar{\boldsymbol{\mathcal{E}}} = E_{\max}^2 + E_{\min}^2$ is proportional to the intensity. P gives the fraction of the energy carried by the linearly polarized part of the wave. Equation (13) shows that the first-order term in P comes from $B_+(1)$ alone, and when Eqs. (5) and (8) are used it reduces to

$$P = |\sigma|\epsilon/(k^a u_a)^2. \tag{14}$$

If an effect on the polarization is going to be observed, Eq. (14) says that the gravitational field must introduce large amounts of shear into the light waves. Two obvious applications are (1) at focal points in Schwarzschild fields and (2) in high shearing cosmologies near the big bang. In the next section the first case is considered.

OPTICS IN THE FIELD OF A DENSE STAR

A dense star is interesting because of its radius R is between $2m$ and $3m$ it can focus its own light (see Fig. 1) and near focus $\sigma \to \infty$. When looking at a point source on the star the expansion and shear θ and σ can be defined in terms of the two dimensions of the wavefronts

$$\theta = \tfrac{1}{2}|\dot{D}_+/D_+ + \dot{D}_-/D_-|, \tag{15}$$

$$\sigma = \tfrac{1}{2}|\dot{D}_+/D_+ - \dot{D}_-/D_-|. \tag{16}$$

D_+ and D_- are given by

$$D_- = \delta\beta r \sin\phi, \tag{17}$$

$$D_+ = \delta l r\{r\}^{1/2}\int_R^r \frac{dr}{r^2\{r\}^{3/2}}, \tag{18}$$

where $\{r\} \equiv \{1-(l^2/r^2)[1-(2m/r)]\}$.

In the above equations, r and ϕ are the usual Schwarzschild coordinates ($\phi = 0, r = R$ is the location of the point source) and l is the impact parameter at ∞. $\delta\beta$ is an isotropy parameter and is defined in Fig. 1. σ (and hence P) becomes large near focus ($\phi \to n\pi$) and can be approximated by

$$\sigma \sim \tfrac{1}{2}\cot\phi(kl/r^2) \sim kl/2\Delta\phi r^2, \tag{19}$$

where k is the affine parameter constant for a null geodesic in Schwarzschild space-time and $\Delta\phi$ is defined in Fig. 1. Near focus P becomes

$$P \sim \frac{[1-(2m/r)]^{1/2}\lambda}{4\pi\Delta\phi r}\left(\frac{l}{r}\right). \tag{20}$$

According to Eq. (20) long wavelengths are the most favorable. To estimate P consider a radio antenna centered on the focus of a point source and calculate P at the edge of the antenna. Some reasonable numbers to try

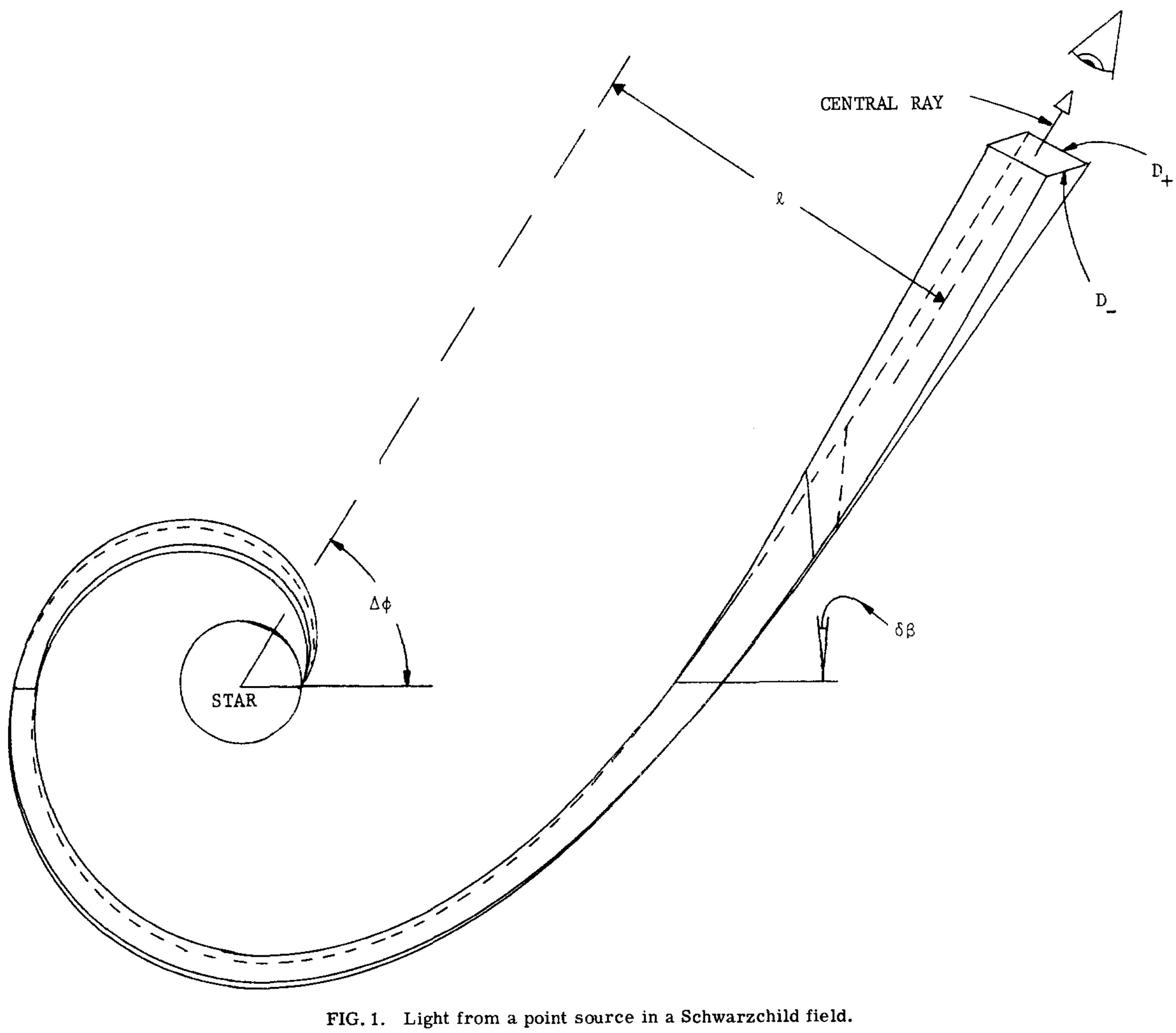


FIG. 1. Light from a point source in a Schwarzchild field.

are $\Delta\phi r \sim 15$ m (antenna size), $\lambda \sim 1$ m, $l \sim 10^4$ m, and $r \sim 10^{18}$ m (star distance). P is then seen to be 10^{-16}—much too small to be detected. The conclusion is that the region of large shear is so small that the antenna would not measure $P \neq 0$ for a single point source. If an extended source is considered, every point of the antenna is a focus of some point on the star; however, the conclusion is the same as before. Only an immeasurable fraction of energy will be seen as high shearing waves. It should be pointed out that P could be increased to 10^{-3} at the edge of the antenna by making observations close to the dense star, e.g., $r \sim 10^5$ m.

CONCLUSION

Other applications can be considered, e.g., (1) polarization in the radiation coming from a collapsing star, (2) polarization in the primeval fire ball due to inhomogeneities, and (3) polarization in the primeval fire ball due to an anisotropy in the Hubble expansion. All three of the above are too small to be seen.

The conclusion is that in spite of the elegance of higher-order optics in a curved space–time, there seems to be no reasonable observation which can detect a correction to g.o.

[1] J. Ehlers, Z. Naturforsch. **229**, 1328 (1967).

[2] I. Robinson and A. Schild, J. Math. Phys. (N.Y.) **4**, 484 (1963).

A new representation of the solution of the Ising model

M. J. Stephen*
Physics Department, Rutgers University, New Brunswick, New Jersey 08903

L. Mittag
Saybrook College, Yale University, New Haven, Connecticut 06520
(Received 3 July 1972)

It is shown that the transfer matrices for various Ising lattices in two dimensions commute with certain linear operators. The problem of finding an explicit form for the largest eigenvector is considerably simplified. The expansion coefficients appearing in the eigenvectors found as the solution of a set of nonlinear difference equations are Pfaffians. The connection between this type of solution and other solutions is clarified. This form for the eigenvector also simplifies the calculation of correlation functions. Some geometrical aspects of the Ising model are discussed.

1. INTRODUCTION

In an early approach to the solution of the two-dimensional Ising model Onsager[1] considered a transfer matrix which extended the lattice along diagonals. The published version[2] analyzes extensions only along columns. It is interesting that the diagonal extension offers many advantages. A simple linear operator which commutes with the diagonal transfer matrix was found by Onsager.[1] This result can be generalized to the triangular, hexagonal, and rectangular lattices. The transfer matrices for these cases are shown to commute with operators linear in the algebra generated by the column and row operators A and B. The eigenfunctions of these linear operators can be constructed by solving a set of difference equations with certain symmetry porperties. Using the diagonal transfer matrix we can write the partition function as a single integral, symmetric in the horizontal and vertical bonds, H_1 and H_2, respectively. The dual transformation is shown to be the diagonal transfer matrix evaluated at $H_1 = -H_2 = \pm \frac{1}{4}\pi i$. Correlations along the diagonal are obtained by a method which uses the dual transform; disorder is easier to calculate than order. The underlying geometry is found to be Euclidean rather than hyperbolic.

Consider an Ising lattice whose sites, n per column, are designated by spin variables $\mu_j = \pm 1$. Let the horizontal and vertical bond strengths be $J_1 = KTH_1$ and $J_2 = KTH_2$ (see Fig. 1). The transfer matrices which represent the columnar and diagonal extensions, are, respectively.

$$W_c(H_1,H_2)_{\mu\mu'} = \prod_j \exp(H_1\mu_j\mu_j' + H_2\mu_j'\mu_{j+1}'), \quad (1.1a)$$

$$W(H_1,H_2)_{\mu\mu'} = \prod_j \exp(H_1\mu_j\mu_j' + H_2\mu_j\mu_{j+1}'). \quad (1.1b)$$

The columnar transfer matrix W_c can be represented as the product of two operators $V_1(H_1)V_2(H_2)$ with

$$V_1(H_1) = (2\sinh 2H_1)^{n/2}\exp(H_1^*B), \qquad B = \sum_j C_j,$$
$$V_2(H_2) = \exp(H_2 A), \qquad A = \sum_j S_j S_{j+1}, \qquad (1.2)$$
$$\tanh H_1^* = \exp(-2H_1), \qquad 1 = \sinh 2H_1 \sinh 2H_1^*.$$

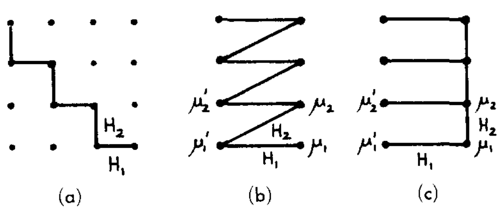


FIG. 1. Bonds appearing in diagonal extension (a), lattice deformed in (b), and the usual columnar extension appears in (c).

The operators C_j and S_j are equivalent to the two Pauli matrices σ_j^x and σ_j^z. They can also be defined as

$$S_j\Phi(\mu_1,\cdots,\mu_j,\cdots) = \mu_j\Phi(\mu_1,\cdots,\mu_j,\cdots). \quad (1.3a)$$

$$C_j\Phi(\mu_1,\cdots,\mu_j,\cdots) = \Phi(\mu_1,\cdots,-\mu_j,\cdots). \quad (1.3b)$$

Cyclic boundary conditions are imposed; $C_{j+n} = C_j$, $S_{j+n} = S_j$. The following single-spin representation is helpful. Generalization to n spins is straightforward:

$$I_{\mu\mu'} = \delta(\mu-\mu') = \tfrac{1}{2}(1+\mu\mu'), \quad (1.4a)$$

$$C_{\mu\mu'} = \delta(\mu+\mu') = \tfrac{1}{2}(1-\mu\mu'), \quad (1.4b)$$

$$S_{\mu\mu'} = \mu\delta(\mu-\mu') = \tfrac{1}{2}(\mu+\mu'). \quad (1.4c)$$

With μ standing for a column of spins $\mu_1\cdots\mu_n$, in this representation one form of the dual transformation is given explicitly by

$$L_{\mu\mu'} = (1/2)^{(n+1)/2}\prod_j[\delta(\mu_j-\mu_{j+1}) + \mu_j'\delta(\mu_j+\mu_{j+1})]. \quad (1.5)$$

This operator is orthogonal in the even space, the subspace left invariant by the projection operator $\Lambda_+ = \frac{1}{2}(I+U)$ where $U = C_1C_2\cdots C_n$. In this subspace L interchanges A and B. Also

$$LV_1(H_1)\tilde{L} = (2\sinh 2H_1)^{n/2}V_2(H_1^*)\Lambda_+, \quad (1.6)$$
$$LV_2(H_2)\tilde{L} = (\tfrac{1}{2}\sinh 2H_2)^{n/2}V_1(H_2^*)\Lambda_+.$$

We will show in Appendix C that the dual transformation is essentially the transfer matrix $W(H_1,H_2)$ evaluated at $H_1 = -H_2 = \pm i\pi/4$. Below we determine some of the eigenvectors and eigenvalues of W and hence of L as well.

2. COMMUTATION RELATIONS

In Appendix A we show that the diagonal transfer matrices $W(H_1,H_2)$ and $W(H_1',H_2')$ commute if they have the same modulus k defined by

$$k = \sinh 2H_1 \sinh 2H_2. \quad (2.1)$$

This modulus plays a multiplicative role under the dual transformation as seen from

$$LW(H_1,H_2)\tilde{L} = k^n W(H_2^*,H_1^*)\Lambda_+. \quad (2.2)$$

At the critical point, $H_1 = H_2^*$; thus $k_c = 1$.

Transfer matrices for the triangular and hexagonal lattices are given by (see Fig. 2)

$$W_\Delta(H_1,H_2,H_3) = V_2(H_3)W(H_1,H_2), \tag{2.3a}$$

$$W_{\text{hex}}(H_1,H_2,H_3) = W(H_1,H_2)V_1(H_3) \tag{2.3b}$$

A relationship exists between hexagonal bonds H_1, H_2, H_3 and the triangular bonds H_{12}, H_{13}, H_{23} called the star-triangle transformation (see Appendix E). It is expressed by the following equation which also holds when the order of operators on both sides is permuted. $N(H_1, H_2, H_3)$ is a function:

$$W(H_1, H_2)V_1(H_3) = N(H_1, H_2, H_3)^n V_2(H_{12})W(H_{13}, H_{23}). \tag{2.4}$$

Onsager[1] has used the star-triangle transformation to derive the important result

$$[B + kA, W] = 0. \tag{2.5}$$

A simple calculation also gives this result. The vanishing of the commutator Equation (2.5) shows that there exists a representation in which both W and $B + kA$ have the same eigenvectors. The latter is obviously Hermitean while the former is normal. This is indicated in Appendix A. The eigenvectors and the spin–spin correlations along a diagonal can depend only on the modulus k.

The simple method used to prove Eq.(2.5) can be generalized to find linear operators which commute with the various transfer matrices. Let W_Δ and $W_s = V_2^{-1/2} W_\Delta V_2^{1/2}$ be two equivalent transfer matrices for the triangular lattice. Then the following equations hold (see Appendix B):

$$[B + k_1A + k_2A_2, W_s] = 0, \quad A_2 = \sum_j S_{j-1}C_jS_{j+1},$$
$$[B + k_1'A + k_2'G_1, W_\Delta] = 0, \quad 4G_1 = [B,A]. \tag{2.6}$$

We introduce a notation to simplify equation with hyperbolic functions

$$c_i = \cosh 2H_i, \quad \hat{c}_i = \cosh H_i,$$
$$s_i = \sinh 2H_i, \quad \hat{s}_i = \sinh H_i. \tag{2.7}$$

Then

$$k_1(\hat{c}_3)^2 = k_1'c_3 = s_1s_2c_3 + c_1c_2s_3$$
$$k_2(\hat{c}_3)^2 = -(\hat{s}_3)^2, \quad k_2'c_3 = -s_3. \tag{2.8}$$

The pseudo-Hamiltonians in Eqs.(2.6) are of course related by a similarity transformation. The hexagonal lattice is included in these results by the use of the star-triangle transformation Eq.(2.4). In the limit of $H_2 \to 0$, a result is obtained closely related to a recent discovery of Suzuki.[3] We find that $V_2(\frac{1}{2}H_3)V_1(H_1)V_2(\frac{1}{2}H_3)$ commutes with

$$B + (2c_1\hat{s}_3/\hat{c}_3)A - (\hat{s}_3/\hat{c}_3)^2A_2. \tag{2.9}$$

Suzuki's result is equivalent to making a similarity transformation of Eq.(2.9) with the operator gL, where L is the dual transformation and $g = 2^{-n/2}\prod_j (C_j + S_j)$. The pseudo-Hamiltonians in (2.5), (2.6), and (2.9) are easily diagonalized by introducing fermion operators. However, it is instructive to proceed in a somewhat different manner.

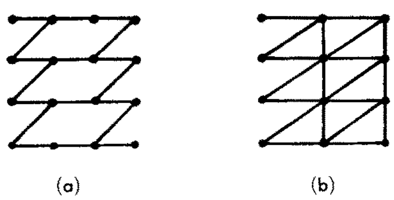


FIG. 2. Hexagonal lattice (a), and triangular lattice (b) deformed so that their transfer matrices can be represented as in the text.

3. SYMMETRIC EIGENFUNCTIONS OF $B + kA$

A. Even space

Since the transfer matrix commutes with U, the space of functions decomposes into even and odd subspaces. A typical member of the even subspace is

$$\Phi(\mu) = 2^{-n/2}\Big(f_0 + \sum_{i>j} f_{ij}\mu_i\mu_j + \sum_{i>j>k>l} f_{ijkl}\mu_i\mu_j\mu_k\mu_l + \cdots\Big). \tag{3.1}$$

Now $B + kA$ may be thought of as a Hamiltonian with k as an interaction constant. (The ground state of $-B - kA$ determines the partition function.) At high temperatures $k \to 0$, so that B is the noninteracting part. Simple considerations imply that $\Phi_0(\mu)$, the ground state, is constant in this limit, yielding the totally disordered state while

$$\sum_{\mu'} L_{\mu\mu'}\Phi_0(\mu') = (1/\sqrt{2})\prod_j \delta(\mu_j - \mu_{j+1}), \quad k = 0, \tag{3.2}$$

is the totally ordered state, L interconverting order and disorder. Let $\Phi_0(\mu, k)$ be the ground state eigenfunction corresponding to a particular modulus k, then generally

$$\sum_{\mu'} L_{\mu\mu'}\Phi_0(\mu', k) = \Phi_0(\mu, k^{-1}). \tag{3.3}$$

From a group theoretic viewpoint, the transfer matrix $W(H_1, H_2)$ is normal and invariant with respect to the cyclic group while $B + kA$ is Hermitean and invariant with respect to the dihedral group. The ground state lies in the symmetric subspace, invariant to all group operations. Thus $f_{ij}(k)$ in Eq.(3.1) is a function of the difference $i - j$. The interpretation of f_{ij} as the probability amplitude of finding spins $j+1, j+2, \ldots, i$ identical while all other spins are opposite can be seen from Eq.(3.3).

The sums present in Eq.(3.1) define f_{ij} for $i > j$. We may extend f_{ij} antisymmetrically for other values.

$$f_{ij} + f_{ji} = 0, \quad f_{ii} = 0. \tag{3.4}$$

Boundary conditions imply the following restrictions, coefficients of the ground state must satisfy

$$f_{21} = f_{32} = \cdots = f_{n+1,n} = f_{n,1}. \tag{3.5}$$

The last equality is satisfied by imposing anticyclic conditions $f_{ij} = -f_{n+i,j}$. Only odd multiples of π/n will then occur in the Fourier series for f_{ij}.

Higher coefficients f_{ijkl}, etc., can also be extended antisymmetrically (determinants are such examples), but in our case we assume that f_{ijkl} and higher coefficients can be represented by Pfaffian forms

$$f_{ijkl} = f_{ij}f_{kl} - f_{ik}f_{jl} + f_{il}f_{jk}. \tag{3.6}$$

This is just sufficient to solve all the resulting difference equations by Fourier analysis. A fermion type solution is recovered.

Let us seek an eigenfunction of $B + kA$ of the form of Eq. (3.1) with $f_0 = 1$, the other coefficients vanishing in the high temperature limit. Consider the equation

$$\sum_{\mu'} (B_{\mu\mu'} + kA_{\mu\mu'})\Phi(\mu') = \lambda\Phi(\mu). \tag{3.7}$$

If we multiply this equation by 1, $\mu_i\mu_j$, $\mu_i\mu_j\mu_k\mu_l, \ldots$, with $i > j > k > l$, and sum over all $\mu_i = \pm 1$, we obtain a series of difference equations, the first two of which are

$$\lambda = n + k\sum_j f_{j+1,j}. \tag{3.8a}$$

$$\lambda f_{ij} = (n-4) f_{ij} + k[\delta_{i,j+1} + \{f_{i\pm 1,j\pm 1}\} + \sum_k f_{i,j,k+1,k}], \tag{3.8b}$$

$$\{f_{i\pm 1,j\pm 1}\} \equiv f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1}. \tag{3.9}$$

The quartic term should have its subscripts ordered; but $k+1, k$ always occur consecutively and ordered so that

$$f_{i,j,k+1,k} = f_{i,k+1,k,j} = f_{k+1,k,i,j}.$$

The three regimes can be written as a single sum over k. If we eliminate λ from Eqs. (3.8a) and (3.8b) and use the Pfaffian decomposition Eq. (3.6), we obtain

$$4f_{ij} = k[\delta_{i,j+1} - \delta_{j,i+1} + \{f_{i\pm 1,j\pm 1}\}] + k\sum_k [f_{i,k} f_{j,k+1} - f_{i,k+1} f_{j,k}]. \tag{3.10}$$

Since the calculation was made for $i > j$, the subtraction of $\delta_{j,i+1}$ preserves the antisymmetry of f_{ij}. Higher terms decompose similarly. The nonlinear equation (3.10) describes the propagation of order along a diagonal.

In the case of interest f_{ij} is a function of $i - j$. The last term in Eq. (3.10) is the difference of two convolutions, so the equation can be solved by Fourier analysis. Choose n even for simplicity and let

$$f_{jk} = (1/n) \sum_q f_q \exp i(j-k)q. \tag{3.11a}$$

$$\delta_{j,k} = (1/n) \sum_q \exp i(j-k)q, \tag{3.11b}$$

$$q = \pm \pi/n, \pm 3\pi/n, \ldots, \pm [(n-1)/n]\pi.$$

Restricting q to odd multiples of π/n satisfies the anticyclic property of f_{jk}. Substitution of Eqs. (3.11a) and (3.11b) into Eq. (3.10) yields a quadratic equation in f_q:

$$k \sin q(1 + f_q^2) = 2i(1 - k\cos q) f_q. \tag{3.12}$$

This equation can be readily solved by introducing an angle ϕ_q and an amplitude R_q such that

$$R_q \sin 2\phi_q = k \sin q, \tag{3.13a}$$

$$R_q \cos 2\phi_q = 1 - k\cos q. \tag{3.13b}$$

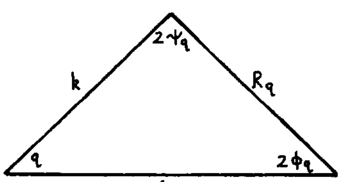


FIG. 3. Euclidean triangle of Eq. (3.13).

The two solutions are

$$f_q = \left\{ \begin{array}{l} -i\tan\phi_q = \dfrac{R_q - 1 + k\cos q}{ik \sin q} \\ i\cot\phi_q = \dfrac{R_q + 1 - k\cos q}{-ik\sin q} \end{array} \right\}. \tag{3.14}$$

The first solution $(-i\tan\phi_q)$ has the proper high temperature limit $f_q \to 0$. The other solution represents excited states within the symmetric subspace.

The relations defining R_q, ϕ_q in Eq. (3.13) are the geometry of the Euclidean triangle in Fig. 3. (A hyperbolic triangle plays a similar role in the analysis of $V_1^{1/2} V_2 V_1^{1/2}$; see Ref. 2, p. 135.)

The wavefunction can be normalized by evaluating

$$N_0^{-2} = \sum_\mu \Phi^2(\mu) = (1 + \sum_{i>j} f_{ij}^2 + \sum_{i>j>k>l} f_{ijkl}^2 + \cdots).$$

But squares of Pfaffians are determinants of skew-symmetric matrices. If F is the $n \times n$ matrix with components f_{ij}, then the normalization constant N_0. satisfies

$$N_0^{-2} = \det(I + F) = |I + F|. \tag{3.15}$$

B. Odd space

A typical member of the odd space is

$$\Phi^-(\mu) = 2^{-n/2}\Big(\sum_i g_i\mu_i + \sum_{i>j>k} g_{ijk}\mu_i\mu_j\mu_k + \sum_{i>j>k>l>m} g_{ijklm}\mu_i\mu_j\mu_k\mu_l\mu_m + \cdots\Big). \tag{3.16}$$

At high temperatures where B dominates, the eigenfunction corresponding to the maximum eigenvalue of B has g_i constant and all other coefficients vanish. To obtain an eigenfunction with this limit we use our previous procedure of extending the coefficients antisymmetrically. We set $g_i = 1$ and let

$$g_{ijk} = g_{ij} - g_{ik} + g_{jk}, \quad g_{ijklm} = g_{ijkl} - g_{ijkm} + g_{ijlm} - g_{iklm} + g_{jklm}, \quad \text{etc.} \tag{3.17}$$

Again we assume that g_{ijkl} can be expressed as a Pfaffian

$$g_{ijkl} = g_{ij}g_{kl} - g_{ik}g_{jl} + g_{il}g_{jk}. \tag{3.18}$$

The boundary condition $g_{n+l,i,j} = g_{ijl}$ and the cyclic symmetry condition $g_{ijk} = g_{i+l,j+l,k+l}$ imply that g_{ij} is a function of the difference $i - j$. But now the boundary condition is satisfied with only even multiples of π/n present in the Fourier series.

Multiplying the eigenvector equation by μ_i, $\mu_i\mu_j\mu_k$, $\mu_i\mu_j\mu_k\mu_l\mu_m, \cdots$ and summing over all $\mu_i = \pm 1$ results in a set of difference equations which as before can be solved by Fourier analysis. Let

$$g_{jk} = (1/n) \sum_q g_q \exp i(j-k)q, \quad q = \pm 2\pi/n \pm 4\pi/n, \cdots. \tag{3.19}$$

then

$$g_q = -i\tan\phi_q \tag{3.20}$$

with ϕ_q determined from the same triangle as before.

C. Triangular lattice

For the triangular and hexagonal lattices the eigenfunctions may again be taken in the form (3.1). Choosing the symmetric form for the transfer matrix of the triangular lattice the pseudo-Hamiltonian is given by $B + k_1 A + k_2 A_2$ in Eq. (2.6). The arguments leading to (3.10) may be repeated exactly leading to the equation determining the coefficients f_{ij}:

$$\begin{aligned} 4f_{ij} &= k_1[\delta_{i,j+1} - \delta_{j,i+1} + \{f_{i\pm 1,j\pm 1}\}] \\ &+ k_2[\delta_{i,j+2} - \delta_{j,i+2} + \{f_{i\pm 2,j\pm 2}\}] \\ &+ k_1 \sum_k (f_{ik} f_{jk+1} - f_{ik+1} f_{jk}) \\ &+ k_2 \sum_k (f_{ik-1} f_{jk+1} - f_{ik+1} f_{jk-1}). \end{aligned} \tag{3.21}$$

This equation can again be solved by introducing the Fourier transforms (3.11) with the result for the triangular lattice

$$f_q = \frac{R_q^T - 1 + k_1 \cos q + k_2 \cos 2q}{i(k_1 \sin q + k_2 \sin 2q)}, \tag{3.22}$$

where

$$R_q^T = [(1 - k_1 \cos q - k_2 \cos 2q)^2 + (k_1 \sin q + k_2 \sin 2q)^2]^{1/2}.$$

At the critical point, $k_1 + k_2 = 1$.

4. CONNECTION WITH OTHER SOLUTIONS

The connection between spinors and fermion operators was used by Schultz, Mattis, and Lieb[4] in their solution of the Ising model. To translate our results into their language, let us regard $\Phi(\mu)$ as the representative $\langle \mu | \Phi \rangle$ of some abstract vector $|\Phi\rangle$. Let the vacuum be denoted by $|0\rangle$, such that $\langle \mu | 0 \rangle = 2^{-n/2}$ is the normalized ground state at high temperatures.

Following Kaufman[5] we introduce the following fermion operators

$$b_j = \tfrac{1}{2} C_1 C_2 \cdots C_{j-1} S_j (I_j - C_j) I_{j+1} \cdots I_n, \tag{4.1a}$$

$$b_j^\dagger = \tfrac{1}{2} C_1 C_2 \cdots C_{j-1} S_j (I_j + C_j) I_{j+1} \cdots I_n, \tag{4.1b}$$

$$\{b_j, b_k^\dagger\} = \delta_{jk}, \quad \{b_j, b_k\} = 0, \quad \text{etc.}$$

Operating on the vacuum $b_j^\dagger$ has the effect of multiplying by μ_j. In fact, a one-particle representation is

$$b_{\mu\mu'} = \tfrac{1}{2}\mu', \quad b_{\mu\mu'}^\dagger = \tfrac{1}{2}\mu.$$

In terms of these operators

$$|\Phi_0\rangle = (1 + \sum_{i>j} f_{ij} b_j^\dagger b_i^\dagger + \sum_{i>j>k>l} f_{ijkl} b_l^\dagger b_k^\dagger b_j^\dagger b_i^\dagger + \cdots)|0\rangle. \tag{4.2}$$

The Pfaffian nature of the coefficients and the fermion commutation rules allow us to express Eq. (4.2) in two ways (Hurst[6] used the following forms to generate Pfaffians):

$$|\Phi_0\rangle = \prod_{i>j} (1 + f_{ij} b_j^\dagger b_i^\dagger)|0\rangle \tag{4.3a}$$

$$= \exp(\sum_{i>j} f_{ij} b_j^\dagger b_i^\dagger)|0\rangle \tag{4.3b}$$

The functional representation of any vector $|\Phi\rangle$ may be recovered by evaluating

$$\langle \mu | \Phi \rangle = 2^{-n/2} \langle 0 | \prod_j (1 + \mu_j b_j) | \Phi \rangle. \tag{4.4}$$

The quadratic form in Eq. (4.3b) is simplified by Fourier analysis. Let us use the transformation introduced by Schultz, Mattis, and Lieb[4]

$$\begin{aligned} b_j &= (1/\sqrt{n}) e^{-i(\pi/4)} \sum_q a_q e^{ijq} \\ a_q &= (1/\sqrt{n}) e^{i(\pi/4)} \sum_j b_j e^{-ijq}, \\ q &= \pm \pi/n, \pm 3\pi/n, \ldots, \pm [(n-1)/n]\pi. \end{aligned} \tag{4.5}$$

The operators $a_q, a_q^\dagger$ satisfy fermion rules also. The quadratic form becomes $-\sum_{q>0} \tan\phi_q a_q^\dagger a_{-q}^\dagger$. The unnormalized ground state is

$$|\Phi_0\rangle = \prod_{q>0} (1 - a_q^\dagger a_{-q}^\dagger \tan\phi_q)|0\rangle. \tag{4.6}$$

With respect to this new vacuum, annihilation and creation operators are easily found.

In Ref. 2, Onsager introduced a Lie algebra as follows:

$$\begin{aligned} &A_0 = -B = -\sum_j C_j, \quad A_{n+k} = -UA_k, \\ &A_1 = A = \sum S_j S_{j+1}, \quad 4G_{k-j} = [A_k, A_j], \\ &A_2 = \sum_j S_{j-1} C_j S_{j+1}, \quad [G_j, A_k] = 2A_{k+j} - 2A_{k-j}, \\ &A_3 = \sum_j S_{j-2} C_{j-1} C_j S_{j+1}, \text{ etc.}, \quad [G_j, G_k] = 0. \end{aligned} \tag{4.7}$$

The operators A_j, C_k were then Fourier-analyzed:

$$\begin{aligned} X_q \pm iY_q &= (2n)^{-1} \sum_{j=1}^{2n} A_j e^{\pm iqj}, \\ Z_q &= (i/2n) \sum_{j=1}^{2n} G_j \sin qj, \\ A_j &= \sum_q X_q \cos qj - Y_q \sin qj. \end{aligned} \tag{4.8}$$

In the (even, odd) space q is an (odd, even) multiple of π/n. In terms of fermion operators we find

$$\begin{aligned} &X_q = n_q + n_{-q} - 1, \quad n_q = a_q^\dagger a_q \\ &Y_q = a_q^\dagger a_{-q}^\dagger + a_{-q} a_q, \quad Z_q = i(a_q^\dagger a_{-q}^\dagger - a_{-q} a_q). \\ &[X_q, Y_q] = -2iZ_q, \quad X_q^2 = Y_q^2 = Z_q^2 = R_q \\ &R_q X_q = X_q, \quad \text{etc.} \end{aligned} \tag{4.9}$$

In the even space the pseudo-Hamiltonian $B + kA = -A_0 + kA_1$ becomes

$$-2 \sum_{q>0} X_q (1 - k \cos q) + Y_q k \sin q$$

or

$$-2 \sum_{q>0} R_q (X_q \cos 2\phi_q + Y_q \sin 2\phi_q)$$

But the coefficient of R_q may be also written as X_q transformed by the unitary operator $\prod_{q>0} \exp(i\phi_q Z_q)$. This is an S matrix which transforms the noninteracting ground state into the new ground state. Using (4.10) we obtain (4.6) as before.

We have thus established the connection between the different representations of the eigenfunction of the transfer matrix corresponding to the explicit Pfaffian solution, the fermion solution, and Onsager's algebraic construction.

5. THE COEFFICIENT f_{ij}

The coefficient f_{ij} is not a correlation function but is related to the existence of long range order in Φ. If f_{ij} has a positive lower bound when site i and j are far apart, then Φ describes an ordered state. We will examine the asymptotic behavior in the case of an infinite lattice. We have

$$f_{r+j,j} = f(r) = (1/n)\sum_q f_q e^{iqr}$$
$$f(r) = \frac{1}{2\pi i}\int_{-\pi}^{\pi} \frac{R(\theta) - 1 + k\cos\theta}{k\sin\theta} e^{ir\theta} d\theta. \tag{5.1}$$

Let $z = e^{i\theta}$, for $k < 1$ the singularities at $\theta = 0, \pi$ are removable, and $f(r)$ can be expressed as a contour integral around the unit circle. This integral can be deformed to surround the branch cut from $z = 0$ to k. Scaling yields

$$f(r) = \frac{k^r}{\pi}\int_0^1 (1-k^2t)^{1/2}(1-t)^{1/2}(1-k^2t^2)^{-1}t^{r-1/2}dt. \tag{5.2}$$

The substitution $t = \exp(-s^2)$ transforms the integral into a canonical form for the study of its asymptotic behavior for large r. The leading terms are

$$f(r) \sim \frac{k^r}{(1-k^2)^{1/2}} \frac{1}{2\sqrt{\pi} r^{3/2}} + O(k^{r+1}), \quad k < 1. \tag{5.3}$$

The vanishing of $f(r)$ for $k < 1$ as $r \to \infty$ is related to the vanishing probability of finding r consecutive spins $(+1)$ surrounded by $(n - r)$ spins (-1) with $n \to \infty$ in the ordered state with modulus k^{-1}.

The case for $k > 1$ differs in that the singularity at $\theta = 0$ is no longer removable. The integral, now considered as a Cauchy principal value, can be expressed by an indented contour plus half the residue at $z = 1$, the contour lying within the unit circle. This integral may also be deformed around branchpoints $z = 0, k^{-1}$. Similar analysis as above yields

$$f(r) \sim 1 - \frac{1}{k} + \frac{k^{-r}}{(1-k^{-2})^{1/2}} \frac{1}{2\sqrt{\pi} r^{3/2}} + O(k^{-r-1}), \quad k > 1. \tag{5.4}$$

Here, of course, there exists a possibility of finding correlated spins in the disordered state. It is interesting to note that the long-range order in $f(r)$ is proportional to $(1 - T/T_c)$ for $T < T_c$.

When $k = 1$, $R(\theta) = 2|\sin(\theta/2)|$, and

$$f(r) = \frac{4}{\pi}\left(\frac{1}{2r+1} - \frac{1}{2r+3} + \frac{1}{2r+5} - \cdots\right) \tag{5.5}$$
$$= \frac{4}{\pi}\int_0^\infty \frac{e^{-2rt}}{\cosh t} dt$$
$$f(r) \sim \frac{1}{\pi}\frac{1}{r} + O\left(\frac{1}{r^3}\right). \tag{5.6}$$

At the critical point, the exponential nature of the amplitudes $f(r)$ changes abruptly.

6. THE PARTITION FUNCTION

It is well known that the partition function Q per site is determined by the maximum eigenvalue Q_n of the transfer matrix

$$Q = \lim_{n\to\infty} (Q_n)^{1/n}, \tag{6.1}$$

with

$$Q_n \Phi_0(\mu) = \sum_{\mu'} W_{\mu\mu'} \Phi_0(\mu'). \tag{6.2}$$

Substituting Eq. (3.1) in Eq. (6.2) with $f_0 = 1$ and summing over all $\mu_i = \pm 1$ we obtain

$$Q_n = 2^{n/2} \sum_\mu \prod_j (\hat{c}_1\hat{c}_2 + \mu_j\mu_{j+1}\hat{s}_1\hat{s}_2)\Phi_0(\mu). \tag{6.3}$$

The product in Eq. (6.3) can be put into a familiar form by introducing

$$\beta\cosh\gamma = \hat{c}_1\hat{c}_2, \quad \beta\sinh\gamma = \hat{s}_1\hat{s}_2. \tag{6.4}$$

The eigenvalue Q_n can then be expressed in vector form

$$Q_n = (2\beta)^n \langle 0|\exp\gamma A|\Phi_0\rangle. \tag{6.5}$$

To evaluate this expression it is convenient to take $|\Phi_0\rangle$ in the form (4.6). From Eq. (4.9) the operator $A = A_1$ takes the following form in the even space:

$$A = 2\sum_{q>0} X_q^*, \tag{6.6}$$

$$X_q^* = X_q \cos q - Y_q \sin q. \tag{6.7}$$

Now X_q^* has the property $(X_q^*)^2|0\rangle = |0\rangle$. Thus the left vector $\langle 0|\exp\gamma A$ becomes

$$\langle 0 \prod_{q>0} (\cosh 2\gamma - \cos q \sinh 2\gamma - a_{-q}a_q \sin q \sinh 2\gamma), \tag{6.8}$$

and

$$Q_n = (2\beta)^n \prod_q (\cosh 2\gamma - \cos q \sinh 2\gamma + \sin q \sinh 2\gamma \tan\phi_q). \tag{6.9}$$

Using (3.13) and (6.4), we get

$$Q_n = \prod_{q>0} 2(c_1c_2 + R_q). \tag{6.10}$$

Taking the infinite limit $n \to \infty$, we can express the partition function as a single integral, symmetric in the bond strengths:

$$\log Q = (2\pi)^{-1}\int_0^\pi d\theta \log 2 [c_1c_2 + (1 + k^2 - 2k\cos\theta)^{1/2}]. \tag{6.11}$$

The transformation connecting Eq. (6.11) with previous integral representations obtained by Onsager is found in Appendix D.

7. THE CORRELATION FUNCTION

The calculation of spin–spin correlation functions $\langle S_i S_j\rangle$ was first accomplished by Kaufman and Onsager.[7] The two spins were located on the same column and the correlation function was obtained as the sum of two Toeplitz determinants. Wick's theorem unknown at that time was later used by Schultz, Mattis, and Lieb[4] in their fermion approach to the Ising model. Correlation functions were also considered by Montroll, Potts, and Ward[8] who used the dimer[9] approach introduced by Kasteleyn[10] and by a number of other authors.[11–13]

The correlation function along a line is expressible either as a single Toeplitz determinant or as the sum of two such determinants depending upon which of the symmetric forms of the columnar transfer matrix is used. In the limit of large separations $i - j \to \infty$, the spin–spin correlation determines the square of the magnetization.

Using the Pfaffian form of the eigenvector Equation (3.1) and the dual transformation, we can give a rather simple derivation of the spin–spin correlations along a diagonal

$$\langle S_j S_i \rangle = \langle \Phi_0(k) | S_j S_i | \Phi_0(k) \rangle, \qquad i > j, \tag{7.1}$$

where Φ_0 is normalized.

In the representation we have chosen the operators $S_1, S_2, \cdots$ are not diagonal but $C_1, C_2, \cdots$ are. The expectation value can be transformed by the dual L to the C operators. Thus, in a sense, the calculation is effected by computing the disorder at the reciprocal modulus. From Eq. (3.3) and the dual L, we have

$$\begin{aligned} L S_j S_i \tilde{L} &= C_{j+1} C_{j+2} \cdots C_i \Lambda_+, \\ L | \Phi_0(k) \rangle &= | \Phi_0(k^{-1}) \rangle. \end{aligned} \tag{7.2}$$

Cyclic invariance implies that we can rewrite the spin-spin correlation as

$$\langle S_j S_i \rangle = \langle \Phi_0(k^{-1}) | C_1 \cdots C_m | \Phi(k^{-1}) \rangle, \qquad m = i - j. \tag{7.3}$$

The operation of C_k is to change the sign of any term containing μ_k. Thus $\langle S_j S_i \rangle$ can be expressed as the ratio of two determinants, the denominator being, of course, the normalization constant $|I + F(k^{-1})|$ from Eq. (3.15). The numerator is also of this form. Thus

$$\langle S_j S_i \rangle = |I + \hat{F}(k^{-1})| / |I + F(k^{-1})|. \tag{7.4}$$

The components of the $n \times n$ matrix $\hat{F}(k^{-1})$ are given by

$$\begin{aligned} \hat{F}_{rs}(k^{-1}) &= \epsilon_r \epsilon_s F_{rs}(k^{-1}), \\ \epsilon_r &= \begin{cases} i, & 1 \leqslant r \leqslant m \\ 1, & m < r < n \end{cases}. \end{aligned} \tag{7.5}$$

Multiplying the m top rows and the m left columns of $I + F$ by $-i$, we find that

$$\langle S_j S_i \rangle = (-1)^m | I - 2 I_m [I + F(k^{-1})]^{-1} |, \tag{7.6}$$

where I_m is an $n \times n$ diagonal matrix whose first m elements are unity, all others vanishing. The $n \times n$ matrix $I + F(k^{-1})$ can be readily inverted by using Eqs. (3.11a) and (3.11b). Its components are

$$[I + F(k^{-1})]^{-1}_{rs} = \tfrac{1}{2}\delta_{rs} + (1/2n) \sum_q \exp i[2\phi_q(k^{-1}) + q(r - s)]. \tag{7.7}$$

The right-hand side of Eq. (7.6) becomes, after substituting Eq. (7.7), an $m \times m$ determinant generated by $\exp 2i\phi_q(k^{-1}) = \exp[2i\psi_q(k)]$. In the limit $n \to \infty$, $\langle S_j S_i \rangle$ is represented by the determinant of an $m \times m$ Toeplitz matrix $T^{(m)}$ whose components are Fourier transforms of $\exp[2i\psi(\theta, k)]$:

$$\langle S_j S_{j+m} \rangle = |T^{(m)}|, \tag{7.8}$$

$$T^{(m)}{}_{rs} = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \ \exp\{ i[(r - s)\theta + 2\psi(\theta)] \}, \qquad 1 \leqslant r, \ s \leqslant m, \tag{7.9}$$

with

$$\exp[2i\psi(\theta)] = \begin{cases} -e^{-i\theta}(1 - ke^{i\theta})^{1/2}(1 - ke^{-i\theta})^{-1/2}, & k < 1 \\ (1 - k^{-1}e^{-i\theta})^{1/2}(1 - k^{-1}e^{-i\theta})^{-1/2}, & k > 1 \end{cases}. \tag{7.10}$$

The generating function for correlations along a diagonal is simpler than the function for the columns which contains four factors instead of two. The angle involved is an element of a hyperbolic triangle.

ACKNOWLEDGMENTS

We would like to thank Professor Lars Onsager for his interest and encouragement. M. J. S. is grateful to Professor F. R. N. Nabarro F.R.S. for hospitality at the University of the Witwatersrand where part of this work was done. We would also like to thank Professor G. Benedek for hospitality at M. I. T. where this work was completed.

APPENDIX A

The product $(WW')_{\mu\mu'}$, pictured in Fig. 4a can be written

$$(WW')_{\mu\mu'} = \sum_{\lambda = \pm 1} \prod_j \exp[\mu_j(H_1\lambda_j + H_2\lambda_{j+1}) + \mu'_{j+1}(H'_2\lambda_j + H'_1\lambda_{j+1})]. \tag{A1}$$

Each factor can also be represented by

$$(\alpha_j + \beta_j\lambda_j)\delta(\lambda_j - \lambda_{j+1}) + (\gamma_j + \delta_j\lambda_j)\delta(\lambda_j + \lambda_{j+1})$$

or in matrix form by

$$T_j = \alpha_j I + \beta_j S + \gamma_j C + \delta_j SC, \tag{A2}$$

with

$$\begin{aligned} \alpha_j &= \cosh[\mu_j(H_1 + H_2) + \mu'_{j+1}(H'_1 + H'_2)], \\ \gamma_j &= \cosh[\mu_j(H_1 - H_2) + \mu'_{j+1}(H'_2 - H'_1)], \\ \beta_j &= \sinh[\mu_j(H_1 + H_2) + \mu'_{j+1}(H'_1 + H'_2)], \\ \delta_j &= \sinh[\mu_j(H_1 - H_2) + \mu'_{j+1}(H'_2 - H'_1)]. \end{aligned} \tag{A3}$$

Thus, WW' can be expressed as a trace:

$$WW' = \operatorname{Tr} \prod_j T_j. \tag{A4}$$

Similarly for $W'W$ let us call the corresponding matrix T'_j whose coefficients $\alpha'_j, \beta'_j, \gamma'_j, \delta'_j$ are related to the preceding ones by

$$\begin{aligned} \alpha'_j &= \alpha_j, & \gamma'_j &= \gamma_j, \\ \beta'_j &= \mu_j\mu'_{j+1}\beta_j, & \delta'_j &= -\mu_j\mu'_{j+1}\delta_j. \end{aligned} \tag{A5}$$

Following a technique similar to Baxter,[14] we seek a nonsingular matrix R independent of j such that

$$T_j R = R T'_j.$$

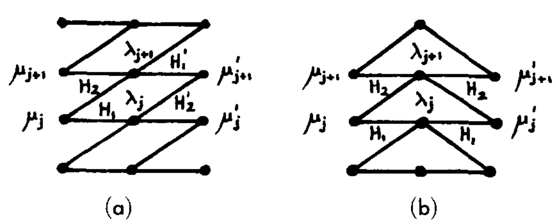


FIG. 4 The products (a) WW' and (b) WW^τ.

If there exists such a matrix R, then the operators W and W' will commute. It is sufficient to examine matrices of the form $R = a + bC$. A pair of linear equations result from this choice of R in Eq. (A6). Eliminating a and b from these equations, we obtain the condition

$$\sinh 2H_1 \sinh 2H_2 = \sinh 2H_1' \sinh 2H_2'.$$

Examination of Fig. 4b indicates that a similar procedure can be used to prove that W is normal:

$$[W, W^+] = 0.$$

APPENDIX B

We will consider the more symmetric form of the triangular transfer matrix

$$W_s(H_1, H_2, H_3) = V_2(\tfrac{1}{2}H_3) W(H_1, H_2) V_2(\tfrac{1}{2}H_3). \tag{B1}$$

Premultiplying W_s by $C_j + k_2 S_{j-1} C_j S_{j+1}$, we obtain

$$(1 + k_2 \mu_{j-1}\mu_{j+1}) \exp[-H_3\mu_j(\mu_{j-1} + \mu_{j+1}) - 2(H_1\mu_j\mu_{j+1}' + H_2\mu_j\mu_{j+1}')](W_s)_{\mu\mu'}.$$

The choice of $k_2 = -\tanh^2 H_3 = -\hat{s}_3{}^2/\hat{c}_3{}^2$ simplifies the factor depending on H_3 which becomes with the notation of Eqs. (2.7)

$$\hat{c}_3{}^{-2}[c_3 - \tfrac{1}{2}\mu_j(\mu_{j-1} + \mu_{j+1})s_3]. \tag{B2}$$

Similarly, postmultiplication yields

$$\hat{c}_3{}^{-2}[c_3 - \tfrac{1}{2}\mu_j'(\mu_{j-1}' + \mu_{j+1}')s_3] \exp[-2H_1\mu_j\mu_j' + H_2\mu_{j-1}\mu_j')(W_s)_{\mu\mu'}]. \tag{B3}$$

Subtracting the two forms and then summing over j we find after some detailed algebra that

$$[B + k_2A_2, W_s] = -\hat{c}_3{}^{-2}(s_1s_2c_3 + c_1c_2s_3)[A_1, W_s]. \tag{B4}$$

which is Eq. (2.6) with k_1 and k_2 given by Eq. (2.8).

Interchanging H_2 and H_3 in Eq. (B4) and then taking the limit $H_3 \to 0$, we obtain a new result that

$$[V_2(H_2)V_1(H_1), B + (c_1s_2/c_2)A_1 + \tfrac{1}{4}(s_2/c_2)[A, B]] = 0. \tag{B5}$$

The operator in Eq. (B5) can be put in a summetric form by multiplying by $2c_2/s_2$ and using Eq. (1.2). The pseudo-Hamiltonian becomes

$$(x_2 + x_2{}^{-1})B + (x_1^* + x_1^{*-1})A + \tfrac{1}{2}[A, B], \qquad x_1^* = \tanh H_1^*, \quad x_2 = \tanh H_2. \tag{B6}$$

APPENDIX C

We will now exhibit the relationship between the dual transformation and the diagonal transfer matrix evaluated at $H_1 = -H_2 = \pm\frac{1}{4}\pi i$. This result is suggested by the fact that the modulus k is unity at the critical point and also at the above values of H_1 and H_2. Consider

$$W(\tfrac{1}{4}\pi i, -\tfrac{1}{4}\pi i) = \prod_j 2^{-1/2}(1 + i\mu_j\mu_j')2^{-1/2}(1 - i\mu_{j-1}\mu_j') = \prod_j [\delta(\mu_j - \mu_{j-1}) + i\mu_j\mu_j'\delta(\mu_j + \mu_{j-1})]. \tag{C1}$$

But unity may be expressed by

$$1 = \prod_j \tfrac{1}{2}(1 - i\mu_j)(1 + i\mu_{j-1}), \qquad 1 = \prod_j [\delta(\mu_j - \mu_{j-1}) - i\mu_j\delta(\mu_j + \mu_{j-1})]. \tag{C2}$$

Thus multiplying Eq. (C1) by Eq. (C2), we find that

$$W(\tfrac{1}{4}\pi i, -\tfrac{1}{4}\pi i) = \prod_j [\delta(\mu_j - \mu_{j-1}) + \mu_j'\delta(\mu_j + \mu_{j-1})], \tag{C3}$$

which aside from a normalization constant is essentially the dual transformation. Complex conjugation of Eq. (C3) completes the demonstration.

APPENDIX D

Two integral representations found by Onsager for the partition function are

$$\log Q = \tfrac{1}{2}\log 2s_1 + \frac{1}{2\pi}\int_0^\pi d\theta \cosh^{-1}(c_1^*c_2 - s_1^*s_2\cos\theta), \tag{D1}$$

$$= \frac{1}{8\pi^2}\int_{-\pi}^{\pi} d\theta_2 \int_0^{2\pi} d\theta_1 \log 4(c_1c_2 - s_1\cos\theta_1 - s_2\cos\theta_2). \tag{D2}$$

Equation (D2) was actually given over a smaller domain; but this is compensated by the numerical factor. Consider the transformation

$$\theta_1 = \tfrac{1}{2}\theta + \omega, \quad \theta_2 = \tfrac{1}{2}\theta - \omega, \tag{D3}$$

which effects a rotation of $\pi/4$ in the coordinate system. Judicious translations by 2π enable us to obtain a domain of integration; $0 \leqslant \theta \leqslant 2\pi$, $0 \leqslant \omega \leqslant 2\pi$. Integration over ω is easily performed by noting that

$$\frac{1}{2\pi}\int_0^{2\pi} d\omega \log(a + b\cos\omega + c\sin\omega) = \log\tfrac{1}{2}[a + (a^2 - b^2 - c^2)^{1/2}]. \tag{D4}$$

Finally we obtain Eq. (6.11):

$$\log Q = \frac{1}{2\pi}\int_0^\pi d\theta \log 2[c_1c_2 + (1 + k^2 - 2k\cos\theta)^{1/2}]. \tag{6.11}$$

APPENDIX E

The star-triangle transformation was mentioned briefly by Onsager in Ref. 2 and later in his talk at the Batelle Institute.[1] We would like to make some observations on the geometrical significance of this transformation. First let us define it. The hexagonal Ising model has a coordination number of three so that the spins at a selected set of sites can be summed to get a triangular lattice. At any one such site we have (see Fig. 5)

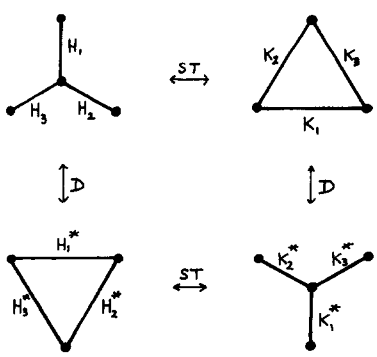


FIG. 5. The star-triangle transformation (ST) combined with the dual transformation (D).

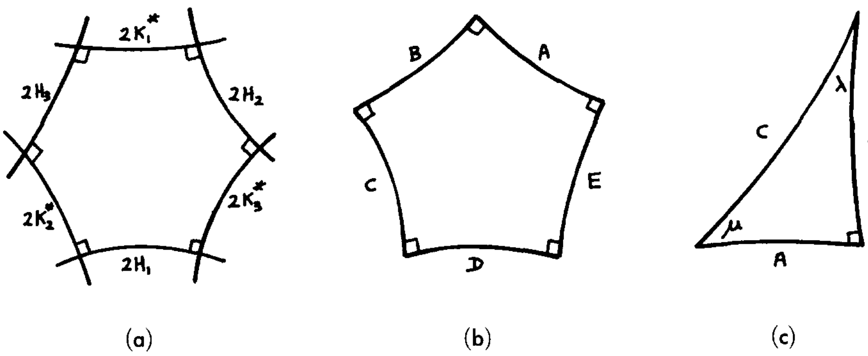


FIG. 6. (a) The right-angled hyperbolic hexagon, (b) right-angled pentagon, and (c) a right triangle.

$$2\cosh(H_1\mu_1 + H_2\mu_2 + H_3\mu_3) = N\exp(K_1\mu_2\mu_3 + K_2\mu_3\mu_1 + K_3\mu_1\mu_2). \qquad \text{(E1)}$$

From (E1) we extract three equations, one of which is

$$x_1x_2 = (y_1y_2 + y_3)/(1 + y_1y_2y_3), \quad x_i = \tanh H_i, \; y_i = \tanh K_i. \qquad \text{(E2)}$$

The others are found by cyclic permutation. We can solve for x_1 by considering $(x_1x_2)(x_3x_1)/(x_2x_3)$; but solving for the triangular bonds is slightly more difficult. The solution is

$$y_1^*y_2^* = (x_1^*x_2^* + x_3^*)/(1 + x_1^*x_2^*x_3^*), \quad \text{etc.}, \quad x_i^* = \tanh H_1^*, \; y_i^* = \tanh K_i^*. \qquad \text{(E3)}$$

Equation (E3) is in the same form as (E2) with x and y interchanged, and the bonds are replaced by their duals. This completes the proof of the transformation implied in Fig. 5. From (E2) we can obtain two sets of equations which are related to the laws of sines and cosines. They are

$$\frac{\sinh 2H_1}{\sinh 2K_1^*} = \frac{\sin 2hH_2}{\sinh 2K_2^*} = \frac{\sin 2H_3}{\sinh 2K_3^*}, \text{[15]} \qquad \text{(E4a)}$$

$$\cosh 2K_1^* = -\cosh 2K_2^* \cosh 2K_3^* + \sinh 2K_2^* \sinh 2K_3^* \cosh 2H_1, \quad \text{etc.} \qquad \text{(E4b)}$$

Onsager[1] noted that the hexagonal bonds $2H_1, 2H_2, 2H_3$ and $2K_1^*, 2K_2^*, 2K_3^*$ were sides of mutually polar hyperbolic (hyp.) triangles. He also noted that the sextuple $2H_1, 2K_3^*, 2H_2, 2K_1^*, 2H_3, 2K_2^*$ in that order formed the sides of a completely right-angled hyp. hexagon. The two representations are, of course, related, but it is to the latter that we address ourselves. If we can derive the relations (E4) from the right-angled hexagon, we will have proven the geometrical equivalence. In hyp. geometry[16] any pair of nonintersecting (this excludes parallel) lines has a unique perpendicular line. Thus, there exists a unique perpendicular to the lines containing $2H_1$ and $2K_1^*$ [see Fig. 6a]. This perpendicular divides the hexagon into two right-angled pentagons. Now a right-angled pentagon has the property that the hyp. cosine of any side is equal to the product of the hyp. cotangents of the adjacent sides and also to the products of the hyp. sines of the opp. sides.[17] Thus in Fig. 6b

$$\cosh A = \coth B \coth E = \sinh C \sinh D, \quad \text{etc.} \qquad \text{(E5)}$$

These equations (E5) may be used to obtain the trigonometry of the hexagon in the same way as a general triangle is analyzed by dividing it into two right triangles; Eqs. (E4a) and (E4b) are obtained. Indeed there is a complete similarity between the rules for a right-angled pentagon (E5) and a right triangle in hyp. geometry. A mapping of the five elements of each exists.

There is a mnemonic device due to Napier which gives the trigonometry of a right spherical triangle. This device can be generalized for hyp. right triangles and are sometimes called the Engel–Napier rules. These rules coincide exactly with those of the pentagon with the following prescription: Let the sides and angles of a right hyp. triangle be A, B, C, and λ, μ as in Fig. 6c; the presence of angles λ, μ is inconvenient so we replace them by hyp. elements L, M for which they are the angles of parallelism.[18] Thus λ is the complement of the gudermannian of L or

$$\cos\lambda = \tanh L, \qquad \cos\mu = \tanh M. \qquad \text{(E6)}$$

Then the elements A^*, C, B^*, L, M related to the right hyp. triangle may be assigned in that order as consecutive sides of a right-angled pentagon. The same trigonometric equations hold!

---

*Supported in part by the National Science Foundation and the Research Council of Rutgers University.

[1] L. Onsager, *Critical Phenomena in Alloys, Magnets and Superconductors*, edited by Mills, Ascher, and Jaffee (McGraw-Hill, New York, 1971), p. 3.

[2] L. Onsager, Phys. Rev. **65**, 117 (1944).

[3] M. Suzuki, Progr. Theor. Phys. **46**, 1337 (1971).

[4] T. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

[5] B. Kaufman, Phys. Rev. **76**, 1232 (1949).

[6] C. A. Hurst, J. Math. Phys. (N.Y.) **7**, 305 (1966).

[7] B. Kaufman and L. Onsager, Phys. Rev. **76**, 1244 (1949).

[8] E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. (N.Y.) **4**, 308 (1963).

[9] The correlations along the diagonal can be obtained by distorting the dimer lattice so that diagonals become columns while rows remain the same.

[10] P. Kasteleyn, J. Math. Phys. (N.Y.) **4**, 308 (1963).

[11] L. P. Kadanoff, Nuovo Cimento B **44**, 276 (1966).

[12] J. Stephenson, J. Math. Phys. (N.Y.) **5**, 1009 (1964).

[13] T. T. Wu, Phys. Rev. **149**, 380 (1966).

[14] R. J. Baxter, Ann. Phys. (N.Y.) **70**, 193 (1972).

[15] At the critical point this ratio is unity.

[16] D. M. Y. Sommerville, *The Elements of Non-Euclidean Geometry* (Dover, New York, 1958), Chap. II.

[17] This theorem is given as a problem on p. 87, Ref. 16.

[18] See Ref. 16, Chap. II, Sec. 27.

On the inverse problem for a hyperbolic dispersive partial differential equation

V. H. Weston

Division of Mathematical Sciences, Purdue University, Lafayette, Indiana 47907

(Received 3 July 1972)

The inverse problem for a two-dimensional (space-time) hyperbolic partial differential equation, with coefficients, functions of the spatial variable only, is considered. Exterior to a region of compact support in the spatial variable, the equation reduces to the wave equation, and, from knowledge of the solution in the exterior region (namely in terms of reflected and transmitted waves for a prescribed incident wave), the problem is to deduce the coefficients in the interior region. This is achieved by treating the problem as a Cauchy initial value problem and using the Riemann function to deduce a dual set of integral equations. The coefficients or linear combinations of them are deduced from the solutions of the integral equations. The question of uniqueness is partially answered, by estimating the domain of convergence of the Neumann series. The application of the analysis to electromagnetic scattering from a slab of varying conductivity and permittivity is indicated.

The inverse problem consists of determining the coefficients of a partial differential equation from the knowledge of the asymptotic behavior of the solution.[1-4] In many of the physical problems involving spatial and time-independent variables, with the coefficients depending upon the spatial variable only, the analysis is based upon the determination of the coefficients from the spectrum of the partial differential operator associated with the spatial variable. The time-dependent approach was considered by Kay,[5,6] who transformed the Gel'fand-Levitan equation[7] into a resulting time-dependent integral equation. Sondi and Gopinath[8] and recently Niznik[9] have examined the time-dependent problem directly.

Here we consider the time-dependent inverse problem directly by making use of the theory of hyperbolic differential equations. The equation to be considered, of two independent variables, has coefficients functions of the spatial variable only, but contains a dispersive term, i.e., a term involving the first derivative of the time variable. It is shown that a dual set of generalized Gel'fand-Levitan type integral equations are obtained, involving the transmission and reflection coefficients. The solution of these equations leads to the determination of the unknown coefficients of the original partial differential equation. Application of the results to electromagnetic scattering is considered.

CAUCHY PROBLEM AND THE SCATTERING OPERATOR

The differential equation to be considered is the following

$$u_{xx} - u_{tt} + A(x)u_x + B(x)u_t + C(x)u = 0, \tag{1}$$

where A, B and their derivatives and C and continuous functions of compact support vanishing outside the domain $0 < x < l$. $B(x)$ will be taken to be negative corresponding to most physical situations where energy is absorbed. An application of the above differential equation to electromagnetic theory will be given below.

For an arbitrary incident wave $u_+^i(x-t)$ propagating in the direction of the positive x axis, such that $u_+^i(s) = 0$ for $s > \lambda$, there is no loss in generality if we take $\lambda = 0$, since this can be achieved by a linear transformation of the variable t, without affecting Eq. (1). Hence we will consider the class of twice continuously differentiable functions $u_+^i(s)$ which vanish for $s > 0$. It follows that Eq. (1) subject to initial conditions

$$u(x,0) = u_+^i(x), \qquad u_t(x,0) = -u_+^{i\prime}(x)$$

can be transformed to a Volterra integral equation for $t \geq 0$, from which it may be deduced that $u(x,t) = 0$ for $x - t > 0$. In addition it can be shown that, for $x \leq 0$,

$$u(x,t) = u_+^i(x-t) + u_+^r(x+t),$$

where the reflected component $u_+^r(s)$ vanishes for $s < 0$ and that, for $x \geq l, u(x,t) = u_+^t(x-t)$, where the transmitted wave satisfies the causality condition $u_+^t(s) = 0, s > 0$.

For an arbitrary incident wave of the form $u_-^i(x+t)$ propagating in the direction of the negative x axis, the class of twice continuously differentiable functions to be considered will be those for which $u_-^i(s) = 0$ for $s < 0$. These will give rise to reflected wave component $u_-^r(x-t)$ in the domain $x \geq l$, such that $u_-^r(s) = 0$ for $s > 2l$ and a transmitted wave $u_-^t(x+t)$ in the domain $x \leq 0$, such that $u_-^t(s) = 0$ for $s < 0$.

In order to obtain the functional relationship between the reflected, transmitted, and incident portions the following lemma is needed.

Lemma: The solution to Eq. (1) subject to conditions at $x = \nu$, where ν lies outside $0 < x < l$,

$$u(\nu,t) = v(\nu - t) + w(\nu + t),$$
$$u_x(\nu,t) = v'(\nu - t) + w'(\nu + t)$$

is given by

$$\begin{aligned} u(x,t) = {} & \exp\left[\tfrac{1}{2}\int_x^\nu [A(\tau) + B(\tau)]d\tau\right] \\ & \times \left(w(x+t) - \int_{2\nu - x}^x K_-(x,y,\nu)w(y+t)dy\right) \\ & + \exp\left[\tfrac{1}{2}\int_x^\nu [A(\tau) - B(\tau)]d\tau\right] \\ & \times \left(v(x-t) - \int_{2\nu-x}^x K_+(x,y,\nu)v(y-t)dy\right), \end{aligned} \tag{2}$$

where $K_\pm(x,t,\nu)$ satisfy the differential equation

$$\left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2} \pm B(x)\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right) + D_\pm(x)\right]K_\pm = 0$$

and boundary conditions

$$K_\pm(x, 2\nu - x, \nu) = 0,$$
$$K_\pm(x,x,\nu) = -\tfrac{1}{2}\int_x^\nu D_\pm(\tau)d\tau,$$

with $D_\pm(x) = C(x) - \frac{1}{2}A'(x) \pm \frac{1}{2}B'(x)\frac{1}{4}(B^2 - A^2)$.

A direct proof is obtained by expressing Eq. (1) in terms of characteristic coordinates (ξ, η), where $\xi = x + t$ and $\eta = x - t$, yielding the following differential equation:

$$L(\xi,\eta)u = 0.$$

Then on employing the Riemann function[10] $g(\xi,\eta,\xi_0,\eta_0)$, which satisfies the adjoint equation, and the following boundary conditions

$$g = \exp\left(\tfrac{1}{4}\int_{\xi_0+\eta_0}^{\xi_0+\eta}[A(\tau/2) + B(\tau/2)]d\tau\right) \quad \text{for } \xi = \xi_0,$$

$$g = \exp\left(\tfrac{1}{4}\int_{\xi_0+\eta_0}^{\xi+\eta_0}[A(\tau/2) - B(\tau/2)]d\tau\right) \quad \text{for } \eta = \eta_0,$$

it follows that

$$u(\xi_0,\eta_0) = [vg]_P + [wg]_Q - \int_P^Q \frac{\partial g}{\partial \xi} w(\xi)d\xi + \int_P^Q \frac{\partial g}{\partial \eta} v(\eta)d\eta,$$

where the integrals are along the line $\xi + \eta = 2\nu$ and P and Q have characteristic coordinates $(2\nu - \eta_0, \eta_0)$ and $(\xi_0, 2\nu - \xi_0)$, respectively. The first integral is reduced by noting that g as a function of ξ_0 and η_0 satisfies the differential equation

$$L(\xi_0,\eta_0)g = 0$$

and boundary conditions the same as above. From this, the differential equation and boundary conditions for $\partial g/\partial\xi$ as a function of (ξ_0,η_0) are easily obtained. Upon transforming of the variables, $\xi = y + t, \eta = 2\nu - y - t$, $\xi_0 = x + t$, and $\eta_0 = x - t$, the resulting expression in terms of $K_-(x,y,\nu)$ is obtained. The second integral is reduced in a similar manner.

Expression (2) with $x \le 0, \nu = l$, may be employed to obtain the following functional relationship between the incident, reflected, and transmitted components:

$$u_+^i(x-t) + u_+^r(x+t) = \exp\left[\tfrac{1}{2}\int_0^l (A-B)d\tau\right]\left(u_+^t(x-t) + \int_x^{2l-x} K_+(x,y,l)u_+^t(y-t)dy\right). \quad (3)$$

From the differential equation and associated boundary conditions for K_+, it follows that K_+ may be decomposed:

$$K_+(x,t,l) = L_+(x-t) + M_+(x+t),$$

$$x \le 0, \ x \le t \le 2l - x,$$

where

$$L_+(s) = L_+(-2l) \quad \text{for } s \le 2l,$$

$$M_+(s) = 0 \quad \text{for } s \le 0,$$

$$M_+(2l) + L_+(-2l) = 0, \quad L_+(0) = \tfrac{1}{2}\int_0^l D_+(\tau)d\tau.$$

If the domain of definition of $M_+(s)$ is extended as follows,

$$M_+(s) = M_+(2l) \quad \text{for } s \ge 2l,$$

Eq. (3) can be partitioned to yield

$$u_+^i(\eta) = \exp\left[\tfrac{1}{2}\int_0^l (A-B)d\tau\right]\left(u_+^t(\eta) + \int_\eta^0 L_+(\eta - s)u_+^t(s)ds\right), \quad (4)$$

$$u_+^r(\xi) = \exp\left(\tfrac{1}{2}\int_0^l (A-B)d\tau\right)\int_{-\xi}^0 M_+(\xi + s)u_+^t(s)ds \quad (5)$$

for transmitted waves $u_+^t(s)$ which vanish for $s > 0$. The above yields $u_+^i(\eta) = 0$ and $u_+^r(\xi) = 0$ for $\eta > 0$ and $\xi < 0$, respectively.

The inversion of (4) yields the forward scattering operator (with direction of incidence along the positive x axis), mapping the incident wave into the transmitted wave

$$u_+^t(\eta) = \exp\left[-\tfrac{1}{2}\int_0^l (A-B)d\tau\right]\left(u_+^i(\eta) + \int_\eta^0 T_+(\eta - s)u_+^i(s)ds\right), \quad (6)$$

where

$$T_+(\eta) + L_+(\eta) + \int_\eta^0 L_+(\eta - y)T_+(y)dy = 0, \quad \eta \le 0. \quad (6)$$

The back-scattering operator (direction of incidence along the positive x axis), mapping the incident wave into the reflected wave is given by

$$u_+^r(\xi) = \int_{-\xi}^0 R_+(\xi + s)u_+^i(s)ds, \quad (7)$$

where

$$R_+(\xi) = M_+(\xi) + \int_{-\xi}^0 M_+(\xi + y)T_+(y)dy. \quad (7')$$

In a similar manner the scattering operator may be obtained for the direction of incidence in the negative x direction. $K_-(x,t,0)$ has the decomposition

$$K_-(x,t,0) = -L_-(x-t) - M_-(x+t), \quad x \ge l, -x \le t \le x,$$

where

$$L_-(s) = L_-(2l) \quad \text{for } s \ge 2l,$$

$$M_-(s) = 0 \quad \text{for } s > 2l,$$

$$M_-(0) + L_-(2l) = 0, \quad L_-(0) = -\tfrac{1}{2}\int_0^l D_-(\tau)d\tau.$$

Extending the domain of definition of $M_-(s)$ as follows,

$$M_-(s) = M_-(0) \quad \text{for } s \le 0,$$

one obtains

$$u_-^i(\xi) = \exp -\left[\tfrac{1}{2}\int_0^l (A+B)d\tau\right]\left(u_-^t(\xi) + \int_0^\xi L_-(\xi - s)u_-^t(s)ds\right) \quad (8)$$

$$u_-^r(\eta) = \exp -\left(\tfrac{1}{2}\int_0^l (A+B)d\tau\right)\int_0^{2l-\eta} M_-(s+\eta)u_-^t(s)ds, \quad (9)$$

for transmitted waves $u_-^t(s)$ which vanish for $s < 0$. The forward- and back-scattering operators are obtained as before:

$$u_-^t(\xi) = \exp\left[\tfrac{1}{2}\int_0^l (A+B)d\tau\right]\left(u_-^i(\xi) + \int_0^\xi T_-(\xi - s)u_-^i(s)ds\right), \quad (10)$$

$$u_-^r(\eta) = \int_0^{2l-\eta} R_-(s+\eta)u_-^i(s)ds, \quad (11)$$

where

$$T_-(\xi) + L_-(\xi) + \int_0^\xi L_-(\xi - s)T_-(s)ds = 0 \quad \text{for } \xi \ge 0, \quad (12)$$

$$R_-(\eta) = M_-(\eta) + \int_0^{2l-\eta} M_-(\eta + y)T_-(y)dy. \quad (13)$$

RELATIONS BETWEEN THE SCATTERING KERNELS

The scattering kernels $R_\pm, T_\pm$ are not completely independent of each other. To develop relationships between them, consider first the case of the incident wave propagating in the positive x direction. Equation (2) with $\nu = 0$, combined with causality, is used to obtain, for $x \ge 0, x > t$,

$$u_+^i(x-t) - \int_{-x}^{x} K_+(x,y,0)u_+^i(y-t)dy + G(x) \times \left(u_+^r(x+t) - \int_{-x}^{x} K_-(x,y,0)u_+^r(y+t)dy\right) = 0, \quad (14)$$

where

$$G(x) = \exp\left(-\int_0^x B(\tau)d\tau\right) \quad (15)$$

Since u_+^r may be expressed in terms of u_+^i which is arbitrary, the above yields

$$K_+(x,t,0) = G(x)\left(R_+(x+t) - \int_{-t}^{x} K_-(x,y,0)R_+(y+t)dy\right) \quad (16)$$

for $-x \le t \le x, x \ge 0$. In particular, for $x \ge l$, it follows from the differential equation that $K_+(x,t,0)$ may be decomposed in the form

$$K_+(x,t,0) = G(l)[P_+(x+t) + Q_+(x-t)],$$

$$P_+(\xi) = R_+(\xi) + \int_0^{\xi} R_+(s)L_-(\xi - s)ds, \quad (17)$$

$$Q_+(\eta) = \int_0^{2l-\eta} M_-(\eta + s)R_+(s)ds, \qquad \eta \ge 0. \quad (18)$$

Note that $P_+(\xi)$ and $Q_+(\eta)$ vanish for $\xi \le 0$ and $\eta \ge 2l$, respectively, and that $P_+(\xi)$ is constant for $\xi \ge 2l$. Using the identity

$$1 - \int_{-x}^{x} K_+(x,y,0)dy = G(l)\left(1 - \int_{-x}^{x} K_-(x,y,0)dy\right), \qquad x \ge l,$$

one finds the value of the constant to be $M_-(0)$; hence

$$P_+(\xi) = M_-(0) \quad \text{for } \xi \ge 2l.$$

On employing Eq. (2), for $x \ge l, t > x$, one obtains an equation similar to Eq. (14), but with the right-hand side replaced by

$$u_+^t(x-t)\exp\left(\tfrac{1}{2}\int_0^l (A-B)d\tau\right).$$

Since u_+^t and u_+^r may be expressed in terms of u_+^i using the scattering operators, one obtains, for arbitrary u_+^i,

$$T_+(\eta) = G(l)[Q_+(\eta) + M_-(0)] \quad \text{for } \eta \le 0. \quad (19)$$

This last equation allows one to compute T_+ from knowledge of the kernels $T_-, R_\pm$.

The case for propagation of the incident wave in the direction of the negative x axis is treated in a similar manner. For $x \le l$, and $x \le t \le 2l - x$, one obtains

$$G(x)K_-(x,t,l) + G(l)\left(R_-(x+t) + \int_x^{2l-t} K_+(x,y,l)R_-(y+t)dy\right) = 0 \quad (20)$$

and, in particular for $x \le 0$,

$$K_-(x,t,l) = -G(l)\{P_-(x+t) + Q_-(x-t)\},$$

where P_- and Q_- are defined as:

$$P_-(\xi) = R_-(\xi) + \int_{\xi}^{2l} R_-(s)L_+(\xi - s)ds, \quad (21)$$

$$Q_-(\eta) = \int_{-\eta}^{2l} R_-(s)M_+(\eta + s)ds, \quad (22)$$

and it can be shown that,

$$P_-(\xi) = M_+(2l) \quad \text{for } \xi \le 0,$$

$$T_-(\eta) = G(l)[Q_-(\eta) + M_+(2l)] \quad \text{for } \eta \ge 0. \quad (23)$$

INVERSE PROBLEM

The inverse problem consists of determining the coefficients A, B, C, when the scattering kernels $R_\pm$, $T_\pm$ and the attenuation factor

$$G(l)^{-1} = \exp\left(+\int_0^l B(\tau)d\tau\right)$$

are given. In actual physical practice, these kernels are measured directly by using incident waves which closely approximate a delta function.

First consider the case where the direction of incidence is in the positive x direction. For a fixed x, lying in the domain $0 \le x \le l$, use Eq. (2) to express $u_+(x,t)$ in terms of the incident and reflected components on the boundary $x = 0$. Similarly express $u_+(x,t)$ in terms of the transmitted component on the boundary $x = l$. Equate these two expressions, and represent u_+^i and u_+^r in terms of u_+^t, by employing the scattering operators. Since u_+^t is arbitrary, one obtains the following integral equations:

$$L_+(x-t) + G(x)S_+(x,t) - G(x)\int_{-x}^{x} K_-(x,y,0)S_+(y,t)dy \quad (24a)$$

$$= \begin{cases} K_+(x,t,l) \\ 0 \end{cases}, \quad (24b)$$

where Eqs. (24a) and (24b) hold for $x \le t \le 2l - x$ and $2l - x \le t$, respectively, and

$$S_+(y,t) = R_+(y+t) + \int_x^t R_+(y+s)L_+(s-t)ds.$$

In a similar manner, the following set of integral equations is obtained for the direction of incidence in the negative x direction:

$$G(x)L_-(x-t) + G(l)S_-(x,t) + G(l)\int_x^{2l-x} K_+(x,y,l)S_-(y,t)dy \quad (25a)$$

$$= \begin{cases} -K_-(x,t,0)G(x) \\ 0 \end{cases}, \quad (25b)$$

where Eqs. (25a) and (25b) hold for $-x \le t \le x$ and $t \le -x$, respectively, and

$$S_-(y,t) = R_-(y+t) + \int_t^x R_-(s+y)L_-(s-t)ds.$$

For fixed x, the above constitutes a set of integral equations, where $L_\pm$ and $S_\pm$ are known functions since they can be determined from $R_\pm$ and $T_\pm$. However, $G(x)$ is unknown, but, for fixed x, can be taken as a parameter which occurs linearly, if one solves for the unknown quantities $K_+(x,t,l)$ and $G(x)K_-(x,t,0)$.

Note that if only R_+ and T_+ are known, then system (24a), (24b) only would be used. However, Eq. (24b) is a Fredholm equation of the first kind (with non-self-adjoint continuous kernel). For $t \ge 2l + x$ it can be shown that this equation is independent of t, and thus need only be considered for t in the interval $2l - x \le t \le 2l + x$. The question of uniqueness for this equation (which remains to be proved or disproved) is extremely critical, since the dimension of the null space of the operator could be infinite, in which case there would be an infinite set of solutions.

However, if R_- and T_- are also known, then the system (24a), (25a) may be used, yielding a Fredholm equation of the second kind (with continuous kernel). At $x = 0$, it is seen that (24a) yields $K_+(0,t,l)$ directly, and, $x = l$, Eq. (25a) yield $K_-(l,t,0)$ directly. This suggests that the Neumann series should converge in some neighborhood of $x = 0$ and $x = l$. Since the kernel depends upon the values of $R_\pm(s)$ for $0 \le s \le 2l$, $T_-(s)$ for $0 \le s \le 2x$, and $T_+(-s)$ for $0 \le s \le 2(l-x)$, by taking a sup norm of these quantities over their respective intervals and estimating the norm of the operator, it can be shown that there is a neighborhood about $x = 0$, and $x = l$, for which the Neumann series converges. For strong enough conditions imposed upon the coefficients A, B, and C, the Neumann series will converge for all values of x, $0 \le x \le l$.

Apart from the above case, the question of uniqueness remains to be answered. Since the solution will be required for x in $0 \le x \le l$, nonuniqueness would not pose a problem (from the practical sense) if it occurs only for a discrete set of values of x.

If system (24a), (25a) is used, will their solution satisfy Eqs. (24b) and (25b)? This can be partially answered as follows. Define the left-hand sides of (24b) and (25b) as $f(x,t)$ and $g(x,t)$, respectively. Then $f(x,t)$ and $g(x,t)$ will be constant for $t > 2l + x$ and $t < -2l + x$, respectively. If $K_+(x,t,l)$ and $K_-(x,t,0)$ are solutions of (24a) and (25a), it can be shown, using the relations between the scattering kernels, that $f(x,t)$ and $g(x,t)$ must satisfy a set of coupled integral equations not containing K_+ or K_-. These can be reduced to a single homogeneous integral equation of the second kind for either $g(x,t)$ or $f(x,t)$. Hence, if it has only the trivial solution, then $f(x,t)$ and $g(x,t)$ must vanish identically. In this case the solution of system (24a), (25a) will automatically satisfy (24b), (25b).

Once the solution of the set of integral equations has been found in the form

$$K_+(x,t,l) = G(x)K_+^1(x,t) + K_+^2(x,t),$$
$$G(x)K_-(x,t,0) = G(x)K_-^1(x,t) + K_-^2(x,t),$$

where $G(x)$ is the unknown parameter, $B(x)$ has to be determined. This is achieved by employing the boundary condition

$$K_+(x,x,l) - K_-(x,x,0) = L_+(0) + \tfrac{1}{2}B(x),$$

which yields the following nonlinear differential equation for $G(x)$ as a function of x:

$$\frac{1}{2}\frac{dG}{dx} = K_-^2(x,x) + G(x)\,[L_+(0) + K_-^1(x,x) - K_+^2(x,x)] - [G(x)]^2 K_+^1(x,x).$$

This equation is linearized upon the substitution

$$G(x) = [2H(x)K_+^1(x,x)]^{-1}\frac{dH(x)}{dx},$$

yielding a second-order linear differential equation for $H(x)$. Since $G(0)$ and $G(l)$ are known, boundary conditions can be imposed to determine any arbitrary constants.

$B(x)$ is then determined from the relation

$$B(x) = -\frac{d}{dx}\ln[G(x)].$$

The remaining coefficients are determined from the boundary condition

$$K_+(x,x,l) + K_-(x,x,0) = \int_0^x [C - \tfrac{1}{4}A^2 + \tfrac{1}{4}B^2]d\tau - \tfrac{1}{2}A(x) + L_+(0).$$

However, this will yield only the following combination of A and C;

$$C(x) - \frac{1}{4}A^2(x) - \frac{1}{2}\frac{d}{dx}A(x),$$

and either $A(x)$ or $C(x)$ must be prescribed initially, to determine the coefficients explicity. In the example, that follows $C(x) = 0$.

APPLICATION TO ELECTROMAGNETIC THEORY

Maxwell's equations for electromagnetic propagation in a direction along the z axis normal to a slab of varying permittivity $\epsilon(z)$ and conductivity $\sigma(z)$, reduce to the following equation for the electric intensity:

$$E_{zz} - \epsilon(z)\mu_0 E_{tt} - \sigma(z)\mu_0 E_t = 0, \tag{26}$$

where the permeability μ_0 is constant. Exterior to the slab of thickness L, i.e., $z < 0$ and $z > L$, the permittivity is constant $\epsilon = \epsilon_0$ and $\sigma = 0$. Both ϵ and σ will be assumed to be sufficiently smooth functions of z so that the preceding analysis holds and that ϵ will be positive.

The equation will be transformed to the form of (1), by a change of variable from z to x as follows (similar analysis was used by Sharpe[11]):

$$x = \int_0^z [\mu_0\epsilon(s)]^{1/2}ds, \qquad l = x(L).$$

Equation (26) reduces to the form

$$E_{xx} - E_{tt} + A(x)E_x + B(x)E_t = 0,$$

where

$$A(x) = -\frac{d}{dz}[\mu_0\epsilon]^{-1/2}, \qquad B(x) = -\sigma/\epsilon,$$

for which the preceding analysis may be employed. Once $A(x)$ and $B(x)$ are known for $0 \le x \le l$, $\epsilon(z)$ and $\sigma(z)$ can be recovered through the following relations:

$$[\mu_0\epsilon_0]^{1/2}z = \int_0^x \exp\left(-\int_0^\tau A(s)ds\right)d\tau,$$
$$[\epsilon(z)/\epsilon_0]^{1/2} = \exp\left(\int_0^x A(\tau)d\tau\right),$$
$$\sigma(z) = -\epsilon(z)B(x(z)).$$

ACKNOWLEDGMENT

This research was supported by National Science Foundation under Grant GP-22583.

[1] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[2] L. D. Faddeyev and B. Seckler, J. Math. Phys. (N.Y.) 4, 72 (1963).
[3] R. G. Newton, SIAM (Soc. Ind. App. Math.) Rev. 2, 346 (1970).
[4] R. G. Newton, *Scattering of Waves and Particles* (McGraw-Hill, New York, 1966).
[5] I. Kay, Commun. Pure Appl. Math. 13, 371 (1960).
[6] I. Kay, "The inverse scattering problem for transmission lines," presented at workshop on the Mathematics of Profile Inversion, July 1971, Ames Research Center.

[7] I. M. Gel'fand, and R. M. Levitan, Am. Math. Soc. Transl. **1**, 253 (1955).
[8] M. M. Sondhi, and B. Gopinath, J. Acoust. Soc. Am. **49**, 1867 (1971).
[9] L. P. Niznik, Sov. Math. **12**, 306 (1971).
[10] R. Courant, and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1966), Vol.II.
[11] C. B. Sharpe, Quart. J. App. Math. **21**, 105 (1963).

On the structure of the canonical tensor operators in the unitary groups. I. An extension of the pattern calculus rules and the canonical splitting in $U(3)$*

L. C. Biedenharn[†] **and J. D. Louck**

Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544

E. Chacón[‡] **and M. Ciftan**[§]

Department of Physics, Duke University, Durham, North Carolina 27706

(Received 8 May 1972; revised manuscript received 25 July 1972)

The structure of the totally symmetric unit tensor operators (and their conjugates) in $U(n)$ is examined from the viewpoint of the pattern calculus and the factorization lemma. The geometrical properties of the arrow patterns of the fundamental projective (tensor) operators are demonstrated to be the origin of the existence of simple structural expressions for a class of reduced matrix elements of the totally symmetric unit projective operators. An extension of the pattern calculus rules is given whereby these matrix elements can be written out directly. This class of reduced matrix elements is sufficient to permit the construction of the general totally symmetric unit tensor operator. The canonical splitting of the multiplicity in $U(3)$ is similarly shown to be implied uniquely by the geometrical properties of the arrow patterns of the fundamental projective operators and their conjugates. This fact is used to construct explicitly the class of $U(3)$ unit tensor operators having maximal null space. Explicit expressions for a large class of Racah coefficients are also given, and the implications of their limit properties discussed.

1. INTRODUCTION

One of the fundamental problems in the application of symmetry techniques to quantum mechanics is the construction of a suitable basis for the set of all operators mapping the set of all unitary irreducible representation spaces into itself. As is well known, such operators may themselves be characterized by representation labels—this is the tensor operator classification. This classification is, however, incomplete for the general case. Specifically, it is incomplete in the sense that there exist, for the general case, several tensor operators which are labeled by the same state vector labels (the so-called Gel'fand patterns which specify the subgroup properties), and which map a specified irreducible representation (irrep) space $[m]_n$ into a specified irrep space $[m']_n$.

It was early shown[1] that a basis for the set of irreducible tensor operators transforming like the state vector labeled[2]

$$\begin{pmatrix} [M]_n \\ (M)_{n-1} \end{pmatrix}$$

could be labeled by a second set of patterns $\{(\Gamma)_{n-1}\}$, which is in one-to-one correspondence with the subgroup labels $\{(M)_{n-1}\}$, i.e., for specified irrep labels $[M]_n$, the two sets $\{(\Gamma)_{n-1}\}$ and $\{(M)_{n-1}\}$ are equal. This led[1] to the designation

$$\left\langle \begin{matrix} (\Gamma)_{n-1} \\ [M]_n \\ (M)_{n-1} \end{matrix} \right\rangle \tag{1.1}$$

for a *unit tensor operator* or, as it is also called, a *Wigner operator*.

Despite the equality of the numerical arrays contained in the two sets of labels $\{(\Gamma)_{n-1}\}$ and $\{(M)_{n-1}\}$, *the structural significance of the set* $\{(\Gamma)_{n-1}\}$ *is completely different*[1,3,4] (unless otherwise proved) from that of $\{(M)_{n-1}\}$, since this latter set of (Gel'fand) labels derives its significance *entirely* from the existence of the Weyl branching law for the subgroup chain

$$U(n) \supset U(n-1) \supset \cdots \supset U(1). \tag{1.2}$$

No such law is known to hold for the set of patterns $\{(\Gamma)_{n-1}\}$. Accordingly, we refer to $(\Gamma)_{n-1}$ as an *operator pattern* and $(M)_{n-1}$ as a *Gel'fand pattern* to emphasize this distinction. [An exception occurs for $n = 2$ where a very special type of transformation between operator patterns–isomorphic to a $U(2)$ transformation–may be defined.[5]]

What then is the significance of an operator pattern? Finding a complete answer to this question comprises, we believe, *the principal unsolved problem in the theory of tensor operators in the unitary groups*. We make this assertion because the *formal algebra* of the $U(n)$ Wigner operators has been given completely, starting with the work of Ref. 1 and continuing through the work of Refs. 3–6. (The significance of the associativity law for the multiplication of Wigner operators was first noted in Ref. 6.) It is a remarkable fact that the structure of this algebra–termed the $U(n)$ Racah-Wigner calculus–can be given, knowing[1] but a single structural feature of the operator pattern $(\Gamma)_{n-1}$: *The Wigner operator* (1.1) *maps an arbitrary vector belonging to the irrep space labeled by* $[m]_n$ *into either the zero vector or into a vector belonging to the irrep space labeled by* $[m]_n + [\Delta(\Gamma)]_n$, *where*

$$[\Delta(\Gamma)]_n = [\Delta_{1n}(\Gamma)\Delta_{2n}(\Gamma)\cdots\Delta_{nn}(\Gamma)], \tag{1.3a}$$

$$\Delta_{in}(\Gamma) = \sum_{j=1}^{i} \Gamma_{ij} - \sum_{j=1}^{i-1} \Gamma_{i-1,j}, \tag{1.3b}$$

$$\Gamma_{in} = M_{in}. \tag{1.3c}$$

Consider now that we select any operator pattern[7]:

$$\begin{pmatrix} [M]_n \\ (\Gamma)_{n-1} \end{pmatrix}. \tag{1.4}$$

Then $[\Delta(\Gamma)]_n$ follows from the rule (1.3). We call each such $[\Delta(\Gamma)]_n$ *a* Δ *pattern belonging to* $[M]_n$. The mapping (1.3) of operator patterns onto Δ patterns is clearly many-to-one, in the general case. (Whenever it is one-to-one, the Wigner operator is uniquely labeled by its Δ pattern. This occurs, for example, for all the totally symmetric Wigner operators, $\langle p00\cdots 0\rangle$.)

Suppose we now select any Δ pattern $[\Delta]_n$ belonging to $[M]_n$. Then the set of Wigner operators

$$\left\{\left\langle\begin{matrix}(\Gamma)_{n-1}\\ [M]_n\\ (M)_{n-1}\end{matrix}\right\rangle : \begin{matrix}\text{all } (\Gamma)_{n-1} \text{ such that}\\ [\Delta(\Gamma)]_n = [\Delta]_n;\ (M)_{n-1} \text{ arbitrary}\end{matrix}\right\} \quad (1.5)$$

defines the multiplicity set of Wigner operators having the prescribed pattern $[\Delta]_n$.

The principal problem alluded to above can now be stated: *To understand and elucidate the structure which differentiates between the Wigner operators $\langle [M]_n \rangle$ belonging to the multiplicity set of a prescribed Δ pattern. This implies that one must also determine that structural property which assigns a definite operator pattern to a specific Wigner operator in the multiplicity set.* (In this motivating discussion, we have assumed implicitly that a solution exists and is unique.)

This program appears to be, and is, a sizeable undertaking, particularly, when we insist that there exists a *canonical solution* (to within equivalence, if necessary), where we use the term canonical in the sense of being free of arbitrary choice[8] aside from phase (±).

Let us now inquire as to what properties of these Wigner operators belonging to a given multiplicity set could possibly distinguish among them. The first property which comes to mind is the null spaces of the operators.[9] The null space of the Wigner operator (1.1) is the set of all irrep spaces which are annihilated by the operator, i.e., the set of all irrep spaces with labels $[m]_n$ such that

$$\left\langle\begin{matrix}(\Gamma)_{n-1}\\ [M]_n\\ (M)_{n-1}\end{matrix}\right\rangle \left|\left(\begin{matrix}[m]_n\\ (m)_{n-1}\end{matrix}\right)\right\rangle = 0 \quad (1.6)$$

for all Gel'fand patterns $(m)_{n-1}$ and $(M)_{n-1}$. The existence of such null spaces is assured by the properties of the intertwining number—the number of times an irrep $[m']_n$ is contained in the direct product $[M]_n \otimes [m]_n$.

Let us be explicit and state the three structural properties which we believe will ultimately be proved, and which will be decisive in establishing the existence of a canonical labeling for all Wigner operators of $U(n)$. Let $(\Gamma_1), (\Gamma_2), \ldots, (\Gamma_{\mathfrak{M}})$ denote the operator patterns belonging to a given multiplicity set, and let $\mathfrak{N}(\Gamma_k)$, $k = 1, 2, \ldots, \mathfrak{M}$ denote the null space of the corresponding Wigner operator:

Conjecture 1: The operator patterns are simply ordered by the inclusion property

$$\mathfrak{N}(\Gamma_1) \supset \mathfrak{N}(\Gamma_2) \supset \cdots \supset \mathfrak{N}(\Gamma_{\mathfrak{M}})$$

of the null spaces.

Conjecture 2: The unique numerical assignment of the labels in an operator pattern is a consequence of *limit properties*.[4]

Conjecture 3: The uniqueness of the canonical labeling is a consequence of the *indecomposability*[10] of the associated Wigner operators.

A proof of these three conjectures would constitute the basis for the canonical resolution of the tensor operator labeling problem.

Let us remark that Conjectures 1–3 have been proved[4] for all the adjoint tensor operators in $U(n)$ (for all $n = 2, 3, \cdots$). It has also been shown[11] that for $U(3)$ there exists a canonical splitting of all multiplicities; that is, the labeling of all Wigner operators in $U(3)$ is unique and free of arbitrary choices, to within equivalence representing the choice of 1, 2, 3 explicit in the Weyl canonical labeling of state vectors within an irrep.

One of the principal aims of the present work is to demonstrate by explicit construction that this canonical splitting in $U(3)$ verifies Conjectures 1–3.

Although a canonical resolution of the multiplicity problem for $U(3)$ has been shown to exist, the explicit construction of the associated set of Wigner operators and Racah invariant operators (both of which now are unique to within phase) is still a formidable task. Complete results have been given only for the tensor operators having the irrep labels [210] and [420], i.e., the adjoint[12] $\langle 210\rangle$ ("octet") and the "27-plet operator" $\langle 420\rangle$.[13]

The construction of the multiplicity free operators $\langle p0\cdots 0\rangle$, even though there is no point of principle involved whatsoever, is itself a sizeable task: This has recently been done by註išauskas *et al.*[14] and by Chacón *et al.*[15]

It is the purpose of the present series of three papers to illustrate and discuss the structural properties of the canonical unit tensor operators. There are two aspects to this program: (1) the elucidation of the structure of the multiplicity-free tensor operators and (2) the verification of Conjectures 1–3 by actual construction of the unit tensor operators in a multiplicity set.

The first part of this program is by no means trivial, although much easier than the second. The mere writing out of complicated matrix elements—even though essential—is but the first step and of itself contributes very little to one's understanding of the origins and significance of such expressions. There are two indispensable tools which we have found can render such otherwise complicated results comprehensible: the factorization lemma[11,3] and the pattern calculus.[16] By exploiting these tools to the fullest extent, one is able to see through the superficial complexity of the individual matrix elements and understand quite directly the *structure* of the answers—a structure which is often both elegant and elementary.

Let us now summarize the plan of this series of papers. In the present paper (I), we give in Sec. 2 a résumé of the basic tools required in the subsequent developments. In Sec. 3, the totally symmetric tensor operators in $U(n)$ are considered, the emphasis, as mentioned, being on the structure of the results. The significant contribution of this section is an extension of the pattern calculus rules. This generalization allows one to read off directly from the arrow patterns the complete matrix element expressions (except for phase) for all totally symmetric projective operators (and their conjugates) having either (1) *arbitrary* upper operator patterns and *extremal* lower patterns or (2) *extremal* upper operator patterns and *arbitrary* lower operator patterns. The totally symmetric Racah functions are given in Sec. 3E, and the implications of the limit properties discussed.

We restrict our attention to $U(3)$ in Sec. 4, demonstrating that *the origin of the canonical splitting has the same geometrical basis in terms of the arrow patterns as occurs for the totally symmetric operators.* [Unfortunately, this feature does not directly generalize to $U(n)$—hence, the reason for considering $U(3)$ only.] We construct in detail the class of $U(3)$ tensor operators having maximal null space; this class necessarily includes all multiplicity free operators. The present paper concludes by giving an algorithm whereby all $U(3)$ tensor operators can be constructed.

While the considerations of $U(3)$ in this paper are logically complete (in the sense that all answers are given fully), we relegate to a second paper (II) the nontrivial task of verifying that our $U(3)$ results prove Conjectures 1–3 for $U(3)$. In Paper III, we will resume our general studies of the structural properties of the canonical tensor operators in $U(n)$.

2. RÉSUMÉ OF BASIC RESULTS[17]

A. The general coupling laws

The formal algebra of Wigner operators has been developed in Refs. 1, 3, and 4, and the reader is referred to these papers for a more detailed explanation of the notations and proofs of the results summarized in this section.

The first basic result is the coupling law for Wigner operators. This law is given symbolically as follows:

$$\left\langle \begin{matrix} \cdot \\ [M'] \\ \cdot \end{matrix} \right\rangle \overset{\{R\}}{\underset{\langle W\rangle}{}} \left\langle \begin{matrix} \cdot \\ [M] \\ \cdot \end{matrix}\right\rangle = \left\langle \begin{matrix} (\Gamma'') \\ [M''] \\ (M'') \end{matrix}\right\rangle, \tag{2.1}$$

where the dots in the left-hand side indicate that the respective patterns are summed over: (1) The lower *Gel' fand patterns* are coupled by Wigner coefficients (indicated by $\langle W\rangle$), and (2) the upper *operator patterns* are coupled by the Racah invariant operators (indicated by $\{R\}$). In detail, Eq. (2.1) takes the form as follows:

$$\delta_{(\Lambda')(\Lambda)} \left\langle \begin{matrix} (\Gamma'') \\ [M]+[\Delta(\Lambda)] \\ (M'') \end{matrix}\right\rangle = \sum_{(M')}$$

$$= \sum_{\substack{(M')(M) \\ (\Gamma')(\Gamma)}} \left\langle \begin{pmatrix} [M]+[\Delta(\Lambda)] \\ (M'') \end{pmatrix} \middle| \left\langle \begin{matrix} (\Lambda) \\ [M'] \\ (M') \end{matrix}\right\rangle \middle| \begin{pmatrix} [M] \\ (M) \end{pmatrix} \right\rangle \times \left\{ \begin{pmatrix} [M]+[\Delta(\Lambda')] \\ (\Gamma'') \end{pmatrix} \begin{pmatrix} (\Lambda') \\ [M'] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [M] \\ (\Gamma) \end{pmatrix} \right\} \left\langle \begin{matrix} (\Gamma') \\ [M'] \\ (M') \end{matrix}\right\rangle \left\langle \begin{matrix} (\Gamma) \\ [M] \\ (M) \end{matrix}\right\rangle. \tag{2.2}$$

The summation over (Γ') and (Γ) is over all operator patterns such that $[\Delta(\Gamma'')]_n = [\Delta(\Gamma)]_n + [\Delta(\Gamma')]_n$. However, it is not necessary to specify this explicitly, since the Racah invariants are, by definition, zero whenever this condition is violated. A similar, but more restricted statement[3,4] applies to the summation over (M') and (M). [We generally adopt the practice of omitting the subscripts on $[M]_n$, $(M)_{n-1}$, etc., whenever they are clearly implied.]

The second basic coupling law follows from Eq. (2.2) upon using the subgroup reduction law for Wigner operators.[1,4] This is the coupling law for $U(n):U(n-1)$ projective operators and is expressed symbolically as follows:

$$\begin{bmatrix} \cdot \\ [M'] \\ \cdot \end{bmatrix} \overset{\{R\}}{\underset{[R]}{}} \begin{bmatrix} \cdot \\ [M] \\ \cdot \end{bmatrix} = \begin{bmatrix} (\Gamma'') \\ [M]+[\Delta] \\ (\gamma'') \end{bmatrix} \tag{2.3}$$

The explicit form of this symbolic coupling is

$$\delta_{(\Lambda')_{n-1}(\Lambda)_{n-1}} \begin{bmatrix} (\Gamma'')_{n-1} \\ [M]_n+[\Delta(\Lambda)]_n \\ (\gamma'')_{n-1} \end{bmatrix}$$

$$= \sum_{\substack{(\gamma)_{n-1}(\gamma')_{n-1} \\ (\Gamma)_{n-1}(\Gamma')_{n-1}}} \left[\begin{pmatrix} [M]_n+[\Delta(\Lambda)]_n \\ (\gamma'')_{n-1} \end{pmatrix} \begin{pmatrix} (\Lambda)_{n-1} \\ [M']_n \\ (\gamma')_{n-1} \end{pmatrix} \begin{pmatrix} [M]_n \\ (\gamma)_{n-1} \end{pmatrix} \right]$$

$$\times \left\{ \begin{pmatrix} [M]_n+[\Delta(\Lambda')]_n \\ (\Gamma'')_{n-1} \end{pmatrix} \begin{pmatrix} (\Lambda')_{n-1} \\ [M']_n \\ (\Gamma')_{n-1} \end{pmatrix} \begin{pmatrix} [M]_n \\ (\Gamma)_{n-1} \end{pmatrix} \right\}$$

$$\times \begin{bmatrix} (\Gamma')_{n-1} \\ [M']_n \\ (\gamma')_{n-1} \end{bmatrix} \begin{bmatrix} (\Gamma)_{n-1} \\ [M]_n \\ (\gamma)_{n-1} \end{bmatrix}, \tag{2.4a}$$

where we have introduced a new object, called a *square-bracket invariant* (in analogy to the curly-bracket, or Racah, invariant):

$$\left[\begin{pmatrix} [M]_n+[\Delta(\Lambda)]_n \\ (\gamma'')_{n-1} \end{pmatrix} \begin{pmatrix} (\Lambda)_{n-1} \\ [M']_n \\ (\gamma')_{n-1} \end{pmatrix} \begin{pmatrix} [M]_n \\ (\gamma)_{n-1} \end{pmatrix} \right]$$

$$\equiv \sum_{(\Lambda')_{n-2}} \left\langle \begin{pmatrix} [M]_n+[\Delta(\Lambda)]_n \\ [\gamma]_{n-1}+[\Delta(\Lambda')]_{n-1} \end{pmatrix} \middle| \begin{bmatrix} (\Lambda)_{n-1} \\ [M']_n \\ (\Lambda')_{n-1} \end{bmatrix} \middle| \begin{pmatrix} [M]_n \\ [\gamma]_{n-1} \end{pmatrix} \right\rangle$$

$$\times \left\{ \begin{pmatrix} [\gamma'']_{n-1} \\ (\gamma'')_{n-2} \end{pmatrix} \begin{pmatrix} (\Lambda')_{n-2} \\ [\gamma']_{n-1} \\ (\gamma')_{n-2} \end{pmatrix} \begin{pmatrix} [\gamma]_{n-1} \\ (\gamma)_{n-2} \end{pmatrix} \right\}, \tag{2.4b}$$

in which

$$(\Lambda')_{n-1} = \begin{pmatrix} [\gamma']_{n-1} \\ (\Lambda')_{n-2} \end{pmatrix}. \tag{2.4c}$$

The first factor on the right-hand side of definition (2.4b) is a $U(n)$: $U(n-1)$ reduced Wigner coefficient; the second factor a $U(n-1)$ Racah invariant–its eigenvalue depends only on the labels $[m]_{n-1}$. Thus, the square-bracket invariant, denoted $[\cdots]$, is a $U(n-1)$ invariant. [Note that, for $n=2$, projective operators become Wigner operators, the square-bracket invariant becomes a Wigner *coefficient*, and Eq. (2.4) reduces (properly) to the coupling law for Wigner operators.]

The coupling law (2.3) for projective operators will be used frequently in one form or another in the work to follow. [Further discussion of Eq. (2.3) can be found in Refs. 3 and 4.]

The following brief notational summary is included to aid the reader in identifying the symbols used to denote the basic quantities which enter into the coupling laws described above:

$\left\langle\begin{matrix}\cdot\\ \cdot\\ \cdot\end{matrix}\right\rangle$: unit tensor operator [cf. Eq. (1.1)];

$\left[\begin{matrix}\cdot\\ \cdot\\ \cdot\end{matrix}\right]$: unit projective operator [cf. Eq. (2.3)];

$\left\{\left(\begin{matrix}\cdot\\ \cdot\end{matrix}\right)\left(\begin{matrix}\cdot\\ \cdot\\ \cdot\end{matrix}\right)\left(\begin{matrix}\cdot\\ \cdot\end{matrix}\right)\right\}$: Racah invariant operator [cf. Eq. (2.2)];

$\left[\left(\begin{matrix}\cdot\\ \cdot\end{matrix}\right)\left(\begin{matrix}\cdot\\ \cdot\\ \cdot\end{matrix}\right)\left(\begin{matrix}\cdot\\ \cdot\end{matrix}\right)\right]$: square-bracket invariant operator [cf. Eq. (2.4)].

B. The pattern calculus rules

It is a remarkable fact that the explicit matrix elements of all *extremal* unit $U(n):U(n-1)$ projective operators can be calculated from a few simple rules of the pattern calculus.[16] In particular, this class of explicitly known projective operators includes all *elementary* operators of the form $[\dot{1}_k\ \dot{0}_{n-k}]$ and $[\dot{0}_{n-k} - \dot{1}_k]$ (a dot over a numeral implies that the numeral is repeated a number of times equal to the subscript), which themselves are a basis for constructing all $U(n)$ tensor operators.

The pattern calculus proceeds by considering a given unit $U(n):U(n-1)$ projective operator

$$\begin{bmatrix}(\Gamma)_{n-1}\\ [M]_n\\ (\gamma)_{n-1}\end{bmatrix}, \tag{2.5}$$

where both $(\Gamma)_{n-1}$ and $(\gamma)_{n-1}$ are extremal patterns. To this operator we assign a Δ pattern of two rows, corresponding to the shifts $[\Delta(\Gamma)]_n$ and $[\Delta(\gamma)]_{n-1}$.[18] From the Δ pattern, we construct an "arrow pattern" and write out the $U(n):U(n-1)$ reduced Wigner coefficient by the following rules.

The arrow-pattern rules

Rule 1: Write out two rows of dots, as shown:

• • • • • • n dots
…
• • • • $n-1$ dots.

Rule 2: Draw arrows between dots as follows: Select a dot i in row n and a dot j in row $n-1$. If $\Delta_{in}(\Gamma) > \Delta_{jn-1}(\gamma)$, draw $\Delta_{in}(\Gamma) - \Delta_{jn-1}(\gamma)$ arrows from dot i to dot j; if $\Delta_{jn-1}(\gamma) > \Delta_{in}(\Gamma)$, draw the arrows from dot j to dot i. Carry out this procedure for all dots in rows n and $n-1$. This yields a *numerator arrow pattern* with arrows going *between* rows.

Carry out this procedure for dots within row n and dots within row $n-1$. This yields a *denominator arrow pattern* with arrows going *within* rows.

Rule 3: In the arrow patterns, assign the partial hook p_{in} to dot i, $i = 1, 2, \ldots, n$, in row n and p_{jn-1} to dot j, $j = 1, 2, \ldots, n-1$, in row $n-1$. ($p_{ij} \equiv m_{ij} + j - i$.)

Rule 4: In general, there will be several arrows going between two dots in the arrow patterns. Assign to the first arrow the factor

$$p(\text{tail}) - p(\text{head}) + e(\text{tail}),$$

to the second arrow, the factor

$$p(\text{tail}) - p(\text{head}) + e(\text{tail}) + 1,$$

etc., until all arrows going between the same two dots have been counted:

$$e(\text{tail}) \equiv 1, \quad \text{if tail of arrow on row } n-1,$$
$$\equiv 0, \quad \text{if tail of arrow on row } n.$$

Rule 5: Write out the products

$N^2 = |$product of all factors for numerator arrow pattern$|$,

$D^2 = |$product of all factors for denominator arrow pattern$|$.

The net effect of these rules is to make the associations

$$\begin{bmatrix}(\Gamma)\\ (M)\\ (\gamma)\end{bmatrix} \longleftrightarrow \Delta \text{ pattern} \longleftrightarrow \begin{matrix}\text{arrow}\\ \text{pattern}\end{matrix} \longleftrightarrow \begin{matrix}\text{algebraic}\\ \text{factor}\end{matrix} \equiv \left|\frac{N}{D}\right|.$$

The arrow-pattern rules clearly yield the same result if we effect an integral shift $\Delta_{in}(\Gamma) \to \Delta_{in}(\Gamma) + \lambda$, $i = 1, 2, \ldots, n$, $\Delta_{in-1}(\gamma) \to \Delta_{in-1}(\gamma) + \lambda$, $i = 1, 2, \ldots, n-1$. Thus, the rules apply to Δ patterns which contain negative integers. In particular, all operators of the form $[\dot{0}_k - \dot{1}_{n-k}]$ are obtained from the rules above.

The value of the matrix element of the extremal projective operator (2.5) then takes the symbolic form

$$\left\langle\begin{matrix}[m]_n + [\Delta(\Gamma)]_n\\ [m]_{n-1} + [\Delta(\gamma)]_{n-1}\end{matrix}\right| \begin{bmatrix}(\Gamma)\\ [M]\\ (\gamma)\end{bmatrix} \left|\begin{matrix}[m]_n\\ [m]_{n-1}\end{matrix}\right\rangle = (\text{phase})\,|N/D|. \tag{2.6}$$

C. Projective functions and Δ pattern functions

The final labels in the reduced matrix element (2.6) are implied by the initial labels. Accordingly, we may interpret a matrix element of a unit projective operator as defining a *unit projective function*. Thus, for arbitrary operator patterns (Γ) and (Γ'), we *define* a unit projective function[4]

$$\begin{bmatrix}(\Gamma)\\ [M]\\ (\Gamma')\end{bmatrix} \tag{2.7a}$$

by giving its value on the set of all $U(n)$ and $U(n-1)$ irrep labels,

$$[m] \equiv [m_{1n} m_{2n} \cdots m_{nn}], \tag{2.7b}$$

$$[m'] \equiv [m_{1,n-1} m_{2,n-1} \cdots m_{n-1,n-1}], \tag{2.7c}$$

which satisfy the Weyl branching law:

$$\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix} \equiv \left\langle \begin{matrix}{[m]}+[\Delta]\\ {[m']}+[\Delta']\end{matrix}\right| \begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix} \left| \begin{matrix}{[m]}\\ {[m']}\end{matrix}\right\rangle, \tag{2.7d}$$

in which

$$[\Delta] = [\Delta(\Gamma)]_n = [\Delta_1\Delta_2\cdots\Delta_n], \tag{2.7e}$$

$$[\Delta'] = [\Delta(\Gamma')]_{n-1} = [\Delta'_1\Delta'_2\cdots\Delta'_{n-1}]. \tag{2.7f}$$

Furthermore, the value of the projective function is defined to be zero unless $[m]+[\Delta]$ and $[m']+[\Delta']$ satisfy the Weyl branching law (the so-called lexical or "betweenness" conditions).

{We have attempted to introduce a notation in Eq. (2.7d) which avoids excessive subscripts n and $n-1$. Thus, the *placement* of a quantity serves to indicate (uniquely) whether it should carry n or $n-1$. In particular, note that $[\Delta']$ is defined in terms of the *lower* operator pattern (Γ') in *exactly* the same way that $[\Delta]$ is defined in terms of the *upper* operator pattern (Γ), the only difference being that Δ'_n is left out of the definition of $[\Delta']$. Note, however, the exceptions for the *extended projective operators*,[4,16] where both $[\Delta']$ and $[m']$ become of length n.}

In order that the functional interpretation (2.7) reflect properly the matrix element multiplication property, it is necessary to define the product of two unit projective functions by the following rule: The product

$$\begin{bmatrix}(\Gamma'')\\ {[M']}\\ (\Gamma''')\end{bmatrix}\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix} \tag{2.8a}$$

is the function whose value is given by

$$\left(\begin{bmatrix}(\Gamma'')\\ {[M']}\\ (\Gamma''')\end{bmatrix}\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\right)\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix} = \begin{bmatrix}(\Gamma'')\\ {[M']}\\ (\Gamma''')\end{bmatrix}\begin{pmatrix}{[m]}+[\Delta]\\ {[m']}+[\Delta']\end{pmatrix} \times \begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix}. \tag{2.8b}$$

It is easily seen that the general rule is: Any unit projective function standing in a string of such functions gets evaluated on the labels $[m]$ and $[m']$ *shifted* by the sum of all the upper and lower Δ patterns, respectively, of the functions standing to the right of it. One verifies immediately that the multiplication defined by Eq. (2.8b) is associative, but, in general, non-commutative.

In an analogous fashion, we define the *Hermitian conjugate unit projective function*

$$\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}^{\dagger} \tag{2.9a}$$

by giving its value

$$\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}^{\dagger}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix} = \left\langle \begin{matrix}{[m]}-[\Delta]\\ {[m']}-[\Delta']\end{matrix}\right| \begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}^{\dagger} \left| \begin{matrix}{[m]}\\ {[m']}\end{matrix}\right\rangle = \begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\begin{pmatrix}{[m]}-[\Delta]\\ {[m']}-[\Delta']\end{pmatrix}. \tag{2.9b}$$

(The matrix elements can always be, and are, chosen to be real.)

In the product rule (2.8), we are allowed, by using Eq. (2.9), to put a dagger on either or both of the projective functions. Note, in particular, that

$$\left(\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}^{\dagger}\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\right)\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix} = \left(\begin{bmatrix}(\Gamma)\\ {[M]}\\ (\Gamma')\end{bmatrix}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix}\right)^2. \tag{2.10}$$

Observe that, for (Γ) and (Γ') extremal, the pattern calculus rules apply directly to the Hermitian conjugate unit projective functions through Eq. (2.9b), i.e., $p_{in} \to p_{in} - \Delta_i$ and $p_{in-1} \to p_{in-1} - \Delta'_i$ in rule 4.

Let us next observe that the pattern calculus rules, in fact, utilize only the two Δ patterns $[\Delta]$ and $[\Delta']$ in drawing the arrow patterns, to which, in turn, there is associated an algebraic factor depending on $[m]$ and $[m']$. These rules associate a perfectly well-defined arrow pattern with arbitrary sets of integers $[\Delta_1\Delta_2\cdots\Delta_n]$ and $[\Delta'_1\Delta'_2\cdots\Delta'_{n-1}]$. However, the only such sets of interest for the unit projective operators are those for which $[\Delta]$ and $[\Delta']$ belong to $[M]$. (We say that $[\Delta'_1\Delta'_2\cdots\Delta'_{n-1}]$ belongs to $[M]$ if $[\Delta'_1\Delta'_2\ldots\Delta'_n]$ does.) Clearly, these two Δ patterns define both a function

$$F\begin{pmatrix}{[\Delta]}\\ {[M]}\\ {[\Delta']}\end{pmatrix}, \tag{2.11a}$$

and an arrow pattern, from which we can read off, using rules 4 and 5, the value of the function at the "point" $\binom{[m]}{[m']}$:

$$F\begin{pmatrix}{[\Delta]}\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix}. \tag{2.11b}$$

We define the product of two such functions in analogy to Eq. (2.8b):

$$\left[F\begin{pmatrix}{[\Delta'']}\\ {[M']}\\ {[\Delta''']}\end{pmatrix}F\begin{pmatrix}{[\Delta]}\\ {[M]}\\ {[\Delta']}\end{pmatrix}\right]\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix} = F\begin{pmatrix}{[\Delta'']}\\ {[M']}\\ {[\Delta''']}\end{pmatrix}\begin{pmatrix}{[m]}+[\Delta]\\ {[m']}+[\Delta']\end{pmatrix} \times F\begin{pmatrix}{[\Delta]}\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}{[m]}\\ {[m']}\end{pmatrix}. \tag{2.12}$$

Again note that the associative law holds, but the commutative law fails, in general.

The calculus of the functions defined by Eqs. (2.11) and the product rule (2.12) is well defined independently of the relation of the two patterns to the operator patterns. We term such a function (2.11) a Δ *pattern function*, since the *function itself* is uniquely represented geometrically by the rules of the pattern calculus.

Observe that we have the following identity between unit projective functions and Δ pattern functions *for both* (Γ) *and* (Γ') *extremal patterns:*

$$\begin{bmatrix}(\Gamma)\\ [M]\\ (\Gamma')\end{bmatrix} = (\text{phase})\, F\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix} \tag{2.13}$$

for $[\Delta] = [\Delta(\Gamma)]$ and $[\Delta'] = [\Delta(\Gamma')]$.

We also define the Hermitian conjugate Δ pattern function

$$F^{\dagger}\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix}, \tag{2.14a}$$

by giving its value

$$F^{\dagger}\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix}\begin{pmatrix}[m]\\ [m']\end{pmatrix} \equiv F\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix}\begin{pmatrix}[m]-[\Delta]\\ [m']-[\Delta']\end{pmatrix}. \tag{2.14b}$$

Again we may place a dagger on either or both of the Δ pattern functions in the product (2.12). We note, in particular, that

$$F^{\dagger}\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix} F\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix} \tag{2.15a}$$

has the value

$$\left[F\begin{pmatrix}[\Delta]\\ [M]\\ [\Delta']\end{pmatrix}\begin{pmatrix}[m]\\ [m']\end{pmatrix}\right]^2. \tag{2.15b}$$

The Gel'fand pattern notation is unwieldy and overly redundant for the elementary operators $\langle 1\dot{0}\rangle$ and $\langle \dot{1}0\rangle$. Accordingly, we introduce a special, more compact notation for these. In this paper, three such notations occur:

$$\begin{pmatrix}[1 & \dot{0}]\\ & \tau\end{pmatrix},\quad \begin{pmatrix}[\dot{1} & 0]\\ & \tilde{\tau}\end{pmatrix},\quad \begin{pmatrix}[\dot{0} & -1]\\ & \bar{\tau}\end{pmatrix}, \tag{2.16}$$

where $\tau = 1, 2, \ldots, n$ in each pattern. (We omit the subscripts on $\dot{0}$ or $\dot{1}$ when it is clear how many times the numeral is repeated.) The integers τ, $\tilde{\tau}$, and $\bar{\tau}$ specify, respectively, the operator patterns which have Δ patterns given by $\Delta(\tau) \equiv [0\cdots010\cdots0]_n$ (1 in position τ), $\Delta(\tilde{\tau}) = [1\cdots101\cdots1]_n$ (0 in position τ), and $\Delta(\bar{\tau}) = -\Delta(\tau)$. Thus, for example,

$$\begin{bmatrix}\tau\\ [1 \quad \dot{0}]\\ \rho\end{bmatrix}, \qquad \tau, \rho = 1, 2, \ldots, n \tag{2.17}$$

designates the fundamental $U(n):U(n-1)$ projective operator (or the unit projective function) having upper operator pattern with Δ pattern $\Delta(\tau)$ and lower operator pattern with Δ pattern $\Delta(\rho)$.

We also extend this abbreviated notation to the following patterns:

$$\begin{pmatrix}[p & \dot{0}]\\ & \tau\end{pmatrix},\quad \begin{pmatrix}[\dot{p} & 0]\\ & \tilde{\tau}\end{pmatrix},\quad \begin{pmatrix}[\dot{0} & -p]\\ & \bar{\tau}\end{pmatrix}, \tag{2.18}$$

where $\tau = 1, 2, \ldots, n$. The integers τ, $\tilde{\tau}$, and $\bar{\tau}$ now specify, respectively, the operator patterns which have Δ patterns given by $p\Delta(\tau)$, $p\Delta(\tilde{\tau})$, and $-p\Delta(\tau)$. Observe that, in each instance, the Δ pattern is a permutation of the irrep labels, i.e., the pattern is extremal. This implies, for example, that

$$\begin{bmatrix}\tau\\ [p \quad \dot{0}]\\ \rho\end{bmatrix} = \begin{bmatrix}\tau\\ [1 \quad \dot{0}]\\ \rho\end{bmatrix}^{p}. \tag{2.19}$$

Particular examples of this notation combined with Eq. (2.13) are

$$\begin{bmatrix}\tau\\ [1 \quad \dot{0}]\\ \rho\end{bmatrix} = S(\rho-\tau)F\begin{pmatrix}\Delta(\tau)\\ [1 \quad \dot{0}]\\ \Delta'(\rho)\end{pmatrix}, \tag{2.20a}$$

$$\begin{bmatrix}\bar{\tau}\\ [\dot{0} \quad -1]\\ \bar{\rho}\end{bmatrix} = (-1)^{\rho-\tau}S(\rho-\tau)F\begin{pmatrix}-\Delta(\tau)\\ [\dot{0} \quad -1]\\ -\Delta'(\rho)\end{pmatrix}, \tag{2.20b}$$

$$\begin{bmatrix}\tilde{\tau}\\ [\dot{1} \quad 0]\\ \tilde{\rho}\end{bmatrix} = (-1)^{\rho-\tau}S(\rho-\tau)F\begin{pmatrix}\Delta(\tilde{\tau})\\ [\dot{1} \quad 0]\\ \Delta'(\tilde{\rho})\end{pmatrix}, \tag{2.20c}$$

$$\begin{bmatrix}\tau\\ [p \quad \dot{0}]\\ \rho\end{bmatrix} = [S(\rho-\tau)]^{p}F\begin{pmatrix}p\Delta(\tau)\\ [p \quad \dot{0}]\\ p\Delta'(\rho)\end{pmatrix}, \tag{2.20d}$$

where $S(\rho-\tau)$ is defined to be $+1$ for $\rho \geqslant \tau$ and -1 for $\rho < \tau$. $\Delta'(\rho)$ is simply $\Delta(\rho)$ with the nth component missing, e.g., $\Delta'(n) = [0\cdots0]$. The values of the Δ pattern functions F(:) are, of course, known explicitly from the pattern calculus rules.

A new development of the pattern calculus rules is obtained in Sec. 3. In anticipation of these results, we now note some additional features of the Δ pattern functions. The pattern calculus factor $|N/D|$ is actually of the more explicit form

$$|(N_{n:n-1} \times N_{n-1:n})/(D_n \times D_{n-1})|, \tag{2.21}$$

where

$N^2_{n:n-1} = |$product of all factors for arrows going from row n to row $n-1|$,

$N^2_{n-1:n} = |$product of all factors for arrows going from row $n-1$ to row $n|$,

$D^2_n = |$product of all factors for arrows going within row $n|$,

$D^2_{n-1} = |$product of all factors for arrows going within row $n-1|$.

Thus, a Δ pattern calculus function is, in fact, the product of four functions, each of which has a well-defined meaning. We introduce now two combinations of these factors (the importance of these particular combinations will subsequently be shown).

Let

$$F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix} \tag{2.22a}$$

denote the *restricted* Δ *pattern function* having the value

$$F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}[m]\\ {[m']}\end{pmatrix}, \tag{2.22b}$$

obtained by deleting the denominator factor D_n from expression (2.21), i.e., *the value obtained from the pattern calculus rules by leaving out all arrows within row n.* Similarly, let

$$F_{R'}\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix} \tag{2.23a}$$

denote the restricted Δ pattern function having the value

$$F_{R'}\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}[m]\\ {[m']}\end{pmatrix}, \tag{2.23b}$$

obtained by deleting the denominator factor D_{n-1} from expression (2.21), i.e., *the value obtained from the pattern calculus rules by leaving out all arrows with row* $n-1$.

A Δ pattern function is now expressed in terms of a restricted pattern function by

$$F\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix} = F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\Bigg/ d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix} = F_{R'}\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\Bigg/ d\begin{pmatrix}[M]\\ {[\Delta']}\end{pmatrix}, \tag{2.24}$$

where the value of each of the quotient functions is defined to be the value of the restricted Δ pattern function divided by $d\binom{[\Delta]}{[M]}([m])$ and $d\binom{[M]}{[\Delta']}([m'])$, respectively. Here $d\binom{[\Delta]}{[M]}$ denotes the $U(n)$ *denominator function* defined by the arrow pattern for row n, and $d\binom{[\Delta]}{[M]}([m])$ denotes its value D_n. Similarly, $d\binom{[M]}{[\Delta']}$ denotes the $U(n-1)$ denominator function defined by the arrow pattern for row $n-1$, and $d\binom{[M]}{[\Delta']}([m'])$ denotes its value D_{n-1}. [Observe that D_{n-1} is *not* just the value of D_n with $n \to n-1$, since the rules for determining the $U(n-1)$ denominator *have an extra shift* $+1$ associated with them.]

The multiplication rule (2.12) now *implies* the multiplication rule for the quotient functions:

$$F_R\begin{pmatrix}[\Delta'']\\ {[M']}\\ {[\Delta''']}\end{pmatrix} F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\Bigg/ d\begin{pmatrix}[\Delta'']\\ {[M']}\end{pmatrix} d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}$$

$$\equiv \left[F_R\begin{pmatrix}[\Delta'']\\ {[M']}\\ {[\Delta''']}\end{pmatrix}\Bigg/ d\begin{pmatrix}[\Delta'']\\ {[M']}\end{pmatrix}\right] \times \left[F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\Bigg/ d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}\right], \tag{2.25a}$$

where the value of the numerator product is defined by the rule (2.12), and the value of the denominator product is defined by

$$\left[d\begin{pmatrix}[\Delta'']\\ {[M']}\end{pmatrix} d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}\right]([m]) = d\begin{pmatrix}[\Delta'']\\ {[M']}\end{pmatrix}([m]+[\Delta]) \times d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}([m]). \tag{2.25b}$$

The second quotient function in Eq. (2.24) obeys these same rules.

The Hermitian conjugate Δ pattern functions have a similar decomposition into a restricted Δ pattern function part and a denominator part—simply place a dagger on $F, F_R, F_{R'}$, and d in Eq. (2.24). The appropriate definitions are

$$F_R^\dagger\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}[m]\\ {[m']}\end{pmatrix} = F_R\begin{pmatrix}[\Delta]\\ {[M]}\\ {[\Delta']}\end{pmatrix}\begin{pmatrix}[m]-[\Delta]\\ {[m]-[\Delta']}\end{pmatrix}, \tag{2.26a}$$

$$d^\dagger\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}([m]) = d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}([m]-[\Delta]). \tag{2.26b}$$

Similar definitions are made for the second decomposition.

Observe then that property (2.15) holds for the restricted Δ pattern functions. In addition, we have a similar property for the denominator functions

$$\left[d^\dagger\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix} d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}\right]([m]) = \left[d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}([m])\right]^2. \tag{2.27}$$

It should be noted very carefully that in all cases of the various types of Δ pattern functions introduced, it is the Hermitian square $f^\dagger f$ which has for its value $[f(x)]^2$. It is convenient to introduce the following notation for the product $f^\dagger f$:

$$|f|^2 \equiv f^\dagger f. \tag{2.28}$$

In particular,

$$\left|d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}\right|^2 = d^\dagger\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix} d\begin{pmatrix}[\Delta]\\ {[M]}\end{pmatrix}. \tag{2.29}$$

We may place a dagger on either or both of the d's in Eq. (2.25b). For example,

$$\left[d\begin{pmatrix}[\Delta'']\\ [M']\end{pmatrix} d^\dagger \begin{pmatrix}[\Delta]\\ [M]\end{pmatrix}\right]([m])$$
$$= d\begin{pmatrix}[\Delta'']\\ [M']\end{pmatrix}([m]-[\Delta])d\begin{pmatrix}[\Delta]\\ [M]\end{pmatrix}([m]-[\Delta]). \qquad (2.30)$$

If f and g are Δ pattern functions, we also define $(fg)^\dagger$ by the rule

$$(fg)^\dagger \equiv g^\dagger f^\dagger. \qquad (2.31)$$

An important special case of these notations and definitions is as follows: Let

$$d_{\Delta(\tau)} \equiv d\begin{pmatrix} & \Delta(\tau) & \\ [1 & & \dot{0}]\end{pmatrix}. \qquad (2.32)$$

Then

$$|d_{\Delta(\tau')}d_{\Delta(\tau)}|^2 = (d_{\Delta(\tau')}d_{\Delta(\tau)})^\dagger d_{\Delta(\tau')}d_{\Delta(\tau)}$$
$$= d^\dagger_{\Delta(\tau)}d^\dagger_{\Delta(\tau')}d_{\Delta(\tau')}d_{\Delta(\tau)}, \qquad (2.33a)$$

and the value of the function at the point $[m]$ is given by

$$|d_{\Delta(\tau')}d_{\Delta(\tau)}|^2([m]) = |d_{\Delta(\tau')}|^2([m]+\Delta(\tau)) \times |d_{\Delta(\tau)}|^2([m])$$
$$= \{d_{\Delta(\tau')}([m]+\Delta(\tau))d_{\Delta(\tau)}([m])\}^2. \qquad (2.33b)$$

D. The factorization lemma

The use of boson variables as a convenient realization for the carrier space of $U(n)$ is very familiar.[19] In order to realize all irreps of $U(n)$, it is necessary to assume n kinematically independent copies of an n-state boson variable; that is, one takes the variables a^i_j, $i,j = 1,2,\ldots,n$, with the commutators

$$[\bar{a}^i_j, a^{i'}_{j'}] = \delta^{i'}_i\delta^{j'}_j, \qquad (2.34)$$

all other commutators defined to be zero. The generators E_{ij} of the group $U(n)$ are defined by the mapping

$$E_{ij} \to \mathcal{E}_{ij} \equiv \sum_{k=1}^n a^k_i \bar{a}^k_j. \qquad (2.35)$$

It is clear, however, that these boson variables admit also of a second, isomorphic $U(n)$ group generated by the operator mapping

$$E^{ij} \to \mathcal{E}^{ij} \equiv \sum_{k=1}^n a^i_k \bar{a}^j_k, \qquad (2.36)$$

and that, moreover, the two sets of operators $\{E_{ij}\}$ and $\{E^{ij}\}$ commute. Thus, this boson realization involves the direct product group $U(n)\times U(n)$.

In fact, one sees at once that this boson realization $\{a^i_j\}$ really involves the group $U(n^2)$ and all totally symmetric irreps thereof. This defines a canonical imbedding of $U(n)$ in the sequence of groups $U(n^2)\supset U(n)\times U(n)\supset U(n)$, in which, moreover, the irrep labels of the two $U(n)$ groups in $U(n)\times U(n)$ coincide [we denote this by $U(n)\star U(n)$]. This structure is precisely the analog to that exhibited by the tensor operators of $U(n)$, and Ref. 11 discusses this canonical imbedding in detail, proving the factorization lemma to which we now turn.

Let

$$\left|\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}\right\rangle \qquad (2.37)$$

denote a normalized basis vector in an irrep space of $U(n)\star U(n)$. In this notation, the first $U(n)$ refers to the $U(n)$ group with generators E_{ij}, the second to the $U(n)$ group with generators E^{ij}. These two $U(n)$ groups are isomorphic but distinct (and commuting); the placement of the indices is merely a reminder as to which group is which ("upper" vs "lower")—there is no other implication.

The star signifies that the Casimir invariants of the irreps of these two groups coincide. Hence, both

$$(M) = \begin{pmatrix}[M]\\ (M)\end{pmatrix} \quad \text{and} \quad (M') = \begin{pmatrix}(M')\\ [M]\end{pmatrix} \qquad (2.38)$$

in Eq. (2.37) are Gel'fand patterns, the second one being inverted. The basis vector (2.37) may also be written in the form

$$\left|\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}\right\rangle = \mathfrak{M}([M])^{-1/2}B\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}(A)|0\rangle, \qquad (2.39)$$

where

$$B\begin{pmatrix}(M')\\ [M]\\ (M)\end{pmatrix}(A) \qquad (2.40)$$

is an operator-valued polynomial in the set of boson creation operators $A=\{a^i_j\}$, the symbol $|0\rangle$ denotes the vacuum ket, and $\mathfrak{M}([M])$ is the measure of the highest weight tableau associated with $[M]$:

$$\mathfrak{M}([M]) \equiv \frac{\prod_{i=1}^n (M_{in}+n-i)!}{\prod_{i<j=1}^n (M_{in}-M_{jn}+j-i)}. \qquad (2.41)$$

The introduction of $\mathfrak{M}^{-1/2}$ into Eq. (2.39) defines the manner in which the boson operators (2.40) are normalized: For example, if (M') and (M) are maximal, i.e.,

$$M'_{ij} = M_{in}, \quad M_{ij} = M_{in}, \quad \text{all } i,j, \qquad (2.42)$$

then

$$B\begin{pmatrix}(\max)\\ [M]\\ (\max)\end{pmatrix}(A) = \prod_{k=1}^n (a^{12\cdots k}_{12\cdots k})^{M_{kn}-M_{k+1n}}, \qquad (2.43)$$

where $a^{1\cdots k}_{1\cdots k}$ is the determinant formed from the k bosons a^i_j, $i,j\leqslant k$.

The boson operator (2.40) is clearly a tensor operator in either its lower or upper Gel'fand pattern with respect to transformations in the respective $U(n)$ subgroup of $U(n)\star U(n)$. As such, it must be bilinear in the canonical Wigner operators which are defined, respectively, on the two $U(n)$ groups. *The factorization lemma asserts that the precise form of this bilinear relation is*

$$B\begin{pmatrix} (M') \\ [M] \\ (M) \end{pmatrix}(A) = \mathfrak{M}^{1/2} \sum_{(\Gamma)} \left\langle \begin{matrix} (\Gamma) \\ [M] \\ (M) \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} (\Gamma) \\ [M] \\ (M') \end{matrix} \right\rangle_{u} \mathfrak{M}^{-1/2}, \quad (2.44)$$

where $\mathfrak{M}$ is an invariant operator of $U(n) \star U(n)$ which has eigenvalue equal to the measure $\mathfrak{M}([m])$ for an arbitrary vector with labels $[m]$. The indices ℓ and u designate the fact that the Wigner operators act, respectively, on the lower and upper Gel'fand patterns of an arbitrary vector of $U(n) \star U(n)$:

$$\left| \begin{pmatrix} (\mu) \\ [m] \\ (m) \end{pmatrix} \right\rangle = \left| \begin{matrix} \begin{pmatrix} (\mu) \\ [m] \end{pmatrix} \\ --- \\ \begin{pmatrix} [m] \\ (m) \end{pmatrix} \end{matrix} \right\rangle. \quad (2.45)$$

Note that when we apply the *individual* Wigner operators in Eq. (2.44) to an arbitrary basis vector (2.45), we should consider the common labels $[m]$ to be two identical sets of labels as indicated. Note also that the two Wigner operators in Eq. (2.44) commute since they act in different spaces and that the application of a *single* Wigner operator carries a vector *outside* the irrep space of $U(n) \star U(n)$ in the general case. [More precisely, these properties serve to *define* the meaning of the product of operators in Eq. (2.44).]

The work in Sec. 3 makes use of the following important special case of Eq. (2.44):

$$a_i^j = \mathfrak{M}^{1/2} \left(\sum_{\tau=1}^{n} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ i \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ j \end{matrix} \right\rangle_{u} \right) \mathfrak{M}^{-1/2}. \quad (2.46)$$

(This special case accounts for the term "boson factorization.")

The boson polynomials (2.40) are of considerable interest from still another point of view: Under the mapping $a_i^j \to u_{ij}$, where u_{ij} is the element in row i and column j of an $n \times n$ unitary matrix U, these polynomials become

$$B\begin{pmatrix} (M') \\ [M] \\ (M) \end{pmatrix}(U). \quad (2.47)$$

These functions are precisely the matrix elements of the unitary matrix irrep $[M]$ *of* $U(n)$.[3,20]

The structural significance of Eq. (2.44) is clear: The matrix elements of the boson operator (2.44) yields, in the boson language, a result which is completely analogous [in $U(n)$] to Wigner's $SU(2)$ result[21] which expresses the integral of three representation functions in terms of two $SU(2)$ Wigner coefficients.

3. THE TOTALLY SYMMETRIC TENSOR OPERATORS IN $U(n)$ AND THEIR CONJUGATES

A. Application of the pattern calculus

Consider the coupling law (2.4) expressed in the inverted form

$$\begin{bmatrix} (\Gamma') \\ [M'] \\ (\gamma') \end{bmatrix} \begin{bmatrix} (\Gamma) \\ [M] \\ (\gamma) \end{bmatrix} = \sum_{(\Lambda)(\Gamma'')(\gamma'')} \left\{ \begin{pmatrix} [M] + [\Delta(\Lambda)] \\ (\Gamma'') \end{pmatrix} \begin{pmatrix} (\Lambda) \\ [M'] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [M] \\ (\Gamma) \end{pmatrix} \right\}$$
$$\times \left[\begin{pmatrix} [M] + [\Delta(\Lambda)] \\ (\gamma'') \end{pmatrix} \begin{pmatrix} (\Lambda) \\ [M'] \\ (\gamma') \end{pmatrix} \begin{pmatrix} [M] \\ (\gamma) \end{pmatrix} \right]$$
$$\times \begin{bmatrix} (\Gamma'') \\ [M] + [\Delta(\Lambda)] \\ (\gamma'') \end{bmatrix}. \quad (3.1)$$

Particularizing to the totally symmetric projective operators of the *same* extremal *lower* patterns—say, those designated by ρ in the notation (2.17) and (2.18)—we obtain the following explicit form:

$$\begin{bmatrix} (\Gamma') \\ [p-1 \quad \dot{0}] \\ \rho \end{bmatrix} \begin{bmatrix} \tau \\ [1 \quad \dot{0}] \\ \rho \end{bmatrix} = \left\{ \begin{pmatrix} [p \quad \dot{0}] \\ (\Gamma) \end{pmatrix} \begin{pmatrix} 1 \\ [p-1 \quad \dot{0}] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [1 \quad \dot{0}] \\ \tau \end{pmatrix} \right\} \begin{bmatrix} (\Gamma) \\ [p \quad \dot{0}] \\ \rho \end{bmatrix}, \quad (3.2)$$

where (Γ) is uniquely determined from its Δ pattern: $[\Delta(\Gamma)] = [\Delta(\Gamma')] + [\Delta(\tau)]$. [The square-bracket invariant in Eq. (3.1) does not appear in Eq. (3.2) because it has value 1 on the relevant extremal patterns.] We next iterate Eq. (3.2), multiplying at each iteration by the appropriate Racah invariant to deduce

$$\begin{bmatrix} \tau_p \\ [1 \quad \dot{0}] \\ \rho \end{bmatrix} \begin{bmatrix} \tau_{p-1} \\ [1 \quad \dot{0}] \\ \rho \end{bmatrix} \cdots \begin{bmatrix} \tau_2 \\ [1 \quad \dot{0}] \\ \rho \end{bmatrix} \begin{bmatrix} \tau_1 \\ [1 \quad \dot{0}] \\ \rho \end{bmatrix} = \mathcal{I}_{\tau_p \cdots \tau_2 \tau_1} \begin{bmatrix} (\Gamma) \\ [p \quad \dot{0}] \\ \rho \end{bmatrix}, \quad (3.3)$$

where $\tau_p \cdots \tau_2 \tau_1$ is any set of integers $(1 \leq \tau_i \leq n)$ such that it contains Δ_1 1's, Δ_2 2's, ..., Δ_n n's, where $\Delta_1 + \Delta_2 + \cdots + \Delta_n = p$, and $\Delta_i \equiv \Delta_{in}(\Gamma)$. Then

$$\sum_{i=1}^{p} \Delta(\tau_i) = [\Delta]. \quad (3.4)$$

$\mathcal{I}_{\tau_p \cdots \tau_2 \tau_1}$ is explicitly given by a string of Racah invariants (p in number); but we will utilize for the present only the fact that it is a $U(n)$ invariant, i.e., $\mathcal{I}_{\tau_p \cdots \tau_2 \tau_1}$ is a function whose values depend only on the irrep labels $[m]$ of an arbitrary state vector.

We next replace each projective function in the left-hand side of Eq. (3.3) by the form (2.20a), where, in addition, we split each fundamental Δ pattern function into the first of the quotient forms (2.24). The left-hand side of Eq. (3.3) becomes

$$\text{(phase) } F_R\begin{pmatrix} & \Delta(\tau_p) & \\ [1 & & \dot{0}] \\ & \Delta'(\rho) & \end{pmatrix} \cdots F_R\begin{pmatrix} & \Delta(\tau_2) & \\ [1 & & \dot{0}] \\ & \Delta'(\rho) & \end{pmatrix}$$
$$\times F_R\begin{pmatrix} & \Delta(\tau_1) & \\ [1 & & \dot{0}] \\ & \Delta'(\rho) & \end{pmatrix} \Big/ d_{\Delta(\tau_p)} \cdots d_{\Delta(\tau_2)} d_{\Delta(\tau_1)}, \quad (3.5a)$$

where we have written

$$d_{\Delta(\tau)} = d\begin{pmatrix} & \Delta(\tau) & \\ [1 & & \dot{0}] \end{pmatrix}. \quad (3.5b)$$

We now point out the following fundamental fact: *The restricted Δ pattern functions appearing in the numerator of Eq.* (3.5a) *multiply by addition of their patterns; the underlying origin of this property is geometrical—the arrow patterns of this string of p functions contain no opposing arrows*. Thus, the numerator of Eq. (3.5a) is simply

$$\text{(phase) } F_R\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \\ & \rho & \end{pmatrix}, \quad (3.6)$$

independently of the particular τ_i which satisfy Eq. (3.4). The denominator functions appearing in Eq. (3.5a) and the invariant factor appearing in Eq. (3.3) can contribute at most a function whose values depend on the $U(n)$ irrep labels $[m]$; the following *form* must hold:

$$\begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & \rho & \end{bmatrix} = \text{(phase) } F_R\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \\ & p\Delta'(\rho) & \end{pmatrix} \Big/ D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \end{pmatrix}, \quad (3.7)$$

where $D\left(\begin{smallmatrix} [\Delta] \\ [p \quad \dot{0}] \end{smallmatrix}\right)$ denotes a *new* $U(n)$ denominator function *with properties yet to be determined*. We know it to be a function which takes its values only on $[m]$. The form *of Eq.* (3.7) *is uniquely implied by the coupling law* (3.2) *and the geometrical properties of the arrow patterns*. [Note that there is a relation implied by Eq. (3.3) between the denominator function $D\left(\begin{smallmatrix} [\Delta] \\ [p \quad \dot{0}] \end{smallmatrix}\right)$, the string of fundamental denominators in Eq. (3.5a), and the invariant $\mathcal{g}_{\tau_p\cdots\tau_2\tau_1}$; but this relation will not be required, since we plan to give a more direct derivation of the function $D\left(\begin{smallmatrix} [\Delta] \\ [p \quad \dot{0}] \end{smallmatrix}\right)$.]

Let us remark that the value of the Δ pattern function appearing in Eq. (3.7) is given explicitly by the pattern calculus rules. Only the phase and the denominator are undetermined. Observe, furthermore, that for $(\Gamma) \to \tau$ [the extremal pattern in the notation (2.18)], then

$$D\begin{pmatrix} & \Delta(\tau) & \\ [p & & \dot{0}] \end{pmatrix} = d\begin{pmatrix} & \Delta(\tau) & \\ [p & & \dot{0}] \end{pmatrix}, \quad (3.8)$$

i.e., the value of the denominator function, in this special case, is obtained from the usual $U(n)$ denominator pattern calculus rules. Our aim is to demonstrate *a more general U(n) denominator pattern calculus rule which yields the value of the denominator function* $D\left(\begin{smallmatrix} [\Delta] \\ [p \quad \dot{0}] \end{smallmatrix}\right)$, this rule reducing, of course, to the usual rule in the particular case (3.8).

Starting from

$$\begin{bmatrix} & \tau & \\ [p-1 & & \dot{0}] \\ & (\Gamma'') & \end{bmatrix} \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho & \end{bmatrix}$$
$$= \left[\begin{pmatrix} [p & & \dot{0}] \\ & (\Gamma') & \end{pmatrix} \begin{pmatrix} & 1 & \\ [p-1 & & \dot{0}] \\ & (\Gamma'') & \end{pmatrix} \begin{pmatrix} [1 & & \dot{0}] \\ & \rho & \end{pmatrix}\right] \begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix} \quad (3.9)$$

in which $[\Delta(\Gamma')]_n = [\Delta(\Gamma'')]_n + [\Delta(\rho)]_n$ and following the same iterative procedure which led to Eq. (3.3), we also deduce the form

$$\begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_p & \end{bmatrix} \cdots \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_2 & \end{bmatrix} \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_1 & \end{bmatrix}$$
$$= [q!\,(p-q)!/p!]^{1/2} \mathcal{g}'_{\rho_p\cdots\rho_2\rho_1} \begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix}, \quad (3.10a)$$

$$\sum_{i=1}^{p} \Delta(\rho_i) = [\Delta']_n = [\Delta'_1 \Delta'_2 \cdots \Delta'_n], \quad (3.10b)$$

where $\mathcal{g}'_{\rho_p\cdots\rho_2\rho_1}$ is a string of p $U(n-1)$ Racah invariants, and where $q \equiv \Gamma'_{1,n-1} = \Delta'_1 + \Delta'_2 + \cdots + \Delta'_{n-1}$. The values of this string of Racah functions depend only on the $U(n-1)$ irrep labels $[m']$ of an arbitrary state vector. {The binomial coefficient (inverted) comes from expressing the string of p square-bracket invariants in terms of the $U(n):U(n-1)$ reduced matrix elements of the relevant projective operators [see Eq. (2.4)] and $U(n-1)$ Racah invariants.} Using the same argument which led to Eq. (3.7), we obtain the following result:

$$\begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix} = \text{(phase) } F_{R'}\begin{pmatrix} & p\Delta(\tau) & \\ [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix} \Big/ D\begin{pmatrix} [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix}, \quad (3.11)$$

where $[\Delta'] = [\Delta']_{n-1}$, and $D\left(\begin{smallmatrix} [p \quad \dot{0}] \\ [\Delta'] \end{smallmatrix}\right)$ is a $U(n-1)$ denominator function with properties yet to be determined. It is a function which takes its values only on the $U(n-1)$ irrep labels $[m']$.

Once again *the origin of the restricted Δ pattern function in Eq.* (3.11) *is the geometrical property expressed by the statement that there are no opposing arrows in the string of p fundamental restricted Δ pattern functions occurring in the left-hand side of Eq.* (3.10a)—*these factors accordingly multiply by addition of their Δ patterns*. The value of this restricted pattern function is, of course, given by the pattern calculus rules.

One suspects (correctly) that forms having a structure identical to that of Eqs. (3.7) and (3.11) should hold also for the $U(n)$ conjugates $\langle \dot{0}\ -p\rangle$ and the $SU(n)$ conjugates $\langle \dot{p}\ 0\rangle$. However, such results can be obtained from the following more general result relating the Hermitian conjugate to the $U(n)$ conjugate, for totally symmetric Wigner operators:

$$\left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ (M) \end{matrix} \right\rangle^{\dagger} = (-1)^{\phi(\Gamma)-\phi(M)} \times \mathfrak{D}^{-1/2} \left\langle \begin{matrix} (\bar{\Gamma}) \\ [\dot{0} \quad -p] \\ (M) \end{matrix} \right\rangle \mathfrak{D}^{1/2}. \tag{3.12a}$$

This relation is an immediate consequence of Eq. (4.7) of Ref. 4. $\mathfrak{D}$ denotes the dimension operator. Specifically,

$$\mathfrak{D}([m]) = \prod_{i<j=1}^{n} (p_{in} - p_{jn}) \Big/ 1!\,2! \cdots (n-1)!. \tag{3.12b}$$

$\phi(\Gamma)$ and $\phi(M)$ are, respectively, the sums of all the entries in the operator pattern (Γ) and the Gel'fand pattern (M):

$$\phi(\Gamma) = \sum_{i=1}^{n} \Gamma_{1i}, \tag{3.12c}$$

$$\phi(M) = \sum_{i=1}^{n} M_{1i}. \tag{3.12d}$$

$(\bar{\Gamma})$ denotes the unique operator pattern determined by

$$[\Delta(\bar{\Gamma})]_n = -[\Delta(\Gamma)]_n, \tag{3.12e}$$

and $(\bar{M})$ has a similar definition, i.e., $(\bar{M})$ denotes the Gel'fand pattern having the negative weight of the pattern (M). [A more general definition of $(\bar{\Gamma})$ and $(\bar{M})$ will be found in Ref. 3.]

Using the $U(n):U(n-1)$ subgroup reduction law, Eq. (3.12a) is easily transcribed into the following relation between projective operators:

$$\begin{bmatrix} (\bar{\Gamma}) \\ [\dot{0} \quad -p] \\ (\bar{\gamma}) \end{bmatrix} = (-1)^{\phi(\Gamma)-\phi(\gamma)} \times \left(\frac{\mathfrak{D}_n}{\mathfrak{D}_{n-1}}\right)^{1/2} \begin{bmatrix} (\Gamma) \\ [p \quad \dot{0}] \\ (\gamma) \end{bmatrix}^{\dagger} \left(\frac{\mathfrak{D}_n}{\mathfrak{D}_{n-1}}\right)^{-1/2}, \tag{3.13}$$

where the subscripts n and $n-1$ on $\mathfrak{D}$ refer, respectively, to the dimension operators in $U(n)$ and $U(n-1)$. Noting that

$$\begin{bmatrix} (\tilde{\Gamma}) \\ [\dot{p} \quad 0] \\ (\bar{\gamma}) \end{bmatrix} = \begin{bmatrix} (1) \\ [\dot{1}] \\ (1) \end{bmatrix}^{p} \begin{bmatrix} (\bar{\Gamma}) \\ [\dot{0} \quad -p] \\ (\bar{\gamma}) \end{bmatrix}, \tag{3.14}$$

we also obtain the relation between the $[\dot{p}\ 0]$ projective operators and the totally symmetric ones, $[p\ \dot{0}]$. $(\tilde{\Gamma})$ denotes the pattern determined (uniquely) by $[\Delta(\tilde{\Gamma})]_n = [\dot{p}]_n + [\Delta(\bar{\Gamma})]_n$.

We now turn to the following section for the determination of the denominator functions appearing in Eqs. (3.7) and (3.11).

B. Application of the factorization lemma

There are several ways of determining the denominator functions appearing in Eqs. (3.7) and (3.11). The factorization lemma provides an extremely useful technique for illuminating the simple structures underlying otherwise very complicated expressions. In view of the fact that one of the principal aims of the present paper is to explain, by actual examples, structural approaches to group theoretic problems, we shall accordingly emphasize, more than is customary, the purely technical and manipulative aspects of our work. The present section, developing properties of the denominator functions, provides an instructive example.

It is first of all clear that the denominator function in Eq. (3.7) cannot depend on ρ. Consider then the following totally symmetric boson polynomial in the irrep space of $U(n) \star U(n)$:

$$B\begin{pmatrix} 1 \\ [p \quad \dot{0}] \\ 1 \end{pmatrix}(A) = (a_1^1)^p, \tag{3.15}$$

where the 1 designates the maximal pattern in the notation (2.18). We now apply the factorization lemma (2.44) directly to this operator and obtain

$$B\begin{pmatrix} 1 \\ [p \quad \dot{0}] \\ 1 \end{pmatrix}(A) = \mathfrak{M}^{1/2} \sum_{(\Gamma)} \left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{u} \mathfrak{M}^{-1/2} \tag{3.16}$$

But also, from Eq. (2.46), we have

$$(a_1^1)^p = \mathfrak{M}^{1/2} \left(\sum_{\tau=1}^{n} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{u} \right)^{p} \mathfrak{M}^{-1/2} \tag{3.17}$$

and, therefore,

$$\sum_{(\Gamma)} \left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{u} = \left(\sum_{\tau=1}^{n} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{\ell} \left\langle \begin{matrix} \tau \\ [1 \quad \dot{0}] \\ 1 \end{matrix} \right\rangle_{u} \right)^{p}. \tag{3.18}$$

Relation (3.18) is an operator identity on the space of state vectors of $U(n) \star U(n)$, and we are at liberty to let Eq. (3.18) act on any selected $U(n) \star U(n)$ state. We choose states which are maximal in the $U(n-2) \star U(n-2)$ subgroup labels and which carry the same $U(n-1)$ labels $[m']$:

$$\left| \begin{pmatrix} (\max) \\ [m'] \\ [m] \\ [m'] \\ (\max) \end{pmatrix} \right\rangle. \tag{3.19}$$

This class of state vectors is precisely the space in which the projective operators act [see Eq. (2.25) of Ref. 4]. Furthermore, the relevant $U(n-1)$ Wigner operators in the subgroup reduction of Eq. (3.18) act like the identity on the subspace of states of the form (3.19). (Their matrix elements have numerical value 1.) Thus, Eq. (3.18) implies the following relation between unit projective operators:

$$\sum_{(\Gamma)} \begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix}_\ell \begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix}_u = \left(\sum_{\tau=1}^{n} \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix}_\ell \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix}_u \right)^p, \tag{3.20}$$

i.e., this is an operator identity on the $U(n) \star U(n)$ states of the type (3.19). [Indeed, we could even choose the upper $U(n-1)$ labels and lower $U(n-1)$ labels to be distinct.]

We now take matrix elements of Eq. (3.20) between the initial state (3.19) and a final state of the same form but with $U(n) \star U(n)$ irrep labels $[m] + [\Delta]$—where $[\Delta]$ is any $[\Delta]$ belonging to $[p\ \dot{0}]$—and with $U(n-1) \star U(n-1)$ irrep labels $[m'] + p\Delta'(1)$. This selects a single term from the left-hand side and gives the square of a reduced matrix element—the term appearing in the left-hand side of Eq. (3.22a) below.

Next consider this same matrix element for the right-hand side. We introduce p summation patterns $\tau_p \cdots \tau_2 \tau_1$ and write the pth power in Eq. (3.20) as a string of p operators (the same operator written p times), denoted symbolically as follows:

$$\left(\sum_{\tau_p}[\ ]_\ell [\ ]_u\right) \cdots \left(\sum_{\tau_2}[\ ]_\ell [\ ]_u\right)\left(\sum_{\tau_1}[\ ]_\ell [\ ]_u\right) = \sum_{\tau_p \cdots \tau_2 \tau_1}\left([\ ]_\ell [\ ]_u\right) \cdots \left([\ ]_\ell [\ ]_u\right)\left([\ ]_\ell [\ ]_u\right). \tag{3.21}$$

The application of the first pair $[\ ]_\ell [\ ]_u$ (the pair on the far right) to the state (3.19) gives a new state of the type (3.19) with $[m] \to [m] + \Delta(\tau_1)$, $[m'] \to [m'] + \Delta'(1)$, multiplied by the *square* of a fundamental reduced matrix element, $\cdots$, the application of the last pair $[\ ]_\ell [\ ]_u$ to the state generated by all the pairs to the right gives a final state of the type (3.19) with $[m] \to [m] + \sum_{i=1}^{p} \Delta(\tau_i)$, $[m'] \to [m'] + p\Delta'(1)$, multiplied by the square of a fundamental reduced matrix element. The result may be expressed as follows:

$$\left\langle \begin{matrix} [m] + [\Delta] \\ [m'] + p\Delta'(1) \end{matrix} \right| \begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix} \left| \begin{matrix} [m] \\ [m'] \end{matrix} \right\rangle^2 = \sum_{\tau_p \cdots \tau_2 \tau_1} \left\langle \begin{matrix} [m] + [\Delta] \\ [m'] + p\Delta'(1) \end{matrix} \right| \begin{bmatrix} & \tau_p & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \cdots \begin{bmatrix} & \tau_2 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \times \begin{bmatrix} & \tau_1 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \left| \begin{matrix} [m] \\ [m'] \end{matrix} \right\rangle^2, \tag{3.22}$$

where the summation is over all $\tau_p \cdots \tau_2 \tau_1$ $(1 \leqslant \tau_i \leqslant n)$ such that

$$\sum_{i=1}^{p} \Delta(\tau_i) = [\Delta]. \tag{3.23}$$

Observe that—although we have used a specific realization (the boson basis) in obtaining Eq. (3.22)—this result is an *abstract general relation*. Indeed, it is precisely the *abstract operator statement*:

$$\begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix}^\dagger \begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix} = \sum_{\tau_p \cdots \tau_2 \tau_1} \left(\begin{bmatrix} & \tau_p & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \cdots \begin{bmatrix} & \tau_2 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \begin{bmatrix} & \tau_1 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \right)^\dagger \times \left(\begin{bmatrix} & \tau_p & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \cdots \begin{bmatrix} & \tau_2 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \begin{bmatrix} & \tau_1 & \\ [1 & & \dot{0}] \\ & 1 & \end{bmatrix} \right). \tag{3.24}$$

We could have derived this result, Eq. (3.24), by purely abstract (algebraic) techniques; but this would have required detailed use of the properties of the Racah invariants. The power of the factorization lemma for directly obtaining abstract and general results is clear. [A similar technique has previously been used[4] to determine the abstract structure of the generators of $U(n)$ in terms of the fundamental Wigner operators and their conjugates.]

Equation (3.20), hence, Eq. (3.24), is also correct for an arbitrary lower extremal pattern ρ, and again the proof can be given from the factorization lemma, starting with $(a_p^\rho)^p$, but now paying more attention to the $U(n-1)$ Wigner operators which occur. Note, however, that for $\rho = n$ the result is immediate.

Next, let us interpret Eq. (3.24) in terms of the projective functions and the restricted Δ pattern functions introduced in Sec. 2C. The result is

$$\left\{ \begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix} \begin{pmatrix} [m] \\ [m'] \end{pmatrix} \right\}^2 = \left\{ F_R \begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \\ & \Delta'(1) & \end{pmatrix} \begin{pmatrix} [m] \\ [m'] \end{pmatrix} \right\}^2 \times \sum_{\tau_p \cdots \tau_2 \tau_1} \frac{1}{|d_{\Delta(\tau_p)} \cdots d_{\Delta(\tau_2)} d_{\Delta(\tau_1)}|^2([m])}, \tag{3.25}$$

that is,

$$\begin{bmatrix} & (\Gamma) & \\ [p & & \dot{0}] \\ & 1 & \end{bmatrix} = (\text{phase})\ F_R \begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \\ & \Delta'(1) & \end{pmatrix} \Big/ D \begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \end{pmatrix}, \tag{3.26a}$$

where

$$\frac{1}{\left| D \begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \end{pmatrix} \right|^2} \equiv \sum_{\tau_p \cdots \tau_2 \tau_1} \frac{1}{|d_{\Delta(\tau_p)} \cdots d_{\Delta(\tau_2)} d_{\Delta(\tau_1)}|^2}, \tag{3.26b}$$

where the summation is over all $\tau_p \cdots \tau_2 \tau_1$ $(1 \leqslant \tau_i \leqslant n)$

satisfying Eq. (3.23). *The complete result, including the determination of the new denominator function, is thus proved directly from the factorization lemma and the geometrical properties of the arrow patterns.*

Let us remark that Eq. (3.26b) completely determines the denominator function $\left|D\left(\begin{smallmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{smallmatrix}\right)\right|^2$, since the right-hand denominator functions are fundamental, and their values are already given by the pattern calculus rules.

We can given an elegant interpretation to Eq. (3.26b) by considering the irrep labels $[m] = [m_{1n} m_{2n} \cdots m_{nn}]$ as specifying a lattice point in n-dimensional Euclidean space R^n. We then consider $\Delta(\tau) = [0\cdots 010 \cdots 0]$ (1 in position τ) as an elementary shift acting along axis τ. Since τ may be $1, 2, \ldots,$ or n, we can make elementary shifts along any of the n perpendicular directions. The label $[m]$ defines a lattice point in R^n as does $[m] + [\Delta]$. A selected set $\tau_p, \ldots, \tau_2, \tau_1$ (such that $\sum_i \Delta(\tau_i) = [\Delta]$) of integers then defines a *path* beginning at $[m]$ and ending at $[m] + [\Delta]$. Our result, Eq. (3.26b), then takes the very suggestive form

$$1\Big/\left|D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{pmatrix}\right|^2([m]) = \sum_{\substack{\text{all paths from} \\ [m] \text{ to } [m]+[\Delta]}} \begin{pmatrix}\text{path} \\ \text{contribution}\end{pmatrix}, \tag{3.27a}$$

where a "path contribution" is defined to be

$$\begin{pmatrix}\text{path} \\ \text{contribution}\end{pmatrix} = 1/|d_{\Delta(\tau_p)} \cdots d_{\Delta(\tau_2)}\, d_{\Delta(\tau_1)}|^2([m]), \tag{3.27b}$$

with the value of the denominator functions $d_{\Delta(\gamma)}$ being given by using Eq. (2.33b).

The expressions (3.27) are clearly suggestive of Feynman's approach to quantum mechanics and indicates that the pattern calculus has ultimately some kind of interpretation as a lattice quantum mechanics. The evident vagueness of this remark is balanced by its equally evident interest.

We defer further discussion of Eq. (3.26b) to the next subsection, and consider now the second denominator function of Eq. (3.11). It is related to the first denominator function of Eq. (3.26) for $n \to n-1$. To establish this relation, we first observe that Eq. (3.24) implies the following property of the $U(n)$ invariant operator of Eq. (3.3):

$$\sum_{\tau_p \cdots \tau_2 \tau_1} \mathcal{G}^\dagger_{\tau_p \cdots \tau_2 \tau_1}\, \mathcal{G}_{\tau_p \cdots \tau_2 \tau_1} = 1, \tag{3.28}$$

where the sum is over all τ_i satisfying Eq. (3.4). Next, consider the $U(n-1)$ invariant operator in Eq. (3.10a). We claim that the property (3.28) implies also

$$\sum_{\rho_p \cdots \rho_2 \rho_1} \mathcal{G}'^\dagger_{\rho_p \cdots \rho_2 \rho_1}\, \mathcal{G}'_{\rho_p \cdots \rho_2 \rho_1} = 1, \tag{3.29}$$

where the sum is over all $\rho_p \cdots \rho_2 \rho_1$ $(1 \leqslant \rho_i \leqslant n)$ such that Eq. (3.10b) holds. We argue as follows: Eq. (3.28) is a statement about a string of p $U(n)$ Racah invariants—*Eq.* (3.29) *is precisely this same property applied to a string of* q $U(n-1)$ *Racah invariants.* At first glance this conclusion appears to be incorrect because the ρ_i in the summation (3.29) can assume the value n. However, closer examination will show that for each ρ_i which assumes the value n, the corresponding $U(n-1)$ operator pattern in the $U(n-1)$ Racah invariant becomes

$$\begin{pmatrix} 0 \cdots 0 \\ \ddots \ \ \dot{} \\ \cdot 0 \cdot \end{pmatrix},$$

and the Racah invariant (for the patterns which actually occur) becomes the identity operator 1. Since each sequence $\rho_p \cdots \rho_2 \rho_1$ contains Δ'_1 1's, Δ'_2 2's, $\ldots$, Δ'_n n's, it follows that each string of p $U(n-1)$ Racah invariants in $\mathcal{G}'_{\rho_p \cdots \rho_2 \rho_1}$ reduces, in fact, to a string of $\Delta'_1 + \Delta'_2 + \cdots + \Delta'_{n-1} = q$ $U(n-1)$ Racah invariants. The summation in Eq. (3.29) assumes the form

$$\sum_{\rho_q \cdots \rho_2 \rho_1} \mathcal{G}'^\dagger_{\rho_q \cdots \rho_2 \rho_1}\, \mathcal{G}'_{\rho_q \cdots \rho_2 \rho_1}, \tag{3.30}$$

where the summation is now over all integers $\rho_q \cdots \rho_2 \rho_1$ $(1 \leqslant \rho_i \leqslant n-1)$ such that

$$\sum_{i=1}^{q} \Delta'(\rho_i) = [\Delta']_{n-1} = [\Delta'_1 \Delta'_2 \cdots \Delta'_{n-1}].$$

Furthermore, the form (3.30) becomes *precisely* the form of the left-hand side of Eq. (3.28) for $p \to q$ and $n \to n-1$. Since property (3.28) is true for all $n = 2, 3, \cdots$ and all $p = 1, 2, \cdots$, the proof of Eq. (3.29) follows.

Using property (3.29), we obtain the following result from Eq. (3.10a):

$$\begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix}^\dagger \begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix} = \begin{pmatrix} p \\ q \end{pmatrix} \sum_{\rho_p \cdots \rho_2 \rho_1} \left(\begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_p & \end{bmatrix} \cdots \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_2 & \end{bmatrix} \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_1 & \end{bmatrix} \right)^\dagger \times \left(\begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_p & \end{bmatrix} \cdots \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_2 & \end{bmatrix} \begin{bmatrix} & \tau & \\ [1 & & \dot{0}] \\ & \rho_1 & \end{bmatrix} \right), \tag{3.31a}$$

in which the sum is over all $\rho_p \cdots \rho_2 \rho_1$ $(1 \leqslant \rho_i \leqslant n)$ such that

$$\sum_{i=1}^{p} \Delta(\rho_i) = [\Delta']_n. \tag{3.31b}$$

Equation (3.31a) now implies

$$\begin{bmatrix} & \tau & \\ [p & & \dot{0}] \\ & (\Gamma') & \end{bmatrix} = (\text{phase})\, F_{R'} \begin{pmatrix} & p\Delta(\tau) & \\ [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix} \Big/ D \begin{pmatrix} [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix}, \tag{3.32a}$$

in which now $[\Delta'] = [\Delta']_{n-1}$. The new denominator function is defined by

$$\frac{1}{\left| D \begin{pmatrix} [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix} \right|^2} \equiv \begin{pmatrix} p \\ q \end{pmatrix} \sum_{\rho_p \cdots \rho_2 \rho_1} \frac{1}{|d'_{\Delta'(\rho_p)} \cdots d'_{\Delta'(\rho_2)}\, d'_{\Delta'(\rho_1)}|^2}, \tag{3.32b}$$

where

$$d'_{\Delta'(\rho_i)} \equiv d\begin{pmatrix}[1 & \dot{0}] \\ & \Delta'(\rho_i)\end{pmatrix}, \tag{3.32c}$$

$$\Delta'(\rho_i) = [0\cdots 010\cdots 0]_{n-1}, \quad 1 \text{ in position } \rho_i, \tag{3.32d}$$

$$\Delta'(n) = [0\cdots 0]_{n-1}, \tag{3.32e}$$

and where the summation is still over all $\rho_p \cdots \rho_2\rho_1$ $(1 \leqslant \rho_i \leqslant n)$, such that $\sum_{i=1}^{p}\Delta(\rho_i) = [\Delta']_n$. However, noting from Eq. (2.24) (for $[M] = [1\ \dot{0}]$) that $d'_{\Delta'(n)} = 1$, we see that Eq. (3.32b) reduces to

$$\frac{1}{\left|D\begin{pmatrix}[p & 0] \\ & [\Delta']\end{pmatrix}\right|^2} = \binom{p}{q} \sum_{\rho_q\cdots\rho_2\rho_1} \frac{1}{|d'_{\Delta'(\rho_q)}\cdots d'_{\Delta'(\rho_2)}\, d'_{\Delta'(\rho_1)}|^2}, \tag{3.33a}$$

in which now the summation is over all $\rho_q \cdots \rho_2\rho_1$ $(1 \leqslant \rho_i \leqslant n-1)$ such that

$$\sum_{i=1}^{q} \Delta'(\rho_i) = [\Delta']_{n-1}. \tag{3.33b}$$

Recall also that

$$q = \Gamma'_{1,n-1} = \Delta'_1 + \Delta'_2 + \cdots \Delta'_{n-1}. \tag{3.33c}$$

This new denominator function also has a sum-over-paths interpretation, where now each path carries equal weight given by the binomial coefficient $\binom{p}{q}$.

Despite the similarity of the sums in Eqs. (3.26b) and (3.33a), they do *not* represent examples of the *same* general sum, the reason being that the pattern calculus rules for forming fundamental $U(n)$ denominators differ from those for forming fundamental $U(n-1)$ denominators. It thus appears that we are confronted with the task of performing two difficult summations, and not just one. This, however, is not the case: It will now be demonstrated how the second sum (3.33a) can be converted into one of the form identical to that of Eq. (3.26b).

Consider the denominator arrow pattern for the fundamental Δ pattern $\Delta(\tau)$. This defines the function $d_{\Delta(\tau)}$. Its value is given by applying the pattern calculus rules for row n:

$$d_{\Delta(\tau)}([m]) = \left(\left|\prod_{\substack{i=1\\ i\neq\tau}}^{n} (p_{\tau n} - p_{in})\right|\right)^{1/2}. \tag{3.34a}$$

Suppose, however, we apply the *rule* appropriate to a $U(n-1)$ denominator, i.e., we evaluate the same function $d_{\Delta(\tau)}$ on the m_{in}, but now use the extra + 1 in each factor appropriate to the rules for a $U(n-1)$ denominator. This defines a new function evaluated at $[m]$:

$$d'_{\Delta(\tau)}([m]) = \left(\left|\prod_{i=1}^{n} (p_{\tau n} - p_{in} + 1)\right|\right)^{1/2}. \tag{3.34b}$$

The relation between these functions is simply

$$d'_{\Delta(\tau)} = \mathfrak{D}_n^{1/2}\, d_{\Delta(\tau)}\, \mathfrak{D}_n^{-1/2}, \tag{3.35a}$$

where $\mathfrak{D}_n$ is the $U(n)$ dimension function. The meaning to be associated with this multiplication of an ordinary scalar function with a Δ pattern function is[4]

$$d'_{\Delta(\tau)}([m]) = \mathfrak{D}_n^{1/2}([m] + \Delta(\tau))\, d_{\Delta(\tau)}([m])\, \mathfrak{D}_n^{-1/2}([m]). \tag{3.35b}$$

[Equation (3.35a) may be viewed as a statement as to how to convert, operationally, the $U(n)$ denominator rule into the $U(n-1)$ denominator rule.]

Replacing n by $n-1$ in Eq. (3.35a), we obtain

$$d'_{\Delta'(\rho)} = \mathfrak{D}_{n-1}^{1/2}\, d_{\Delta'(\rho)}\, \mathfrak{D}_{n-1}^{-1/2}, \tag{3.36}$$

where $\Delta'(\rho) = [\Delta(\rho)]_{n-1}$, $d'_{\Delta'(\rho)}([m'])$ is the value of the $U(n-1)$ denominator function obtained by the *usual* (+1) *pattern calculus rules*, and $d_{\Delta'(\rho)}([m'])$ is the value assigned to a $U(n-1)$ denominator function by using the factors (no + 1 is added) appropriate to the pattern calculus rules for $U(n)$ denominators.

We now use relation (3.36) in the right-hand side of Eq. (3.33a) to obtain

$$|d'_{\Delta'(\rho_q)}\cdots d'_{\Delta'(\rho_1)}|^2 = |\mathfrak{D}_{n-1}^{1/2}\, d_{\Delta'(\rho_q)}\cdots d_{\Delta'(\rho_1)}\mathfrak{D}_n^{-1/2}|^2$$
$$= (d_{\Delta'(\rho_q)}\cdots d_{\Delta'(\rho_1)})^{\dagger}\mathfrak{D}_{n-1}(d_{\Delta'(\rho_q)}\cdots d_{\Delta'(\rho_1)})\mathfrak{D}_{n-1}^{-1}. \tag{3.37}$$

Thus, we can write

$$1\Big/D\begin{pmatrix}[p & \dot{0}]_n \\ & [\Delta']_{n-1}\end{pmatrix} = \binom{p}{q}^{1/2}\Big/\mathfrak{D}_{n-1}^{1/2}\, D\begin{pmatrix} & [\Delta']_{n-1} \\ [q & \dot{0}]_{n-1} & \end{pmatrix}\mathfrak{D}_{n-1}^{-1/2}, \tag{3.38a}$$

where

$$\frac{1}{\left|D\begin{pmatrix} & [\Delta']_{n-1} \\ [q & \dot{0}]_{n-1} & \end{pmatrix}\right|^2} = \sum_{\rho_q\cdots\rho_2\rho_1} \frac{1}{|d_{\Delta'(\rho_q)}\cdots d_{\Delta'(\rho_1)}|^2}. \tag{3.38b}$$

The last denominator is now of precisely the same form as the one of Eq. (3.26b).

With this simplification in hand, we now turn, in the following section, to the problem of evaluating explicitly the complicated appearing sum of Eq. (3.26b).

C. The generalized denominator function pattern calculus

The defining equation for the denominator function under discussion, Eq. (3.26b), is useful conceptually but needs to be implemented in practice. We have succeeded in obtaining a remarkably concise and explicit form for these functions which represents a further development in the pattern calculus. This result is a generalization of the answer for $U(2)$ given earlier[5] (without proof).

Since the $U(2)$ result is much more easily understood, we shall discuss it in detail. Consider at first the special $U(2)$ denominator function

$$\frac{1}{\left|D\begin{pmatrix}1 & 1\\ 2 & 0\end{pmatrix}\right|^2} = \frac{1}{|d_{[1\ 0]}d_{[0\ 1]}|^2} + \frac{1}{|d_{[0\ 1]}d_{[1\ 0]}|^2}. \tag{3.39}$$

The value of $|d_{[1\ 0]}d_{[0\ 1]}|^2$ is $\{(d_{[1\ 0]}d_{[0\ 1]})(m_{12}m_{22})\}^2 = \{d_{[1\ 0]}(m_{12}, m_{22}+1)\}^2 \times \{d_{[0\ 1]}(m_{12}m_{22})\}^2 = (p_{12}-p_{22}-1)(p_{12}-p_{22})$. Similarly, the value of the second denominator is $(p_{12}-p_{22}+1)(p_{12}-p_{22})$. Thus,

$$\left\{D\begin{pmatrix}1 & 1\\ 2 & 0\end{pmatrix}(m_{12}m_{22})\right\}^2 = \tfrac{1}{2}(p_{12}-p_{22}-1)(p_{12}-p_{22}+1). \tag{3.40}$$

We now seek to understand the structure of the result (3.40). First, we decompose the Δ pattern [1 1] into the form [1 1] = [1 0] + [0 1]. We next draw the arrow-patterns for [1 0] and [0 1]:

Over each of these arrow patterns, we write the shift of the other arrow pattern:

Next, we associate with each of these arrow patterns an algebraic factor

$$[p(\text{tail}) + \delta_1] - [p(\text{head}) + \delta_2], \tag{3.41}$$

where δ_1 designates the numeral sitting over the tail of the arrow and δ_2 designates the numeral sitting over the head of the arrow. This yields the two algebraic factors $p_{12} - p_{22} - 1$ and $p_{22} - p_{12} - 1$, the absolute value of which is just the denominator (3.40), except for the factor 2, which is associated with the number of paths.

Encouraged by this simple structural interpretation of $D(\begin{smallmatrix}1 & 1\\ 2 & 0\end{smallmatrix})$, we try to generalize to $D(\begin{smallmatrix}\Delta_1 & \Delta_2\\ p & 0\end{smallmatrix})$, having the Δ pattern $[\Delta_1 \, \Delta_2]$, where $\Delta_1 + \Delta_2 = p$. Following the same procedure, we are led to the patterns

Using rule (3.41) for the first arrow in each pattern, adding 1 for the second arrow, etc., we write out the two algebraic factors associated with the two patterns above: $(p_{22} - p_{12} - \Delta_1)(p_{22} - p_{12} - \Delta_1 + 1) \cdots (p_{22} - p_{12} - \Delta_1 + \Delta_2 - 1)$ and $(p_{12} - p_{22} - \Delta_2) \times (p_{12} - p_{22} - \Delta_2 + 1) \cdots (p_{12} - p_{22} - \Delta_2 + \Delta_1 - 1)$. Since the number of distinct paths in the sum (3.26b) (for $n = 2$) is $p!/(\Delta_1)!(\Delta_2)!$, we are led to conjecture the following general form:

$$\left\{D\begin{pmatrix}\Delta_1 & \Delta_2\\ p & 0\end{pmatrix}(m_{12}m_{22})\right\}^2 = \frac{\Delta_1!\Delta_2!}{p!} \times \frac{(p_{12} - p_{22} - \Delta_2 + \Delta_1 - 1)!}{(p_{12} - p_{22} - \Delta_2 - 1)!} \times \frac{(p_{12} - p_{22} + \Delta_1)!}{(p_{12} - p_{22} + \Delta_1 - \Delta_2)!}$$
$$= \frac{\Delta_1!\Delta_2!}{p!} \times \frac{(\Delta_1 + \Delta_2 + 1)!}{(p_{12} - p_{22} + \Delta_1 - \Delta_2)} \times \begin{pmatrix}p_{12} - p_{22} + \Delta_1\\ \Delta_1 + \Delta_2 + 1\end{pmatrix}, \tag{3.42}$$

in which $\Delta_2 = p - \Delta_1$.

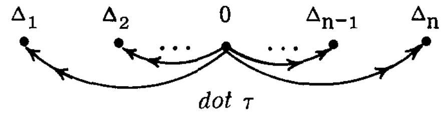
Observe that Eq. (3.42) is correct for the two extremal patterns $\Delta_1 = 0$ or $\Delta_1 = p$, in which case it reduces (properly) to the ordinary pattern calculus factor for the $U(2)$ denominator.

It is easy to prove by induction that the conjecture, Eq. (3.42), is indeed correct. It is, however, essentially as easy to prove the general result for $U(n)$, and this we now do.

Let us first give the generalization of the pattern calculus rules for the $U(n)$ denominator function $D(\begin{smallmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{smallmatrix})$ which has the Δ pattern $[\Delta_1\Delta_2 \cdots \Delta_n]$, where $\sum_i \Delta_i = p$. We decompose $[\Delta]$ in the form

$$[\Delta_1\Delta_2 \cdots \Delta_n] = [\Delta_1 0 \cdots 0] + [0\Delta_2 0 \cdots 0] + \cdots + [0 \cdots 0\Delta_n]. \tag{3.43}$$

Next, the arrow pattern is drawn for the Δ pattern $[0 \cdots 0\Delta_\tau 0 \cdots 0]$:



(3.44)

In this arrow pattern, there are Δ_τ arrows going from dot τ to each of the remaining dots. Above each dot appears the shift associated with the *sum* of all the remaining Δ patterns in the decomposition (3.43). We now associate an algebraic factor with each arrow according to the pattern calculus rules (applied to row n), shifting, however, each p(head) by the Δ_i which sits above dot i. Thus, the algebraic factor associated with the arrow pattern (3.44) is

$$\left|\prod_{\substack{i=1\\ i\neq\tau}}^{n} (p_{\tau n} - p_{in} - \Delta_i)(p_{\tau n} - p_{in} - \Delta_i + 1) \cdots (p_{\tau n} - p_{in} - \Delta_i + \Delta_\tau - 1)\right|. \tag{3.45}$$

We draw the n arrow patterns (3.44) corresponding to $\tau = 1, 2, \ldots, n$ and take the product of all the algebraic factors. Finally, we put in the number of paths factor $p!/\Delta_1!\Delta_2! \cdots \Delta_n!$ to obtain the following conjectured form:

$$\left\{D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{pmatrix}([m])\right\}^2 = \left[\frac{\Delta_1!\Delta_2! \cdots \Delta_n!}{p!}\right] \times \prod_{\tau=1}^{n} \{d_{\Delta(\tau)}([m] + [\Delta] - \Delta_\tau \Delta(\tau))\}^2$$
$$= \left[\frac{\Delta_1!\Delta_2! \cdots \Delta_n!}{p!}\right] \times \prod_{\tau=1}^{n} \prod_{\substack{i=1\\ i\neq\tau}}^{n} |(p_{\tau n} - p_{in} - \Delta_i)(p_{\tau n} - p_{in} - \Delta_i + 1) \times \cdots (p_{\tau n} - p_{in} + \Delta_\tau - \Delta_i - 1)|$$
$$= \frac{\Delta_1!\Delta_2! \cdots \Delta_n!}{p!\,\mathfrak{D}'([m] + [\Delta])} \times \prod_{i<j=1}^{n} (\Delta_i + \Delta_j + 1)! \begin{pmatrix}p_{in} - p_{jn} + \Delta_i\\ \Delta_i + \Delta_j + 1\end{pmatrix}, \tag{3.46a}$$

where $\mathfrak{D}'$ is the dimension operator $\mathfrak{D}$ multiplied by the numerical factors $1!2! \cdots (n-1)!$:

$$\mathfrak{D}'([m]) = \mathfrak{D}([m]) \times (1!2! \cdots (n-1)!) = \prod_{i<j=1}^{n} (p_{in} - p_{jn}). \tag{3.46b}$$

[Note that in making the last step in Eq. (3.46a) we have been careful to arrange the factors so that each term in the binomial coefficient is nonnegative.]

The proof of Eq. (3.46a) will be given by induction on p. Note that the result is correct for $p = 1$, since each allowed operator pattern (Γ) is extremal, and the result reduces correctly to the usual pattern calculus factor for the denominator function associated with row n.

In order to give the induction proof of Eq. (3.46), we first obtain a recursion relation. This is easily done directly from the sum-over-paths formulation (3.26b):

$$\left\{\frac{1}{D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{pmatrix}([m])}\right\}^2 = \sum_{\tau=1}^{n}\left\{\frac{1}{d_{\Delta(\tau)}([m]+[\Delta'])}\right\}^2 \times \left\{\frac{1}{D\begin{pmatrix} & [\Delta'] & \\ [p-1 & & \dot{0}]\end{pmatrix}([m])}\right\}^2, \quad (3.47a)$$

where

$$[\Delta'] = [\Delta']_n = [\Delta] - \Delta(\tau). \quad (3.47b)$$

This relation is just the geometrical statement that there are n points $[m] - \Delta(\tau), \tau = 1, 2, 3, \ldots, n$, which are one unit displaced from the final point of the path from $[m]$ to $[m] + [\Delta]$.

The proof of Eq. (3.46) now follows upon demonstrating that it satisfies Eq. (3.47a). But we easily determine that

$$\left\{\frac{1}{d_{\Delta(\tau)}([m]+[\Delta'])}\right\}^2 \left\{\frac{D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{pmatrix}([m])}{D\begin{pmatrix} & [\Delta'] & \\ [p-1 & & \dot{0}]\end{pmatrix}([m])}\right\}^2 = -\frac{1}{p}\,\frac{\prod_{i=1}^{n}(p'_{in}-p'_{\tau n}-\Delta_i)}{\prod_{\substack{i=1\\ i\neq\tau}}^{n}(p'_{in}-p'_{\tau n})}, \quad (3.48a)$$

where $p'_{in} = p_{in} + \Delta_i\ (i = 1, 2, \ldots, n)$. Thus, we must demonstrate that

$$\sum_{\tau=1}^{n}\frac{\prod_{i=1}^{n}(p'_{in}-p'_{\tau n}-\Delta_i)}{\prod_{\substack{i=1\\ i\neq\tau}}^{n}(p'_{in}-p'_{\tau n})} = -p. \quad (3.48b)$$

But this result is an easy consequence of the general summation formula (A1) of Ref. 4. The general validity of Eq. (3.46) has thus been established.

This result shows that there exists a significant extension of the pattern calculus rules in which *patterns act on patterns* in new and different ways. (These examples above are very likely only a first step in this program.) We find it quite striking that these simple geometrical rules [of the generalized $U(n)$ denominator pattern calculus] can effect completely the summation of the very complicated sum-over-paths relation, Eq. (3.26b).

D. Summary of results

Let us now summarize the results so far obtained by giving the complete answers, including the phases.

The determination of the phase in Eq. (3.7) proceeds as follows. Since the denominator function is positive (we always take the positive solution to $x^2 = a$ from the pattern calculus results), the phase is uniquely determined by the fundamental projective operators appearing in the left-hand side of Eq. (3.3). Since $\tau_p \cdots \tau_2\tau_1$ contains 1 Δ_1 times, ..., n Δ_n times, it follows from Eq. (2.20a) that the phase of Eq. (3.7) is

$$[S(\rho-1)]^{\Delta_1}[S(\rho-2)]^{\Delta_2}\cdots[S(\rho-n)]^{\Delta_n} = \begin{cases}(-1)^{\Delta_{\rho+1}+\Delta_{\rho+2}+\cdots+\Delta_n} & \text{for } \rho < n\\ 1 \quad \text{for } \rho = n\end{cases}. \quad (3.49)$$

Our first complete result is

$$\begin{bmatrix} (\Gamma) & \\ [p & \dot{0}] \\ \rho & \end{bmatrix} = (-1)^{\delta_\rho}F_R\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}] \\ & p\Delta'(\rho) & \end{pmatrix}\Big/ D\begin{pmatrix} & [\Delta] & \\ [p & & \dot{0}]\end{pmatrix}, \quad (3.50a)$$

where

(a) $\delta_\rho = \sum_{i=\rho+1}^{n}\Delta_i, \quad \delta_n = 0,$ (3.50b)

(b) the value of the restricted Δ pattern function F_R is obtained directly from the usual pattern calculus rules, and

(c) the value of the denominator function is given by the positive root of Eq. (3.46a), i.e., is now read off directly from the generalized $U(n)$ denominator pattern calculus rules [henceforth called simply the $U(n)$ *denominator pattern calculus* rules, since they always reduce to the usual rules for extremal patterns].

We also remark that for $(\Gamma) \to \tau$ the phase of Eq. (3.50) becomes +1 and agrees with our general "phase rule" given in Ref. 4.

Our second complete result is

$$\begin{bmatrix} \tau & \\ [p & \dot{0}] \\ (\Gamma') & \end{bmatrix} = (-1)^{\delta'_\tau}F_{R'}\begin{pmatrix} & p\Delta(\tau) & \\ [p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix}\Big/ D\begin{pmatrix}[p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix}, \quad (3.51a)$$

in which $[\Delta'] = [\Delta']_{n-1}$ and

$$\delta'_\tau = \sum_{i=1}^{\tau-1}\Delta'_i, \quad \delta'_1 = 0. \quad (3.51b)$$

The value of the denominator function on the $U(n-1)$ labels $[m'] = [m_{1n-1} \cdots m_{n-1n-1}]$ is obtained from Eq. (3.38a):

$$\frac{1}{D\begin{pmatrix}[p & & \dot{0}] \\ & [\Delta'] & \end{pmatrix}([m'])} = \binom{p}{q}^{1/2}\left[\frac{\mathfrak{D}_{n-1}([m'])}{\mathfrak{D}_{n-1}([m']+[\Delta'])}\right]^{1/2} \times \frac{1}{D\begin{pmatrix} & [\Delta'] & \\ [q & & \dot{0}]\end{pmatrix}([m'])}, \quad (3.51c)$$

where

(a) $q = \Gamma'_{1,n-1} = \sum_{i=1}^{n-1}\Delta'_i,$ (3.51d)

and (b) the value of the denominator function is given by the $U(n)$ denominator rules applied to $n-1$:

$$D\begin{pmatrix} & [\Delta'] & \\ [q & & \dot{0}]\end{pmatrix}([m']) = \left[\frac{(\Delta'_1)!(\Delta'_2)!\cdots(\Delta'_{n-1})!}{q!\,\mathfrak{D}'_{n-1}([m']+[\Delta'])}\right.$$

$$\times \prod_{i<j=1}^{n-1} (\Delta_i' + \Delta_j' + 1)! \begin{pmatrix} p_{in-1} - p_{jn-1} + \Delta_i' \\ \Delta_i' + \Delta_j' + 1 \end{pmatrix}\Bigg]^{1/2}. \tag{3.51e}$$

(This is obtained from Eq. (3.46a) by the appropriate substitutions.)

In addition to these results, we also have four similar expressions, which are implied by relations (3.13) and (3.14). We note only the following one:

$$\begin{bmatrix} (\bar{\Gamma}) \\ [\dot{0} - p] \\ \bar{\rho} \end{bmatrix} = (-1)^{\delta_{\bar{\rho}}} F_R \begin{pmatrix} -[\Delta] \\ [\dot{0} - p] \\ -p\Delta' \ (\rho) \end{pmatrix} \Bigg/ D \begin{pmatrix} -[\Delta] \\ [\dot{0} \quad -p] \end{pmatrix}, \tag{3.52a}$$

where

$$\text{(a)}\quad \delta_{\bar{\rho}} = \delta_\rho + \sum_{i=1}^{n} (\rho - i)\Delta_i, \tag{3.52b}$$

and

$$\text{(b)}\quad D\begin{pmatrix} -[\Delta] \\ [\dot{0} - p] \end{pmatrix}([m]) = \left[\frac{\Delta_1!\Delta_2!\cdots\Delta_n!}{p!\,\mathfrak{D}'([m]-[\Delta])} \times \prod_{i<j=1}^{n} (\Delta_i + \Delta_j + 1)! \begin{pmatrix} p_{in} - p_{jn} + \Delta_j \\ \Delta_i + \Delta_j + 1 \end{pmatrix}\right]^{1/2}. \tag{3.52c}$$

We remark that this denominator is given directly by the (generalized) $U(n)$ denominator rules where the relevant decomposition of $-[\Delta]$ is $-[\Delta] = [-\Delta_1 0 \cdots 0] + [0 - \Delta_2 0 \cdots 0] + \cdots + [0 \cdots 0 - \Delta_n]$. The typical arrow pattern (in the sequence of n arrow patterns) is obtained by reversing the direction of the arrows in the arrow pattern (3.44) and placing a minus sign in front of each Δ_i.

Finally, we remark that the reduced matrix element obtained from Eq. (3.50a) for $\rho = n$ (all zeroes in the lower pattern) is also the matrix element

$$\left\langle \begin{matrix} [m]_n + [\Delta]_n \\ (m)_{n-1} \end{matrix} \middle| \left\langle \begin{matrix} (\Gamma) \\ [p \quad \dot{0}] \\ n \end{matrix} \right\rangle \middle| \begin{matrix} [m]_n \\ (m)_{n-1} \end{matrix} \right\rangle, \tag{3.53}$$

of the Wigner operator of the same labels since the $U(n-1)$ matrix element has the numerical value unity. One may then use the $U(n)$ generators to obtain[14,15] from this result the matrix elements of the general symmetric tensor operator. Equation (3.12a) then gives immediately the matrix elements of the $U(n)$ conjugate operator $\langle \dot{0} - p \rangle$, and by an appropriate shifting of labels also of the $SU(n)$ conjugate operator $\langle \dot{p} \ 0 \rangle$.

E. A class of $U(n)$ Racah coefficients and their limits

A certain class of Racah coefficients of $U(n)$ can be written down immediately from the results of the preceding subsection. The following coupling law is a special case of the general coupling law, Eq. (3.1):

$$\left\{ \begin{pmatrix} [p+q \quad \dot{0}] \\ (\Gamma'') \end{pmatrix} \begin{pmatrix} 1 \\ [q \quad \dot{0}] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [p \quad \dot{0}] \\ (\Gamma) \end{pmatrix} \right\} \times \begin{bmatrix} (\Gamma'') \\ [p+q \quad \dot{0}] \\ 1 \end{bmatrix} = \begin{bmatrix} (\Gamma') \\ [q \quad \dot{0}] \\ 1 \end{bmatrix} \begin{bmatrix} (\Gamma) \\ [p \quad \dot{0}] \\ 1 \end{bmatrix} \tag{3.54}$$

for $[\Delta(\Gamma'')] = [\Delta(\Gamma')] + [\Delta(\Gamma)]$.

Since

$$F_R\begin{pmatrix} [\Delta''] \\ [p+q \ \dot{0}] \\ (p+q)\ \Delta\ (1) \end{pmatrix} = F_R\begin{pmatrix} [\Delta'] \\ [q \ \dot{0}] \\ q\Delta\ (1) \end{pmatrix} F_R\begin{pmatrix} [\Delta] \\ [p \ \dot{0}] \\ p\Delta\ (1) \end{pmatrix}, \tag{3.55}$$

it follows from Eqs. (3.50) that

$$\left\{ \begin{pmatrix} [p+q \quad \dot{0}] \\ (\Gamma'') \end{pmatrix} \begin{pmatrix} 1 \\ [q \quad \dot{0}] \\ (\Gamma) \end{pmatrix} \begin{pmatrix} [p \quad \dot{0}] \\ (\Gamma) \end{pmatrix} \right\} ([m] + [\Delta'']) = \frac{D\begin{pmatrix} [\Delta''] \\ [p+q \quad \dot{0}] \end{pmatrix}([m])}{D\begin{pmatrix} [\Delta'] \\ [q \quad \dot{0}] \end{pmatrix}([m] + [\Delta])\, D\begin{pmatrix} [\Delta] \\ [p \quad \dot{0}] \end{pmatrix}([m])}, \tag{3.56a}$$

where $[\Delta''] = [\Delta(\Gamma'')]$, etc., hence,

$$[\Delta''] = [\Delta'] + [\Delta]. \tag{3.56b}$$

We may, of course, write out similar results for the Racah coefficient having irrep labels all of the type $[\dot{0} - k]$ or all of the type $[\dot{k}0]$. More significant is the fact that *the Racah coefficients of these types can be written out completely from the $U(n)$ denominator pattern calculus rules*. [This was first pointed out for $U(2)$ in Ref. 5.]

It was proved in Ref. 4 that each *unique* Racah coefficient *must* become a square-bracket coefficient in the limit $m_{nn} = p_{nn} \to -\infty$. It is, nonetheless, satisfying to show directly that the Racah coefficient (3.56) exhibits this property. Isolating the terms containing p_{nn} in Eq. (3.56), one easily verifies that these terms have the limit 1: Thus,

$$\lim_{m_{nn}\to-\infty} \left\{ \begin{pmatrix} [p+q \quad \dot{0}] \\ (\Gamma'')_{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ [q \quad \dot{0}] \\ (\Gamma')_{n-1} \end{pmatrix} \begin{pmatrix} [p \quad \dot{0}] \\ (\Gamma)_{n-1} \end{pmatrix} \right\} ([m]_n + [\Delta'']_n)$$

$$= \left[\frac{p!q!}{(p+q)!} \times \frac{(r+s)!}{r!s!} \times \frac{(p+q-r-s)!}{(p-r)!(q-s)!}\right]^{1/2} \left\{ \begin{pmatrix} [r+s \quad \dot{0}] \\ (\Gamma'')_{n-2} \end{pmatrix} \begin{pmatrix} 1 \\ [s \quad \dot{0}] \\ (\Gamma')_{n-2} \end{pmatrix} \begin{pmatrix} [r \quad \dot{0}] \\ (\Gamma)_{n-2} \end{pmatrix} \right\} ([m]_{n-1} + [\Delta'']_{n-1}), \tag{3.57}$$

where $r = \Gamma_{1,n-1} = \Delta_1 + \Delta_2 + \cdots + \Delta_{n-1}$ and $s = \Gamma'_{1,n-1} = \Delta'_1 + \Delta'_2 + \cdots \Delta'_{n-1}$.
One easily verifies that

$$\left\langle \begin{matrix} [p+q \quad \dot{0}]_n \\ [r+s \quad \dot{0}]_{n-1} \end{matrix} \middle| \begin{bmatrix} & 1 & \\ [q & & \dot{0}]_n \\ [s & & \dot{0}]_{n-1} \\ & 1 & \end{bmatrix} \middle| \begin{matrix} [p \ \dot{0}]_n \\ [r \ \dot{0}]_{n-1} \end{matrix} \right\rangle = \left[\frac{p!q!}{(p+q)!} \times \frac{(r+s)!}{r!s!} \times \frac{(p+q-r-s)!}{(p-r)!(q-s)!} \right]^{1/2}. \tag{3.58}$$

Using this result, the right-hand side of Eq. (3.57) is seen to be precisely the square-bracket coefficient

$$\left[\begin{pmatrix} [p+q & & \dot{0}] \\ & (\Gamma'')_{n-1} & \end{pmatrix} \begin{pmatrix} & 1 & \\ [q & & \dot{0}] \\ & (\Gamma')_{n-1} & \end{pmatrix} \begin{pmatrix} [p & & \dot{0}] \\ & (\Gamma)_{n-1} & \end{pmatrix} \right] ([m]_{n-1} + [\Delta'']_{n-1}). \tag{3.59}$$

Thus, the limit relation is verified directly for this special case.

We may continue to take the limits $m_{n-1,n} \to -\infty$, $m_{n-2,n} \to -\infty, \ldots, m_{2n} \to -\infty$, in turn, of Eq. (3.57). The final limit yields precisely the *Wigner coefficient*

$$\left\langle \begin{matrix} [p+q & & \dot{0}] \\ & (\Gamma'') & \end{matrix} \middle| \left\langle \begin{matrix} & 1 & \\ [q & & \dot{0}] \\ & (\Gamma') & \end{matrix} \right\rangle \middle| \begin{matrix} [p & & \dot{0}] \\ & (\Gamma) & \end{matrix} \right\rangle. \tag{3.60}$$

Aside from their intrinsic interest, such limits imply important structural properties of operator patterns.[4] Indeed, one can now proceed to demonstrate (by the procedures of Ref. 4) that the limit properties of the totally symmetric projective operators induce uniquely the *complete* upper operator pattern labels from the lower operator pattern labels. [This property has already been proved for all operator patterns which are uniquely determined by their Δ patterns.[4]] Thus, starting from $U(2)$, the operator patterns are induced, by limits, from the Gel'fand patterns. The upper operator patterns of the symmetric projective operators in $U(3)$ are then induced, by limits, from the lower operator patterns, which are $U(2)$ operator patterns, etc. *All operator patterns of the totally symmetric unit projective operators may be considered to originate from the limit properties of the associated reduced matrix elements.* (A similar statement applies, of course, to the $[\dot{0} - p]$ and $[\dot{p}\ 0]$ operators.)

This brief discussion is intended only to indicate how one can verify the implications of limit properties for the unique operators. Of considerably more interest is the explicit verification that it is this same limiting property which labels the unit projective operators in a *multiplicity set*. To achieve this goal, it is first necessary to demonstrate the property fully for *all* $U(3)$ projective operators. Since we know the canonical splitting of a multiplicity set in $U(3)$, this first step is within our reach. We accordingly now direct our attention to $U(3)$ solely.

4. THE CANONICAL SPLITTING OF THE MULTIPLICITY IN $U(3)$

A. The origin of the splitting

Once we have obtained the Wigner operators $\langle p - q\ 00\rangle$ and $\langle qq0\rangle$, we can obtain more general operators by using the Wigner coupling

$$\left\langle \begin{matrix} \cdot \\ [p-q \quad 0\ 0] \\ \cdot \end{matrix} \right\rangle_{(W)} \left\langle \begin{matrix} \cdot \\ [q\ q\ 0] \\ \cdot \end{matrix} \right\rangle. \tag{4.1a}$$

This Gel'fand pattern coupling is known, since the coupling coefficient is

$$\left\langle \begin{matrix} [p \quad q \quad 0] \\ (M'') \end{matrix} \middle| \left\langle \begin{matrix} & (\max) & \\ p-q & & 0\ 0 \\ & (M') & \end{matrix} \right\rangle \middle| \begin{matrix} [q \quad q \quad 0] \\ (M) \end{matrix} \right\rangle, \tag{4.1b}$$

and this matrix element is determined from knowledge of the operator $\langle p-q\quad 0\ 0\rangle$ itself.

The problem is, of course, that this coupling defines not a single Wigner operator, but rather a linear combination (with invariant Racah functions as coefficients) of the Wigner operators

$$\left\langle \begin{matrix} (\Gamma) \\ p \quad q \quad 0 \\ (M'') \end{matrix} \right\rangle, \tag{4.2}$$

where the sum is over all (Γ).

The key point in the canonical splitting[11] is to recognize that *if the coupled operator $\langle p-q\quad 0\ 0\rangle_{(W)}\langle q\ q\ 0\rangle$ is restricted to a maximal shift in $U(2)$, then the Wigner operators belonging to a multiplicity set collapse to but a single term.* More precisely this is the statement

$$\left\langle \begin{pmatrix} [m] + [\Delta] \\ [m'] + [p\ 0] \\ m_{11} + M''_{11} \end{pmatrix} \middle| \left\langle \begin{matrix} \cdot \\ [p-q \quad 0\ 0] \\ \cdot \end{matrix} \right\rangle_{(W)} \left\langle \begin{matrix} \cdot \\ [q\ q\ 0] \\ \cdot \end{matrix} \right\rangle \middle| \begin{pmatrix} [m] \\ [m'] \\ m_{11} \end{pmatrix} \right\rangle$$

$$= (\text{invariant coefficient})$$

$$\times \left\langle \begin{pmatrix} [m] + [\Delta] \\ [m'] + [p\ 0] \\ m_{11} + M''_{11} \end{pmatrix} \middle| \left\langle \begin{matrix} (\Gamma_s) \\ [p \quad q \quad 0] \\ (M'') \end{matrix} \right\rangle \middle| \begin{pmatrix} [m] \\ [m'] \\ m_{11} \end{pmatrix} \right\rangle, \tag{4.3}$$

where $[m] = [m_{13}m_{23}m_{33}]$, $[m'] = [m_{12}m_{22}]$, and $[\Delta] = [\Delta_1\Delta_2\Delta_3]$ is any preselected $[\Delta]$ belonging to $[p\ q\ 0]$. Of course, relation (4.3) always holds whenever $[\Delta]$ uniquely determines the operator pattern denoted by (Γ_s). The essential contribution of Ref. 11 was the demonstration that Eq. (4.3) *still holds* even when $[\Delta]$ determines a multiplicity set of operator patterns. (Γ_s) then designates precisely *one* of the operator patterns in the multiplicity set determined by $[\Delta]$. (Just which one remains to be determined.[22]) Since the $U(2)$ pattern $[p\ 0]$ is the unique Δ pattern such that $[\Delta_1'\Delta_2']$ has $\Delta_1' - \Delta_2'$ equal to the largest possible value for the $\langle p-q\ \ 0\ 0\rangle_{\langle W\rangle}\langle q\ q\ 0\rangle$ coupling, we designate the pattern by (Γ_s), where s denotes "stretched." While we can guess from previous experience with the adjoint operators and the 27-plet of operators that, depending on the specific shift values $[\Delta_1\Delta_2\Delta_3]$, the operator pattern (Γ_s) has either the form

$$\begin{pmatrix} p & & q & & 0 \\ & \Gamma_{12} & & 0 & \\ & & \Gamma_{11} & & \end{pmatrix} \tag{4.4a}$$

or the form

$$\begin{pmatrix} p & & q & & 0 \\ & p & & \Gamma_{22} & \\ & & \Gamma_{11} & & \end{pmatrix}, \tag{4.4b}$$

this guess remains to be proved. Until this guess is shown to be correct, (Γ_s) *denotes a single, but unspecified, operator pattern belonging to the multiplicity set determined by* $[\Delta]$.

The property expressed by Eq. (4.3) is entirely equivalent to the following property of the $U(3)$: $U(2)$ projective operators

$$\begin{bmatrix} & & (\Gamma) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{bmatrix} = 0, \qquad (\Gamma) \neq (\Gamma_s), \tag{4.5}$$

for all (Γ) belonging to the multiplicity set determined by $[\Delta]$, except for a single upper operator pattern—the one denoted by (Γ_s). Property (4.5) is a very strong statement: It asserts that all but one $U(3)$: $U(2)$ projective operator (in any multiplicity set) having maximally stretched lower pattern is the *zero operator*.

This splitting is a fundamental property of the $U(3)$ operator system, yet it is hardly obvious *a priori*. How does this remarkable property come about? To see this, let us consider representing the projective operator

$$\begin{bmatrix} & & (\Gamma) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{bmatrix} \tag{4.6}$$

in terms of the elementary projections [100] and [110]. One easily determines that *each* such operator (4.6) must have the structure as follows:

$$\begin{bmatrix} & & (\Gamma) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{bmatrix} = (U(3) \text{ invariant factor})$$

$$\times \begin{bmatrix} & & \tau_{p-q} & & \\ 1 & & 0 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix} \cdots \begin{bmatrix} & & \tau_{2} & & \\ 1 & & 0 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix} \begin{bmatrix} & & \tau_{1} & & \\ 1 & & 0 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix}$$

$$\times \begin{bmatrix} & & \tilde\tau_{q} & & \\ 1 & & 1 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix} \cdots \begin{bmatrix} & & \tilde\tau_{2} & & \\ 1 & & 1 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix} \begin{bmatrix} & & \tilde\tau_{1} & & \\ 1 & & 1 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{bmatrix}, \tag{4.7a}$$

where $\tau_{p-q} \cdots \tau_2\tau_1$ $(1 \leq \tau_i \leq 3)$ and $\tilde\tau_q \cdots \tilde\tau_2\tilde\tau_1$ $(\tilde 1 \leq \tilde\tau_i \leq \tilde 3)$ are any assignment of integers such that

$$\sum_{i=1}^{p-q} \Delta(\tau_i) + \sum_{i=1}^{q} \Delta(\tilde\tau_i) = [\Delta]. \tag{4.7b}$$

It is important to observe that, although we have indicated a specific ordering in Eq. (4.7a), our intent is that this string of p elementary projective operators can be put together in any *arbitrary fashion*, it being necessary only to preserve property (4.7b).

Next, let us decompose each elementary projection in the right-hand side of Eq. (4.7a) into the quotient form (2.24). We then observe that in the string of p elementary restricted Δ pattern functions *there are no opposing arrows in the string of p corresponding elementary arrow patterns*. This means that these elementary restricted Δ pattern functions, *in any ordering whatsoever*, multiply by addition of their corresponding Δ patterns. Thus, in any rearrangement of the elementary operators, the right-hand side of Eq. (4.7a) yields the factor

$$F_R\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\quad q\quad 0] \\ [p\quad 0] \end{pmatrix}. \tag{4.8}$$

The denominator factor together with the normalization factor can contribute only a function whose values depend on $[m_{13}m_{23}m_{33}]$, i.e., an over-all normalization, which depends on the specific ordering. *This implies that there exists but one operator (4.6) in the multiplicity set determined by* $[\Delta]$—this unique operator being the one we have designated by the operator pattern (Γ_s) (all others being therefore zero). Furthermore, this operator has the structure as follows:

$$\begin{bmatrix} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{bmatrix}$$

$$= (\text{phase})\, F_R\begin{pmatrix} [\Delta_1\ \ \Delta_2\ \ \Delta_3] \\ [p\quad q\quad 0] \\ [p\quad 0] \end{pmatrix} \Big/ D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\quad q\quad 0] \end{pmatrix}, \tag{4.9a}$$

where

$$D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\quad q\quad 0] \end{pmatrix} \tag{4.9b}$$

denotes a new denominator function, whose values depend only on the irrep labels $[m]$, which has further properties yet to be determined. The value of the restricted Δ pattern function in the numerator is, of course, given completely by the pattern calculus rules.

Let us emphasize again that the results (4.9) and (4.5) are *uniquely* implied by the geometrical properties of the arrow patterns (the same property of no opposing arrows used in the previous sections). *There are no free choices in this structure.* [We establish in II that even the operator pattern itself, (Γ_s), is uniquely assigned by limits.]

B. Application of the factorization lemma

We wish now to apply this splitting to determine explicitly the projective operators

$$\left[\begin{array}{ccccc} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right]. \tag{4.10}$$

The most instructive procedure makes use once again of the factorization lemma.

Consider then the following boson operator

$$B\left(\begin{array}{ccccc} & & p & & \\ & p & & 0 & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right)(A) = (a^1_1)^{p-q}\,(a^{1\,3}_{1\,3})^q. \tag{4.11}$$

Following exactly the procedures, whereby Eq. (3.20) was deduced from the boson operator of Eq. (3.15), we obtain

$$\sum_{(\Gamma_s)}\left[\begin{array}{ccccc} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right]_l \left[\begin{array}{ccccc} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right]_u$$
$$= \left(\sum_{\tau=1}^{3}\left[\begin{array}{ccccc} & & \tau & & \\ 1 & & 0 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{array}\right]_l \left[\begin{array}{ccccc} & & \tau & & \\ 1 & & 0 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{array}\right]_u\right)^{p-q}$$
$$\times\left(\sum_{\tau=1}^{3}\left[\begin{array}{ccccc} & & \tilde{\tau} & & \\ 1 & & 1 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{array}\right]_l \left[\begin{array}{ccccc} & & \tilde{\tau} & & \\ 1 & & 1 & & 0 \\ & 1 & & 0 & \\ & & 1 & & \end{array}\right]_u\right)^{q}. \tag{4.12}$$

We can proceed with this expression in two distinct and instructive ways. In the first method, we recognize that the operator which is raised to the $(p-q)$th power is just

$$\sum_{(\Gamma')}\left[\begin{array}{ccccc} & & (\Gamma') & & \\ p-q & & 0 & & 0 \\ & p-q & & 0 & \\ & & p-q & & \end{array}\right]_l \left[\begin{array}{ccccc} & & (\Gamma') & & \\ p-q & & 0 & & 0 \\ & p-q & & 0 & \\ & & p-q & & \end{array}\right]_u, \tag{4.13a}$$

and the operator which is raised to the qth power is just

$$\sum_{(\Gamma'')}\left[\begin{array}{ccccc} & & (\Gamma'') & & \\ q & & q & & 0 \\ & q & & 0 & \\ & & q & & \end{array}\right]_l \left[\begin{array}{ccccc} & & (\Gamma'') & & \\ q & & q & & 0 \\ & q & & 0 & \\ & & q & & \end{array}\right]_u \tag{4.13b}$$

This corresponds to recognizing in the original boson operator (4.11) that the first factor is

$$B\left(\begin{array}{ccccc} & & p-q & & \\ & p-q & & 0 & \\ p-q & & 0 & & 0 \\ & p-q & & 0 & \\ & & p-q & & \end{array}\right)(A), \tag{4.14a}$$

and the second factor is

$$B\left(\begin{array}{ccccc} & & q & & \\ & q & & 0 & \\ q & & q & & 0 \\ & q & & 0 & \\ & & q & & \end{array}\right)(A). \tag{4.14b}$$

We next use **Eqs.** (4.13) in the right-hand side of Eq. (4.12) and take the matrix elements of the expression between the following $U(n) \star U(n)$ states,

$$\left\langle\left(\begin{array}{c} (\max) \\ [m'] + [p0] \\ [m] \;\; + [\Delta] \\ [m'] + [p0] \\ (\max) \end{array}\right)\right| \cdots \left|\left(\begin{array}{c} (\max) \\ [m'] \\ [m] \\ [m'] \\ (\max) \end{array}\right)\right\rangle, \tag{4.15}$$

where Δ is any Δ pattern belonging to $[pq0]$. This selects *one* term from the left-hand side, since there is but one (Γ_s) in each multiplicity set [if the multiplicity is 1 then (Γ_s) becomes the unique pattern]. The result is just the expression of the abstract operator identity as follows:

$$\left[\begin{array}{ccccc} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right]^\dagger \left[\begin{array}{ccccc} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & p & & 0 & \\ & & p & & \end{array}\right]$$
$$= \sum_{(\Gamma'),(\Gamma'')}\left[\begin{array}{ccccc} & & (\Gamma') & & \\ p-q & & 0 & & 0 \\ & p-q & & 0 & \\ & & p-q & & \end{array}\right]^\dagger \left[\begin{array}{ccccc} & & (\Gamma') & & \\ p-q & & 0 & & 0 \\ & p-q & & 0 & \\ & & p-q & & \end{array}\right]$$

$$\times \begin{bmatrix} & (\Gamma'') & \\ q & q & 0 \\ & q \quad 0 & \\ & q & \end{bmatrix}^{\dagger} \begin{bmatrix} & (\Gamma'') & \\ q & q & 0 \\ & q \quad 0 & \\ & q & \end{bmatrix}, \tag{4.16a}$$

where the sum is over all patterns (Γ') and (Γ'') such that

$$[\Delta'] + [\Delta''] = [\Delta] = [\Delta(\Gamma_s)]. \tag{4.16b}$$

We, in turn, infer from Eq. (4.16a) the *complete result* (except for phase)

$$\begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} = (-1)^{p-\Delta_1} F_R\begin{pmatrix} [\Delta_1 \quad \Delta_2 \quad \Delta_3] \\ [p \quad q \quad 0] \\ [p \quad 0] \end{pmatrix} \Big/ D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\ q\ 0] \end{pmatrix}, \tag{4.17a}$$

where the denominator function is now given explicitly by

$$1 \Big/ \left| D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\ q\ 0] \end{pmatrix}\right|^2 = \sum_{[\Delta'],[\Delta'']} 1 \Big/ \left| D\begin{pmatrix} [\Delta_1' \quad \Delta_2' \quad \Delta_3'] \\ [p-q \quad 0 \quad 0] \end{pmatrix} D\begin{pmatrix} [\Delta_1''\ \Delta_2''\ \Delta_3''] \\ [q \quad q \quad 0] \end{pmatrix}\right|^2, \tag{4.17b}$$

where the summation is over all Δ patterns $[\Delta']$ and $[\Delta'']$, such that Eq. (4.16b) holds. In order to avoid rewriting Eq. (4.17a), we have identified the phase factor by the technique used in obtaining Eq. (3.49).

The value of the restricted Δ pattern function in Eq. (4.17a) is, of course, given completely by the pattern calculus rules. Equation (4.17b) is also explicitly known from Sec. 3D. For completeness, we note the values of the two denominator functions appearing in Eq. (4.17b):

$$D\begin{pmatrix} [\Delta_1' \quad \Delta_2' \quad \Delta_3'] \\ [p-q \quad 0 \quad 0] \end{pmatrix}([m]) = \left[\frac{(\Delta_1')!\,\Delta_2')!\,(\Delta_3')!}{(p-q)!\,\mathfrak{D}'([m]+[\Delta'])} \times \prod_{i<j=1}^{3} (\Delta_i' + \Delta_j' + 1)! \begin{pmatrix} p_{i3} - p_{j3} + \Delta_i' \\ \Delta_i' + \Delta_j' + 1 \end{pmatrix}\right]^{1/2}, \tag{4.18a}$$

$$D\begin{pmatrix} [\Delta_1''\ \Delta_2''\ \Delta_3''] \\ [q \quad q \quad 0] \end{pmatrix}([m]) = \left[\frac{(q-\Delta_1'')!\,(q-\Delta_2'')!\,(q-\Delta_3'')!}{q!\,\mathfrak{D}'([m]+[\Delta''])} \times \prod_{i<j=1}^{3} (2q - \Delta_i'' - \Delta_j'' + 1)! \begin{pmatrix} p_{i3} - p_{j3} + q - \Delta_j'' \\ 2q - \Delta_i'' - \Delta_j'' + 1 \end{pmatrix}\right]^{1/2} \tag{4.18b}$$

{The denominator function (4.18b) is obtained from Eq. (3.52c) for $n = 3$ upon replacing Δ_i by $q - \Delta_i''$ and p by q. We also remark that the value of the denominator function (4.18a) which occurs in Eq. (4.17b) is obtained by shifting the labels $[m]$ in Eq. (4.18a) to $[m] \to [m] + [\Delta'']$, i.e., $p_{i3} \to p_{i3} + \Delta_i''$.}

It is a remarkable fact that the boson factorization lemma together with the geometrical properties of the Δ pattern functions have led us uniquely to the result, Eqs. (4.17). Of course, the expression (4.17b) for the denominator function, while explicit, is very complicated, and one can not be too satisfied with it. The *structure* of the result, Eq. (4.17a), is however, quite elegant, and as we shall see in II the denominator function itself possesses symmetry properties and structures of an unexpected nature. Yet it is these very properties which the denominator *must have* if our program relating to null spaces is to be fulfilled.

We now turn to a further structural interpretation of the denominator function in terms of the sum-over-paths concept.

C. The sum-over-paths formulation of the denominator function

Equation (4.12) has a second interpretation. Observe that in the boson expression (4.11) we can write the $(p-q)a_1^1$'s and the $q\,a_{13}^{12}$'s as a string of p factors in $p!$ ways, there being $\binom{p}{q}$ distinct arrangements.

This commuting property of bosons implies that in Eq. (4.12) we may similarly write the $p - q$ factors

$$\sum_{\tau=1}^{3} \begin{bmatrix} & \tau & \\ 1 & 0 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}_l \begin{bmatrix} & \tau & \\ 1 & 0 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}_u \tag{4.19a}$$

and the q factors

$$\sum_{\tau=1}^{3} \begin{bmatrix} & \tilde{\tau} & \\ 1 & 1 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}_l \begin{bmatrix} & \tilde{\tau} & \\ 1 & 1 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}_u \tag{4.19b}$$

in any order whatsoever, without changing the equality of the resulting expression to the left-hand side. Again there are $\binom{p}{q}$ distinct arrangements.

Let us now introduce the notation

$$P(\tau) = \begin{bmatrix} & \tau & \\ 1 & 0 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}, \quad P(\tilde{\tau}) = \begin{bmatrix} & \tilde{\tau} & \\ 1 & 1 & 0 \\ & 1 \quad 0 & \\ & 1 & \end{bmatrix}. \tag{4.20}$$

Then we can write Eq. (4.12) in the form

$$\sum_{(\Gamma_s)} \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix}_l \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix}_u$$

$$= \binom{p}{q}^{-1} \sum_{\lambda_p \cdots \lambda_2\lambda_1} \sum_{\mathcal{P}} \mathcal{P}[P_\ell(\lambda_p)P_u(\lambda_p) \cdots P_\ell(\lambda_1)P_u(\lambda_1)], \quad (4.21a)$$

where

$$(\lambda_p \cdots \lambda_2\lambda_1) \equiv (\tau_{p-q} \cdots \tau_2\tau_1\tilde{\tau}_q \cdots \tilde{\tau}_2\tilde{\tau}_1), \quad (4.21b)$$

and where $\mathcal{P}$ denotes a permutation of the p symbols $\tau\tau \cdots \tau\tilde{\tau}\tilde{\tau} \cdots \tilde{\tau}$ containing $(p-q)\tau$'s and $q\tilde{\tau}$'s, and the summation over $\mathcal{P}$ is over all $\binom{p}{q}$ distinct permutations of these symbols. Note also that the summation over any τ_i index is from 1 to 3, that of any $\tilde{\tau}_i$ index from $\tilde{1}$ to $\tilde{3}$.

If we now take the matrix element (4.15) of Eq. (4.21a), we obtain the following abstract operator identity:

$$\begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix}^{\dagger} \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} = \binom{p}{q}^{-1} \sum_{\lambda_p \cdots \lambda_2\lambda_1} \sum_{\mathcal{P}} \mathcal{P}[P^\dagger(\lambda_p)P(\lambda_p) \cdots P^\dagger(\lambda_1)P(\lambda_1)], \quad (4.22)$$

where the summation is over all $(\lambda_p \cdots \lambda_2\lambda_1) \equiv (\tau_{p-q} \cdots \tau_1\tilde{\tau}_q \cdots \tilde{\tau}_1)$ such that Eq. (4.7b) holds. We now see that the summation over $\mathcal{P}$ in Eq. (4.22) may be dropped by the following simple device: Let any λ_i denote *either* a τ_j *or* a $\tilde{\tau}_j$. Then we may write

$$\begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix}^{\dagger} \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} = \binom{p}{q}^{-1} \sum_{\lambda_p \cdots \lambda_2\lambda_1} P^\dagger(\lambda_p)P(\lambda_p) \cdots P^\dagger(\lambda_1)P(\lambda_1), \quad (4.23a)$$

where the summation is now over *all* sets $(\lambda_p \cdots \lambda_2\lambda_1)$ containing $(p-q)\tau_i$'s $(1 \le \tau_i \le 3)$ and $q\tilde{\tau}_i$'s $(\tilde{1} \le \tilde{\tau}_i \le \tilde{3})$, such that

$$\sum_{i=1}^{p} \Delta(\lambda_i) = [\Delta]. \quad (4.23b)$$

Finally, upon defining

$$d_\lambda = d_{\Delta(\lambda)}, \quad \lambda = \tau \text{ or } \tilde{\tau}, \quad (4.24)$$

we obtain from Eq. (4.23a) precisely the result, Eq. (4.17a), where now the denominator function is expressed by

$$1 \Big/ \left| D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\quad q\quad 0] \end{pmatrix} \right|^2 = \binom{p}{q}^{-1} \sum_{\lambda_p \cdots \lambda_2\lambda_1} \frac{1}{|d_{\lambda_p} \cdots d_{\lambda_2} d_{\lambda_1}|^2}. \quad (4.25)$$

Let us note the following features of the sum-over-paths interpretation of this result:

(1) Each set of integers satisfying Eq. (4.23b) describes a path from the lattice point $[m_{13}m_{23}m_{33}]$ of R^3 to the lattice point $[m_{13}+\Delta_1 m_{23}+\Delta_2 m_{33}+\Delta_3]$, and the value of the corresponding term defines the path contribution to the sum, where the sum is now simply the expression of all path contributions from admissible paths between $[m]$ and $[m]+[\Delta]$.

(2) The paths are defined by *six* elementary steps (shifts) [100], [010], [001], [011], [101], and [110]. The normalization $\binom{p}{q}^{-1}$ is a direct consequence of the occurrence of two types of elementary steps.

(3) The number N of $\Delta(\tau)$-type steps and the number $\tilde{N}$ of $\Delta(\tilde{\tau})$-type steps is uniquely determined by the shift values $[\Delta_1\Delta_2\Delta_3]$ and the total number of steps $N+\tilde{N}=p$. [Hence, it is only necessary to specify that the $\lambda_p \cdots \lambda_2\lambda_1$ in the summation satisfy Eq. (4.23b).]

(4) For $q=0$, Eq. (4.25) reduces to Eq. (3.26b) (for $n=3$). [Similarly, for $p=q$ it agrees with the result implied by our results of Sec. 3, but not noted explicitly.]

(5) In view of items (1)–(4), we see that the sum-over-paths interpretation of Eq. (4.25) generalizes the usual path formulation in two ways. An overall normalization occurs to account for the two types of steps; and the admissible paths in the sum are restricted to a fixed number p. {Observe that in Eq. (3.26b) the shift $[\Delta]$ itself fixes the number of admissible paths.}

D. Recursion relations for the denominator function

The defining relation for the denominator function, Eq. (4.25), is conceptually quite simple, but for determining the properties of the denominator directly this relation proves to be both complicated and difficult. We are therefore led to approach the study of this function through the use of recursion relations which it satisfies.

The simplest derivation of such recursion relations proceeds directly from the sum-over-paths formulation, Eq. (4.25). First, let us observe that the $\binom{p}{q}$ sums in Eq. (4.25), each corresponding to a definite ordering of $(\tau\tau \cdots \tau)(\tilde{\tau} \cdots \tilde{\tau})$, are, in fact, all equal. (This property is clear from the derivation.) Thus, the right-hand side of Eq. (4.25) may also be written in the form

$$\frac{1}{\left| D\begin{pmatrix} [\Delta_1\Delta_2\Delta_3] \\ [p\quad q\quad 0] \end{pmatrix} \right|^2} = \sum_{\lambda_p \cdots \lambda_2\lambda_1} \frac{1}{|d_{\lambda_p} \cdots d_{\lambda_2} d_{\lambda_1}|^2}, \quad (4.26)$$

where $\lambda_p \cdots \lambda_2\lambda_1$ is a *definite* arrangement of $(\tau_{p-q} \cdots \tau_1)(\tilde{\tau}_q \cdots \tilde{\tau}_1)$ satisfying, of course, Eq. (4.23b). In Eq. (4.26), let us replace p by $p-1$ and q by $q-1$ and denote the corresponding $[\Delta]$ by $[\Delta_1'\Delta_2'\Delta_3']$, and choose an arrangement of $(\lambda_{p-1} \cdots \lambda_2\lambda_1)$ such that λ_1 is a $\tilde{\tau}$. The relation

$$\frac{1}{\left| D\begin{pmatrix} [\Delta_1\Delta_2\Delta_3] \\ [p\quad q\quad 0] \end{pmatrix} \right|^2} = \sum_{\tau=1}^{3} \frac{1}{\left| D\begin{pmatrix} [\Delta_1' & \Delta_2' & \Delta_3'] \\ [p-1 & q-1 & 0] \end{pmatrix} d_{\Delta(\tilde{\tau})} \right|^2}, \quad (4.27)$$

where $[\Delta'] = [\Delta] - \Delta(\tilde{\tau})$, then follows immediately.

In order to simplify the recursion relation (4.27), it is convenient to introduce some auxiliary quantities. We first define for all $i,j = 1,2,3$ the quantities

$$x_{ij} = p_{i3} - p_{j3}, \quad (4.28a)$$

noting that $x_{ii}=0$ and $x_{ji} = -x_{ij}$. In terms of the x_{ij}, we also introduce

$$x_i = x_{jk}, \quad (i,j,k \text{ cyclic in } 1,2,3), \quad (4.28b)$$

noting that

$$x_1 + x_2 + x_3 = 0. \tag{4.28c}$$

Finally, we introduce the auxiliary function $G_q(\Delta;x) = G_q(\Delta_1\Delta_2\Delta_3;x_1x_2x_3)$ by the definition as follows:

$$\left[\frac{1}{D\begin{pmatrix}[\Delta_1\Delta_2\Delta_3]\\ [p\ q\ 0]\end{pmatrix}([m])}\right]^2 = \left[(p-q)!\,\mathfrak{D}'([m]+[\Delta])\Big/\Delta_1!\Delta_2!\Delta_3!\right] \times\left[1\Big/\prod_{i<j=1}^{3}(\Delta_i+\Delta_j+1)!\begin{pmatrix}x_{ij}+\Delta_i\\ \Delta_i+\Delta_j+1\end{pmatrix}\right]G_q(\Delta;x). \tag{4.29}$$

Observe that since $\Delta_1 + \Delta_2 + \Delta_3 = p + q$, the label p is implicitly defined in $G_q(\Delta;x)$.

Using the explicit value of $d_{\Delta(\tilde{\tau})}([m])$ given by the pattern calculus rules and introducing the preceding definitions, Eq. (4.27) yields the following explicit recursion relation for $G_q(\Delta;x)$:

$$\begin{aligned} x_1x_2x_3G_q(\Delta_1\Delta_2\Delta_3;x_1x_2x_3) &= \Delta_1\Delta_2x_3(\Delta_3-x_1)(\Delta_3+x_2)(\Delta_1+x_3)(\Delta_2-x_3)\\ &\times G_{q-1}(\Delta_1-1,\Delta_2-1,\Delta_3;x_1+1,x_2-1,x_3)\\ &+\Delta_2\Delta_3x_1(\Delta_1-x_2)(\Delta_1+x_3)(\Delta_2+x_1)(\Delta_3-x_1)\\ &\times G_{q-1}(\Delta_1,\Delta_2-1,\Delta_3-1;x_1,x_2+1,x_3-1)\\ &+\Delta_3\Delta_1x_2(\Delta_2-x_3)(\Delta_2+x_1)(\Delta_3+x_2)(\Delta_1-x_2)\\ &\times G_{q-1}(\Delta_1-1,\Delta_2,\Delta_3-1;x_1-1,x_2,x_3+1). \end{aligned} \tag{4.30}$$

Observe that for $q = 0$ the factor in Eq. (4.29) in front of $G_0(\Delta;x)$ is just the value of the denominator function (4.18a); hence, the boundary condition for our recursion relation is

$$G_0(\Delta;x) = 1. \tag{4.31}$$

An equally valid recursion relation is obtained from Eq. (4.27) simply by shifting $d_{\Delta(\tilde{\tau})}$ to the left of the denominator function in the sum. The resulting recursion relation for $G_q(\Delta;x)$ reads

$$\begin{aligned} &(x_1+\Delta_2-\Delta_3)(x_2+\Delta_3-\Delta_1)(x_3+\Delta_1-\Delta_2)\\ &\quad\times G_q(\Delta_1\Delta_2\Delta_3;x_1x_2x_3)\\ &\quad=\Delta_1\Delta_2(x_3+\Delta_1-\Delta_2)(\Delta_2+x_1)(\Delta_1-x_2)(\Delta_1+x_3)\\ &\quad\times(\Delta_2-x_3)G_{q-1}(\Delta_1-1,\Delta_2-1,\Delta_3;x_1x_2x_3)\\ &\quad+\Delta_2\Delta_3(x_1+\Delta_2-\Delta_3)(\Delta_3+x_2)(\Delta_2-x_3)(\Delta_2+x_1)\\ &\quad\times(\Delta_3-x_1)G_{q-1}(\Delta_1,\Delta_2-1,\Delta_3-1;x_1x_2x_3)\\ &\quad+\Delta_3\Delta_1(x_2+\Delta_3-\Delta_1)(\Delta_1+x_3)(\Delta_3-x_1)(\Delta_3+x_2)\\ &\quad\times(\Delta_1-x_2)G_{q-1}(\Delta_1-1,\Delta_2,\Delta_3-1;x_1x_2x_3). \end{aligned} \tag{4.32}$$

It is noteworthy that in this second recursion relation the x_i variables are not shifted. The boundary condition is, of course, as before: $G_0(\Delta;x) = 1$.

Two additional recursion relations for $G_q(\Delta;x)$ may also be obtained from a relation of type (4.27): Simply replace $\tilde{\tau}$ by τ. We will, however, not require these relations.

The great advantage in using the recursion relations derived above is that this approach makes evident many essential properties of the denominator function. It is already evident that the functions $G_q(\Delta;x)$ must exhibit a great deal of symmetry. We defer, however, the discussion of these properties to II, where the properties of the functions $G_q(\Delta;x)$ will be developed systematically and fully.

Let us note here the results for $q = 1$ and 2, obtained by the direct iteration of Eq. (4.30). The result for $q = 1$ is easily found:

$$\begin{aligned} G_1(\Delta;x) &= -\Delta_1\Delta_2(\Delta_1+x_3)(\Delta_2-x_3)\\ &-\Delta_2\Delta_3(\Delta_2+x_1)(\Delta_3-x_1)-\Delta_3\Delta_1(\Delta_3+x_2)(\Delta_1-x_2)\\ &-\Delta_1\Delta_2\Delta_3(\Delta_1+\Delta_2+\Delta_3), \end{aligned} \tag{4.33}$$

where, in obtaining this form, it is essential to make use of relation (4.28c). The result for $q = 2$ already presents a formidable calculation. It can, however, be accomplished directly from Eqs. (4.30) and (4.33) with the result

$$\begin{aligned} &G_2(\Delta;x)\\ &=\{\Delta_1(\Delta_1-1)\Delta_2(\Delta_2-1)(\Delta_1+x_3)(\Delta_1+x_3-1)\\ &\quad\times(\Delta_2-x_3)(\Delta_2-x_3-1)\\ &\quad+(\text{cyclic permutations of }1,2,3)\}\\ &\quad+\{2\Delta_1\Delta_2\Delta_3(\Delta_3-1)(\Delta_2+x_1)(\Delta_3-x_1)(\Delta_3+x_2)\\ &\quad\times(\Delta_1-x_2)+(\text{cyclic permutations of }1,2,3)\}\\ &\quad+2(\Delta_1+\Delta_2+\Delta_3-1)\{\Delta_1(\Delta_1-1)\Delta_2(\Delta_2-1)\\ &\quad\times\Delta_3(\Delta_1+x_3)(\Delta_2-x_3)\\ &\quad+(\text{cyclic permutations of }1,2,3)\}\\ &\quad+\Delta_1(\Delta_1-1)\Delta_2(\Delta_2-1)\Delta_3(\Delta_3-1)(\Delta_1+\Delta_2+\Delta_3)\\ &\quad\times(\Delta_1+\Delta_2+\Delta_3-1). \end{aligned} \tag{4.34}$$

The *direct* calculation of $G_3(\Delta;x)$ is an almost impossible task. Fortunately, it is not required. Further detailed calculations of this type would lend little to one's understanding of the *structure* of the result. Since the discussion of structure is the central theme of the present work, our indirect rederivation [to be given in II] of $G_q(\Delta;x)$, *using only its structural properties* is certainly more instructive, and probably more important, than the mere fact that this rederivation shows that an alternative form for the denominator function exists.

For completeness of the present paper, however, let us note this alternative form for G_q, in detail. We will, however, give our initial intuitive arguments leading to this form, deferring a complete proof to II.

Suppose there were *no* shifting of the variables (either in Δ_i or x_i) in the right-hand side of Eq. (4.30). Clearly, the general solution to this simpler recursion relation is just

$$[G_1(\Delta;x)]^q, \tag{4.35}$$

where q is an ordinary power.

It is evident that some sort of "factorial rule" is operative in obtaining the actual solution to Eq. (4.30), which has the role of accounting for the shifts in the variables.

This type of behavior is quite familiar from Gel'fand's[23] symbolic interpretation of the (known) $SU(2)$ Wigner coefficients in terms of the Jacobi polynomials. Indeed, this type of behavior can be understood in much simpler terms. Consider the binomial theorem written in the form

$$(x+y)^q/q! = \sum_{\substack{s,t\\ s+t=q}} x^s y^t/s!t!. \tag{4.36}$$

It is a remarkable fact that under the substitutions

$$x^s/s! \to \binom{x}{s}, \quad y^t/t! \to \binom{y}{t}, \tag{4.37a}$$

$$(x+y)^q/q! \to \binom{x+y}{q}, \tag{4.37b}$$

we obtain from Eq. (4.36) the following correct *general binomial relation*

$$\binom{x+y}{q} = \sum_{\substack{s,t\\ s+t=q}} \binom{x}{s}\binom{y}{t}. \tag{4.38}$$

(This procedure generalizes, in fact, to the multinomial theorem.)

It is even more remarkable that this same procedure with only slight modifications may be used to conjecture the solution to recursion relation (4.30): Using the quadrinomial theorem, we expand $[G_1(\Delta;x)]^q$, collecting together the powers of $\Delta_1, \Delta_2, \Delta_3$, but leaving the powers in (Δ_1+x_3), etc. as well as $(\Delta_1+\Delta_2+\Delta_3)$ undisturbed:

$$\begin{aligned}
&[G_1(\Delta;x)]^q \\
&= (-1)^q q! \sum_{(k)} \Delta_1^{q-k_1}\Delta_2^{q-k_2}\Delta_3^{q-k_3}(\Delta_1+\Delta_2+\Delta_3)^{k_4} \\
&\quad \times (\Delta_2+x_1)^{k_1}(\Delta_3-x_1)^{k_1}(\Delta_3+x_2)^{k_2}(\Delta_1-x_2)^{k_2} \\
&\quad \times (\Delta_1+x_3)^{k_3}(\Delta_2-x_3)^{k_3}/(k_1)!(k_2)!(k_3)!(k_4)!,
\end{aligned} \tag{4.39}$$

where the sum is over all nonnegative integers $(k) = (k_1k_2k_3k_4)$ which add to q, i.e., $k_1+k_2+k_3+k_4 = q$. In this result, we now make the replacements

$$\Delta_i^{q-k_i}/(q-k_i)! \to \binom{\Delta_i}{q-k_i},$$

$$(\Delta_2+x_1)^{k_1}/k_1! \to \binom{\Delta_2+x_1}{k_1}, \text{ etc.,}$$

with the single *exception:* We replace $(\Delta_1+\Delta_2+\Delta_3)^{k_4}/(k_4)!$ by

$$\binom{\Delta_1+\Delta_2+\Delta_3-k_1-k_2-k_3}{k_4}.$$

The result is the following polynomial in x_1, x_2, x_3, which we denote by $G_q^c(\Delta;x)$ (c denotes "conjectured"):

$$\begin{aligned}
G_q^c(\Delta;x) &= (-1)^q q! \sum_{(k)} \binom{\Delta_1+\Delta_2+\Delta_3-k_1-k_2-k_3}{k_4} \\
&\quad \times (k_1)!(q-k_1)! \binom{\Delta_1}{q-k_1}\binom{\Delta_2+x_1}{k_1}\binom{\Delta_3-x_1}{k_1} \\
&\quad \times (k_2)!(q-k_2)! \binom{\Delta_2}{q-k_2}\binom{\Delta_3+x_2}{k_2}\binom{\Delta_1-x_2}{k_2} \\
&\quad \times (k_3)!(q-k_3)! \binom{\Delta_3}{q-k_3}\binom{\Delta_1+x_3}{k_3}\binom{\Delta_2-x_3}{k_3}.
\end{aligned} \tag{4.40}$$

One then easily verifies

$$G_q(\Delta;x) = G_q^c(\Delta;x) \tag{4.41}$$

for $q = 0, 1, 2$. The proof that this result is correct for arbitrary q requires a careful analysis of the properties of $G_q(\Delta;x)$, and we defer this study to II. [A direct proof that $G_q^c(\Delta;x)$ satisfies, say, recursion relation (4.30), in general, has so far eluded us.]

E. The "stretched" Racah coefficients

We are now able to give explicitly the Racah coefficients which occur in the coupling

$$\left\langle \begin{matrix} \cdot \\ [p-q\ \ 0\ \ 0] \\ \cdot \end{matrix} \right\rangle^{(R)}_{(W)} \left\langle \begin{matrix} \cdot \\ [q\ \ q\ \ 0] \\ \cdot \end{matrix} \right\rangle = \left\langle \begin{matrix} (\Gamma_s) \\ [p\ \ q\ \ 0] \\ (M) \end{matrix} \right\rangle. \tag{4.42}$$

These operator pattern couplings are uniquely determined by the following relation [which is a special case of Eq. (3.1)]:

$$\begin{aligned}
&\begin{bmatrix} & (\Gamma') & \\ p-q & 0 & 0 \\ & p-q\ \ 0 & \\ & p-q & \end{bmatrix} \begin{bmatrix} & (\Gamma'') & \\ q & q & 0 \\ & q\ \ 0 & \\ & q & \end{bmatrix} \\
&= \left\{ \begin{pmatrix} [p\ \ q\ \ 0] \\ (\Gamma_s) \end{pmatrix} \begin{pmatrix} (\max) \\ [p-q\ \ 0\ \ 0] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [q\ \ q\ \ 0] \\ (\Gamma'') \end{pmatrix} \right\} \\
&\quad \times \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p\ \ 0 & \\ & p & \end{bmatrix}.
\end{aligned} \tag{4.43a}$$

In this expression, (Γ') and (Γ'') are any arbitrary operator patterns appropriate to $[p-q\ \ 0\ \ 0]$ and $[q\ \ q\ \ 0]$, respectively. (Γ_s) is then a unique (but as yet undetermined) operator pattern belonging to the multiplicity set determined by the Δ pattern

$$[\Delta] = [\Delta'] + [\Delta'']. \tag{4.43b}$$

Next, we use Eqs. (3.50), (3.52), and (3.14) (for $n = 3$) together with Eq. (4.17a) to obtain

$$\begin{aligned}
&\left\{ \begin{pmatrix} [p\ \ q\ \ 0] \\ (\Gamma_s) \end{pmatrix} \begin{pmatrix} (\max) \\ [p-q\ \ 0\ \ 0] \\ (\Gamma') \end{pmatrix} \begin{pmatrix} [q\ \ q\ \ 0] \\ (\Gamma'') \end{pmatrix} \right\}([m]+[\Delta]) \\
&= \frac{D\begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p\ \ q\ \ 0] \end{pmatrix}([m])}{D\begin{pmatrix} [\Delta_1'\ \ \Delta_2'\ \Delta_3'] \\ [p-q\ \ 0\ \ 0] \end{pmatrix}([m]+[\Delta'']) \, D\begin{pmatrix} [\Delta_1''\ \Delta_2''\ \Delta_3''] \\ [q\ \ q\ \ 0] \end{pmatrix}([m])},
\end{aligned} \tag{4.44}$$

where the denominator functions are given, respectively, by Eqs. (4.29) (positive square root) and (4.18), and where, of course, $[\Delta] = [\Delta'] + [\Delta'']$.

Since the lower Gel'fand pattern couplings (Wigner coefficients) are explicitly known, Eq. (4.42) now becomes an equally explicit relation for obtaining all the $U(3)$ Wigner operators of the type

$$\left\langle \begin{matrix} (\Gamma_s) \\ [p \quad q \quad 0] \\ (M) \end{matrix} \right\rangle . \tag{4.45}$$

Let us also note that the coefficient (4.44) is not the most general Racah coefficient which we can obtain from Eq. (4.17a). We can use the orthogonality of the Racah invariants[4] to move the Racah invariant in Eq. (2.4a) to the left-hand side. Upon particularizing to $n = 3$ and choosing $[M'] = [p'\ q'\ 0]$, $[M] = [p''\ q''\ 0]$, and $\gamma'' = (\begin{smallmatrix} p'+p''\,0 \\ p'+p'' \end{smallmatrix})$, we obtain

$$\left\{ \begin{pmatrix} [p \quad q \quad 0] \\ (\Gamma_s) \end{pmatrix} \begin{pmatrix} (\text{max}) \\ [p' \quad q' \quad 0] \\ (\Gamma'_s) \end{pmatrix} \begin{pmatrix} [p'' \quad q'' \quad 0] \\ (\Gamma''_s) \end{pmatrix} \right\}$$

$$\times \begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} = \left\langle \begin{pmatrix} p & q & 0 \\ & p \quad 0 & \end{pmatrix} \middle\| \begin{bmatrix} & (\text{max}) & \\ p' & q' & 0 \\ & p' \quad 0 & \\ & p' & \end{bmatrix} \middle\| \begin{pmatrix} p'' & q'' & 0 \\ & p'' \quad 0 & \end{pmatrix} \right\rangle$$

$$\times \begin{bmatrix} & (\Gamma'_s) & \\ p' & q' & 0 \\ & p' \quad 0 & \\ & p' & \end{bmatrix} \begin{bmatrix} & (\Gamma''_s) & \\ p'' & q'' & 0 \\ & p'' \quad 0 & \\ & p'' & \end{bmatrix}, \tag{4.46}$$

where $p = p' + p''$ and $q = q' + q''$. [A reduced matrix element appears in this result in place of a square-bracket invariant because the $U(2)$ Racah invariant part of the square-bracket invariant is unity.] Using Eq. (4.17a), we now obtain

$$\left\{ \begin{pmatrix} [p \quad q \quad 0] \\ (\Gamma_s) \end{pmatrix} \begin{pmatrix} (\text{max}) \\ [p' \quad q' \quad 0] \\ (\Gamma'_s) \end{pmatrix} \begin{pmatrix} [p'' \quad q'' \quad 0] \\ (\Gamma''_s) \end{pmatrix} \right\} ([m] + [\Delta])$$

$$= F_R \begin{pmatrix} [p'\ q'\ 0] \\ [p'\ q'\ 0] \\ [\ p' \ \ 0\ ] \end{pmatrix} \begin{pmatrix} [p'' & q'' & 0] \\ & [p'' & 0] \end{pmatrix} \Big/ D \begin{pmatrix} [p'\ q'\ 0] \\ [p'\ q'\ 0] \end{pmatrix} ([p''\ q''\ 0])$$

$$\times D \begin{pmatrix} [\Delta_1\ \Delta_2\ \Delta_3] \\ [p \quad q \quad 0] \end{pmatrix} ([m]) \Big/ D \begin{pmatrix} [\Delta'_1\ \Delta'_2\ \Delta'_3] \\ [p' \quad q' \quad 0] \end{pmatrix} ([m] + [\Delta'']) \, D \begin{pmatrix} [\Delta''_1\ \Delta''_2\ \Delta''_3] \\ [p'' \quad q'' \quad 0] \end{pmatrix} ([m]). \tag{4.47}$$

Equation (4.44) is a special case of this more general result. The extra generality afforded by this coefficient is, however, not required in the following sections.

F. The general projective operator

Let us now outline an explicit procedure which could be used, in principle, to determine the general projective operator

$$\begin{bmatrix} (\Gamma) \\ p \quad q \quad 0 \\ (\gamma) \end{bmatrix}. \tag{4.48}$$

We first note that the canonical splitting proved in Ref. 11 implies the existence of the following *zero operators:*

$$\begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & p - \alpha + 1 & \end{bmatrix} = 0 \tag{4.49}$$

for all $k = 1, 2, 3, \ldots, \mathfrak{M}$, and all $\alpha = 1, 2, \ldots, k-1$, where $(\Gamma_1), (\Gamma_2), \ldots, (\Gamma_{\mathfrak{M}})$ denote the $\mathfrak{M}$ operator patterns which are determined by a specified $[\Delta_1 \Delta_2 \Delta_3]$ belonging to $[p\ q\ 0]$. Observe that (Γ_1) now denotes the pattern previously denoted by (Γ_s). *We cannot as yet make any definite assignment of* $(\Gamma_1), (\Gamma_2), \ldots, (\Gamma_{\mathfrak{M}})$ *onto the* $\mathfrak{M}$ *operator patterns having the prescribed* Δ *pattern. We only know that such an assignment exists.*

Next, consider the coupling law (2.4) for $U(3)$ specialized as follows:

$$\begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix} = \begin{bmatrix} & \cdot & \\ p-q & 0 & 0 \\ & p-q \quad 0 & \\ & \cdot & \end{bmatrix}^{\{R\}_3}_{\{R\}_2} \begin{bmatrix} & \cdot & \\ q & q & 0 \\ & q \quad 0 & \\ & \cdot & \end{bmatrix}, \tag{4.50}$$

where we note that for the relevant labels the square-bracket invariant coupling reduces to a $U(2)$ Racah invariant operator coupling, as indicated, on the lower operator patterns.

In order to give this coupling explicitly, it is convenient to introduce the following abbreviated notations in which we suppress the labels p and q:

$$\Theta_{\gamma(\Gamma')(\Gamma'')} = \sum_{\alpha = q+\gamma-p}^{q} \left\{ \begin{pmatrix} p & 0 \\ & \gamma \end{pmatrix} \begin{pmatrix} & p-q & \\ p-q & & 0 \\ & \gamma - \alpha & \end{pmatrix} \begin{pmatrix} q & 0 \\ & \alpha \end{pmatrix} \right\}$$

$$\times \begin{bmatrix} & (\Gamma') & \\ p-q & 0 & 0 \\ & p-q \quad 0 & \\ & \gamma-\alpha & \end{bmatrix} \begin{bmatrix} & (\Gamma'') & \\ q & q & 0 \\ & q \quad 0 & \\ & \alpha & \end{bmatrix}, \tag{4.51a}$$

$$R_{(\Gamma_k)(\Gamma')(\Gamma'')} = \left\{ \begin{pmatrix} [p \; q \; 0] \\ (\Gamma_k) \end{pmatrix} \begin{pmatrix} & (\max) & \\ [p-q & 0 & 0] \\ & (\Gamma') & \end{pmatrix} \begin{pmatrix} [q \; q \; 0] \\ (\Gamma'') \end{pmatrix} \right\}. \tag{4.51b}$$

Throughout this discussion, $(\Gamma_1), (\Gamma_2), \ldots, (\Gamma_{\mathfrak{M}})$ denote the distinct operator patterns determined by a *specified* $\Delta = [\Delta_1 \Delta_2 \Delta_3]$ belonging to $[p\; q\; 0]$. Similarly, (Γ') and (Γ'') always denote operator patterns belonging to $[p-q\; 0\; 0]$ and $[q\; q\; 0]$, respectively, which satisfy $[\Delta'] + [\Delta''] = [\Delta]$ ($[\Delta]$ specified).

With these notations and conventions, we may write Eq. (4.50) in the form

$$\begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix} = \sum_{(\Gamma')(\Gamma'')} R_{(\Gamma_k)(\Gamma')(\Gamma'')} \mathcal{O}_{\gamma(\Gamma')(\Gamma'')}. \tag{4.52a}$$

Similarly, we obtain

$$\mathcal{O}_{\gamma(\Gamma')(\Gamma'')} = \sum_{k=1}^{\mathfrak{M}} R_{(\Gamma_k)(\Gamma')(\Gamma'')} \begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix}, \tag{4.52b}$$

which, in turn, implies the following relation:

$$\sum_{k=1}^{\mathfrak{M}} \begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix}^{\dagger} \begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix} = \sum_{(\Gamma')(\Gamma'')} \mathcal{O}^{\dagger}_{\gamma(\Gamma')(\Gamma'')} \mathcal{O}_{\gamma(\Gamma')(\Gamma'')}. \tag{4.52c}$$

We remark that the operator $\mathcal{O}_{\gamma(\Gamma')(\Gamma'')}$ is completely known, the Racah invariants (4.15b) are unknown (except for $k = 1$), and the projective operators defined by Eq. (4.50) are likewise unknown (except for $k = 1$ and $\gamma = p$).

We assert: *The general structural relation, Eq. (4.50), and the canonical splitting conditions, Eq. (4.49), uniquely determine all U(3) projective operators of the form*

$$\begin{bmatrix} & (\Gamma_k) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix}, \tag{4.53}$$

except for phase; furthermore, they determine all the Racah invariants of the type given in Eq. (4.51b).

Let us give this construction. We begin by choosing $\gamma = p$ in Eq. (4.52c). The summation on the left-hand side then reduces to a single term $k = 1$. Thus, the operator

$$\begin{bmatrix} & (\Gamma_1) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} \tag{4.54a}$$

is uniquely determined except for phase [Equation (4.52c) reduces, in fact, to Eq. (4.16a).] We choose the phase[24] and proceed to Eq. (4.52b) for $\gamma = p$. The right-hand side of Eq. (4.52b) reduces to a single term $k = 1$. We use this equation to determine the Racah invariant $R_{(\Gamma_1)(\Gamma')(\Gamma'')}$. Using this Racah invariant in Eq. (4.52a), we obtain the projective operators

$$\begin{bmatrix} & (\Gamma_1) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix} \tag{4.54b}$$

for all $\gamma = 0, 1, \ldots, p$.

We now start the procedure all over, beginning with $\gamma = p - 1$ in Eq. (4.52c). The left-hand side reduces to two terms, $k = 1$ and 2. But the (Γ_1) operator is known from Eq. (4.54b). Hence, Eq. (4.52c) and our phase convention determines

$$\begin{bmatrix} & (\Gamma_2) & \\ p & q & 0 \\ & p \quad 0 & \\ & p-1 & \end{bmatrix}. \tag{4.55a}$$

We next proceed to Eq. (4.52b), setting $\gamma = p - 1$. The right-hand side reduces to two terms, $k = 1$ and 2. The $k = 1$ term is completely known as is the projective operator part for $k = 2$. This equation uniquely determines $R_{(\Gamma_2)(\Gamma')(\Gamma'')}$. Going to Eq. (4.52a), we now determine

$$\begin{bmatrix} & (\Gamma_2) & \\ p & q & 0 \\ & p \quad 0 & \\ & \gamma & \end{bmatrix} \tag{4.55b}$$

for $\gamma = 0, 1, \ldots, p - 1$.

One easily sees that continuing this procedure leads to the proof of the assertion made above. Let us remark that since $[\Delta]$ was arbitrary, our proof applies to the operators in *each* multiplicity set, including multiplicity one.

We can now make the final statement: *The general coupling laws and the canonical splitting in U(3) determine uniquely all projective operators except for phase. The same statement applies to all Wigner operators.*

Proof: The $U(3)$ projective operator $[p\; q\; 0]$ is obtained from the coupling

$$\begin{bmatrix} (\Gamma) \\ [p \; q \; 0] \\ (\gamma) \end{bmatrix} = \begin{bmatrix} \cdot \\ [p-q \;\; 0 \;\; 0] \\ \cdot \end{bmatrix}^{\{R\}}_{[R]} \begin{bmatrix} \cdot \\ [q \; q \; 0] \\ \cdot \end{bmatrix}. \tag{4.56}$$

But the lower operator pattern couplings are known, and the upper operator pattern couplings are just the $R_{(\Gamma)(\Gamma')(\Gamma'')}$ obtained above. (A similar proof applies directly to the Wigner operators.) Finally, the general projective operator $[M]$ is obtained from those of the type $[p\ q\ 0]$ by the following relation:

$$\begin{bmatrix} & \Gamma_{11} & \\ \Gamma_{12} & & \Gamma_{22} \\ M_{13} & M_{23} & M_{33} \\ \gamma_{12} & & \gamma_{22} \\ & \gamma_{11} & \end{bmatrix} = \begin{bmatrix} & 1 & \\ 1 & & 1 \\ 1 & 1 & 1 \\ 1 & & 1 \\ & 1 & \end{bmatrix}^{M_{33}} \times \begin{bmatrix} & \Gamma_{11}-M_{33} & \\ \Gamma_{12}-M_{33} & & \Gamma_{22}-M_{33} \\ M_{13}-M_{33} & M_{23}-M_{33} & 0 \\ \gamma_{12}-M_{33} & & \gamma_{22}-M_{33} \\ & \gamma_{11}-M_{33} & \end{bmatrix} \tag{4.57}$$

for arbitrary integers $M_{13} \geq M_{23} \geq M_{33}$.

The algebraic procedure described in this section is logically complete; but it is a major task to implement these techniques to obtain explicit results. Indeed, the only explicit operator we have given is the important one which occurs at the very first step, (4.54a). Here it was the underlying geometrical properties of the arrow patterns which led to a relatively simple interpretation of the structure of this operator. Our principal program is to uncover additional structures which will make these seemingly complicated algebraic manipulations explicable.

In this connection, we would like to note[25] that in $U(3)$ there is precisely one projective operator which is also a Wigner operator—the isoscalar:

$$\begin{bmatrix} & (\Gamma) & \\ p & q & 0 \\ q & & q \\ & q & \end{bmatrix} = \left\langle \begin{matrix} & (\Gamma) & \\ p & q & 0 \\ q & & q \\ & q & \end{matrix} \right\rangle . \tag{4.58}$$

Thus, one could construct this operator from the coupling, Eq. (4.56), and proceed to determine the general Wigner operator

$$\left\langle \begin{matrix} (\Gamma) \\ [p \quad 0] \\ (M) \end{matrix} \right\rangle \tag{4.59}$$

by using the generators.

G. The null space of the Wigner operator (Γ_s)

As remarked in the preceding section, the only explicit operator which we have constructed is the one labeled by $(\Gamma_1) = (\Gamma_s)$. This operator already exhibits considerable complexity, and clearly one must understand this structure before embarking on the more general program. Accordingly, we will now determine the conditions which specify the null space of the *Wigner operator* labeled by (Γ_s)

If one examines the general expression [Eq. (2.46) of Ref. 4] relating Racah invariants to Wigner operators, it is clear that the null space of a Wigner operator corresponds to the vanishings of a Racah coefficient. In particular, if we denote the null space of the Wigner operator (4.45) by $\mathfrak{N}(\Gamma_s)$, then

$$\mathfrak{N}(\Gamma_s) = \left\{ \begin{matrix} \text{all irrep spaces with labels } [m]: \\ D\begin{pmatrix} [\Delta_1 & \Delta_2 & \Delta_3] \\ [p & q & 0] \end{pmatrix}([m]) = 0 \end{matrix} \right\}. \tag{4.60}$$

Thus, this null space is determined by the zeros of the denominator function

$$D\begin{pmatrix} [\Delta_1 & \Delta_2 & \Delta_3] \\ [p & q & 0] \end{pmatrix}. \tag{4.61}$$

Examining Eq. (4.29), one sees that the determination of the zeros of the denominator function requires detailed knowledge of the properties of the function $G_q(\Delta; x)$. These properties are developed fully in II, where it is proved that $\mathfrak{N}(\Gamma_s) = \mathfrak{N}(\Gamma_1)$, where $\mathfrak{N}(\Gamma_1)$ is the *maximal* null space occurring in the series given in Conjecture 1 of the Introduction.

5. CONCLUDING REMARKS

We began our discussion of the structural properties of the canonical tensor operators by examining the totally symmetric operators in $U(n)$. The underlying structural property which accounts for the simplicity of a class of these operators (and their conjugates) was shown to be geometrical in origin—the no opposing arrow property. [The factorization lemma was demonstrated to be a useful tool for obtaining explicit abstract results without requiring the more technical manipulations of the algebraic method (Racah invariants, etc.).] We were then led to the discovery of the more general arrow pattern analysis for the denominator functions as well as the sum-over-paths formulation.

Turning to $U(3)$, it was demonstrated that the origin of the canonical splitting was again geometrical. However, thus far, this property leads directly to the construction of but one operator in each multiplicity set, although the canonical splitting determines, in principle, all operators.

The outstanding problems of most immediate interest are for $U(3)$: We must demonstrate that the null space of the Wigner operator labeled by (Γ_s) is, as asserted, the maximal null space. The explicit form (numerical array) of this operator pattern must be obtained. These two tasks require further development of the properties of the function $G_q(\Delta; x)$ and are carried out in the second paper of this series. Finally, there is the very difficult task of implementing the construction of the general projective operator in $U(3)$ in accordance with the canonical splitting of the multiplicity. While some progress has been made in this direction, the task remains, for the most part, incomplete.

One can still proceed a great deal further with the discussion of the general structures of tensor operators in $U(n)$. We shall continue this analysis in the third paper of this series, obtaining in the process new structural results for $U(3)$.

*Work supported by the National Science Foundation, the Army Research Office (Durham), and the Atomic Energy Commission.

†Visiting Staff Member (LASL) from Department of Physics, Duke University, Durham, North Carolina 27706.

‡Present address: Instituto de Física, Universidad Nacional Autonoma de México, México 20, D.F.

§AROD, Duke University, Durham, North Carolina 27706.

[1] G. E. Baird and L. C. Biedenharn, J. Math. Phys. (N.Y.) **5**, 1730 (1964).

[2] See Refs. 1, 3, and 4 for a detailed explanation of the notations of this paper.

[3] J. D. Louck, Am. J. Phys. **38**, 3 (1970).

[4] J. D. Louck and L. C. Biedenharn, J. Math. Phys. (N.Y.) 2, 2368 (1970).

[5] L. C. Biedenharn, in *Spectroscopic and Group Theoretical Methods in Physics (Racah Memorial Volume)*, edited by F. Block *et al.* (North-Holland, Amsterdam, 1968), p. 59.

[6] J. D. Louck and L. C. Biedenharn, J. Math. Phys. (N.Y.) **12**, 173 (1971).

[7] For convenience of expression, we do not always invert operator patterns. Note that in the notation for a projective operator the upper pattern is inverted, but the lower one is not. One must be very careful, however, not to confuse operator patterns with Gel'fand patterns despite their one-to-one correspondence.

[8] E. Artin, *Geometric Algebra* (Interscience, New York, 1964).

[9] The importance of the null space has been emphasized in Refs. 3 and 4, where a more complete discussion will be found.

[10] The concept of indecomposability is discussed in some detail in Ref. 4.

[11] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. (N.Y.) **8**, 691 (1967).

[12] G. E. Baird and L. C. Biedenharn, J. Math. Phys. (N.Y.) **6**, 1847 (1965).

[13] J. A. Castilho Alcarás, L. C. Biedenharn, K. T. Hecht, and G. Neely, Ann. Phys. (N.Y.) **60**, 85 (1970).

[14] S. J. Ališauskas, A.-A.A. Jucys, and A. P. Jucys, J. Math. Phys. (N.Y.) **13**, 1349 (1972).

[15] E. Chacón, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. (N.Y.) **13**, 577 (1972).

[16] L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).

[17] For the convenience of the reader, this résumé repeats some previously published material (Secs. 2A, 2B, 2D) in order to summarize in a reasonably self-contained way the three basic techniques on which the present work depends.

[18] The shift $[\Delta(\gamma)]_{n-1}$ is read off from the lower operator pattern in the same way in which $[\Delta(\Gamma)]_n$ is read off from the upper operator pattern [cf. Eqs. (1.3)]. However, the nth component of the shift of the lower pattern is not included in the definition of $[\Delta(\gamma)]_{n-1}$. Let us also note that an extremal operator pattern is one whose Δ pattern is a permutation of the irrep labels $[M]_n$ (see Ref. 16).

[19] Papers which relate most directly to our presentation include: J. Schwinger, *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), p. 229; V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962); T. A. Brody, M. Moshinsky, and I. Renero, J. Math. Phys. **6**, 1540 (1965) (and references therein).

[20] I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR Ser. Mat. **29**, 1329 (1965).

[21] E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra*, translated by J. J. Griffin (Academic, New York, 1959), p. 190.

[22] It is proved in II that (Γ_s) is the operator pattern (4.4a) if $q \leq \Delta_3 \leq p$, while it is the operator pattern (4.4b) if $0 \leq \Delta_3 \leq q$.

[23] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, Oxford, England, and Macmillan, New York, 1963), p. 362.

[24] See Footnote 27 of Ref. 4 for a detailed description of our phase convention.

[25] This fact was brought to our attention long ago by Dr. Mark Bolsterli of the Los Alamos Scientific Laboratory.

On the structure of the canonical tensor operators in the unitary groups. II. The tensor operators in $U(3)$ characterized by maximal null space*

L. C. Biedenharn

Department of Physics, Duke University, Durham, North Carolina 27706

J. D. Louck

Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544

(Received 8 May 1972)

The canonical splitting of the multiplicities of the unit tensor (Wigner) operators in $U(3)$ was used in I to determine explicitly one Wigner operator in each (arbitrary) multiplicity set. The denominator function whose zeroes define the null space of this Wigner operator is presented in a new form from which the complete identification of the null space is made. Using the properties of the intertwining number of $U(3)$, the null spaces of all the $U(3)$ Wigner operators are determined, and it is demonstrated that the null spaces of the operators belonging to a multiplicity set are simply ordered by inclusion. The Wigner operator previously obtained from the canonical splitting is shown to be the one having the maximal null space for its multiplicity set.

1. INTRODUCTION AND RÉSUMÉ

The present paper continues with the study, initiated in the first paper of this series,[1] of the structural properties of the canonical tensor operator labeling in the unitary groups. We now direct our attention to the symmetry group $U(3)$, since the existence of the canonical splitting of *all* multiplicities [for $U(3)$] has been previously proved.[2]

A general procedure for constructing these unique (to within phase) operators implied by the canonical splitting was outlined in Paper I. It was emphasized that, while the procedure is definitive, the explicit construction of these unit tensor (Wigner) operators is a difficult task. Furthermore, even if accomplished, the resulting matrix elements are likely[3] to be too complicated to understand unless one undertakes simultaneously the study of the structure of each result.

We were able to demonstrate,[1] in particular, that each multiplicity set of unit projective operators contains one operator whose structure is uniquely and simply determined to within a normalization by the geometrical properties of the arrow-patterns of the fundamental projective operators and their conjugates. (Indeed, it is quite likely that the origin of all the unit projective operators in a multiplicity set will ultimately be related to the geometrical properties of these arrow-patterns.) This led to the explicit form

$$\begin{bmatrix} & (\Gamma_s) & \\ p & q & 0 \\ & p \quad 0 & \\ & p & \end{bmatrix} = (-1)^{p-\Delta_1} F_R \begin{pmatrix} [\Delta_1 & \Delta_2 & \Delta_3] \\ [p & q & 0 \\ & [p \quad 0] & \end{pmatrix} \Bigg/ D\begin{pmatrix} [\Delta_1 & \Delta_2 & \Delta_3] \\ [p & q & 0\,] \end{pmatrix}, \tag{1.1}$$

where F_R is a restricted arrow-pattern function whose value is read off directly from the pattern calculus rules. The denominator function D is a seemingly very complicated function, which nonetheless could be determined explicitly through the use of the Factorization Lemma. The resulting form was, however, too complicated to understand directly. Turning to the sum-over-paths formulation (which again used the Factorization Lemma), we were able to derive in a simple way two recursion relations satisfied by the denominator. In this paper, we will approach the study of the properties of this denominator function through these recursion relations.

The properties of the denominator function D are far more important than its occurrence in Eq. (1.1) would seem to indicate. This is true because it is the zeroes of this denominator function—the set of irrep labels $\{[m]\}$ such that the denominator function vanishes—which determine completely the null space of the Wigner operator labeled by (Γ_s).[1] Since the principal aim of this series of papers is to illustrate and discuss structural properties of the canonical tensor operators, a complete elucidation of the properties of the denominator function is central to this purpose.

Let us note from Ref. 1 that the denominator function may be written in the following form:

$$D\begin{pmatrix} [\Delta_1 & \Delta_2 & \Delta_3] \\ [p & q & 0\,] \end{pmatrix}([m]) = \left[\frac{\Delta_1!\,\Delta_2!\,\Delta_3!}{(p+q)!\,\mathfrak{D}'([m]+[\Delta])} \prod_{i<j=1}^{3} (\Delta_i+\Delta_j+1)! \times \begin{pmatrix} x_{ij}+\Delta_i \\ \Delta_i+\Delta_j+1 \end{pmatrix}\right]^{1/2} \left(\frac{(p+q)!}{(p-q)!\,G_q(\Delta;x)}\right)^{1/2}, \tag{1.2a}$$

where

$$x_{ij} = p_{i3} - p_{j3}, \tag{1.2b}$$

and

$$G_q(\Delta;x) = G_q(\Delta_1\Delta_2\Delta_3;x_1x_2x_3), \tag{1.2c}$$

$$x_i = x_{jk} \quad (i,j,k \text{ cyclic in } 1,2,3). \tag{1.2d}$$

Let us observe: *The first square root factor in Eq.* (1.2a) *is precisely the result of applying the generalized denominator pattern calculus rules to the* Δ *pattern* $[\Delta_1\Delta_2\Delta_3]$, *using a path weight appropriate to* $[p+q\ 0\ 0]$. We insist that this factor has not been artifically introduced, since for $q=0$,

$$G_0(\Delta;x) = 1, \tag{1.3}$$

and the occurrence of the factor is essential.

Let us now summarize the plan of this paper. In Sec. 2, the properties of the function $G_q(\Delta;x)$ are developed in detail. The proof is given that

$$G_q(\Delta;x) = G_q^c(\Delta;x), \tag{1.4}$$

where the function $G_q^c(\Delta; x)$ was given explicitly in Paper I. These properties include the determination of the symmetries of $G_q(\Delta; x)$ and the determination of the zeroes of $G_q(\Delta; x)$. As previously noted, knowledge of these zeroes is essential to the discussion of the null space properties of the Wigner operator labeled by (Γ_s).

Since the role of the null spaces in the characterization of the canonical tensor operators is basic to our emphasis on structure, we give in Sec. 3 both algebraic and graphical representations of the intertwining number of three irreps of $U(3)$. This intertwining number, in turn, is then shown to determine *a priori* the null spaces of the Wigner operators and the nesting property of the null spaces of the Wigner operators in a multiplicity set is demonstrated. (Hence, the null spaces are simply ordered.)

In Sec. 4, the uniquely determined (by the zeroes of the denominator function) null space of the Wigner operator labeled by Γ_s is shown to be precisely the *maximal null space* determined by the abstract properties of the intertwining number. Observe that *we have no choice in the determination of these null spaces.*

Up to this point, the operator pattern (Γ_s) has been a mere label, although from general principles we know that it belongs to a multiplicity set of operator patterns having Δ pattern $[\Delta_1\Delta_2\Delta_3]$. In Sec. 5, we show how the specific numerical assignment of this pattern is induced by taking limits.

2. DETERMINATION OF THE DENOMINATOR FUNCTION

A. Recursion relations

We wish to determine the function $G_q(\Delta; x)$ through properties which are implied by the two recursion relations which it has been shown to satisfy.[1] We gain greater generality (thereby simplifying some of the proofs to follow) by replacing the integral parameters $\Delta_1, \Delta_2, \Delta_3$ by arbitrary parameters ξ_1, ξ_2, ξ_3, respectively, where $\xi = (\xi_1\xi_2\xi_3)$ may be any point of R^3. We furthermore regard the point $(x_1x_2x_3)$ as an arbitrary point satisfying the barycentric condition

$$x_1 + x_2 + x_3 = 0, \tag{2.1}$$

i.e., $x = (x_1x_2x_3)$ is an arbitrary point in the Möbius plane.[4]

We seek the functions $G_q(\xi; x)$, $q = 0, 1, \ldots,$ which satisfy the two recursion relations as follows [cf. Eqs. (4.30)–(4.32) of Ref. 1]:

$$\begin{aligned} x_1x_2x_3G_q(\xi_1\xi_2\xi_3; x_1x_2x_3) &= \xi_1\xi_2x_3(\xi_3 - x_1)(\xi_3 + x_2)(\xi_1 + x_3)(\xi_2 - x_3)\,G_{q-1}(\xi_1 - 1, \xi_2 - 1, \xi_3; x_1 + 1, x_2 - 1, x_3) \\ &+ \xi_2\xi_3x_1(\xi_1 - x_2)(\xi_1 + x_3)(\xi_2 + x_1)(\xi_3 - x_1)\,G_{q-1}(\xi_1, \xi_2 - 1, \xi_3 - 1; x_1, x_2 + 1, x_3 - 1) \\ &+ \xi_3\xi_1x_2(\xi_2 - x_3)(\xi_2 + x_1)(\xi_3 + x_2)(\xi_1 - x_2)\,G_{q-1}(\xi_1 - 1, \xi_2, \xi_3 - 1; x_1 - 1, x_2, x_3 + 1), \end{aligned} \tag{2.2a}$$

$$\begin{aligned} &(x_1 + \xi_2 - \xi_3)(x_2 + \xi_3 - \xi_1)(x_3 + \xi_1 - \xi_2)\,G_q(\xi_1\xi_2\xi_3; x_1x_2x_3) \\ &= \xi_1\xi_2(x_3 + \xi_1 - \xi_2)(\xi_2 + x_1)(\xi_1 - x_2)(\xi_1 + x_3)(\xi_2 - x_3)\,G_{q-1}(\xi_1 - 1, \xi_2 - 1, \xi_3; x_1x_2x_3) \\ &+ \xi_2\xi_3(x_1 + \xi_2 - \xi_3)(\xi_3 + x_2)(\xi_2 - x_3)(\xi_2 + x_1)(\xi_3 - x_1)\,G_{q-1}(\xi_1, \xi_2 - 1, \xi_3 - 1; x_1x_2x_3) \\ &+ \xi_3\xi_1(x_2 + \xi_3 - \xi_1)(\xi_1 + x_3)(\xi_3 - x_1)(\xi_3 + x_2)(\xi_1 - x_2)\,G_{q-1}(\xi_1 - 1, \xi_2, \xi_3 - 1; x_1x_2x_3), \end{aligned} \tag{2.2b}$$

where

$$G_0(\xi; x) = 1. \tag{2.2c}$$

We note that either of these recursion relations determines

$$\begin{aligned} G_1(\xi; x) = &- \xi_1\xi_2(\xi_1 + x_3)(\xi_2 - x_3) \\ &- \xi_2\xi_3(\xi_2 + x_1)(\xi_3 - x_1) \\ &- \xi_3\xi_1(\xi_3 + x_2)(\xi_1 - x_2) \\ &- \xi_1\xi_2\xi_3(\xi_1 + \xi_2 + \xi_3), \end{aligned} \tag{2.3}$$

where relation (2.1) must be used to obtain this form. One could, of course, continue to iterate, say, Eq. (2.2a) directly to obtain $G_q(\xi; x)$. However, this direct iteration does not lead easily to the answer we seek, namely, that $G_q(\xi; x)$ is a polynomial of degree $2q$ in the variables x (Proposition 2 below). We therefore follow the course of studying the recursion relations directly.

B. Symmetry relations

The first property of the function G_q is almost self-evident:

Lemma 1: The function $G_q(\xi; x)$ has the symmetry: $G_q(P\xi; Px) = G_q(\xi, \delta_P x)$, where P denotes a permutation of the indices 1, 2, 3 ($P \in S_3$) and δ_P is the signature of P.

Proof: The function $G_0(\xi; x) = 1$ certainly possesses this symmetry. It is also evident, by inspection, that the recursion relation (2.2a) shows that G_q possesses this property, if G_{q-1} does. The result follows.

To obtain the next property, it is convenient to note that the factors on the right-hand side of Eq. (2.2a)—the factors that multiply the G_{q-1}—are but cyclic permutations of a single function. That is, we define the function $g(\xi, x)$ to be

$$g(\xi; x) \equiv \xi_1\xi_2x_3(\xi_3 - x_1)(\xi_3 + x_2)(\xi_1 + x_3)(\xi_2 - x_3).$$

One next observes that $g(\xi; x)$ has the symmetry

$$g(\xi_1\xi_2\xi_3; x_1x_2x_3) = g(\xi_1, \xi_3 - x_1, \xi_2 + x_1; x_1x_2x_3).$$

(Verification of this symmetry requires use of the relation $x_1 + x_2 + x_3 = 0$.)

Using the recursion relation (2.2a), one now sees easily that this same symmetry extends to the

$G_q(\xi, x)$, since the symmetry is obviously true for $G_0 = 1$. Hence, we have proved:

Lemma 2: The function $G_q(\xi, x)$ has the symmetry

$$G_q(\xi_1, \xi_3 - x_1, \xi_2 + x_1; x_1 x_2 x_3) = G_q(\xi; x).$$

The symmetry given in Lemma 2 has the feature that the x's stay fixed while the ξ's are changed. A similar feature occurs in the second form of the recursion relation (2.2b). By using both recursion relations, we will obtain a symmetry for G_q in which the x's vary and ξ's stay fixed.

Lemma 3: The function $G_q(\xi, x)$ has the symmetry

$$G_q(\xi_1 \xi_2 \xi_3; -x_1 - \xi_2 + \xi_3, -x_2 - \xi_3 + \xi_1, -x_3 - \xi_1 + \xi_2) = G_q(\xi; x).$$

Proof: Let us assume that this symmetry is true for G_{q-1}. Next in the recursion relation (2.2a), we make the substitutions

$$x_1 \to -x_1 - \xi_2 + \xi_3, \quad x_2 \to -x_2 - \xi_3 + \xi_1, \quad x_3 \to -x_3 - \xi_1 + \xi_2. \tag{2.4}$$

(Note that this preserves the relation $x_1 + x_2 + x_3 = 0$.) It will now be observed that the right-hand side of Eq. (2.2a), after the substitution is made, becomes precisely the right-hand side of Eq. (2.2b)! [Let us be explicit and note that there is an over-all minus sign that will cancel out. Moreover, we should note that we have used our assumption that the symmetry holds for G_{q-1} in making this identification with the right-hand side of Eq. (2.2b).] It is furthermore seen that the left-hand side of Eq. (2.2a), after the substitutions of Eq. (2.4) (and canceling the minus sign), now shows, by Eq. (2.2b), that the symmetry must hold for G_q. The symmetry is obviously true for $G_0 = 1$; hence it holds in general.

The symmetry given by Lemma 3 may be put more perspicuously if we introduce new x_i variables:

$$x_1' \equiv x_1 + \tfrac{1}{2}(\xi_2 - \xi_3), \quad x_2' \equiv x_2 + \tfrac{1}{2}(\xi_3 - \xi_1), \quad x_3' \equiv x_3 + \tfrac{1}{2}(\xi_1 - \xi_2). \tag{2.5}$$

The subsidiary relation $\Sigma x_i = 0$ now implies that $\Sigma x_i' = 0$.

In the variables $\{x_i'\}$, the symmetry given by Lemma 3 becomes the statement: $G_q'(\xi; x') = G_q'(\xi; -x')$. That is: *The function $G_q'(\xi; x') \equiv G_q(\xi; x)$ expressed as a function in the barycentric (Möbius) plane $\{x_i'\}$ shows central symmetry in the origin $x_i' = 0$.*

The symmetry properties given in Lemmas 1–3—and combination of these symmetries—may be put in a very elegant form. Let us write the function G_q in the form

$$G_q(\xi; x) = G_q \begin{pmatrix} \xi_1 & \xi_3 - x_1 & \xi_2 + x_1 \\ \xi_2 & \xi_1 - x_2 & \xi_3 + x_2 \\ \xi_3 & \xi_2 - x_3 & \xi_1 + x_3 \end{pmatrix}. \tag{2.6}$$

Then all the symmetries implied by Lemmas 1–3 are contained in the following statement.

Proposition 1: $G_q(\xi; x)$ written in the form given by Eq. (2.6) is invariant under all permutations of rows and columns, and under transposition.

Proof: Lemma 3 is the statement of invariance under exchange of columns 2 and 3.

Lemma 2 asserts the invariance under transposition (upon noting the subsidiary condition $\Sigma x_i = 0$).

Lemma 1 is equivalent to the invariance under all permutations of the rows, combined with an exchange of columns 2 and 3 if $\delta_P = -1$.

Transposition, followed by exchange of rows 1 and 2, and again transposing, is equivalent to exchange of columns 1 and 2. Thus, all permutations of the columns can be generated.

C. The polynomial property

The importance of these symmetry relations, established in B above, is that they enable us to prove an essential property of G_q:

Proposition 2: The function $G_q(\xi; x)$ is a polynomial of degree $2q$ in the variables x_1, x_2, x_3.

Proof: Consider the right-hand side of Eq. (2.2a), and let $x_3 = 0$, so that the first of the three terms vanishes. The right-hand side thus takes the form ($x_3 = 0 \Rightarrow x_1 = -x_2 \equiv x$):

$$\begin{aligned} \text{RHS} = \xi_1 \xi_2 \xi_3 x (x + \xi_1)(x + \xi_2)(x - \xi_3) \\ \times \{G_{q-1}(\xi_1 - 1, \xi_2, \xi_3 - 1; x - 1, -x, 1) \\ - G_{q-1}(\xi_1, \xi_2 - 1, \xi_3 - 1; x, -x + 1, -1)\}. \end{aligned}$$

Using Proposition 1, one sees that the two G_{q-1} above are equal; hence, the right-hand side vanishes for $x_3 = 0$. Therefore, the right-hand side has x_3 as a factor and thus, by symmetry, the factor $x_1 x_2 x_3$.

That the degree is at most $2q$ follows easily by induction. That the degree is precisely $2q$ follows by examining a special case given below [Eq. (2.7)].

We now turn to developing explicit special cases for G_q. The simplest such case occurs for $\xi_3 = 0$. It follows easily from Eq. (2.2a) that $G_q(\xi_1 \xi_2 0; x)$ has the form

$$G_q(\xi_1 \xi_2 0; x_1 x_2 x_3) = (-1)^q (q!)^4 \times \binom{\xi_1}{q} \binom{\xi_2}{q} \binom{\xi_1 + x_3}{q} \binom{\xi_2 - x_3}{q}, \tag{2.7a}$$

where

$$\binom{z}{q} = z(z-1) \ldots (z - q + 1)/q! \tag{2.7b}$$

denotes the binomial function.

Using Proposition 1, this result can be given a wide variety of forms.

Let us note, for completeness, that this result suffices to establish that the G_q polynomial possesses precisely degree $2q$.

More interesting results obtain for special values of the x_i variables. Let us take $x_1 = \xi_3$. It follows easily once again from the recursion relation (2.2a), that we have the result

$$G_q(\xi_1\xi_2\xi_3;\xi_3,x,-\xi_3-x)=(-1)^q(q!)^4 \times\binom{\xi_1}{q}\binom{\xi_2+\xi_3}{q}\binom{\xi_1-x}{q}\binom{\xi_2+\xi_3+x}{q} \quad (2.8)$$

independent of the result given in Eq. (2.7a). Actually both results may be given as special cases of an elegant general result:

Lemma 4: Let G_q be written in the form

$$G_q\begin{pmatrix}\xi_1 & \xi_3-x_1 & \xi_2+x_1\\ \xi_2 & \xi_1-x_2 & \xi_3+x_2\\ \xi_3 & \xi_2-x_3 & \xi_1+x_3\end{pmatrix}.$$

and define the nine entries in this 3×3 array to be (g_{ij}). If one of the entries is zero, say g_{12}, then

$$G_q=(-)^q(q!)^4\binom{g_{11}}{q}\binom{g_{13}}{q}\binom{g_{22}}{q}\binom{g_{32}}{q}.$$

Note that the entries for the binomial functions in this result are those elements of the 3×3 array which occur in the row and column containing the zero element.

The proof is obvious from Eqs. (2.7), (2.8), and Proposition 1.

Let us obtain one more general property of the function G_q.

Lemma 5: Assume that $q-\xi_i$ is a nonnegative integer. Then G_q has the factor

$$\binom{\xi_j+x_i}{q-\xi_i}\binom{\xi_k-x_i}{q-\xi_i},$$

where (ijk) is a positive permutation of (123).

Proof: Assume this property is true for G_{q-1}. Then the recursion formula (2.2a) shows that the property is then true for G_q. Since this property can be verified to be true for G_1, it is therefore true in general.

D. The zeroes of G_q

The general properties of the function G_q, which have been demonstrated above, are the tools by which we will seek to understand more of the nature of the function. Since this function is now known to be a polynomial (Proposition 2), it is natural to inquire about the set of points $\{(x_1,x_2,x_3)\}$ on which the polynomial has the value zero. The significant result which is required to understand the zeroes of $G_q(\xi;x)$ is

Lemma 6: Let ξ_3-x_1 and ξ_1-x_2 be nonnegative integers. Then $G_q(\xi;x)$ has the value zero for all such integers satisfying $(\xi_3-x_1)+(\xi_1-x_2)\le q-1$.

Proof: Assume the property is true for G_{q-1}. Then the recursion relation shows the property to be true for G_q. Since the property can be verified to be true for G_1, it is true in general.

The set of points defined in Lemma 6 is just the set of lattice points lying on the boundary of, and interior to, an equilateral triangle in the Möbius plane. The vertices of the triangle are located at the points $(\xi_3,\xi_1,-\xi_1-\xi_3)$, $(\xi_3,\xi_1-q+1,-\xi_3-\xi_1+q-1)$, and $(\xi_3-q+1,\xi_1,-\xi_1-\xi_3+q-1)$. Thus, the triangle has q lattice points on each of its sides and contains $q(q+1)/2$ lattice points in all (if one vertex point is taken as the origin, then by lattice points of the triangle we mean the set of points on the boundary of, and interior to, the triangle which have integral coordinates).

If we now consider the positive permutations of a point (x_1,x_2,x_3) together with the central symmetry of Lemma 3, we obtain *six triangles* in the Möbius plane, the lattice points of which are zeroes of the polynomial function $G_q(\xi;x)$. If we suppress the $x_3=-x_1-x_2$ coordinate, then the polynomial $G_q(\xi;x)$ has value zero on the following set of points Z:

$$\begin{aligned}Z=\{&(\xi_3-a_1,\xi_1-b_1),(-\xi_2+a_2,-\xi_3+b_2),\\ &(\xi_1+\xi_3-q+1+a_3,-\xi_3+b_3),\\ &(q-1-\xi_1-\xi_2-a_4,\xi_1-b_4),\\ &(-\xi_2+a_5,\xi_1+\xi_2-q+1+b_5),\\ &(\xi_3-a_6,q-1-\xi_2-\xi_3-b_6):(a_i,b_i)\\ &\text{are nonnegative integers such that } a_i+b_i\\ &\le q-1 \text{ and } (\xi_1\xi_2\xi_3) \text{ is an arbitrary, but fixed}\\ &\text{point of } R^3\}.\end{aligned} \quad (2.9)$$

For a fixed point $\xi\in R^3$, there are, in general, $3q(q+1)$ distinct points in the set Z.

We are now in a position to assert a most remarkable result:

Proposition 3: The set Z, on which $G_q(\xi;x)$ vanishes, uniquely determines $G_q(\xi;x)$ up to a multiplicative factor which depends at most on ξ.

Proof: Let us suppress the ξ dependence and write simply

$$f(x,y)=G_q(\xi;x,y,-x-y).$$

Then $f(x,y)$ is a polynomial of degree $2q$ (Proposition 2) having the form

$$f(x,y)=\sum_{\substack{s,t\\ s+t\le 2q}} a_{st}x^sy^t. \quad (2.10)$$

Since $f(x,y)$ vanishes on the set Z, we have

$$\sum_{\substack{s,t\\ s+t\le 2q}} a_{st}x^sy^t=0,\qquad \text{each } (x,y)\in Z. \quad (2.11)$$

These equations comprise a system of $3q(q+1)$ homogeneous algebraic equations in the $(q+1)(2q+1)$ unknown coefficients $\{a_{st}\}$. Let $M=(x^sy^t)$ denote the matrix in which the columns are enumerated by the $(q+1)(2q+1)$ values which s and t may assume and in which the $3q(q+1)$ rows are enumerated by the points $(x,y)\in Z$ (one row for each point). The system of equations (2.11) can now be written as

$$MA=0, \quad (2.12)$$

where A is the column matrix of $(q+1)(2q+1)$ rows having the $\{a_{st}\}$ as elements. Then the necessary and sufficient condition that Eq. (2.12) has *exactly one* linearly independent solution is

$$\text{rank } M = q(2q + 3), \tag{2.13}$$

where this condition is to hold for generic points ξ.

Subsequently, we will give *explicitly* a polynomial of degree $2q$ which vanishes on the set Z *for each* $\xi \in R^3$. The existence of such a solution implies

$$\text{rank } M \le q(2q + 3) \quad \text{for each} \quad \xi \in R^3. \tag{2.14}$$

A direct demonstration of property (2.13) is quite difficult. It is, however, sufficient to demonstrate that

$$\text{rank } M = q(2q + 3) \quad \text{for} \quad \xi_1 = 0 \quad \text{and generic } \xi_2 \text{ and } \xi_3. \tag{2.15}$$

The proof of this restricted result establishes, in fact, the general result by the argument as follows: If Eq. (2.15) is valid for generic ξ_2, ξ_3, then also Eq. (2.13) is valid for generic ξ_1, ξ_2, ξ_3, since the rank of M can *at most* be decreased by setting $\xi_1 = 0$, i.e., rank $M = q(2q + 3)$ for $\xi_1 = 0$ implies rank $M = q(2q + 3)$, generally. [The fact that there may exist certain points $\xi \in R^3$ for which rank $M < q(2q+3)$ is of no importance—we need only demonstrate that there exists a determinant of M of order $q(2q + 3)$ which is not *identically* zero in ξ.]

The proof of Eq. (2.15) will be given by a direct demonstration that

$$f(x,y) = a(\xi_2\xi_3)\binom{\xi_2 + x}{q}\binom{\xi_3 - x}{q} \tag{2.16}$$

is the *unique polynomial* of degree $2q$ which vanishes on the set of points

$$Z' \equiv Z \quad \text{for} \quad \xi_1 = 0. \tag{2.17}$$

For this purpose, we introduce certain subsets $Z'_k \subset Z'$ for $k = 1, 2, \ldots, q/2$ or $(q + 1)/2$. Z'_k contains the points as follows:

$(a)_k$: $(\xi_3 - a_1, -k + 1), a_1 = k - 1, k, \ldots, q - k,$
$(q - 1 - \xi_2 - a_4, -k + 1), a_4 = 0, 1, \ldots, q - k;$

$(b)_k$: $(-\xi_2 + a_2, -\xi_3 + k - 1),$
$a_2 = k - 1, k, \ldots, q - k,$
$(\xi_3 - q + 1 + a_3, -\xi_3 + k - 1),$
$a_3 = 0, 1, \ldots, q - k;$

$(c)_k$: $(-\xi_2 + k - 1, -\xi_3 + b_2),$
$b_2 = k, k + 1, \ldots, q - k,$
$(-\xi_2 + k - 1, \xi_2 - q + 1 + b_5),$
$b_5 = 0, 1, \ldots, q - k;$

$(d)_k$: $(\xi_3 - k + 1, -b_1), b_1 = k, k + 1, \ldots, q - k,$
$(\xi_3 - k + 1, q - 1 - \xi_2 - \xi_3 - b_6),$
$b_6 = 0, 1, \ldots, q - k.$

[For $k = (q + 1)/2$ (q *odd*), we do not require (c) and (d).]

First consider Z'_1. Since $f(x,y)$ is a polynomial of degree $2q$ which vanishes on the $2q$ distinct points of $(a)_1$, it follows that

$$f(x,0) = a(\xi_2\xi_3)\binom{\xi_2 + x}{q}\binom{\xi_3 - x}{q}.$$

Observe that it also now follows that

$$f(x,0) = 0, \quad \text{each } (x,y) \in Z'.$$

We next write

$$f(x,y) = f(x,0) - y g_{2q-1}(x,y),$$

where $g_{2q-1}(x,y)$ is a polynomial of at most degree $2q - 1$. $g_{2q-1}(x,y)$ vanishes on the $2q$ distinct points $(b)_1$ of Z'_1. Therefore, $g_{2q-1}(x, -\xi_3)$ vanishes identically in x, and $g_{2q-1}(x,y)$ must have the form $g_{2q-1}(x,y) = (\xi_3 + y)g_{2q-2}(x,y)$, where $g_{2q-2}(x,y)$ is a polynomial of degree at most $2q - 2$. It vanishes on the $2q - 1$ distinct points $(c)_1$ of Z'_1. Therefore, $g_{2q-2}(\xi_2, y)$ vanishes identically in y, and $g_{2q-2}(x,y)$ must have the form $g_{2q-2}(x,y) = (\xi_2 + x)g_{2q-3}(x,y)$, where $g_{2q-3}(x,y)$ is of degree at most $2q-3$. It vanishes on the $2q - 1$ distinct points $(d)_1$ of Z'_1. Therefore, $g_{2q-3}(\xi_3, y)$ vanishes identically in y, and $g_{2q-3}(x,y) = (\xi_3 - x)g_{2q-4}(x,y)$, where $g_{2q-4}(x,y)$ is at most of degree $2q - 4$. Thus, the conclusion at the end of step 1 of our proof is that $f(x,y)$ has the form

$$f(x,y) = f(x,0) + (-y)(\xi_3 + y)(\xi_2 + x)(\xi_3 - x)g_{2q-4}(x,y).$$

We continue this procedure to step $2, \ldots,$ step $k, \ldots,$ where we assert that the conclusion at the end of step k (in which the points of Z'_k are considered) is

$$f(x,y) = f(x,0) + \binom{-y}{k}\binom{\xi_3 + y}{k}\binom{\xi_2 + x}{k}\binom{\xi_3 - x}{k}g_{2q-4k}(x,y), \tag{2.18}$$

where g_{2q-4k} is of degree at most $2q - 4k$.

The proof of Eq. (2.18) is by induction on k. Thus, we assume the validity of Eq. (2.18) for $k \to k - 1$ and consider the implications of the vanishing of $f(x,y)$ on the points of Z'_k. Considering the points $(a)_k$ of Z'_k, we see that $g_{2q-4k+4}(x, -k + 1)$, which is a polynomial of degree at most $2q - 4k + 4$ in x, vanishes on the $2q - 3k + 3 > 2q - 4k + 4 (k > 1)$ distinct points of $(a)_k$. Therefore, $g_{2q-4k+4}(x, -k + 1)$ vanishes identically in x, and $g_{2q-4k+4}(x,y)$ must have the form

$$g_{2q-4k+4}(x,y) = (-y - k + 1)g_{2q-4k+3}(x,y).$$

Similarly, we conclude, in turn, from the vanishings on the points $(b)_k$, $(c)_k$, and $(d)_k$ of Z'_k, the results

$$g_{2q-4k+3}(x,y) = (\xi_3 + y - k + 1)g_{2q-4k+2}(x,y),$$
$$g_{2q-4k+2}(x,y) = (\xi_2 + x - k + 1)g_{2q-4k+1}(x,y),$$
$$g_{2q-4k+1}(x,y) = (\xi_3 - x - k + 1)g_{2q-4k}(x,y).$$

The linear factors arising from step k are precisely the factors required to carry Eq. (2.18) for $k \to k - 1$ into the same form for k. Since the form (2.18) is true for $k = 1$, it is true generally.

For the final step, we proceed as follows. We choose $k = q/2$ for q even and select any point $(x,y) \in Z'$

such that the factors preceding g_0 = constant in Eq. (2.18) are nonvanishing, e.g., $(x, y) = (\xi_3 - q + 1, q - 1 - \xi_2 - \xi_3)$. The vanishing of $f(x, y)$ on this point then requires $g_0 = 0$, i.e., $f(x, y) = f(x, 0)$ for q even. Similarly, for q odd, we choose $k = (q-1)/2$ in Eq. (2.18) and note that the polynomial $g_2(x, y)$ of degree at most two must vanish on the points $(a)_{(q+1)/2}$ and $(b)_{(q+1)/2}$ of $Z'_{(q+1)/2}$. These vanishings require that $g_2(x, y) = [-y - \frac{1}{2}(q-1)][\xi_3 + y - \frac{1}{2}(q-1)]g_0$. Finally, we select the point $(x, y) \in Z'$ above to show that $g_0 = 0$, i.e., $f(x, y) = f(x, 0)$ for q odd.

We have now completed the proof of the result: $f(x, y) = f(x, 0)$ *given by* Eq. (2.16) *is the unique polynomial of degree* $2q$ *which vanishes on the set* Z'. Proposition 3 now also follows.

E. The proof of $G_q(\xi; x) = G_q^c(\xi; x)$

Let us now turn to the proof of one of the principal results of this section. Namely, that the unique solution to the recursion relations (2.2) is the conjectured form given in I. We rewrite this conjectured form in terms of the polynomials in three variables x, y, and z defined as follows:

$$f_{q,r}(xyz) \equiv (q-r)!\, r! \binom{x}{q-r}\binom{y}{r}\binom{z}{r}. \qquad (2.19)$$

Note that this function is invariant under the interchange of y and z. The conjectured solution may now be written in the form

$$G_q^c(\xi; x) = (-1)^q q! \sum_{(k)} \binom{\xi_1 + \xi_2 + \xi_3 - k_1 - k_2 - k_3}{k_4}$$
$$\times f_{q,k_1}(\xi_1, \xi_2 + x_1, \xi_3 - x_1)$$
$$\times f_{q,k_2}(\xi_2, \xi_3 + x_2, \xi_1 - x_2)$$
$$\times f_{q,k_3}(\xi_3, \xi_1 + x_3, \xi_2 - x_3), \qquad (2.20)$$

where the sum is over all nonnegative integers k_1, k_2, k_3, k_4 which add to q: $k_1 + k_2 + k_3 + k_4 = q$.

The following result is immediately evident.

Lemma 7: $G_q^c(\xi; x)$ obeys the symmetries stated in Lemmas 1 and 3.

[It is, however, far from obvious that the symmetry of Lemma 2 is obeyed (It is, nonetheless, true). Indeed, a direct proof of this would be quite difficult.]

Observing that

$$f_{q,r}(0, y, z) = \delta_{q,r}\, q! \binom{y}{q}\binom{z}{q}, \qquad (2.21)$$

we obtain from Eq. (2.20) the result

$$G_q^c(\xi_1\xi_2 0; x) = G_q(\xi_1\xi_2 0; x) \qquad (2.22)$$

[cf. Eq. (2.7a)]. However, because the symmetry of Lemma 2 is not manifest in the form (2.20), neither is the property (2.8). This property is, however, correct as will soon be proved.

We digress for a moment to establish two important properties of the polynomials (2.19):

$$\sum_{r=0}^{q-n} \binom{x+y+z-n-r}{q-n-r} f_{q,r}(x, y, z) = f_{q,q-n}(x, x+y-n, x+z-n) \qquad (2.23a)$$

for $n = 0, 1, \ldots, q$;

$$\sum_{r=0}^{q} f_{q,r}(x, y, z) f_{q,q-r}(w, x+y+w-r, z-r) = f_{2q,q}(x+y, x+w, z) \qquad (2.23b)$$

for arbitrary variables x, y, z, w.

Relation (2.23a) is simply a polynomial interpretation of Saalschütz's formula.[5] Relation (2.23b) follows easily from Eq. (2.23a) upon noting that the product under the summation can be written as

$$q! \binom{z}{q}\binom{x+y+w-r}{q-r} f_{q,r}(x, y, w),$$

using

$$\binom{z}{r}\binom{z-r}{q-r} = \binom{q}{r}\binom{z}{q}.$$

Now let us return to our objective: the proof that $G_q = G_q^c$. Relation (2.23a) may be used to carry out one of the summations in Eq. (2.20); one obtains three equivalent results, depending on which summation k_i one elects to eliminate. Thus, combining the binomial factor with f_{q,k_3} and using relation (2.23a) leads to

$$G_q^c(\xi; x) = (-1)^q q! \sum_{(k)} f_{q,k_1}(\xi_1\xi_2 + x_1, \xi_3 - x_1)$$
$$\times f_{q,k_2}(\xi_2, \xi_3 + x_2, \xi_1 - x_2)$$
$$\times f_{q,k_3}(\xi_3, \xi_1 + \xi_3 + x_3 - q + k_3,$$
$$\xi_2 + \xi_3 - x_3 - q + k_3), \qquad (2.24)$$

in which the sum is now over all nonnegative integers $(k) = (k_1k_2k_3)$ which sum to q. Observe that we lose "obvious symmetries" in making this reduction. These symmetries are, of course, contained in the equalities of the forms obtained by reducing Eq. (2.20) in the three possible ways.

Particularizing the result, Eq. (2.24), to the case $x_1 = \xi_3$, using the property

$$f_{q,k_1}(\xi_1, \xi_2 + \xi_3, 0) = \delta_{k_1,0}\, q! \binom{\xi_1}{q},$$

and relation (2.23b), we obtain

$$G_q^c(\xi; \xi_3, x, -\xi_3 - x) = (-1)^q (q!)^2 \binom{\xi_1}{q}$$
$$\times f_{2q,q}(\xi_2 + \xi_3 + x, \xi_2 + \xi_3, \xi_1 - x)$$
$$= G_q(\xi; \xi_3, x, -\xi_3 - x). \qquad (2.25)$$

From Lemma 7 and Eqs. (2.22) and (2.25), we have proved:

Lemma 8: $G_q^c(\xi; x)$ obeys the result given by Lemma 4, namely, when one of the nine entries in the 3×3 array

$$\begin{matrix} \xi_1 & \xi_3 - x_1 & \xi_2 + x_1 \\ \xi_2 & \xi_1 - x_2 & \xi_3 + x_2 \\ \xi_3 & \xi_2 - x_3 & \xi_1 + x_3 \end{matrix}$$

is zero, G_q^c assumes the form given in Lemma 4.

Lemmas 7 and 8 indicate, but do not prove, that G_q and G_q^c are identical. Indeed, we have not yet established the full symmetry of Proposition 1 for G_q^c. The essential lemma which is required to accomplish this fully is the following.

Lemma 9: $G_q^c(\xi;x)$ has the value zero on the set Z given by Eq. (2.9).

Proof: In consequence of the symmetries given by Lemma 7, it is sufficient to prove that $G_q^c(\xi;x)$ obeys Lemma 6. Consider the factors

$$\binom{\xi_3 - x_1}{k_1}\binom{\xi_1 - x_2}{k_2}\binom{-\xi_1 - \xi_3 - x_3 + q - 1}{k_3} \tag{2.26}$$

arising under the summation in Eq. (2.24).

It follows that for nonnegative integers $\xi_3 - x_1$ and $\xi_1 - x_2$ satisfying $(\xi_3 - x_1) + (\xi_1 - x_2) \leq q - 1$, we also have that $-\xi_1 - \xi_3 - x_3 + q - 1 = -(\xi_1 - x_2) - (\xi_3 - x_1) + q - 1$ is a nonnegative integer less than or equal to $q - 1$. Hence, at least one of the factors in (2.26) vanishes unless $k_1 \leq \xi_3 - x_1$, $k_2 \leq \xi_1 - x_2$, and $k_3 \leq -\xi_1 - \xi_3 - x_3 + q - 1$. But this implies $k_1 + k_2 + k_3 \leq q - 1$, which violates $k_1 + k_2 + k_3 = q$ in the summation in Eq. (2.24). Thus, in the form (2.24), $G_q^c(\xi;x)$ vanishes *termwise*, and Lemma 9 is proved.

We are now able to state a principal result:

Proposition 4: The polynomials $G_q(\xi;x)$ and $G_q^c(\xi;x)$ are identically equal.

Proof: Using Lemma 9 and Proposition 3, we must have

$$G_q(\xi;x) = a(\xi)G_q^c(\xi;x).$$

Setting $x_1 = \xi_3$ and using Eq. (2.25), we find $a(\xi) = 1$.

Corollary: $G_q^c(\xi;x)$ obeys Proposition 1.

Using only the symmetries of the polynomial $G_q(\xi;x)$ and its zeroes, we have been able to demonstrate that the solution to the recursion relations (2.2) is uniquely given by $G_q^c(\xi;x)$.

We still have made no use of Lemma 5. The significance of this lemma will become clear in the determination of the null space of the Wigner operator (Γ_s). We now turn to the general discussion of the null spaces of the $U(3)$ Wigner operators.

3. NULL SPACES OF THE $U(3)$ WIGNER OPERATORS

A careful development of the concept of the null space of a Wigner operator is essential to the present work since one of our goals is to understand fully the vanishings of a Wigner coefficient. Before entering into this discussion, we require detailed knowledge of two numbers: the multiplicity $\mathfrak{M}$ of a prescribed Δ pattern belonging to a set of irrep labels $[M] \equiv [M_{13}M_{23}M_{33}]$ and the intertwining number $\mathscr{I}$ which is the number of occurrences of an irrep $[m']$ in the direct product $[M] \otimes [m]$.

A. The multiplicity of a Δ pattern

We first consider the determination of the number m.[6] The Δ pattern of the $U(3)$ Wigner operator specified by the operator pattern

$$\begin{pmatrix} M_{13} & & M_{23} & & M_{33} \\ & \Gamma_{12} & & \Gamma_{22} & \\ & & \Gamma_{11} & & \end{pmatrix} \tag{3.1}$$

is by definition the triplet of integers $[\Delta] = [\Delta_1\Delta_2\Delta_3]$, where

$$\Delta_1 = \Gamma_{11}, \quad \Delta_2 = \Gamma_{12} + \Gamma_{22} - \Gamma_{11}, \\ \Delta_3 = M_{13} + M_{23} + M_{33} - \Gamma_{12} - \Gamma_{22}. \tag{3.2}$$

Thus, for a given operator pattern, one simply reads off the corresponding Δ pattern. Letting Γ_{12}, Γ_{22}, and Γ_{11} run over the set of all integers which satisfy the "betweenness" conditions (the irrep labels $[M]$ being specified), we then obtain the set of Δ patterns belonging to irrep $[M]$. Clearly, those operator patterns having the same value of $\Gamma_{12} + \Gamma_{22}$ correspond to the same Δ pattern.

The inverse problem is: (a) Determine when a specified triplet of numbers $[\Delta_1\Delta_2\Delta_3]$ is the Δ pattern belonging to irrep $[M]$, and (b) determine which operator patterns having irrep labels $[M]$ correspond to this Δ pattern.

The solution to part (a) of the inverse problem is easily given: *The necessary and sufficient conditions that $[\Delta_1\Delta_2\Delta_3]$ be the Δ pattern belonging to irrep $[M]$ are: $\Delta_1 + \Delta_2 + \Delta_3 = M_{13} + M_{23} + M_{33}$ and $M_{13} \geq \Delta_i \geq M_{33}$ for each $i = 1, 2, 3$.*

The solution to part (b) of the inverse problem can be solved by direct enumeration. Thus, if $[\Delta]$ is a specified Δ pattern belonging to $[M]$, then the set of operator patterns (3.1) which corresponds to this Δ pattern is obtained by setting $\Gamma_{11} = \Delta_1$ and enumerating all values of Γ_{12} and Γ_{22} such that $\Gamma_{12} + \Gamma_{22} = \Delta_1 + \Delta_2$, and such that the betweenness conditions are not violated.

The counting problem described above gives rise to eight distinct cases corresponding to the eight possible ways of distributing the three integers $\Delta_1, \Delta_2, \Delta_3$ into the two closed (disjoint) intervals $S_1 \equiv [M_{23}, M_{13}]$ and $S_2 \equiv [M_{33}, M_{23} - 1)$, where S_2 is defined to be empty for $M_{23} = M_{33}$. The eight solutions to the eight counting problems are conveniently given by determining the largest and smallest values of Γ_{12} which can occur for the *specified* Δ pattern. Thus, if we define

$$\Gamma_{12}^{>}(\Delta_1\Delta_2\Delta_3) \equiv \Gamma_{12}\,(\max), \; \Gamma_{12}^{<}(\Delta_1\Delta_2\Delta_3) \equiv \Gamma_{12}\,(\min), \tag{3.3}$$

then, the multiplicity $\mathfrak{M}$ can be verified directly to be

$$\mathfrak{M} = \Gamma_{12}^{>}(\Delta_1\Delta_2\Delta_3) - \Gamma_{12}^{<}(\Delta_1\Delta_2\Delta_3) + 1. \tag{3.4}$$

The eight possible cases are given in Table I below.

B. The intertwining number

The determination of the *intertwining number* $\mathscr{I}$ may be accomplished by various methods (e.g., Little-

TABLE I. The multiplicity of a Δ pattern. $[M_{13}M_{23}M_{33}]$ denotes an arbitrary set of irrep labels. $[\Delta_1\Delta_2\Delta_3]$ denotes any specified triplet of integers satisfying $\Delta_1+\Delta_2+\Delta_3=M_{13}+M_{23}+M_{33}$ and either $\Delta_i \in S_1=[M_{23},M_{13}]$ or $\Delta_i \subset S_2=[M_{33},M_{23}-1]$. Depending on the eight possible distributions of $\Delta_1,\Delta_2,\Delta_3$ into the two intervals S_1 and S_2 given in columns 1–3, there is defined a set of operator patterns having irrep labels $[M_{13}M_{23}M_{33}]$ and Δ pattern $[\Delta_1\Delta_2\Delta_3]$. The values of Γ_{12} which occur in these operator patterns range from the largest value given in column 4 to the smallest value given in column 5. The total number of operator patterns having irrep labels $[M_{13}M_{23}M_{33}]$ and the specified Δ pattern is the multiplicity $\mathfrak{M}$ given in column 6.

Δ_1	Δ_2	Δ_3	$\Gamma_{12}^{>}(\Delta_1\Delta_2\Delta_3)$	$\Gamma_{12}^{<}(\Delta_1\Delta_2\Delta_3)$	$\mathfrak{M}$
S_1	S_1	S_1	$\Delta_1+\Delta_2-M_{33}$	$\Delta_1+\Delta_2-M_{23}$	$M_{23}-M_{33}+1$
S_2	S_2	S_2	M_{13}	M_{23}	$M_{13}-M_{23}+1$
S_1	S_2	S_2	M_{13}	Δ_1	$M_{13}-\Delta_1+1$
S_2	S_1	S_1	$\Delta_1+\Delta_2-M_{33}$	Δ_2	$\Delta_1-M_{33}+1$
S_2	S_1	S_2	M_{13}	Δ_2	$M_{13}-\Delta_2+1$
S_1	S_2	S_1	$\Delta_1+\Delta_2-M_{33}$	Δ_1	$\Delta_2-M_{33}+1$
S_2	S_2	S_1	$\Delta_1+\Delta_2-M_{33}$	M_{23}	$M_{13}-\Delta_3+1$
S_1	S_1	S_2	M_{13}	$\Delta_1+\Delta_2-M_{23}$	$\Delta_3-M_{33}+1$

wood's tableau methods[7]), but we choose to use the method of reducing direct products described in detail in Refs. 8 and 9, since the technique is again just a counting problem on Gel'fand patterns. The counting is, however, highly redundant. Nonetheless, a careful analysis of the procedure allows one to deduce the following much simpler statement: *the intertwining number $\mathcal{I}$ belonging to our triple mMm', i.e., the multiplicity of $[m]+[\Delta]$ in $[M]\otimes[m]$, where $[\Delta]$ is a Δ pattern belonging to $[M]$ is given by*

$$\mathcal{I}=\mathfrak{M}_{123}+\mathfrak{M}_{231}+\mathfrak{M}_{312}-\mathfrak{M}_{213}-\mathfrak{M}_{132}-\mathfrak{M}_{321}, \tag{3.5}$$

where $\mathfrak{M}_{ijk}$ is the multiplicity of the Δ pattern

$$[p_{i3}+\Delta_i, p_{j3}+\Delta_j, p_{k3}+\Delta_k]-[p_{13}p_{23}p_{33}] \tag{3.6}$$

belonging to $[M]$.

In the Δ pattern (3.6), the p_{i3} are the partial hooks defined by $p_{i3}=m_{i3}+3-i$, $i=1,2,3$; furthermore, *$\mathfrak{M}_{ijk}$ is defined to be zero whenever the triplet of integers* (3.6) *fails to be a Δ pattern belonging to $[M]$.*

Equation (3.5) is a rather remarkable formula in that it expresses the intertwining number $\mathcal{I}$ directly in terms of the multiplicity of six Δ patterns belonging to the same irrep labels $[M]$.[10] Observe that $\mathfrak{M}_{123}$ is just $\mathfrak{M}$ of Table I; but the remaining $\mathfrak{M}_{ijk}$ depend on the labels $[m]$.

Equation (3.5) is a very useful form for determining the components $[m']$ which appear in the reduction of $[M]\otimes[m]$ for specific numerical assignments of $[M]$ and $[m]$. It may also be used to determine completely the intertwining number $\mathcal{I}$ as a function $\mathcal{I}([M],[m],[\Delta])$ of $[M]$, $[m]$ and $[\Delta]$. The procedure for accomplishing this is described in the following paragraphs.

Consider the determination of $\mathfrak{M}_{213}$. This number is the multiplicity of $[\Delta_2-x_{12},\Delta_1+x_{12},\Delta_3]$ in $[M]$; equivalently, it is the multiplicity of $[\Delta_1+x_{12},\Delta_2-x_{12},\Delta_3]$ in $[M]$, where $[\Delta_1\Delta_2\Delta_3]$ is a specified Δ pattern of $[M]$ having the multiplicity

$$\mathfrak{M}_{123}=\mathfrak{M} \tag{3.7}$$

appearing in Table I. Suppose $(\Delta_1\Delta_2\Delta_3)\in(S_1S_1S_1)$ (line 1 of Table I). Then we can have either $(\Delta_1+x_{12},\Delta_2-x_{12},\Delta_3)\in(S_1S_1S_1)$ or $(\Delta_1+x_{12},\Delta_2-x_{12},\Delta_3)\in(S_1S_2S_1)$. Referring to Table I again, we find

$$\mathfrak{M}_{213}=\begin{cases}\mathfrak{M}_{123} \text{ for } x_{12}^0\le x_{12}\le\beta_1-\Delta_1,\\ (\alpha_1-\Delta_1+1)-x_{12} \quad \text{for } \beta_1-\Delta_1<x_{12}\le\alpha_1-\Delta_1,\\ 0 \text{ for } x_{12}>\alpha_1-\Delta_1,\end{cases} \tag{3.8}$$

where for brevity the following notations have been introduced:

$$\alpha_1\equiv\Gamma_{12}^{>}(\Delta_1\Delta_2\Delta_3),\quad \beta_1\equiv\Gamma_{12}^{<}(\Delta_1\Delta_2\Delta_3),\quad x_{12}^0=\max(1,\Delta_2-\Delta_1+1). \tag{3.9}$$

We next progress through Table I, considering for each distribution of $(\Delta_1\Delta_2\Delta_3)$ into S_1 and S_2, all possible distributions of $(\Delta_1+x_{12},\Delta_2-x_{12},\Delta_3)$ into S_1 and S_2. (Certain cases violate the lexical conditions $x_{12}\ge x_{12}^0$ and may be discarded.) The result is: *Equation* (3.8) *obtains in each instance, i.e., when the restrictions on x_{12} are expressed in terms of the numbers α_1 and β_1 (the numbers appearing in Table I), then $\mathfrak{M}_{213}$ assumes the single form for any distribution of $(\Delta_1\Delta_2\Delta_3)$ into S_1 and S_2.*

Recognizing that $\mathfrak{M}_{132}$ is the multiplicity of $[\Delta_2+x_{23},\Delta_3-x_{23},\Delta_1]$ in $[M]$, we obtain the following equation for $\mathfrak{M}_{132}$ from Eq. (3.8) by letting $[\Delta_1\Delta_2\Delta_3]\to[\Delta_2\Delta_3\Delta_1]$ and $x_{12}\to x_{23}$:

$$\mathfrak{M}_{132}=\begin{cases}\mathfrak{M}_{123} \text{ for } x_{23}^0\le x_{23}\le\beta_2-\Delta_2,\\ (\alpha_2-\Delta_2+1)-x_{23} \quad \text{for } \beta_2-\Delta_2<x_{23}\le\alpha_2-\Delta_2,\\ 0 \text{ for } x_{23}>\alpha_2-\Delta_2,\end{cases} \tag{3.10}$$

where

$$\alpha_2\equiv\Gamma_{12}^{>}(\Delta_2\Delta_3\Delta_1),\quad \beta_2\equiv\Gamma_{12}^{<}(\Delta_2\Delta_3\Delta_1),\quad x_{23}^0\equiv\max(1,\Delta_3-\Delta_2+1). \tag{3.11}$$

Similarly, we obtain

$$\mathfrak{M}_{321}=\begin{cases}\mathfrak{M}_{123} \text{ for } x_{13}^0\le x_{13}\le\beta_3-\Delta_1,\\ (\alpha_3-\Delta_1+1)-x_{13} \quad \text{for } \beta_3-\Delta_1<x_{13}\le\alpha_3-\Delta_1,\\ 0 \text{ for } x_{13}>\alpha_3-\Delta_1,\end{cases} \tag{3.12}$$

where

$$\alpha_3\equiv\Gamma_{12}^{>}(\Delta_1\Delta_3\Delta_2),\quad \beta_3\equiv\Gamma_{12}^{<}(\Delta_1\Delta_3\Delta_2),\quad x_{13}^0\equiv x_{12}^0+x_{23}^0=\max(2,\Delta_3-\Delta_1+2). \tag{3.13}$$

One should be very careful to note that while certain permutations on $[\Delta_1\Delta_2\Delta_3]$ have been utilized in obtaining Eqs. (3.10) and (3.12), all three sets of formulas, Eqs. (3.8), (3.10) and (3.12), refer to a *common specified* $[\Delta]=[\Delta_1\Delta_2\Delta_3]$. In forming the sum $-\mathfrak{M}_{213}-\mathfrak{M}_{132}-\mathfrak{M}_{321}$, for example, there are eight distinct results corresponding to the eight distributions of

$[\Delta_1\Delta_2\Delta_3]$ into S_1 and S_2. The most direct way to form these sums would be to write out two more tables from Table I (one for $[\Delta_1\Delta_2\Delta_3] \to [\Delta_2\Delta_3\Delta_1]$ and one for $[\Delta_1\Delta_2\Delta_3] \to [\Delta_1\Delta_3\Delta_2]$) in which one then rearranges the first three columns to the order $\Delta_1, \Delta_2, \Delta_3$ followed by a rearrangement of rows of the full table such that the first three columns of all three tables agree. The rows of columns four and five of the three tables then list, respectively, the values $\alpha_1\beta_1, \alpha_2\beta_2$, and $\alpha_3\beta_3$ which are relevant to forming the sum $-\mathfrak{M}_{213} - \mathfrak{M}_{132} - \mathfrak{M}_{321}$ for each specified $[\Delta_1\Delta_2\Delta_3]$. (The last column is, of course, the same in all three tables, i.e., $\mathfrak{M}_{123} = \mathfrak{M} = \alpha_1 - \beta_1 + 1 = \alpha_2 - \beta_2 + 1 = \alpha_3 - \beta_3 + 1$). Actually, it is not necessary to form directly the aforementioned sum; but the fact that one must refer the various $\mathfrak{M}_{ijk}$ to a common $[\Delta] = [\Delta_1\Delta_2\Delta_3]$ before combining must be carefully observed.

The determination of $\mathfrak{M}_{231}$ proceeds along similar lines: $\mathfrak{M}_{231}$ is the multiplicity of $[\Delta_1 + x_{13}, \Delta_2 - x_{12}, \Delta_3 - x_{23}]$ in $[M]$. By examining all distributions of $[\Delta_1\Delta_2\Delta_3]$ into S_1 and S_2 and all corresponding distributions of $[\Delta_1 + x_{13}, \Delta_2 - x_{12}, \Delta_3 - x_{23}]$ into S_1 and S_2 which do not violate the lexical conditions $x_{12} \geq x_{12}^0, x_{23} \geq x_{23}^0, x_{13} \geq x_{13}^0$, we have been able to derive the required explicit formulas. In order to give a concise description of the results, let us refer to the following restrictions on x_{12}, x_{23}, and x_{13} as conditions I and II, respectively.

The following restrictions hold simultaneously:

$$\begin{aligned} &x_{13}^0 \leq x_{13} \leq (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2),\\ &x_{12}^0 \leq x_{12} \leq \alpha_1 - \Delta_1,\\ &x_{23}^0 \leq x_{23} \leq \alpha_2 - \Delta_2. \end{aligned} \tag{I}$$

At *least one* of the following restrictions holds:

$$\begin{aligned} &x_{12} > \alpha_1 - \Delta_1, \quad x_{23} > \alpha_2 - \Delta_2,\\ &x_{13} > (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2). \end{aligned} \tag{II}$$

Then

$$\mathfrak{M}_{231} = \begin{cases} \mathfrak{M}_{321} & \text{for } \Delta_2 \in S_2 \\ 0 & \text{for } \Delta_2 \in S_1 \text{ and conditions II.} \end{cases} \tag{3.14}$$

$$\mathfrak{M}_{231} = -\mathfrak{M}_{123} + \mathfrak{M}_{213} + \mathfrak{M}_{132} \quad \text{for } \Delta_2 \in S_1 \text{ and conditions I.} \tag{3.15}$$

A similar procedure yields

$$\mathfrak{M}_{312} = \begin{cases} \mathfrak{M}_{321} & \text{for } \Delta_2 \in S_1, \\ 0 & \text{for } \Delta_2 \in S_2 \text{ and conditions II,} \end{cases} \tag{3.16}$$

$$\mathfrak{M}_{312} = -\mathfrak{M}_{123} + \mathfrak{M}_{213} + \mathfrak{M}_{132} \quad \text{for } \Delta_2 \in S_2 \text{ and conditions I.} \tag{3.17}$$

Equations (3.15) and (3.17) can be replaced by the single relation

$$\mathfrak{M}_{123} + \mathfrak{M}_{231} + \mathfrak{M}_{312} = \mathfrak{M}_{213} + \mathfrak{M}_{132} + \mathfrak{M}_{321} \quad \text{for conditions I,} \tag{3.18}$$

since for $\Delta_2 \in S_1$ it reduces to Eq. (3.15), and for $\Delta_2 \in S_2$ it reduces to Eq. (3.17).

The five relations, Eqs. (3.8), (3.10), (3.14), (3.16), and (3.18), yield the complete determination of the intertwining number $\mathcal{I}$ of Eq. (3.5). (The *explicit* form of $\mathfrak{M}_{321}$ given by Eq. (3.12) never enters into the calculation of $\mathcal{I}$—it is always canceled by the $\mathfrak{M}_{321}$ piece of $\mathfrak{M}_{231}$ or $\mathfrak{M}_{312}$.) For example, consider $x_{12} > \alpha_1 - \Delta_1$ and $x_{23} > \alpha_2 - \Delta_2$. Then $\mathfrak{M}_{213} = 0$, $\mathfrak{M}_{132} = 0$; $\mathfrak{M}_{231} = 0$ for $\Delta_2 \in S_1$; $\mathfrak{M}_{231} = \mathfrak{M}_{321}$ for $\Delta_2 \in S_2$; $\mathfrak{M}_{312} = \mathfrak{M}_{321}$ for $\Delta_2 \in S_1$; $\mathfrak{M}_{312} = 0$ for $\Delta_2 \in S_2$. Thus,

$$\mathcal{I} = \mathfrak{M}_{123} + \mathfrak{M}_{312} - \mathfrak{M}_{321} = \mathfrak{M}_{123} \quad \text{for} \quad \Delta_2 \in S_1,$$
$$\mathcal{I} = \mathfrak{M}_{123} + \mathfrak{M}_{231} - \mathfrak{M}_{321} = \mathfrak{M}_{123} \quad \text{for} \quad \Delta_2 \in S_2,$$

that is, $\mathcal{I} = \mathfrak{M}_{123}$. Continuing in this manner, we obtain the following explicit set of values of the intertwining number $\mathcal{I}$:

$$\mathcal{I} = \mathfrak{M}_{123} \text{ for } x_{12} > \alpha_1 - \Delta_1,\ x_{23} > \alpha_2 - \Delta_2; \tag{3.19a}$$

$$\mathcal{I} = \mathfrak{M}_{123} - [(\alpha_1 - \Delta_1 + 1) - x_{12}] \quad \text{for } \beta_1 - \Delta_1 < x_{12} \leq \alpha_1 - \Delta_1,\ x_{23} > \alpha_2 - \Delta_2; \tag{3.19b}$$

$$\mathcal{I} = \mathfrak{M}_{123} - [(\alpha_2 - \Delta_2 + 1) - x_{23}] \quad \text{for } x_{12} > \alpha_1 - \Delta_1,\ \beta_2 - \Delta_2 < x_{23} \leq \alpha_2 - \Delta_2; \tag{3.19c}$$

$$\begin{aligned} \mathcal{I} = \mathfrak{M}_{123} &- [(\alpha_1 - \Delta_1 + 1) + (\alpha_2 - \Delta_2 + 1) - x_{13}] \\ &\text{for } x_{12} \leq \alpha_1 - \Delta_1,\ x_{23} \leq \alpha_2 - \Delta_2,\\ &x_{13} \geq (\beta_1 - \Delta_1 + 1) + (\alpha_2 - \Delta_2 + 1) \\ &= (\alpha_1 - \Delta_1 + 1) + (\beta_2 - \Delta_2 + 1); \end{aligned} \tag{3.19d}$$

$\mathcal{I} = 0$ if at least one of the following conditions obtains:

$$\begin{aligned} &\text{(a)}\ x_{12} \leq \beta_1 - \Delta_1, \quad \text{(b)}\ x_{23} \leq \beta_2 - \Delta_2,\\ &\text{(c)}\ x_{13} \leq (\beta_1 - \Delta_1 + 1) + (\alpha_2 - \Delta_2) \\ &\qquad = (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2 + 1). \end{aligned} \tag{3.19e}$$

All $\mathcal{I} = 0$ cases are included in the last equation (for lexical labels $[m]$ and $[m] + [\Delta]$).

In obtaining the last two results above, one must take careful note of the implications of the conditions: For example, $x_{12} \leq \alpha_1 - \Delta_1$, $x_{23} \leq \alpha_2 - \Delta_2$, $x_{13} \geq (\alpha_1 - \Delta_1 + 1) + (\beta_2 - \Delta_2 + 1)$ imply, in fact, that $\beta_1 - \Delta_1 < x_{12} \leq \alpha_1 - \Delta_1$, $\beta_2 - \Delta_2 < x_{23} \leq \alpha_2 - \Delta_2$, $x_{13} \geq (\alpha_1 - \Delta_1 + 1) + (\beta_2 - \Delta_2 + 1)$. The expression (3.19d) given for $\mathcal{I}$ then obtains upon combining Eqs. (3.8), (3.10), (3.14), (3.16), and (3.18). Note that the same result would obtain for $x_{13} = (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2 + 1)$, but this gives $\mathcal{I} = 0$, and this zero has been included in the last equation.

To demonstrate that the $\mathcal{I} = 0$ equation is correct, one must show that these zeroes are *precisely* the ones which obtain from Eqs. (3.8), (3.10), (3.14), (3.16), and (3.18). All possible nonzero values of $\mathcal{I}$ (for lexical labels $[m]$ and $[m] + [\Delta]$) are already given by the first four relations of Eqs. (3.19). This implies that we can *only* get the value zero for $\mathcal{I}$ (for lexical labels) if at least one of the conditions $x_{12} \leq \beta_1 - \Delta_1$, $x_{23} \leq \beta_2 - \Delta_2$, $x_{13} \leq (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2 + 1)$ obtains (since all other possibilities are contained in the first four relations). From Eqs. (3.8) and (3.10), respectively, we obtain $\mathcal{I} = 0$ for either $x_{12} \le \beta_1 - \Delta_1$, $x_{23} > \alpha_2 - \Delta_2$ or $x_{12} > \alpha_1 - \Delta_1$, $x_{23} \le \beta_2 - \Delta_2$. Furthermore, all other possible zeroes are now subsumed under the single condition $x_{13} \le (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2 + 1)$. We have already noted that $\mathcal{I} = 0$ for $x_{13} = (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2 + 1)$. For $x_{13} \le (\alpha_1 - \Delta_1) + (\beta_2 - \Delta_2) = (\beta_1 - \Delta_1) + (\alpha_2 - \Delta_2)$, we have the following possibilities: $x_{12} > \alpha_1 - \Delta_1$, $x_{23} < \beta_2 - \Delta_2$; $x_{12} < \beta_1 - \Delta_1$, $x_{23} > \alpha_2 - \Delta_2$; $x_{12} \le \alpha_1 - \Delta_1$, $x_{23} \le \alpha_2 - \Delta_2$. But each of these possibilities yields $\mathcal{I} = 0$. The conclusion is $\mathcal{I} = 0$ if and only if at least one of the conditions stated in Eq. (3.19e) holds.

We need to note one final property of $\mathcal{I}$ before Eqs. (3.19) are complete: *If any condition on the x_{ij} for a particular branch of $\mathcal{I}$ fails to be satisfied in consequence of imposing the lexical conditions $x_{ij} \ge x_{ij}^0$, then $\mathcal{I}$ has value zero on that branch.*

C. The intertwining number-null space diagram

The derivation of the algebraic expressions (3.19) for the intertwining number has been quite detailed and intricate. It is therefore quite satisfying to observe that the results expressed by Eqs. (3.19) assume a very elegant form when represented geometrically in the Möbius plane: Each point in the plane having integral coordinates

$$(x_1 x_2 x_3) = (x_{23} x_{31} x_{12}) \tag{3.20}$$

has associated with it an intertwining number. The three points

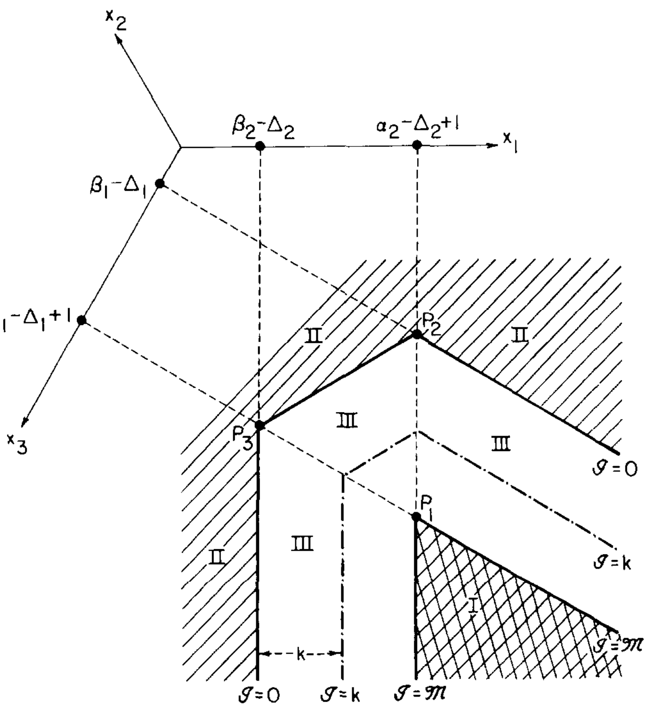


FIG. 1. The intertwining number-null space diagram. The intertwining number $\mathcal{I}$ is defined at each lattice point (points having integral coordinates) of the Möbius plane. At each lattice point in the cross-hatched region I, including the bent solid line, the value of $\mathcal{I}$ is $\mathfrak{M}$; at each lattice point in the shaded region II, including the bent solid line, the value of $\mathcal{I}$ is zero; at each lattice point in the region III between the two bent solid lines, the value of $\mathcal{I}$ is $1, 2, \ldots, \mathfrak{M} - 1$, its value being k at the lattice points on the bent dash-dot line designated by $\mathcal{I} = k$.

$$\begin{aligned} P_1 &= (\alpha_2 - \Delta_2 + 1, -\alpha_1 - \alpha_2 + \Delta_1 + \Delta_2 - 2, \\ &\quad \alpha_1 - \Delta_1 + 1), \\ P_2 &= (\alpha_2 - \Delta_2 + 1, -\alpha_2 - \beta_1 + \Delta_1 + \Delta_2 - 1, \\ &\quad \beta_1 - \Delta_1), \\ P_3 &= (\beta_2 - \Delta_2, -\alpha_1 - \beta_2 + \Delta_1 + \Delta_2 - 1, \\ &\quad \alpha_1 - \Delta_1 + 1) \end{aligned} \tag{3.21}$$

define the vertices of an equilateral triangle (see Fig. 1). This triangle defines a partitioning of the plane into three disjoint sets:

I. The set of lattice points on the solid line designated by $\mathcal{I} = \mathfrak{M}$ in Fig. 1 and all lattice points in the pie-shaped region for which the line $\mathcal{I} = \mathfrak{M}$ is the boundary (the cross-hatched region).

II. The set of lattice points on the solid line designated by $\mathcal{I} = 0$ in Fig. 1 and all lattice points in the shaded region.

III. The set of lattice points lying between the solid lines $\mathcal{I} = 0$ and $\mathcal{I} = \mathfrak{M}$ of Fig. 1.

On the set I, the intertwining number has value $\mathfrak{M}$; on the set II, it has value zero; and on the set III, it has a value which ranges from 1 to $\mathfrak{M} - 1$, its value being $k (1 \le k \le \mathfrak{M} - 1)$ on those lattice points joined by the dash-dot line designated by $\mathcal{I} = k$ in Fig. 1.

We propose to call this diagram the Intertwining Number-Null Space Diagram. This dual nature of the diagram is discussed in detail subsequently.

The Intertwining Number-Null Space Diagram assigns an intertwining number to each point of the Möbius plane. However, only a portion of the diagram corresponds to the actual problem of determining the number of times $[m] + [\Delta]$ occurs in the reduction of $[M] \otimes [m]$, since, *by definition*, $\mathcal{I}$ is zero whenever $[m] + [\Delta]$ is nonlexical, i.e., fails to satisfy $m_{13} + \Delta_1 \ge m_{23} + \Delta_2 \ge m_{33} + \Delta_3$. The lexical conditions are

$$x_1 \ge x_{23}^0, \quad x_2 \le -x_{13}^0, \quad x_3 \ge x_{12}^0, \tag{3.22}$$

where the numbers x_{ij}^0 are defined by Eqs. (3.9), (3.11), and (3.13), respectively. We call the point

$$P_0 = (x_{23}^0, -x_{13}^0, x_{12}^0) \tag{3.23}$$

the *lexical point* of the diagram. The *lexical region* of the diagram is then the set of lattice points which satisfy the lexical conditions (3.22).

The lexical point P_0 can be *one* of six possible points depending on $[\Delta]$: It is $(1, -2, 1)$ for $\Delta_1 \ge \Delta_2 \ge \Delta_3$; $(1, -\Delta_2 + \Delta_1 - 2, \Delta_2 - \Delta_1 + 1)$ for $\Delta_2 \ge \Delta_3 \ge \Delta_1$; $(\Delta_3 - \Delta_2 + 1, -\Delta_3 + \Delta_2 - 2, 1)$ for $\Delta_3 \ge \Delta_1 \ge \Delta_2$; $(1, -\Delta_2 + \Delta_1 - 2, \Delta_2 - \Delta_1 + 1)$ for $\Delta_2 \ge \Delta_1 \ge \Delta_3$; $(\Delta_3 - \Delta_2 + 1, -\Delta_3 + \Delta_2 - 2, 1)$ for $\Delta_1 \ge \Delta_3 \ge \Delta_2$; or $(\Delta_3 - \Delta_2 + 1, \Delta_1 - \Delta_3 - 2, \Delta_2 - \Delta_1 + 1)$ for $\Delta_3 \ge \Delta_2 \ge \Delta_1$.

For example, the lexical point for the Intertwining Number-Null Space Diagram for $[m] + [q\ q\ q]$ contained in $[2q\ q\ 0] \otimes [m]$ is $(1, -2, 1)$. The points P_1, P_2, P_3 become, respectively,

$$\begin{aligned} P_1 &= (q + 1, -2q - 2, q + 1), \\ P_2 &= (q + 1, -q - 1, 0), \\ P_3 &= (0, -q - 1, q + 1). \end{aligned}$$

TABLE II. Coordinate points of the intertwining number-null space diagram.

Δ_1	Δ_2	Δ_3	$\alpha_1 - \Delta_1 + 1$	$\beta_1 - \Delta_1$	$\alpha_2 - \Delta_2 + 1$	$\beta_2 - \Delta_2$
S_1	S_1	S_1	$\Delta_2 - M_{33} + 1$	$\Delta_2 - M_{23}$	$\Delta_3 - M_{33} + 1$	$\Delta_3 - M_{23}$
S_2	S_2	S_2	$M_{13} - \Delta_1 + 1$	$M_{23} - \Delta_1$	$M_{13} - \Delta_2 + 1$	$M_{23} - \Delta_2$
S_1	S_2	S_2	$M_{13} - \Delta_1 + 1$	0	$\Delta_3 - M_{33} + 1$	$M_{23} - \Delta_2$
S_2	S_1	S_1	$\Delta_2 - M_{33} + 1$	$\Delta_2 - \Delta_1$	$M_{13} - \Delta_2 + 1$	$\Delta_3 - M_{23}$
S_2	S_1	S_2	$M_{13} - \Delta_1 + 1$	$\Delta_2 - \Delta_1$	$M_{13} - \Delta_2 + 1$	0
S_1	S_2	S_1	$\Delta_2 - M_{33} + 1$	0	$\Delta_3 - M_{33} + 1$	$\Delta_3 - \Delta_2$
S_2	S_2	S_1	$\Delta_2 - M_{33} + 1$	$M_{23} - \Delta_1$	$M_{13} - \Delta_2 + 1$	$\Delta_3 - \Delta_2$
S_1	S_1	S_2	$M_{13} - \Delta_1 + 1$	$\Delta_2 - M_{23}$	$\Delta_3 - M_{33} + 1$	0

Thus, in the lexical region of the diagram, we have $\mathcal{I} = 0$ whenever $-q - 1 \leq x_2 \leq -2$.

Since the four numbers $\alpha_1 - \Delta_1 + 1$, $\beta_1 - \Delta_1$, $\alpha_2 - \Delta_2 + 1$, and $\beta_2 - \Delta_2$ play a crucial role in the Intertwining Number-Null Space Diagram, it is convenient to give the explicit tabulation of them in Table II (these results are read off directly from Table I).

D. Determination of the null spaces from the properties of the intertwining number

Let us now discuss the dual nature of the Intertwining Number-Null Space Diagram, i.e., we wish to justify the appellation "null space."

Consider the set of $\mathfrak{M}$ Wigner operators of irrep labels $[M]$ which belong to the multiplicity set having a prescribed Δ pattern $[\Delta]$. The operators in this set are enumerated by operator patterns of the type (3.1). Let us denote these $\mathfrak{M}$ operator patterns by (Γ_1), $(\Gamma_2), \ldots, (\Gamma_{\mathfrak{M}})$, making, however, no specific assignment of the (Γ_k) to the patterns. Thus, the set of Wigner operators under consideration is

$$\left\{ \left\langle \begin{matrix} (\Gamma_k) \\ [M] \\ \cdot \end{matrix} \right\rangle : [\Delta(\Gamma_k)] = [\Delta] \right\}. \tag{3.24}$$

More generally, we do not even consider the (Γ_k) to be operator patterns, but rather only *symbols* which enumerate a set of *orthogonal unit tensor operators*, each of which effects the mapping $[m] \rightarrow [m] + [\Delta]$ of a generic irrep space $[m]$.

The coefficients

$$\left\langle \begin{pmatrix} [m] + [\Delta] \\ (m') \end{pmatrix} \middle| \left\langle \begin{matrix} (\Gamma_k) \\ [M] \\ (m) \end{matrix} \right\rangle \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle \tag{3.25}$$

are then the coupling coefficients such that the vectors defined by (cf. Ref. 9)

$$\left| \begin{pmatrix} [m] + [\Delta] \\ (m') \end{pmatrix} ; (\Gamma_k) \right\rangle = \sum_{(M),(m)} \left\langle \begin{pmatrix} [m] + [\Delta] \\ (m') \end{pmatrix} \middle| \left\langle \begin{matrix} (\Gamma_k) \\ [M] \\ (M) \end{matrix} \right\rangle \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle \times \left| \begin{pmatrix} [M] \\ (M) \end{pmatrix} \right\rangle_2 \left| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle_1 \tag{3.26}$$

have the following properties: For a *specified k*, $1 \leq k \leq \mathfrak{M}$, *either* the coupled vectors are *orthornormal* in the labels (m'), in which case the vectors are a basis of a carrier space of irrep $[m] + [\Delta]$ of $U(3)$, *or* each vector corresponding to any Gel'fand pattern (m') is the zero vector. Furthermore, the carrier spaces corresponding to distinct values of k are perpendicular.

Consider next the implications of the Intertwining Number-Null Space Diagram. If the labels $[m]$ belong to the region R of the Möbius plane[11] for which $\mathcal{I} = \mathfrak{M}$, then Eq. (3.26) must provide us with precisely $\mathfrak{M}$ perpendicular carrier spaces of irrep $[m] + (\Delta)$, i.e., one for each k.

This implies that

$$\left\langle \begin{matrix} (\Gamma_k) \\ [M] \\ (M) \end{matrix} \right\rangle \begin{matrix} [m] \\ (m) \end{matrix} \Bigg\rangle \neq 0, \quad k = 1, 2, \ldots, \mathfrak{M}, [m] \in R, \tag{3.27}$$

for all (M), (m). This is just the statement that there exists $\mathfrak{M}$ Wigner operators.[8] But now consider the set L_1 of irrep labels $\{[m]\}$ such that (x_{23}, x_{31}, x_{12}) is a point on the line for which $\mathcal{I} = \mathfrak{M} - 1$. Then *precisely one* Wigner operator, call it (Γ_1), must have the property

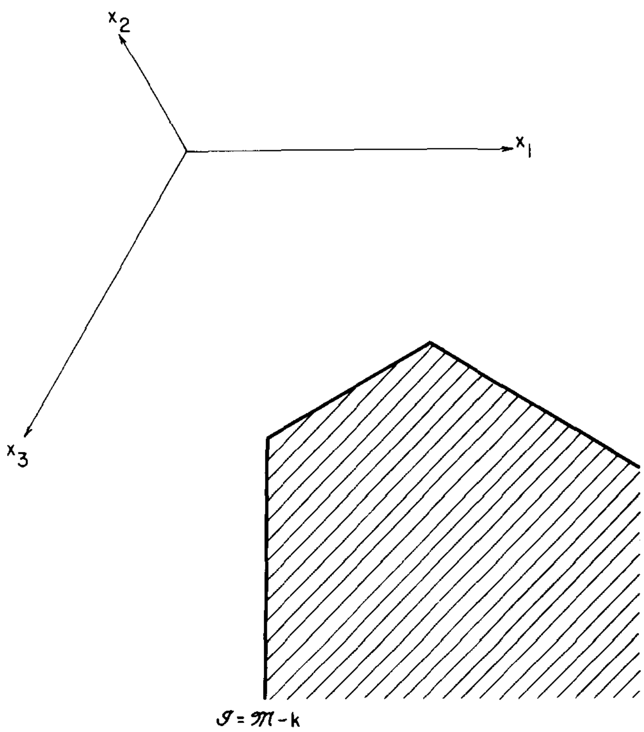


FIG. 2. The null space of the $U(3)$ Wigner operator (Γ_k). Lattice points on the bent solid line $\mathcal{I} = \mathfrak{M} - k$ and exterior to the shaded region define the set of irrep labels $\{[m]\}$ which belong to the null space $\mathfrak{N}_k$ of the Wigner operator designated by (Γ_k). The exact positioning of the solid line is determined from Fig. 1.

$$\left\langle \begin{matrix} (\Gamma_1) \\ [M] \\ (M) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_1, \tag{3.28}$$

for all (M) and (m), for otherwise, we would obtains $\mathfrak{M}$ perpendicular carrier spaces of irrep $[m] + [\Delta]$, and only $\mathfrak{M} - 1$ such spaces exist. Next, consider the set L_2 of irrep labels $\{[m]\}$ such that (x_{23}, x_{31}, x_{12}) is a point on the line for which $\mathfrak{I} = \mathfrak{M} - 2$. Then *precisely two* Wigner operators must annihilate all irrep spaces having $[m] \in L_2$. These can only be the operator designated by (Γ_1) and one more, call it (Γ_2):

$$\begin{aligned} &\left\langle \begin{matrix} (\Gamma_1) \\ [M] \\ (M) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_1 \cup L_2, \\ &\left\langle \begin{matrix} (\Gamma_2) \\ [M] \\ (M) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_2, \end{aligned} \tag{3.29}$$

for all (M) and (m). We continue in this manner, designating by (Γ_3) the new Wigner operator which must annihilate those irrep space $[m] \in L_3$, where L_3 is the set of labels $\{[m]\}$ such that (x_{23}, x_{31}, x_{12}) is a point on the line for which $\mathfrak{I} = \mathfrak{M} - 3$, i.e.,

$$\begin{aligned} &\left\langle \begin{matrix} (\Gamma_1) \\ [M] \\ (m) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_1 \cup L_2 \cup L_3, \\ &\left\langle \begin{matrix} (\Gamma_2) \\ [M] \\ (M) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_2 \cup L_3, \\ &\left\langle \begin{matrix} (\Gamma_3) \\ [M] \\ (M) \end{matrix} \middle| \begin{pmatrix} [m] \\ (m) \end{pmatrix} \right\rangle = 0, \quad [m] \in L_3. \end{aligned} \tag{3.30}$$

The general conclusion is: Let

$$\mathfrak{N}_k = \mathfrak{N}(\Gamma_k) \tag{3.31}$$

denote the null space of the Wigner operator designated by (Γ_k) in the above enumeration. Then the Möbius plane of the Intertwining Number-Null Space Diagram is separated into two regions by the line on which $\mathfrak{I} = \mathfrak{M} - k$, as shown in Fig. 2. The nesting property of these null space is obvious:

$$\mathfrak{N}_1 \supset \mathfrak{N}_2 \supset \cdots \supset \mathfrak{N}_{\mathfrak{M}}. \tag{3.32}$$

4. PROOF THAT THE NULL SPACE $\mathfrak{N}(\Gamma_S)$ IS MAXIMAL

Sections 2 and 3 have been developed quite independently of one another. The aim of this section is to demonstrate precisely the elegant manner in which the zeroes of the denominator function, Eq. (1.2a), of Sec. 1 fit into the more general scheme of null spaces developed in Sec. 3.

This necessary meshing of structures is a consequence of the next proposition. (Throughout this discussion, we impose the lexical restrictions $x_1 = x_{23} \leq 1$, $x_2 = x_{31} \leq -2$, $x_3 = x_{12} \geq 1$, unless otherwise noted.

Proposition 5: After all linear factors in x_{23}, x_{31}, x_{12} are removed from $G_q(\Delta; x)$, the polynomial which remains vanishes on the lattice points of the boundary and those interior to the equilateral triangle in the Möbius plane which has vertex points as follows:

$$\begin{aligned} P_1' &= (\alpha_2 - \Delta_2, -\alpha_1 - \alpha_2 + \Delta_1 + \Delta_2, \alpha_1 - \Delta_1), \\ P_2' &= (\alpha_2 - \Delta_2, -\alpha_2 - \beta_1 + \Delta_1 + \Delta_2 - 1, \beta_1 - \Delta_1 + 1), \\ P_3' &= (\beta_2 - \Delta_2 + 1, -\alpha_1 - \beta_2 + \Delta_1 + \Delta_2 - 1, \alpha_1 - \Delta_1). \end{aligned} \tag{4.1}$$

The numbers appearing as the coordinates of the points P_k' are obtained from those tabulated in Table II for $[M] = [p\ q\ 0]$. (The x_2 coordinates of P_2' and P_3' agree in consequence of the relation $\alpha_1 - \alpha_2 = \beta_1 - \beta_2$.) *Observe that the triangle $P_1P_2P_3$ of Fig. 1 and the triangle $P_1'P_2'P_3'$ share the common line $P_2'P_3'$, but the remaining two sides of $P_1'P_2'P_3'$ lie one unit interior to $P_1P_2P_3$.*

The proof of Proposition 5 is given by using (Lemma 5) the fact that $G_q(\Delta; x)$ contains the factor

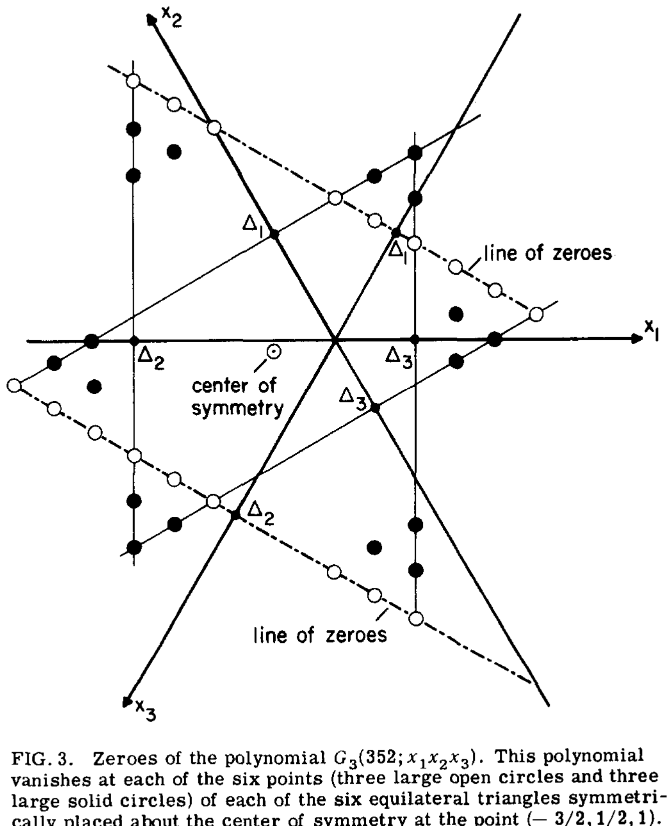


FIG. 3. Zeroes of the polynomial $G_3(352; x_1x_2x_3)$. This polynomial vanishes at each of the six points (three large open circles and three large solid circles) of each of the six equilateral triangles symmetrically placed about the center of symmetry at the point $(-3/2, 1/2, 1)$. This set of points is the set Z of Eq. (2.9) (for $\xi_1 = 3, \xi_2 = 5, \xi_3 = 2$). The linear factors of the polynomial are $(x_3 + 3)(x_3 - 5)$. Hence, the polynomial also vanishes on the lines $x_3 = -3$ and $x_3 = 5$ (the dash-dot lines). Removing these linear factors from the polynomial leaves a new polynomial which still vanishes at each of three points (the large solid circles) of each of six equilateral triangles which are still symmetrically placed about the center of symmetry.

$$\binom{\Delta_1 + x_{12}}{q - \Delta_3} \binom{\Delta_2 - x_{12}}{q - \Delta_3} \quad (4.2)$$

for $0 \le \Delta_3 \le q$. [In general, $G_q(\xi; x)$, for arbitrary ξ, does not contain linear factors; but it may (by Lemma 5) when ξ is particularized to Δ. This implies that $G_q(\Delta; x)$ vanishes not only on the finite set of points Z of the six triangles of Sec. 2D, but also on infinite lines which intersect these triangles in a most interesting way (see Fig. 3).] Using the basic result (4.2), together with Lemma 1, we will verify Proposition 5 for each of the eight cases listed in Table II. Since we are restricting our attention to lexical values of the variables $(x_1 x_2 x_3) = (x_{23} x_{31} x_{12})$, let us note that of the six general triangles on which $G_q(\xi; x)$ vanishes, the one corresponding to the lexical region of the Intertwining Number-Null Space Diagram is the one which has vertices at the points (we now set $\xi_i = \Delta_i$)

$$\begin{aligned} Q_1 &= (\Delta_3, -\Delta_2 - \Delta_3, \Delta_2), \\ Q_2 &= (\Delta_3, -\Delta_2 - \Delta_3 + q - 1, \Delta_2 - q + 1), \quad (4.3) \\ Q_3 &= (\Delta_3 - q + 1, -\Delta_2 - \Delta_3 + q - 1, \Delta_2). \end{aligned}$$

Our procedure is to prove Proposition 5 by verifying explicitly that when the zeroes of the linear factors of $G_q(\Delta; x)$ are removed from the triangle defined by the points $Q_1 Q_2 Q_3$, we are left, in each of the eight possible cases, with the triangle defined by the points $P'_1 P'_2 P'_3$. We give the proof only for two cases, the remaining six cases being established in a similar manner.

(1) $(\Delta_1 \Delta_2 \Delta_3) \in S_1$. From Table II (for $[M] = [p\ q\ 0]$), we find $\alpha_1 - \Delta_1 = \Delta_2$, $\beta_1 - \Delta_1 = \Delta_2 - q$, $\alpha_2 - \Delta_2 = \Delta_3$, and $\beta_2 - \Delta_2 = \Delta_3 - q$. There are no linear factors in $G_q(\Delta; x)$, and we see that the set of points (4.1) becomes the set of points (4.3).

(2) $(\Delta_1 \Delta_2 \Delta_3) \in S_2$. From Table II, we find $\alpha_1 - \Delta_1 = p - \Delta_1$, $\beta_1 - \Delta_1 = q - \Delta_1$, $\alpha_2 - \Delta_2 = p - \Delta_2$, and $\beta_2 - \Delta_2 = q - \Delta_2$. The set of points (4.1) becomes

$$\begin{aligned} P'_1 &= (p - \Delta_2, \Delta_1 + \Delta_2 - 2p, p - \Delta_1), \\ P'_2 &= (p - \Delta_2, \Delta_1 + \Delta_2 - p - q - 1, q - \Delta_1 + 1), \\ P'_3 &= (q - \Delta_2 + 1, \Delta_1 + \Delta_2 - p - q - 1, p - \Delta_1). \end{aligned} \quad (4.4)$$

The linear factors of $G_q(\Delta; x)$ are

$$\binom{\Delta_2 + x_1}{q - \Delta_1}\binom{\Delta_3 - x_1}{q - \Delta_1}\binom{\Delta_3 + x_2}{q - \Delta_2}\binom{\Delta_1 - x_2}{q - \Delta_2} \times \binom{\Delta_1 + x_3}{q - \Delta_3}\binom{\Delta_2 - x_3}{q - \Delta_3}. \quad (4.5)$$

These factors yield a zero in the lexical region of the Möbius plane whenever at *least one* of the following conditions obtains: $p - \Delta_2 + 1 \le x_1 \le \Delta_3$, $p - \Delta_1 + 1 \le -x_2 \le \Delta_3$, $p - \Delta_1 + 1 \le x_3 \le \Delta_2$. Now consider the *intersection* of all points satisfying at least one of these conditions with the triangle defined by the three points $Q_1 Q_2 Q_3$ of Eq. (4.3). Removing this intersection from the triangle $Q_1 Q_2 Q_3$, we are left with precisely the triangle defined by the points $P'_1 P'_2 P'_3$ of Eq. (4.4). Hence, the polynomial factor of $G_q(\Delta; x)$ which remains after removing the factor (4.5) vanishes on the lattice points on the boundary of and interior to the triangle $P'_1 P'_2 P'_3$.

In order to illustrate these remarkable properties of $G_q(\Delta; x)$, we have displayed in Fig. 3 six triangles (ignoring now the lexical conditions) of zeroes (six zeroes in each triangle) of the polynomial $G_3(352; x_1 x_2 x_3)$, i.e., for $q = 3$ and $[\Delta_1 \Delta_2 \Delta_3] = [352]$. In this case, G_3 contains the linear factors $(x_3 + 3)(x_3 - 5)$ given by Eq. (4.2). Observe the remarkable geometrical positioning of these triangles: The line of zeroes $x_3 = 5$ (the lower dot-dash line) includes three zeroes from each of the three lower triangles, while the line of zeroes $x_3 = -3$ (the upper dot-dash line) includes three zeroes from each of the three upper triangles. Furthermore, upon removing these linear factors from G_3, we are left with a polynomial which has zeroes on six equilateral triangle (still symmetrically positioned) of three points each. [The lexical triangle which remains is, of course, just the one defined by the three vertex points of Proposition 5.]

Proposition 5 will now be used to establish a principal result of this paper.

Let us recall that the null space $\mathfrak{N}(\Gamma_s)$ of the Wigner operator labeled (Γ_s) is the set of all irrep spaces with labels $[m]$ such that

$$D\begin{pmatrix} [\Delta_1 \Delta_2 \Delta_3] \\ [p\ \ q\ \ 0] \end{pmatrix}([m]) = 0. \quad (4.6)$$

(This was demonstrated in I.) Since, by assumption, the labels $[m] + [\Delta]$ are lexical, we obtain from Eq. (1.2a) the result

$$\mathfrak{N}(\Gamma_s) = \left\{ \begin{array}{l} \text{All irrep spaces with labels } [m_{13} m_{23} m_{33}]: \\ \prod_{i<j}^{3} \binom{\Delta_i + x_{ij}}{\Delta_i + \Delta_j + 1} \Big/ G_q(\Delta; x) = 0 \end{array} \right\}. \quad (4.7)$$

It is our aim to show that $\mathfrak{N}(\Gamma_s) = \mathfrak{N}_1$, where $\mathfrak{N}_1$ is the maximal null space determined in Sec. 3. To establish this result, we first show that the function appearing in Eq. (4.7) can be written in the form

$$\prod_{i<j}^{3} \binom{\Delta_i + x_{ij}}{\Delta_i + \Delta_j + 1} \Big/ G_q(\Delta; x) = \frac{L_\alpha(\Delta; x)}{F_\alpha(\Delta; x)}, \qquad \alpha = 1, 2, \ldots, 8, \quad (4.8)$$

where $F_\alpha(\Delta; x)$ is that part of $G_q(\Delta; x)$ which remains after separating off the linear terms (Proposition 5). $L_\alpha(\Delta; x)$ is a product of linear factors obtained by combining the linear factors of the left-hand side of Eq. (4.8) with those separated off from G_q (each factor which separates off from G_q is canceled by a corresponding factor in the numerator). The eight ways of writing this result correspond to the eight cases of Table II. While one can list explicitly the eight forms of $L_\alpha(\Delta; x)$ and $F_\alpha(\Delta; x)$, we will not do so; but note instead the essential properties of the right-hand side of Eq. (4.8) which are obtained.

(1) In each instance, the linear factor $L_\alpha(\Delta; x)$ vanishes for each set of labels $[m]$ such that (x) belongs to the null space $\mathfrak{N}_1$, and on no other lexical points.

(2) In each instance, the denominator polynomial $F_\alpha(\Delta; x)$ vanishes (Proposition 5) on each set of labels $[m]$ (a finite number) such that (x) is a lattice point belonging to the triangle defined by the three points $P'_1 P'_2 P'_3$ of Eq. (4.1), and on no other lexical points.

(3) At each lexical point (x) where $F_\alpha(\Delta;x)$ vanishes, precisely two linear factors of the form $x_{12}-a$ and $x_{23}-b$ of $L_\alpha(\Delta;x)$ vanish, and no linear factor in x_{13} vanishes; in each case, $L_\alpha(\Delta;x)/F_\alpha(\Delta;x)=0$.

The above properties are precisely the statement:

Proposition 6: The null space $\mathfrak{N}(\Gamma_s)$ is precisely the maximal null space $\mathfrak{N}_1$.

[Let us also remark that the properties (1)–(3) described above also apply (See Fig. 3) to the *nonlexical* regions of the Möbius plane corresponding to the symmetries of G_q.]

We have now accomplished a major goal: The Wigner operator labeled (Γ_s), the construction of which was uniquely determined by the geometrical properties of the arrow-patterns (since the relevant Racah invariant operators were uniquely determined), is the one in the multiplicity set determined by $[\Delta]$ which has the maximal null space.

There still remains the problem of identifying the operator label (Γ_s) with a definite numerical array, i.e., a definite *operator pattern*. We next describe the method of making this identification.

5. ASSIGNMENT OF THE OPERATOR PATTERN Γ_s

Our procedure for assigning a specific numerical array (operator pattern) to the Wigner operator designated by the symbol (Γ_s) is based on certain limit properties of the Racah functions and the projective operators.[12]

Consider the $U(3)$ Racah functions which effect the following upper operator pattern coupling:

$$\begin{bmatrix} & (\Gamma_s) & \\ [p & q & 0] \\ & (\gamma) & \end{bmatrix} = \begin{bmatrix} & \cdot & \\ [p-q & 0 & 0] \\ & \cdot & \end{bmatrix} \begin{matrix}\{R\}\\ \\ [R]\end{matrix} \begin{bmatrix} & \cdot & \\ [q & q & 0] \\ & \cdot & \end{bmatrix}. \tag{5.1}$$

These Racah coefficients were determined explicitly in Ref. 1. We note again their explicit form:

$$\left\{\begin{pmatrix} [p & q & 0] \\ & (\Gamma_s) & \end{pmatrix}\begin{pmatrix} & (\max) & \\ [p-q & 0 & 0] \\ & (\Gamma') & \end{pmatrix}\begin{pmatrix} [q & q & 0] \\ & (\Gamma'') & \end{pmatrix}\right\}([m]+[\Delta]) = \left(q!\prod_{i=1}^{3}\frac{(\Delta_i)!}{(\Delta_i')!(q-\Delta_i'')!}\right)^{1/2}$$
$$\times\left(\prod_{i<j=1}^{3}\frac{(x_{ij}+\Delta_i''-\Delta_j'')(\Delta_i+\Delta_j+1)!\binom{x_{ij}+\Delta_i}{\Delta_i+\Delta_j+1}}{(\Delta_i'+\Delta_j'+1)!\binom{x_{ij}+\Delta_i-\Delta_j''}{\Delta_i'+\Delta_j'+1}(2q-\Delta_i''-\Delta_j''+1)!\binom{x_{ij}+q-\Delta_j''}{2q-\Delta_i''-\Delta_j''+1}}\right)^{1/2}\frac{1}{[G_q(\Delta;x)]^{1/2}}, \tag{5.2a}$$

where

(a) $[\Delta']=[\Delta(\Gamma')]$, $\quad [\Delta'']=[\Delta(\Gamma'')]$,

(b) $[\Delta]$ is any arbitrarily selected Δ pattern belonging to $[p\ q\ 0]$,

(c) the Δ pattern of the *label*

$$\begin{pmatrix} [p & q & 0] \\ & (\Gamma_s) & \end{pmatrix}$$

is $[\Delta]$,

(d) and the Δ patterns satisfy

$$[\Delta]=[\Delta']+[\Delta'']. \tag{5.2b}$$

Note that since $[\Delta]$ is prescribed, condition (5.2b) is a constraint on the patterns (Γ') and (Γ'').

We next take the limit of Eq. (5.2a) as $m_{33}\to-\infty$. This is easily accomplished upon noting that

$$k!\binom{x+a}{k}\approx x^k \tag{5.3}$$

for fixed a and k and for large positive x.

The factor in Eq. (5.2a) preceding $[G_q(\Delta;x)]^{-1/2}$ assumes the following form for sufficiently large $-m_{33}$:

$$\left(q!\prod_{i=1}^{3}\frac{(\Delta_i)!}{(\Delta_i')!(q-\Delta_i'')!}\right)^{1/2}[x_{12}+\Delta_1''-\Delta_2'']^{1/2}$$
$$\times\left[\frac{(x_{12}+\Delta_1)!(x_{12}+\Delta_1''-\Delta_2-1)!}{(x_{12}-\Delta_2-1)!(x_{12}+\Delta_1-\Delta_2'')!}\times\frac{(x_{12}-q+\Delta_1''-1)!}{(x_{12}+q-\Delta_2'')!}\right]^{1/2}(-m_{33})^{\Delta_3''}. \tag{5.4}$$

To obtain the form of $G_q(\Delta;x)$ for large $-m_{33}$, we use Eq. (2.24) [for $\xi_i=\Delta_i$ and $x_i=x_{jk}(i,j,k$ cyclic)]. For sufficiently large $-m_{33}$, we may write

$$G_q(\Delta_1\Delta_2\Delta_3;x_{23},-x_{13},x_{12})$$
$$\approx q!\sum_{k_3}k_3!(q-k_3)!\binom{\Delta_3}{q-k_3}\binom{x_{12}-\Delta_2-\Delta_3+q-1}{k_3}$$
$$\times\binom{x_{12}+\Delta_1+\Delta_3-q+k_3}{k_3}(-m_{33})^{2q-2k_3}\Delta_1!\Delta_2!$$
$$\times\sum_{k_1+k_2=q-k_3}\frac{1}{k_1!k_2!(\Delta_1-q+k_1)!(\Delta_2-q+k_2)!}$$
$$=q!\sum_{k_3}(k_3!)^3\binom{\Delta_1}{k_3}\binom{\Delta_2}{k_3}\binom{\Delta_3}{q-k_3}\frac{(\Delta_1+\Delta_2-2k_3)!}{(\Delta_1+\Delta_2-q-k_3)!}$$
$$\times\binom{x_{12}-\Delta_2-\Delta_3+q-1}{k_3}\binom{x_{12}+\Delta_1+\Delta_3-q+k_3}{k_3}$$
$$\times(-m_{33})^{2q-2k_3}. \tag{5.5}$$

For $\Delta_3\geqslant q$, the $k_3=0$ term in this summation dominates the others, i.e.,

$$G_q(\Delta_1\Delta_2\Delta_3; x_{23}, -x_{13}, x_{12}) \approx (q!)^2 \binom{\Delta_3}{q}\binom{\Delta_1+\Delta_2}{q} \times (-m_{33})^{2q} \quad (5.6a)$$

for $\Delta_3 \in S_1$ and sufficiently large $-m_{33}$. For $0 \leqslant \Delta_3 \leqslant q$, the term in the summation (5.5) having $k_3 = q - \Delta_3$ dominates the others, i.e.,

$$G_q(\Delta_1\Delta_2\Delta_3; x_{23}, -x_{13}, x_{12}) \approx q!\,[(q-\Delta_3)!]^3 \times \binom{\Delta_1}{q-\Delta_3}\binom{\Delta_2}{q-\Delta_3}\frac{(\Delta_1+\Delta_2+2\Delta_3-2q)!}{(\Delta_1+\Delta_2+\Delta_3-2q)!} \times \binom{x_{12}-\Delta_2-\Delta_3+q-1}{q-\Delta_3}\binom{x_{12}+\Delta_1}{q-\Delta_3}(-m_{33})^{2\Delta_3} \quad (5.6b)$$

for $\Delta_3 \in S_2$ and sufficiently large $-m_{33}$.

Combining Eqs. (5.4) and (5.6), we obtain the following two explicit forms for the limit of the Racah coefficient (5.2a):

$$\lim_{m_{33}\to-\infty}\left\{\begin{pmatrix}[p & q & 0]\\ & (\Gamma_s) & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right\}([m]+[\Delta])$$
$$= \delta_{q,\Delta_3''}\left(\frac{(\Delta_1)!(\Delta_2)!}{(\Delta_1')!(\Delta_2')!(\Delta_1'')!(\Delta_2'')!}\;\frac{(\Delta_1'+\Delta_2')!(\Delta_1''+\Delta_2'')!}{(\Delta_1+\Delta_2)!} \times \frac{(x_{12}+\Delta_1''-\Delta_2'')(x_{12}+\Delta_1)!(x_{12}+\Delta_1''-\Delta_2-1)!(x_{12}-\Delta_2''-1)!}{(x_{12}-\Delta_2-1)!(x_{12}+\Delta_1-\Delta_2'')!(x_{12}+\Delta_1'')!}\right)^{1/2} \quad (5.7a)$$

for $\Delta_3 \in S_1$;

$$\lim_{m_{33}\to-\infty}\left\{\begin{pmatrix}[p & q & 0]\\ & (\Gamma_s) & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right\}([m]+[\Delta])$$
$$= \delta_{0,\Delta_3'}\left(\frac{(p-q)!(p-\Delta_1)!(p-\Delta_2)!\Delta_3!}{(p-q+\Delta_3)!(\Delta_1')!(\Delta_2')!(q-\Delta_1'')!(q-\Delta_2'')!} \times \frac{(x_{12}+\Delta_1''-\Delta_2'')(x_{12}+p-\Delta_2)!(x_{12}+\Delta_1''-\Delta_2-1)!(x_{12}+\Delta_1''-q-1)!}{(x_{12}+\Delta_1-\Delta_2'')!(x_{12}+\Delta_1-p-1)!(x_{12}+q-\Delta_2'')!}\right)^{1/2} \quad (5.7b)$$

for $\Delta_3 \in S_2$.

We have given a detailed derivation of the limits, Eqs. (5.7a) and (5.7b), because of the considerable importance of these limits for inducing upper operator patterns.[12] The significance of the *explicit forms* (5.7a) and (5.7b) becomes apparent upon recognizing that the *right-hand sides of these equations are square-bracket coefficients of definite labels*. Namely, Eqs. (5.7a) and (5.7b) are expressions of the following relations:

$$\lim_{m_{33}\to-\infty}\left\{\begin{pmatrix}[p & q & 0]\\ & (\Gamma_s) & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right\}([m]+[\Delta])$$
$$= \left[\begin{pmatrix}p & & q & & 0\\ & \Delta_1+\Delta_2 & & 0 & \\ & & \Delta_1 & & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right](m_{13}+\Delta_1, m_{23}+\Delta_2) \quad (5.8a)$$

for $\Delta_3 \in S_1$;

$$\lim_{m_{33}\to-\infty}\left\{\begin{pmatrix}[p & q & 0]\\ & (\Gamma_s) & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right\}([m]+[\Delta])$$
$$= \left[\begin{pmatrix}p & & q & & 0\\ & p & & \Delta_1+\Delta_2-p & \\ & & \Delta_1 & & \end{pmatrix}\overset{(\max)}{\begin{pmatrix}[p-q & 0 & 0]\\ & (\Gamma') & \end{pmatrix}}\begin{pmatrix}[q & q & 0]\\ & (\Gamma'') & \end{pmatrix}\right](m_{13}+\Delta_1, m_{23}+\Delta_2) \quad (5.8b)$$

for $\Delta_3 \in S_2$. The proofs of Eqs. (5.8a) and (5.8b) are given by evaluating the $U(3)$ square-bracket coefficients *directly* from their definitions [Eqs. (2.4b) and (2.4c) of Ref. 1] in terms of the *known* reduced matrix elements of the $[p-q\ \ 0\ \ 0]$ projective operators and the *known* $U(2)$ Racah coefficients, thus showing directly the equality of these square-bracket coefficients to the right-hand sides of Eqs. (5.7a) and (5.7b), respectively. [For the relation of the notation for $U(2)$ Racah coefficients used here to Racah's W notation, see Eq. (4.12) of Ref. 9.]

Properties (5.8a) and (5.8b) may now be used directly to assign unambiguously the following operator pattern to (Γ_s):

$$\begin{pmatrix} p & q & 0 \\ & (\Gamma_s) & \end{pmatrix} = \begin{pmatrix} p & & q & & 0 \\ & \Delta_1+\Delta_2 & & 0 & \\ & & \Delta_1 & & \end{pmatrix} \qquad (5.9a)$$

for $\Delta_3 \in S_1$;

$$\begin{pmatrix} p & q & 0 \\ & (\Gamma_s) & \end{pmatrix} = \begin{pmatrix} p & & q & & 0 \\ & p & & \Delta_1+\Delta_2-p & \\ & & \Delta_1 & & \end{pmatrix} \qquad (5.9b)$$

for $\Delta_3 \in S_2$.

Observe from Table I that these operator patterns correspond to

$$\Gamma_{12} = \Gamma_{12}^{>}(\Delta_1\Delta_2\Delta_3), \qquad (5.10)$$

that is, the "stretched" pattern (Γ_s) is the one having the difference $\Gamma_{12} - \Gamma_{22}$ equal to the maximal value which is compatible with the prescribed Δ pattern.[2]

An alternative procedure for inducing the operator pattern assignment of (Γ_s) uses the coupling law (5.1) and properties (5.8a) and (5.8b). One can now prove directly the following limit properties of the projective functions.

(a) $\Delta_3 \in S_1$:

$$\lim_{m_{33}\to-\infty} \begin{bmatrix} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & \gamma_{12} & & \gamma_{22} & \\ & & \gamma_{11} & & \end{bmatrix} \begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \end{pmatrix}$$

$$= \delta_{\Delta_1+\Delta_2,\gamma_{12}}\,\delta_{0,\gamma_{22}} \begin{bmatrix} & (\Gamma_s)_{11} & \\ \Delta_1+\Delta_2 & & 0 \\ & \gamma_{11} & \end{bmatrix}_{\text{ext}} \begin{pmatrix} m_{13} & & m_{23} \\ & m_{12} & & m_{22} \end{pmatrix}, \qquad (5.11a)$$

(b) $\Delta_3 \in S_2$:

$$\lim_{m_{33}\to-\infty} \begin{bmatrix} & & (\Gamma_s) & & \\ p & & q & & 0 \\ & \gamma_{12} & & \gamma_{22} & \\ & & \gamma_{11} & & \end{bmatrix} \begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \end{pmatrix}$$

$$= \delta_{p,\gamma_{12}}\,\delta_{\Delta_1+\Delta_2-p,\gamma_{22}} \begin{bmatrix} & (\Gamma_s)_{11} & \\ p & & \Delta_1+\Delta_2-p \\ & \gamma_{11} & \end{bmatrix}_{ext} \begin{pmatrix} m_{13} & & m_{23} \\ & m_{12} & & m_{22} \end{pmatrix}, \qquad (5.11b)$$

where "ext" denotes an extended $U(2)$ projective function.[13] The significance of the limit relations in this form is to demonstrate the manner in which *an upper operator pattern is induced by limits from the set of lower operator patterns*, e.g., in Eq. (5.11a), the limit is zero unless $\gamma_{12} = \Delta_1 + \Delta_2$, $\gamma_{22} = 0$, and these are the values which we assign to $(\Gamma_s)_{12}$ and $(\Gamma_s)_{22}$ in the upper operator pattern. (A detailed description of this procedure is given in Ref. 12.)

6. CONCLUDING REMARKS

In this paper, the properties of a class of unique, nontrivial $U(3):U(2)$ projective functions have been developed in considerable detail. The purpose of this analysis has been to demonstrate, by giving explicit results, the elegant structures which are implied by the canonical splitting of the multiplicities of the unit tensor operators in $U(3)$. In particular, we have given explicitly a truly remarkable polynomial form $G_q(\Delta; x)$, having precisely the properties required to describe the intricate null space vanishings of the $U(3)$ Wigner operator having maximal null space.

Finally, let us emphasize again that the coupling coefficients appearing in the left-hand side of Eq. (5.1) are completely known as are the $U(3):U(2)$ projective operators $[p-q\ \ 0\ \ 0]$ and $[q\ q\ 0]$. Thus, all $U(3)$: $U(2)$ projective operators for which the upper operator pattern is stretched and the lower operator pattern is arbitrary have been completely determined. Using now the $U(3):U(2)$ subgroup reduction law, we see that we have indeed *constructed for each multiplicity set* the unique $U(3)$ Wigner operator in the set having the maximal null space.

ACKNOWLEDGMENT

We would like to thank Dr. E. Chacón of the Instituto de Física, Universidad de México for his help and participation in numerous discussions leading to this paper.

*Supported in part by the U.S. Army Research Office (Durham), the National Science Foundation, and the U.S. Atomic Energy Commission.

[1] L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. (N.Y.) **13**, 1957 (1972).

[2] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. (N.Y.) **8**, 691 (1967).

[3] J. A. Castilho Alcaras, L. C. Biedenharn, K. T. Hecht, and G. Neely, Ann. Phys. (N.Y.) **60**, 85 (1970).

[4] Three unit coplanar vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ emanating from a common origin O at a mutual angle of 120° define the three positive coordinate axes of a reference frame for the Möbius plane. The ith coordinate axis is the line which extends from $-\infty$ to $+\infty$ and contains vector $\mathbf{a}_i$. Each point P in the plane is represented by an ordered triplet of real numbers $(x_1 x_2 x_3)$, where $|x_i|$ is the distance from the origin O to the point P_i determined by the intersection of axis i with the line through P perpendicular to axis i. x_i is positive (negative) if the direction of the vector $\overrightarrow{OP}_i$, is the same as (opposite to) that of $\mathbf{a}_i$.

[5] A. Erdelyi, *Higher transcendental functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), vol. 1, p. 66.

[6] Numerically, $\mathfrak{M}$ is the same as the so-called *inner multiplicity*. $\mathcal{S}$ is also called the *outer multiplicity*. See B. Gruber, J. Math. Phys. (N.Y.) **11** 1783, 3077 (1970), for a different approach to this subject.

[7] D. E. Kittlewood, *The theory of group characters and matrix representations of groups* (Oxford U.P., London, 1950), 2nd ed., p. 98.

[8] G. E. Baird and L. C. Biedenharn, J. Math. Phys. (N.Y.) **5**, 1730 (1964).

[9] J. D. Louck, Am. J. Phys. **38**, 3 (1970).

[10] It generalizes in the obvious way to $U(n)$.

[11] By the statement that $[m_{13} m_{23} m_{33}]$ belongs to a region R of the Möbius plane, we mean, more precisely, that $(x_{23} x_{31} x_{12}) \in R$, where the x_{ij} are defined in terms of the m_{i3} through Eq. (1.2b) and $p_{i3} = m_{i3} + 3 - i$.

[12] J. D. Louck and L. C. Biedenharn, J. Math. Phys. (N.Y.) **11**, 2368 (1970).

[13] In III of this series, we prove that the extended $U(2)$ Wigner coefficients are, in fact, $U(2)$ Racah coefficients.

Gauge transformations of second type and their implementation. I. Fermions

J. F. Gille* and J. Manuceau†
Centre de Physique Théorique, C.N.R.S., 13-Marseille 9^e^, France
(Received 12 May 1972)

A necessary and sufficient condition for implementation of some local gauge transformations in a class of irreducible representations of the CAR algebra is proved. Some particular results on the unitary group of implementation are then given. Not all of the pure states induced by these representations are unitarily equivalent to "quasifree" states of the class we consider; it is shown that such a state is unitarily equivalent to a quasifree state if and only if a certain property (characterizing the "discrete" states) holds.

I. PRELIMINARIES

A. The fermion C^*-algebra and some of its gauge transformations of second type

Let (H, s) be a real separable Hilbert space. Consider the CAR algebra $\mathfrak{A} \equiv \overline{\mathfrak{A}(H,s)}$ built on (H,s), i.e., the C^*-algebra generated by the elements $B(\psi)$, where B is a one-to-one linear map of H into $\mathfrak{A}$ such that

$$[B(\psi), B(\varphi)]_+ = 2s(\psi, \varphi)I \quad \forall \psi, \varphi \in H$$

(I the identity element on $\mathfrak{A}$).

Suppose Λ is a linear operator on H such that

(i) $\dim(\ker \Lambda)$ is not odd (this is not a restriction).

(ii) $|\Lambda|$ is diagonalizable (where $\Lambda = J_0|\Lambda|$ in the polar decomposition).

We choose a complex structure J of H such that

$$\begin{cases} J|(\ker \Lambda)^\perp = J_0|(\ker \Lambda)^\perp, \\ J|\ker \Lambda \text{ an arbitrary complex structure of } \ker \Lambda. \end{cases}$$

We shall write

$$|\Lambda| = \sum_{k \in \mathbb{N}} \lambda_k P_{H_k} \lambda_k \in \mathbb{R}$$

where P_{H_k} are the orthogonal projections on H_k and H_k a two-dimensional real subspace of H which is invariant by J and such that $H = \bigoplus_{k \in \mathbb{N}} H_k$. We remark that some λ_k are possibly not different. (From now we denote by $\bigoplus$ the Hilbert sum and by $\oplus$ the weak sum). Λ is the infinitesimal generator of a one-parameter strongly continuous orthogonal group $\{T_\theta\}_{\theta \in \mathbb{R}}$ on H. By Ref. 1 we can define an automorphism τ_θ of $\mathfrak{A}$ with

$$\tau_\theta(B(\psi)) = B(T_\theta \psi).$$

B. The problem

We look for irreducible representations of $\mathfrak{A}$ for which τ_θ is implementable.

This problem was approached by Dell'Antonio.[2] We give here full proofs of the results announced by him and we generalize some of them.

II. THE CLASS OF REPRESENTATIONS WE CONSIDER

Let $\{\psi_k^1, \psi_k^2\}$ be an orthonormal basis of H_k; then $\Theta_k = -iB(\psi_k^1)B(\psi_k^2)$ verifies

$$[\Theta_k, B(\varphi)]_+ = 0 \quad \forall \varphi \in H_k, \qquad \Theta_k^2 = 1. \tag{II. 1}$$

The center of $\mathfrak{A}_k \equiv \mathfrak{A}(H_k, s)$ is reduced to the scalars, and therefore any solution of (II. 1) is Θ_k or $-\Theta_k$.

Let π'_k be an arbitrary irreducible representation of $\mathfrak{A}_k$ into $\mathcal{H}_k = C^2$.

We construct the representation π' of $\mathfrak{A}$ into $\mathcal{H} = \bigotimes_{k \in \mathbb{N}} \mathcal{H}_k$ from the following: $\forall k \in \mathbb{N}, j = 1, 2$,

$$\pi'(B(\psi_k^j)) = \bigotimes_{j\,1}^{k-1} \pi'_j(\epsilon_j \Theta_j) \otimes \pi'_k(B(\psi_k^j)) \otimes \bigotimes_{j\,k+1}^{\infty} \times I_j (I_j = I_{C^2}),\ \epsilon_j = \pm 1.$$

It is well known that each $\Omega = \bigotimes_{k \in \mathbb{N}} \Omega_k$, Ω_k being a unitary vector of $\mathcal{H}_k$, determines an incomplete tensor product $\mathcal{H}^\Omega = \bigotimes_{k \in \mathbb{N}}^{\mathcal{C}(\Omega)} \mathcal{H}_k$ with $\mathcal{C}(\Omega)$ the equivalence class of Ω for the relation $\Omega \approx \Omega'$ iff

$$\sum_{k \in \mathbb{N}} |(\Omega_k | \Omega'_k) - 1| < +\infty.$$

It is not difficult to see that the $\mathcal{H}^\Omega$ are invariant subspaces of π' and that the restrictions of π' to those subspaces denoted by π'_Ω are irreducible and therefore π' is the direct sum of the set of the π'_Ω.

Let π be the representation of $\mathfrak{A}$ into $\mathcal{H}$ defined by

$$\pi(B(\psi_k^j)) \bigotimes_{l\,1}^{k-1} \pi_l(\Theta_l) \otimes \pi_k(B(\psi_k^j)) \otimes \bigotimes_{l\,k+1}^{\infty} I_l, \quad j = 1, 2,$$

where

$$\pi_l(\Theta_l) = \sigma_l^3, \qquad \pi_k(B(\psi_k^j)) = \sigma_k^j$$

and

$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the matrix of σ_l^3 in the canonical basis of $\mathcal{H}_l$,

$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the matrix of σ_l^1 in the canonical basis of $\mathcal{H}_l$,

$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ is the matrix of σ_l^2 in the canonical basis of $\mathcal{H}_l$.

Accordingly we shall write $\pi = \mathring{\bigotimes}_{k \in \mathbb{N}} \pi_k$.[3] It is clear that for each $l \in \mathbb{N}$, a unitary operator U_l on $\mathcal{H}_l$ exists such that $\forall x \in \mathfrak{A}_l$, $\pi_l(x) = U_l \pi'_l(x) U_l^*$. If $U = \bigotimes_{l \in \mathbb{N}} U_l$, we construct

$$\pi''(x) = U\pi'(x)U^*, \quad \forall x \in \mathfrak{A};$$

hence

$$\pi''(B(\psi_k)) = \bigotimes_{l\,1}^{k-1} \epsilon_l \sigma_l^3 \otimes \pi_k(B(\psi_k)) \otimes \bigotimes_{l\,k+1}^{\infty} I_l.$$

$V = \bigotimes_{l \in \mathbb{N}} V_l$ with $V_l = \sigma_l^3$ if the number of $k < l$ such that $\epsilon_k = -1$ is odd and $V_l = I_l$ otherwise.

Clearly $\pi(x) = V\pi''(x)V^*$, $\forall x \in \mathfrak{A}$; hence $\pi'(x) = W\pi(x)W^*$, $\forall x \in \mathfrak{A}$, where W is a unitary operator on $\mathcal{H}$.

Any irreducible subrepresentation π'_Ω of π' is unitary equivalent to the subrepresentation $\pi_{W^*\Omega}$ of π. Therefore we can restrict our attention to the study of the irreducible subrepresentations of π.

Proposition: π_Ω is unitarily equivalent to $\pi_{\Omega'}$ if and only if Ω and Ω' are weakly equivalent.

Proof: Recall that $\Omega = \bigotimes_{k\in\mathbb{N}} \Omega_k$ and $\Omega' = \bigotimes_{k\in\mathbb{N}} \Omega'_k$ are weakly equivalent iff $\sum_{k\in\mathbb{N}} |(|(\Omega_k|\Omega'_k)| - 1)| < +\infty$.

Suppose that Ω and Ω' are weakly equivalent. By Ref. 4, one can find for each $k \in \mathbb{N}$, $\nu_k \in \mathbb{R}$ such that

$$(\Omega'_k)_{k\in\mathbb{N}} \approx (e^{i\nu_k}\Omega_k)_{k\in\mathbb{N}}.$$

Let $U = \bigotimes_{k\in\mathbb{N}} e^{i\nu_k} I_k$. Then $U\Omega \in \mathcal{H}^{\Omega'}$ and we have

$$\pi_{\Omega'}(x) = U\pi_\Omega(x)U^*, \quad \forall x \in \mathfrak{A}.$$

Conversely, if Ω and Ω' are not weakly equivalent, let us denote

$$\omega_\Omega(x) = (\Omega|\pi_\Omega(x)\Omega), \quad x \in \mathfrak{A},$$

and

$$\omega_{\Omega'}(x) = (\Omega'|\pi_{\Omega'}(x)\Omega').$$

Let $U_k \in \mathcal{L}(\mathcal{H}_k)$ be an unitary operator such that

$$U_k\Omega'_k = \Omega_k,$$

and let

$$U'_k = \bigotimes_{j\,1}^{k-1} I_j \otimes U_k \otimes \bigotimes_{j\,k+1}^{\infty} I_j,$$

$$u_k = \pi^{-1}(U'_k).$$

The proof will be continued in the same way as in Sec. IIIA2.

III. THE THEOREM

We note

$$\Omega = \bigotimes_{k\in\mathbb{N}} \begin{pmatrix}\alpha_k\\ \beta_k\end{pmatrix} \quad \text{and} \quad x_k = |\alpha_k|^2.$$

A. Statement

A one-particle evolution τ_θ is implementable for the representation π_Ω if and only if the following condition holds:

$$\text{(A)} \qquad \sum_{k\in\mathbb{N}} x_k(1-x_k)\inf(1,\lambda_k^2) < +\infty.$$

If this occurs, a strongly continuous one-parameter group of unitary operators (we shall call such groups SCOPUG) $\{W_\theta\}_{\theta\in\mathbb{R}}$, $W_\theta \in \pi_\Omega(\mathfrak{A})'' = \mathcal{L}(\mathcal{H}^\Omega)$, exists such that

$$\forall x \in \mathfrak{A},\ \forall\theta \in \mathbb{R} \quad \pi_\Omega(\tau_\theta(x)) = W_\theta\pi_\Omega(x)W_{-\theta}.$$

B. Proof

1. Sufficiency

Suppose

$$\sum_{k\in\mathbb{N}} x_k(1-x_k)\inf(1,\lambda_k^2) < +\infty.$$

Let

$$U_{k,\theta} = \begin{pmatrix}1 & 0\\ 0 & e^{i\lambda_k\theta}\end{pmatrix}.$$

It is a unitary operator on $\mathcal{H}_k$. $U_\theta = \bigotimes_{k\in\mathbb{N}} U_{k,\theta}$ is a unitary operator on $\mathcal{H}$.[5] Clearly

$$\pi(\tau_\theta(B(\psi_k^i))) = U_\theta\pi(B(\psi_k^i))U_\theta^{-1}, \quad i = 1, 2,\ k \in \mathbb{N};$$

hence U_θ implements τ_θ for the representation π. Changing $U_{k,\theta}$ into $V_{k,\theta} = e^{i\mu_k}U_{k,\theta}$, $\mu_k \in \mathbb{R}$, $V_\theta = \bigotimes_{k\in\mathbb{N}} V_{k,\theta}$ implements τ_θ.

We choose μ_k such that

$$\arg(V_{k,\theta}\Omega_k|\Omega_k) = 0.$$

We get

$$(V_{k,\theta}\Omega_k|\Omega_k)^2 = |(U_{k,\theta}\Omega_k|\Omega_k)|^2 = 1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2);$$

from the hypothesis

$$\sum_{k\in\mathbb{N}} x_k(1-x_k)\sin^2(\lambda_k\theta/2) < +\infty;$$

hence

$$\sum_{k\in\mathbb{N}} |(V_{k,\theta}\Omega_k|\Omega_k)^2 - 1| < +\infty.$$

$\prod_{k\in\mathbb{N}}(V_{k,\theta}\Omega_k|\Omega_k)^2$ converges and $V_\theta\mathcal{H}^\Omega \subset \mathcal{H}^\Omega$. We note now V_θ its restriction to $\mathcal{H}^\Omega$. Hence

$$\pi_\Omega(\tau_\theta(x)) = V_\theta\pi_\Omega(x)V_\theta^*, \quad \forall x \in \mathfrak{A}, \quad \text{holds.}$$

It is important to remark that V_θ has been calculated for each $\theta \in \mathbb{R}$ so that $\{V_\theta\}_{\theta\in\mathbb{R}}$ is *not* a group in the general case. By Ref. 6 there exists a SCOPUG $\{W_\theta\}_{\theta\in\mathbb{R}}$ in $\mathcal{L}(\mathcal{H}^\Omega)$ such that

$$\forall x \in \mathfrak{A},\ \forall\theta \in \mathbb{R}, \quad \pi_\Omega(\tau_\theta(x)) = W_\theta\pi_\Omega(x)W_{-\theta}.$$

2. Necessity

Condition (A) is equivalent to the both following conditions:

$$\text{(i)} \quad \sum_{k,|\lambda_k|\geq 1} x_k(1-x_k) < +\infty,$$

$$\text{(ii)} \quad \sum_{k,|\lambda_k|\leq 1} x_k(1-x_k) < +\infty.$$

Suppose condition (A) is false. Then either (i) or (ii) is false. The following two lemmas prove that in the both cases $\exists\theta \in \mathbb{R}$ and $\sum_{k\in\mathbb{N}} x_k(1-x_k)\sin^2(\lambda_k\theta/2) = +\infty$.

Lemma 2.1: Let $(r_k)_{k\in\mathbb{N}}$, $0 \leq r_k \leq 1$, and let $(\lambda_k)_{k\in\mathbb{N}}$, $\lambda_k \in \mathbb{R}$, $|\lambda_k| \geq 1$. Then

$$\left(\sum_{k\in\mathbb{N}} r_k\sin^2(\lambda_k\theta) < +\infty,\ \forall\theta \in \mathbb{R}\right) \Rightarrow \sum_{k\in\mathbb{N}} r_k < +\infty.$$

Proof: In our case we have, for $r_k = 4x_k(1-x_k)$,

$$\sum_{k\in\mathbb{N}} r_k\sin^2(\lambda_k\theta) < +\infty, \quad \forall\theta \in \mathbb{R}.$$

In the proof of the sufficient condition we saw that the convergence of this series implies the existence of an SCOPUG $\{W_\theta\}_{\theta\in\mathbb{R}}$, $W_\theta \in \mathcal{L}(\mathcal{H}^\Omega)$ such that

$$\forall x \in \mathfrak{A},\ \forall \theta \in \mathbb{R},\quad \pi_\Omega(\tau_\theta(x)) = W_\theta \pi_\Omega(x) W_\theta^*.$$

Now we constructed a set of unitary operators $\{V_\theta\}_{\theta\in\mathbb{R}}$ such that

$$\forall x \in \mathfrak{A},\ \forall \theta \in \mathbb{R},\quad \pi_\Omega(\tau_\theta(x)) = V_\theta \pi_\Omega(x) V_\theta^*.$$

π_Ω being an irreducible representation,

$$W_\theta = \chi(\theta) V_\theta;\qquad \chi : \mathbb{R} \to \mathbb{C},\ |\chi(\theta)| = 1;$$

hence

$$|(W_\theta\Omega|\Omega)| = |\chi(\theta)|\,|(V_\theta\Omega|\Omega)| = |(V_\theta\Omega|\Omega)|$$

and

$$|(W_\theta\Omega|\Omega)|^2 = \prod_{k\,1}^{\infty} [1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2)].$$

Now $\{W_\theta\}_{\theta\in\mathbb{R}}$ is strongly continuous in θ; therefore

$$\theta \to |(W_\theta\Omega|\Omega)|^2 = \prod_{k\,1}^{\infty} [1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2)].$$

is continuous $\forall\theta \in \mathbb{R}$. Let us call

$$f_k(\theta) = 1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2),\quad P(\theta) = \prod_{k\,1}^{\infty} f_k(\theta).$$

We have $P(0) = 1$ and $\theta \to P(\theta)$ is continuous $\forall\theta\in\mathbb{R}$,

$$0 \le P(\theta) \le f_k(\theta) \le 1,\quad \forall k \in \mathbb{N},\ \forall\theta\in\mathbb{R},$$

and

$$\tfrac{2}{3}[1 - f_k(\theta)] \le |\mathrm{Log} f_k(\theta)|,\quad \forall k\in\mathbb{N},\ \forall\theta\in\mathbb{R},$$

(Log is Neper logarithm)

and

$$\mathrm{Log}\,P(\theta) = \sum_{k\,1}^{\infty} \mathrm{Log} f_k(\theta)\quad \text{for small } \theta\text{'s}.$$

Let us call

$$S(\theta) = \sum_{k\,1}^{\infty} [1 - f_k(\theta)] < +\infty \quad \text{for } P(\theta) \ne 0;$$

i.e., in a neighborhood of 0

$$\tfrac{2}{3}S(\theta) \le -\mathrm{Log}\,P(\theta)\quad \text{for } |\theta| \le \theta_0 < 1;$$

moreover,

$$\tfrac{2}{3}S_n(\theta) = \tfrac{2}{3}\sum_{k\,1}^{n} [1 - f_k(\theta)] \le \tfrac{2}{3}S(\theta) \le -\mathrm{Log}\,P(\theta).$$

Now, on $[-\theta_0, +\theta_0]$, $\theta \to -\mathrm{Log}\,P(\theta)$ is an integrable function, and S is measurable as a pointwise limit of measurable functions. Hence S is integrable on $[-\theta_0, +\theta_0]$. We take now $\theta \in [-\theta_0, +\theta_0]$,

$$S(\theta) = \sum_{k\,1}^{\infty} r_k\left(\frac{1-\cos\lambda_k\theta}{2}\right) < \infty,$$

$$F_n(\theta) = \int_0^\theta S_n(t)\,dt = \sum_{k\,1}^{n}\left(\frac{r_k\theta}{2} - \frac{\sin(\lambda_k\theta)}{2\lambda_k} r_k\right) \le \int_0^\theta S(t)\,dt.$$

Let

$$F(\theta) = \int_0^\theta S(t)\,dt,$$

$$\int_0^\theta F_n(t)\,dt = \sum_{k\,1}^{n}\left(\frac{r_k\theta^2}{4} + r_k\frac{\cos(\lambda_k\theta)-1}{2\lambda_k^2}\right) \le \int_0^\theta F(t)\,dt < +\infty.$$

(B) $\sum_{k\,1}^{\infty} r_k\left(\frac{1-\cos(\lambda_k\theta)}{2\lambda_k^2}\right) < +\infty$ since $|\lambda_k| \ge 1$.

(C) $\sum_{k\,1}^{\infty}\left(\frac{r_k\theta^2}{4} + r_k\frac{\cos(\lambda_k\theta)-1}{2\lambda_k^2}\right)$

absolutely converges, since

$$\sum_{k\,1}^{n}\left|\frac{r_k\theta^2}{4} + r_k\frac{\cos(\lambda_k\theta)-1}{2\lambda_k^2}\right| = \sum_{k\,1}^{n}\int_0^\theta |g_k(t)|\,dt \le \int_0^\theta F(t)\,dt < +\infty.$$

with $g_k(\theta) = \int_0^\theta [1 - f_k(t)]\,dt$

Obviously the sum of (B) and (C) shows that

$$\tfrac{1}{4}\theta^2 \sum_{k\,1}^{\infty} r_k < +\infty \quad \text{with } \theta \ne 0;$$

hence $\sum_{k\,1}^{\infty} r_k < +\infty$. ∎

Lemma 2.2: If $f : \mathbb{R} \to \mathbb{R}$, $f(0) = 0$, f differentiable at 0 and $f'(0) = 1$, $u_k \in \mathbb{R}$, $(u_k)_{k\in\mathbb{N}}$ bounded, $r_k \ge 0\ \forall k\in\mathbb{N}$, then

$$(\exists I \in \mathcal{V}_{\mathbb{R}}(0) \text{ and } \forall\theta\in I,\ \sum_{k\,1}^{\infty} r_k[f(u_k\theta)]^2 < +\infty) \iff \sum_{k\,1}^{\infty} r_k u_k^2 < +\infty.$$

Proof: If $J \in \mathcal{V}_{\mathbb{R}}(0)$ is such that $x \in J \Rightarrow |[f(x)/x] - 1| < \frac{1}{2}$,

$$\tfrac{1}{2}x < f(x) < \tfrac{3}{2}x,$$

and $I \in \mathcal{V}_{\mathbb{R}}(0)$ is such that $\forall\theta\in I, \theta u_k \in J$, $\forall k\in\mathbb{N}$, then

$$\tfrac{1}{4}\theta^2\sum_{k\,1}^{\infty} r_k u_k^2 \le \sum_{k\,1}^{\infty} r_k[f(u_k\theta)]^2 \le \tfrac{9}{4}\theta^2\sum_{1}^{\infty} r_k u_k^2. \ ∎$$

Now, we return to the proof of necessity. Let $\theta\in\mathbb{R}$ such that $\sum_{k\in\mathbb{N}} x_k(1-x_k)\sin^2(\lambda_k\theta/2) = +\infty$.

Let $u_k(\theta) = B(\cos(\lambda_k\theta/2)\psi_k^1 - \sin(\lambda_k\theta/2)\psi_k^2)B(\psi_k^1)$.

$$E_{n,m} = \bigoplus_{k=n}^{m} H_k,\qquad u_{n,m}(\theta) = \prod_{k=n}^{m} u_k(\theta).$$

$$\omega_\Omega(x) = (\Omega|\pi_\Omega(x)\Omega),\quad \forall x\in\mathfrak{A} \tag{2.1}$$

We have

$$\forall x \in \mathfrak{A}(E_{n,m}, s),\quad \omega_\Omega(x) = \omega_\Omega \circ \tau_\theta(u_{n,m}(\theta)\,x\,u_{n,m}^*(\theta)).$$

Since

$$B(\psi)B(\varphi)B(\psi) = B(2s(\varphi,\psi)\psi - \varphi) = B(S_\psi\varphi),$$

S_ψ the symmetry with regard to ψ ($\|\psi\| = 1$).

For any $\varphi\in H_k$, $\tau_\theta(B(\varphi)) = B(e^{i\lambda_k\theta}\varphi) = B(R_{\lambda_k\theta}\varphi)$, $R_{\lambda_k\theta}$ the rotation of the argument $\lambda_k\theta$. Hence (2.1) holds.

Let us consider $\Theta_{n,m} = \prod_{k\,n}^{m}\Theta_k$; we shall note $\overline{\mathfrak{A}_e(H,s)}$ (resp. $\overline{\mathfrak{A}_0(H,s)}$) the C^*-algebra (resp. the closed vector-subspace) of $\overline{\mathfrak{A}(H,s)}$, generated by products of even (resp. odd) number of $B(\psi)$'s. Let us denote

$$\omega_{n,m} = \omega_\Omega|\mathfrak{A}_e(E_{n,m},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n,m},s),$$

$\pi_{n,m} = \overset{\circ}{\bigotimes}{}_n^m\, \pi_k$ (tensor product "à la Powers"[3]),

$\Omega_{n,m} = \bigotimes_n^m \Omega_k$.

It is not difficult to see that

$\zeta: x + \Theta_{1,n-1}y \in \mathfrak{A}_e(E_{n,m},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n,m},s) \to x + y \in \mathfrak{A}(E_{n,m},s)$ is a C^*-isomorphism and that $\omega_{n,m}(z) = (\Omega_{n,m}|\pi_{n,m}(\zeta(z))\Omega_{n,m})$, $\forall z \in \mathfrak{A}_e(E_{n,m},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n,m},s)$. $\pi_{n,m}$ is an irreducible representation, and hence $\omega_{n,m}$ is a pure state. Lemma 2.4 of Ref. 7 implies ($\mathfrak{A}_e(E_{n,m},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n,m},s)$ is a C^*-algebra[8]):

$$\|(\omega_\Omega - \omega_\Omega \circ \tau_\theta)|\mathfrak{A}_e(E_{n,m},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n,m},s)\|$$
$$= 2[1 - |\omega_\Omega(u_{n,m}(\theta)|^2]^{1/2}$$
$$= 2\left(1 - \prod_{k=n}^m [1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta)/2)]\right)^{1/2}.$$

Now

$$\sum_{k\in N} x_k(1-x_k)\sin^2(\lambda_k\theta/2) = +\infty$$

implies

$$\prod_{i=n}^{\infty} [1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2)] = 0,$$

i.e.,

$$\lim_{m,\infty} \prod_{i=n}^{m} [1 - 4x_k(1-x_k)\sin^2(\lambda_k\theta/2)] = 0.$$

Denote by $\mathfrak{A}(E_n,s)^c$ the commutant of $\mathfrak{A}(E_n,s)$ in $\mathfrak{A}$. Then

$$\left(E_n = \bigoplus_{k=1}^n H_k\right), \quad E_n^\perp = \bigoplus_{k=n+1}^\infty H_k,$$

$$\mathfrak{A}(E_n,s)^c = \mathfrak{A}_e(E_n^\perp,s) \oplus \Theta_{1,n}\mathfrak{A}_0(E_n^\perp,s),^9$$
$$\mathfrak{A}_e(E_n^\perp,s) \oplus \Theta_{1,n}\mathfrak{A}_0(E_n^\perp,s) \supset \bigcup_{k\ n+1}^{\infty} [\mathfrak{A}_e(E_{n+1,k},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n+1,k},s)].$$

Thus

$$\|(\omega_\Omega - \omega_\Omega \circ \tau_\theta)|\mathfrak{A}(E_n,s)^c\|$$
$$= \|(\omega_\Omega - \omega_\Omega \circ \tau_\theta)|\mathfrak{A}_e(E_n^\perp,s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_n^\perp,s)\|$$
$$\geq \lim_{k,\infty} \|(\omega_\Omega - \omega_\Omega \circ \tau_\theta)|\mathfrak{A}_e(E_{n+1,k},s) \oplus \Theta_{1,n-1}\mathfrak{A}_0(E_{n+1,k},s)\|$$
$$= 2.$$

Now $E_{n+1} \supset E_n$ and $\bigcup_{k\in N} E_k = \bigoplus_{k\in N} H_k$, $\overline{\bigcup_{k\in N} E_k} = H$.

Hence, by Lemma 2.1 of Ref. 7, ω_Ω and $\omega_\Omega \circ \tau_\theta$ are not unitarily equivalent; therefore, no unitary $U_\theta \in \mathfrak{L}(\mathcal{H}^\Omega)$ can exist such that, $\forall x \in \mathfrak{A}$,

$$\pi_\Omega(\tau_\theta(x)) = U_\theta\tau_\Omega(x)U_\theta^*;$$

τ_θ is not implementable for the representation π_Ω. ■

IV. OTHER PROPOSITIONS AND REMARKS

(1) Fix $\theta \in \mathbf{R}$; there exists a unitary operator $U_\theta \in \mathfrak{L}(\mathcal{H}^\Omega)$ such that

$$\pi_\Omega(\tau_\theta(x)) = U_\theta\pi_\Omega(x)U_\theta^* \quad \forall x \in \mathfrak{A}$$

if and only if

$$\sum_{k\in N} x_k(1-x_k)\sin^2(\lambda_k\theta/2) < +\infty. \tag{IV.1}$$

Proof: If (IV.1) is true, the existence of U_θ is checked (see the beginning of Sec. III).

If such a U_θ exists, $U_\theta = e^{i\rho}V_\theta$, V_θ is the operator constructed (Sec. IIIA)

$$U_\theta \in \mathfrak{L}(\mathcal{H}^\Omega), \quad U_\theta = e^{i\rho} \bigotimes_{k\in N} V_{k,\theta}.$$

$U_\theta\Omega = e^{i\rho}(V_{k,\theta}\Omega_k)_k \in \mathcal{H}^\Omega$, therefore $(V_{k,\theta}\Omega_k)_k \approx (\Omega_k)_k$ which implies $\sum_{k\in N} |(V_{k,\theta}\Omega_k|\Omega_k) - 1| < +\infty$.

Recall that $\arg(V_{k,\theta}\Omega_k|\Omega_k) = 0$; hence

$$(V_{k,\theta}\Omega_k|\Omega_k) = |(V_{k,\theta}\Omega_k|\Omega_k)|$$

and

$$\sum_{k\in N} x_k(1-x_k)\sin^2(\lambda_k\theta/2) < +\infty. \qquad ■$$

(2) Let

$$\mathfrak{N}_\Omega = \left\{\theta \in \mathbf{R} \;\middle|\; \begin{array}{l}\text{there exists a unitary operator}\\ U_\theta \in \mathfrak{L}(\mathcal{H}^\Omega) \text{ such that: } \forall x \in \mathfrak{A}\\ \pi_\Omega(\tau_\theta(x)) = U_\theta\pi_\Omega(x)U_\theta^*.\end{array}\right\}$$

$\mathfrak{N}_\Omega$ is an additive subgroup of $\mathbf{R}$.

Proof: Let $\theta_1, \theta_2 \in \mathfrak{N}_\Omega$. Then

$$\sum_{k\in N} x_k(1-x_k)\sin^2(\lambda_k\theta_1/2) < +\infty$$

and

$$\sum_{k\in N} x_k(1-x_k)\sin^2(\lambda_k\theta_2/2) < +\infty.$$

Let us set $r_k = x_k(1-x_k)$, $\varphi_k^1 = \lambda_k\theta_1/2$, $\varphi_k^2 = \lambda_k\theta_2/2$; $\sum_{k\in N} r_k \sin^2(\varphi_k^1 + \varphi_k^2)$ converges, for

$$M = \sum_{k\in N} r_k \sin^2(\varphi_k^1)\cos^2(\varphi_k^2) \leq \sum_{k\in N} r_k \sin^2\varphi_k^1 < +\infty.$$

$$N = \sum_{k\in N} r_k \sin^2(\varphi_k^2)\cos^2(\varphi_k^2) \leq \sum_{k\in N} r_k \sin^2\varphi_k^2 < +\infty.$$

$$|L| \leq \sum_{k\in N} 2r_k|\sin(\varphi_k^1)\sin(\varphi_k^2)\cos(\varphi_k^1)\cos(\varphi_1^2)| \leq \sum_{k\in N} r_k[\sin^2(\varphi_k^1) + \sin^2(\varphi_k^2)] < +\infty.$$

Now

$$M + N + L = \sum_{k\in N} r_k \sin^2(\varphi_k^1 + \varphi_k^2).$$

Obviously $\theta \in \mathfrak{N}_\Omega$ and $\theta \in \mathfrak{N}_\Omega \Longrightarrow -\theta \in \mathfrak{N}_\Omega$.

(3) If $\sum_{k\in N} x_k(1-x_k) < +\infty$ we shall say that representation π_Ω is a *discrete* one. Sec. IIIB1 implies that all the monoparticular evolutions are implementable for every discrete representation.

Statement: If π_Ω is *not* a discrete representation (i.e., $\sum_{k\in N} x_k(1-x_k) = +\infty$) and if $\{\lambda_k\}_{k\in N}$ has neither 0 nor infinite as accumulation points, then $\mathfrak{N}_\Omega$

$= a\mathbb{Z}_+, a \in \mathbb{R}_+$ ($\mathbb{Z}$ the additive group of the relative integers).

Proof: Except for a finite number of k's we have

$$\lambda_k \in [a'', b''] \cup [a', b'] \quad \text{with } a'' < b'' < 0 < a' < b'.$$

We can omit a finite number of k's without changing $\mathfrak{N}_\Omega$ which is determined by the convergence of some series. The convergence of which is not changed by the suppression of a finite number of terms. Let us build a dividing decomposition of those intervals.

Let $a'_n = \frac{4}{3} a'_{n-1} = (\frac{4}{3})^n a'$, $I'_n = [a'_n = [a'_n, a'_{n+1}]$. A finite number of I'_n overlaps $[a', b']$.

Let $[r'_n, s'_n] = [\pi/3a'_n, \pi/2a'_{n+1}]$ which is a proper interval.

If $\mu \in I'_n$ and $\theta \in [r'_n, s'_n]$, then $\mu\theta \in [\pi/3, \pi/2]$. In the same way let us write

$$a''_n = \tfrac{3}{4} a''_{n-1} = (\tfrac{3}{4})^n a'' I''_n = [a''_n, a''_{n+1}].$$

A finite number of I''_n overlaps $[a'', b'']$.

Let $[r''_n, s''_n] = [\pi/2a''_n, \pi/3a''_{n+1}]$ which is a proper interval.

If $\mu \in I''_n$ and $\theta \in [r''_n, s''_n]$, then $\mu\theta \in [\pi/3, \pi/2]$. Let us denote $\{I_p\}_{1 \le p \le m}$ and $\{[r_p, s_p]\}_{1 \le p \le m}$ those intervals and let

$$L_p = \{k \in \mathbb{N} | \lambda_k \in I_p\}.$$

Then

$$\sum_{k \in \mathbb{N}} x_k(1 - x_k) = \sum_{k \in \bigcup_p{}_1^m L_p} x_k(1 - x_k) = +\infty.$$

If $k \in L_p$, then $\lambda_k \in I_p$, $\lambda_k\theta/2 \in [\pi/3, \pi/2]$ as soon as $\theta \in [2r_p, 2s_p]$; hence $\sin^2(\lambda_k\theta/2) \in [\frac{3}{4}, 1]$ and

$$\sum_{k \in \mathbb{N}} x_k(1 - x_k) \sin^2(\lambda_k\theta/2) = +\infty.$$

Thus not any U_θ can exist for not any $\theta \in [r_p, s_p]$.

From that we conclude that $\mathfrak{N}_\Omega = a\mathbb{Z}$ for some $a \in \mathbb{R}_+$. ■

(4.1) Definitions: As in Sec. IIIB2, we shall denote $\overline{\mathfrak{A}_0(H,s)}$ the closed vector subspace of $\mathfrak{A}$ generated by products of odd number of $B(\psi)$'s.

A state ω on $\mathfrak{A}$ will be called *quasifree*[3,10,1] when $\omega | \overline{\mathfrak{A}_0(H,s)} = 0$,

$$\omega\left(\prod_{i\,1}^{2n} B(\varphi_i)\right) = \sum_{\substack{i_1 < i_2 < \cdots < i_n \\ i_k < j_k}} \epsilon_\sigma \prod_{k\,1}^{n} \omega(B(\varphi_{i_k})B(\varphi_{j_k})),$$

ϵ_σ being the parity of the permutation σ

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & 2n-1 & 2n \\ i_1 & j_1 & \cdots & i_n & j_n \end{pmatrix}.$$

Let us call $\omega_\Omega(x) = (\Omega | \pi_\Omega(x)\Omega)$, $x \in \mathfrak{A}$, with

$$\Omega = \bigotimes_{k \in \mathbb{N}} \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}.$$

Accordingly with Sec. 4.3 below we call ω_Ω a *discrete state* iff $\sum_{k\in\mathbb{N}} x_k(1 - x_k) < +\infty$ ($x_k = |\alpha_k|^2$).

(4.2) Lemma: ω_Ω is quasifree if and only if $\alpha_k\beta_k = 0$, $\forall k \in \mathbb{N}$.

Proof: Suppose ω_Ω is quasifree.

$$\left.\begin{aligned} \omega_\Omega(B(\psi_1^1)) &= 2\,\mathrm{Re}(\alpha_1\overline{\beta_1}) = 0 \\ \omega_\Omega(B(\psi_1^2)) &= -2\,\mathrm{Im}(\alpha_1\overline{\beta_1}) = 0 \end{aligned}\right\}$$

$\Longrightarrow \alpha_1\beta_1 = 0$ and $|\alpha_1|^2 - |\beta_1|^2 = \pm 1$;

hence

$$\left.\begin{aligned} \omega_\Omega(B(\psi_k^1)) &= \pm 2\,\mathrm{Re}(\alpha_k\overline{\beta_k}) = 0 \\ \omega_\Omega(B(\psi_k^2)) &= \mp 2\,\mathrm{Im}(\alpha_k\overline{\beta_k}) = 0 \end{aligned}\right\} \Longrightarrow \alpha_k\beta_k = 0.$$

Conversely, suppose $\forall k \in \mathbb{N}$, $\alpha_k\beta_k = 0$. Let $y = \prod_{k\,1}^{n} y_k$ with $y_k = B(\psi_k^1)B(\psi_k^2)$ or $y_k = B(\psi_k^1)$ or $y_k = B(\psi_k^2)$; if $y \in \overline{\mathfrak{A}_0(H,s)}$, at least there exists a $k_0 \in \mathbb{N}$ such that $y_{k_0} = B(\psi_{k_0}^j)$, $j = 1$ or $j = 2$, and

$$\omega_\Omega(y) = (\Omega | \pi_\Omega(y)\Omega) = \prod_{l \in \mathbb{N}} (\Omega_l | \pi_l(y_l)\Xi_l\Omega_l),$$
$$\Xi_l = I_l \text{ or } \sigma_l^3.$$

From

$$\left.\begin{aligned} (\Omega_{k_0} | \pi_{k_0}(B(\psi_{k_0}^j))\Omega_{k_0}) &= \pm 2\,{}^{\mathrm{Re}}_{\mathrm{Im}}(\alpha_{k_0}\overline{\beta_{k_0}}) = 0 \\ (\Omega_{k_0} | \pi_{k_0}(B(\psi_{k_0}^j))\sigma_{k_0}^3\Omega_{k_0}) &= 2i\,{}^{\mathrm{Im}}_{\mathrm{Re}}(\alpha_{k_0}\overline{\beta_{k_0}}) = 0 \end{aligned}\right\}$$
$$\times \begin{pmatrix} j = 1 \text{ higher position} \\ j = 2 \text{ lower position} \end{pmatrix}$$

we deduce $\omega_\Omega | \overline{\mathfrak{A}_0(H,s)} = 0$.

Moreover,

$$\omega_\Omega\left(\prod_{k\,1}^{n} B(\psi_k^1)B(\psi_k^1)\right) = \prod_{k\,1}^{n} \omega_\Omega(B(\psi_k^1)B(\psi_k^2)). \qquad ■$$

(4.3) Proposition: There exists a quasifree state $\omega_{\Omega'}$ unitarily equivalent to ω_Ω iff ω_Ω is a discrete state.

Proof: Suppose ω_Ω is unitarily equivalent to a quasifree state $\omega_{\Omega'}$, with

$$\Omega' = \bigotimes_{k \in \mathbb{N}} \begin{pmatrix} \alpha'_k \\ \beta'_k \end{pmatrix}, \quad \alpha'_k\beta'_k = 0, \quad \forall k \in \mathbb{N}.$$

Recall that ω_Ω and $\omega_{\Omega'}$ are unitarily equivalent iff (Sec. II, Proposition)

$$\sum_{k \in \mathbb{N}} [1 - |(\Omega_k | \Omega'_k)|] < +\infty,$$

which is equivalent to $\exists\, M, L \subset \mathbb{N} |$, $M \cup L = \mathbb{N}$, $M \cap L = \emptyset$

$$\sum_{k \in M} (1 - |\alpha_k|) + \sum_{k \in L} (1 - |\beta_k|) < +\infty,$$

$$\sum_{k \in M} (1 - \sqrt{x_k}) + \sum_{k \in L} (1 - \sqrt{1 - x_k}) < +\infty$$

which implies that: $\prod_{k \in M} \sqrt{x_k}$ converges and is different from 0, therefore so does $\prod_{k \in M} x_k$, $\sum_{k \in M} (1 - x_k) < +\infty$; $\prod_{k \in L} \sqrt{1 - x_k}$ converges and is different from 0, therefore so does $\Pi_{k\in L}(1-x_k)$, $\sum_{k\in L} x_k < +\infty$; so $\sum_{k\in \mathbb{N}} x_k(1-x_k) < +\infty$.

Conversely, if $\sum_{k\in \mathbb{N}} x_k(1-x_k) < +\infty$, let

$$M = \{k \in \mathbb{N} \mid x_k > \tfrac{1}{2}\} \qquad \sum_{k\in M} (1-x_k) < +\infty,$$

$$L = \mathbb{N} - M, \qquad \sum_{k\in L} x_k < +\infty,$$

which implies $\Pi_{k\in M} x_k$ converges and is different from 0 such as $\Pi_{k\in L}(1-x_k)$ and hence $\Pi_{k\in M}\sqrt{x_k}$ and $\Pi_{k\in L}\sqrt{1-x_k}$. In other words

$$\sum_{k\in M} (1-\sqrt{x_k}) + \sum_{k\in L} (1-\sqrt{1-x_k}) < +\infty.$$

Calling

$$\Omega' = \bigotimes_{k\in\mathbb{N}} \begin{pmatrix}\alpha'_k\\ \beta'_k\end{pmatrix} \quad \text{with} \begin{cases}\alpha'_k = 1,\ \beta'_k = 0 & \text{if } k \in M\\ \alpha'_k = 0,\ \beta'_k = 1 & \text{if } k \in L'\end{cases}$$

we have that $\sum_{k\in\mathbb{N}} [1 - |(\Omega_k|\Omega')|] < +\infty$ and the quasifree state $\omega_{\Omega'}$ is unitarily equivalent to ω_{Ω}. ■

ACKNOWLEDGMENTS

The authors are very indebted to H. Faure, G. Rauzy, G. Rideau, R. Seiler, M. Sirugue, D. Testard, J. C. Trotin, and A. Verbeure for illuminating discussions. The authors would like also to thank M. A. Grossmann for reading the manuscript.

---

*Attaché de Recherche–C.N.R.S.
†Maître de Conférence–Université de Provence.

[1] E. Balslev, J. Manuceau, and A. Verbeure, Commun. Math. Phys. 8, 315 (1968), Sec. 1.
[2] G. F. Dell'Antonio, J. Math. Phys. 12, 148 (1971).
[3] R. T. Powers, thesis (Princeton University, 1967).
[4] J. von Neumann, "On Infinite Direct Products," in *Collected Works* (Pergamon, New York, 1961), Vol. III, Def. 6.1.1 and Lemma 6.1.1.
[5] Y. Nakagami, Kōdai Math. Sem. Rept. 22, 341 (1970), Lemma 3.1, Def. 3.1.
[6] R. R. Kallman, J. Functional Anal. 7, 43 (1971), Theorem 0.7.
[7] R. T. Powers and E. Størmer, Commun. Math. Phys. 16, 1 (1970).
[8] J. Manuceau, F. Rocca, and D. Testard, Commun. Math. Phys. 12, 43 (1969), see (2.1).
[9] Reference 8, Lemma 2.3.1.
[10] E. Balslev and A. Verbeure, Commun. Math. Phys. 8, 315 (1968).

On the divergent perturbation expansion for the vacuum polarization by an external field

S. Graffi*†
Department of Physics, New York University, New York, New York 10003

V. Grecchi
Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Bologna, Italy
(Received 5 June 1972)

The divergent perturbation expansion of the exactly solvable vacuum polarization by an external constant electromagnetic field is examined. It is proven that the Stieltjes method, known to be valid for the vacuum polarization by a pure electric or a pure magnetic field, fails in the general case to sum the perturbation expansion to the exact solution. This implies the nonconvergence of the Padé approximants to the solution under these conditions. The validity of the Borel summation method and the convergence of the related approximation procedures are also proven.

1. INTRODUCTION

The vacuum polarization by a prescribed external constant electromagnetic field is one of the few phenomena in quantum electrodynamics for which an exact solution is known.[1] Furthermore, this solution gives rise to a divergent expansion in powers of the fine structure constant. Now in recent years there has been considerable interest in proving the applicability of the classical summation methods for divergent power series[2] to many problems in quantum mechanics as well as in quantum field theory, for which the perturbation expansion is known to be divergent:[3] it is, therefore, interesting to show the validity of these methods in the present context.

For the particular case of the vacuum polarization by a pure magnetic or a pure electric constant field it has been proven that the divergent perturbation expansion is Borel[4] and also Stieltjes[5] summable to the exact solution, this last result implying the convergence of the Padé approximants.

In the present paper the general case in which both the electric and the magnetic fields are present will be examine, and it will be proven that the perturbation expansion does not sum to the solution under the Stieltjes method (while it still does under the Borel one), thus showing the nonconvergence of the Padé approximants under these conditions.

However, the convergence to the solution of the generalized Padé approximants[6] will be shown and also that of the approximation method consisting of applying the Padé approximants to the appropriate generalized Borel transform.[7]

2. INAPPLICABILITY OF THE STIELTJES METHOD AND NONCONVERGENCE OF THE PADÉ APPROXIMANTS

The complete Lagrangian due to the vacuum polarization by an external constant electromagnetic field, as it has been computed by Schwinger,[1] reads

$$L = -F - \frac{1}{8\pi^2}\int_0^\infty \frac{e^{-s}}{s^3} \times \left((es)^2 G\,\frac{\operatorname{Re}\cos(esX)}{\operatorname{Im}\cosh(esX)} - 1 - \tfrac{2}{3}(es)^2 F\right), \qquad (2.1)$$

where e is the electron charge, $F = \frac{1}{4}F_{\mu\nu}^2 = \frac{1}{2}(\mathbf{H}^2 - \mathbf{E}^2)$ is the free electromagnetic field Lagrangian, $G = \frac{1}{4}\epsilon_{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} = \mathbf{E}\cdot\mathbf{H}$ is the pseudoscalar electromagnetic field invariant, $X = [2(F+iG)]^{1/2}$, and the electron mass has been put equal to 1.

Putting $\operatorname{Re}(esX) = x$, $\operatorname{Im}(esX) = y$,

we have

$$4\pi\alpha s^3 G = xy, \qquad 4\pi\alpha s^2 F = \tfrac{1}{2}(x^2 - y^2), \qquad (2.2)$$

and

$$x^2 = 4\pi\alpha s^2[F + (F^2 + G^2)^{1/2}],$$

$$y^2 = 4\pi\alpha s^2[-F + (F^2 + G^2)^{1/2}],$$

where $\alpha = e^2/4\pi$ is the fine structure constant. The Lagrangian (2.1) takes now the form

$$L = -F - L_I,$$

where the interaction part L_1 is given by

$$L_I = -\frac{1}{8\pi^2}\int_0^\infty \frac{e^{-s}}{s^3} \times [xy\,\coth(x)\cot(y) - 1 - \tfrac{1}{3}(x^2 - y^2)]\,ds \qquad (2.3)$$

and use of the identity $\cosh(x+iy) = \cos y \cosh x + i \sin y \sinh x$ has been made. Notice that L_I is a function only of α through (2.3), since F and G are constant quantities.] The first step in proving our statements is the following:

Lemma 2.1: The Lagrangian (2.1) may be written in the form

$$L_I(\alpha) = 2\alpha^2 \int_{-\infty}^{+\infty} \frac{\psi(t)dt}{1+\alpha t} \qquad (2.4)$$

$\psi(t)$ being a function positive in $(-\infty, +\infty)$ with finite moments in the same interval, i.e.

$$\int_{-\infty}^{+\infty} t^n \psi(t)\,dt < +\infty, \qquad n = 0, 1, 2 \cdots.$$

Proof: From (2.3), using the expansions

$$x\coth x = 1 + 2x^2\sum_{n=1}^\infty \frac{1}{x^2 + n^2\pi^2},$$

$$y\cot y = 1 + 2y^2\sum_{n=1}^\infty \frac{1}{y^2 - n^2\pi^2}, \quad \frac{\pi^2}{6} = \sum_{n=1}^\infty \frac{1}{n^2},$$

we get

$$L_1(\alpha) = -\frac{1}{4\pi^2}\int_0^\infty \frac{e^{-s}}{s^3}\, x^2y^2\left(\frac{y^2}{x^2}\sum_{n=1}^\infty \frac{1}{(y^2 - n^2\pi^2)n^2\pi^2} - \frac{x^2}{y^2}\sum_{n=1}^\infty \frac{1}{(x^2 + n^2\pi^2)n^2\pi^2} + 2\sum_{n,m=1}^\infty \frac{1}{(x^2 + n^2\pi^2)(y^2 - n^2\pi^2)}\right)ds$$

Taking into account the expansion of $\coth x$ and performing an elementary rearrangement, this formula becomes

$$L_I(\alpha) = -\frac{1}{4\pi^2}\int_0^\infty \frac{e^{-3}}{s^3}\, x^2y^2\Big[\frac{y^2}{x^2}\sum_{n=1}^{\infty}\frac{1}{(y^2-n^2\pi^2)n^2\pi^2}$$

$$-\frac{x^2}{y^2}\sum_{n=1}^{\infty}\frac{1}{(x^2+n^2\pi^2)n^2\pi^2} - 2\sum_{n=1}^{\infty}\frac{1}{x^2+n^2\pi^2}\, f_1\!\left(\frac{y}{x}\,n\pi\right)$$

$$+ 2\sum_{n=1}^{\infty}\frac{1}{y^2-n^2\pi^2}\, f_1\!\left(\frac{x}{y}\,n\pi\right)\Big]ds,$$

where $f_1(z) = (z\coth z)/2z^2$, and putting $f(z) = \coth z/z$ we have

$$L_I(\alpha) = \frac{1}{4\pi^2}\int_0^\infty \frac{e^{-s}}{s^3}\, x^2y^2\Big[\sum_{n=1}^{\infty}\frac{1}{x^2+n^2\pi^2}\, f\!\left(\frac{y}{x}\,n\pi\right)$$

$$+\sum_{n=1}^{\infty}\frac{1}{-y^2+n^2\pi^2}\, f\!\left(\frac{x}{y}\,n\pi\right)\Big]ds. \qquad (2.5)$$

Let us now introduce the following variables:

$$s' = s^2,\quad x' = x^2/\alpha s^2 = 4\pi[F + (F^2+G^2)^{1/2}],$$
$$y' = y^2/\alpha s^2 = 4\pi[-F + (F^2+G^2)^{1/2}],$$

so that (2.5) becomes

$$L_I(\alpha) = 2\alpha^2\int_0^\infty e^{-s'}G^2\Big[\sum_{n=1}^{\infty}\frac{1}{\alpha s'x' + n^2\pi^2}\, f\!\left(\frac{y'}{x'}\right)^{1/2} n\pi$$

$$+\sum_{n=1}^{\infty}\frac{1}{n^2\pi^2-\alpha s'y'}\, f\left(\left(\frac{x'}{y'}\right)^{1/2}\right)\Big]ds' \qquad (2.6)$$

and can be rewritten in the following way:

$$L_I(\alpha) = 2\alpha^2\Big(\frac{G^2}{y'}$$

$$\times\int_{-\infty}^{0}\frac{\sum_{n=1}^{\infty} f((x'/y')^{1/2}n\pi)\exp[-n\pi(|t|)^{1/2}/y'^{1/2}]}{1+\alpha t}\,dt$$

$$+\frac{G^2}{x'}\int_0^\infty\frac{\sum_{n=1}^{\infty} f((y'/x')^{1/2}n\pi)\exp(-n\pi\, t^{1/2}/x'^{1/2})}{1+\alpha t}\,dt\Big)$$

$$= 2\alpha^2\int_{-\infty}^{+\infty}\frac{\psi(t)\,dt}{1+\alpha t},$$

where

$$\psi(t) = \frac{G^2}{y'}\sum_{n=1}^{\infty} f\left(\left(\frac{x'}{y'}\right)^{1/2} n\pi\right)$$

$$\times\exp[-n\pi(-t)^{1/2}/(y^2)^{1/2}]\,\theta(-t) + \frac{G^2}{x'}$$

$$\times\sum_{n=1}^{\infty} f\left(\left(\frac{y'}{x'}\right)^{1/2} n\pi\right)\exp(-n\pi\, t^{1/2}/x'^{1/2})\,\theta(t), \qquad (2.7)$$

the moments being

$$A_m = \int_{-\infty}^{+\infty} t^m\,\psi(t)\,dt$$

$$= 2(2m+1)!\sum_{n=1}^{\infty}\frac{G^2}{n^2\pi^2}\left[f\left(\left(\frac{x'}{y'}\right)^{1/2} n\pi\right)\left(\frac{x'}{n^2\pi^2}\right)^m\right]. \qquad (2.8)$$

The following statements are now immediate consequences of Lemma 2.1.

(a) The perturbation expansion of $L_I(\alpha)$ in power series of α has vanishing radius of convergence. It is indeed enough to remark that, from (2.4), the formal expansion of $L_I(\alpha)$ is given by

$$L_I(\alpha) = 2\alpha^2\sum_{m=1}^{\infty} A_m(-\alpha)^m \qquad (2.9)$$

and from (2.8) it follows that

$$\lim_{n\to\infty} A_n^{1/n} = +\infty.$$

(b) An integral of the type (2.4) defines, *a priori*, two analytic functions: one in the upper half-plane $\mathrm{Im}\alpha < 0$, and another in the lower half-plane $\mathrm{Im}\alpha > 0$. In our case, however, one function can be analytically continued into the other, since $\psi(t)$ defined by (2.7) is piecewise analytic, and this implies that the real axis is not a natural boundary.

To prove our statements about the failure of the Stieltjes method in summing the divergent expansion (2.9) and the nonconvergence of the Padé approximants to the exact solution, we need also the following:

Lemma 2.2: The Hamburger moment problem

$$A_n = \int_{-\infty}^{+\infty} t^n\,\phi(t)\,dt, \qquad (2.10)$$

where A_n are given by (2.8), and $\phi(t)$ is a nonnegative function in $(-\infty, +\infty)$, is indeterminate.

Proof: It will be shown by explicit construction that there exists at least one function other than $\psi(t)$ defined by (2.7) whose moments coincide with A_n. It is indeed easily seen that there exists $A > 0$ such that $\psi(t) > Ae^{-B(|t|)^{1/2}}$, where $B = \pi/\min(\sqrt{y'}, \sqrt{x'})$. Now, taking into account the well-known integrals[8]

$$\int_{-\infty}^{+\infty} t^n e^{-kt^{2/3}}\cos(k\sqrt{3}\,t^{2/3})\,dt = 0,\quad n = 0,1,2,\ldots,$$
$$k > 0,$$

we have that $\psi_1(t) = \psi(t) + Ae^{-B}e^{-Bt^{2/3}}\cos(\sqrt{3}\,Bt^{2/3})$ is positive in $(-\infty, +\infty)$ and

$$\int_{-\infty}^{+\infty}\psi_1(t)\,t^n dt = \int_{-\infty}^{+\infty}\psi(t)\,t^n dt = A_n.$$

As a direct consequence of the former lemmas, we have now the following:

Theorem 2.1: The divergent perturbation expansion (2.9) does not sum to $L_1(\alpha)$ under the Stieltjes method, i.e., the Stieltjes type continued fraction associated with the power series (2.9) does not converge to $L_1(\alpha)$. This implies that no diagnonal sequences of Padé approximants to (2.9) converge to the exact solution.

Proof: We know that $L_1(\alpha)$ has the representation (2.4), where $\psi(t)$ is a particular solution of the indeterminate Hamburger moment problem (2.10). Then, by a well-known theorem of Hamburger,[9] for the Stieltjes type continued fraction associated with the series (2.9) which exists because the moment problem has solutions, one of the two following cases holds:

(i) The continued fractions is divergent, or
(ii) the continued fraction converges to a meromorphic function in the whole complex α plane.

Since we know that $L_1(\alpha)$ is not meromorphic, in any case the continued fraction does not converge to it, and neither do the odd and even approximants of the continued fraction, the $[N,N]$ and $[N,N-1]$ sequences of Padé approximants. The same result holds for any other diagonal sequence by a theorem of Wall.[10]

Let us close this section by indicating how it can be recovered in this framework the Stieltjes summability obtained in Ref. 5 with a different method, for the particular case of a pure electric or pure magnetic field.

A pure magnetic field is obtained when $G=0$, $F=\frac{1}{2}\mathbf{H}^2>0$, i.e., through (2.2), $y=0, x=esH$. Alternatively one has a pure electric field when $x=0, y=esE$. Consider now only the pure magnetic field, since for the pure electric one our considerations are the same.

By taking the constant H equal to 1, formula (2.3) becomes

$$L_I^H = -\frac{1}{8\pi^2}\int_0^\infty \frac{e^{-s}}{s^3}\left[(es)\coth(es) - 1 - \tfrac{1}{3}(es)^2\right]ds, \tag{2.11}$$

and its divergent Taylor expansion in powers of α is given by

$$L_I^H(\alpha) = -\frac{1}{8\pi^2}\sum_{n=2}^{\infty}(8\pi)^{2n}B_{2n}\frac{(2n-3)!}{(2n)!}\alpha^n, \tag{2.12}$$

where B_{2n} are the Bernoulli numbers. The same procedure worked out for $L_I(\alpha)$ shows that $(L_I^H(\alpha))/(\alpha^2)$ may be written under the form

$$\frac{L_I^H(\alpha)}{\alpha^2} = -\int_0^\infty \frac{\sigma(t)\,dt}{1+\alpha t}, \tag{2.13}$$

where

$$\sigma(t) = \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n^2\pi^2}e^{-n\sqrt{\pi}t/2} > 0, \qquad 0 \le t < \infty, \tag{2.14}$$

its moments being of course given by

$$a_n = \int_0^\infty t^n\sigma(t)\,dt = \frac{(-1)^{n+1}}{8\pi^2}(8\pi)^{2(n+2)}B_{2(n+2)}\frac{(2n+1)!}{[2(n+2)]!}$$

Now, since $B_{2n} \sim (-1)^{n-1}(2n)!/(2^{2n-1}\pi^{2n})$ as $n\to\infty$,[11] the Carleman criterion[8] $\sum_{n=0}^{\infty} 1/\sqrt[2n]{a_n} = \infty$ is satisfied so that the Stieltjes moment problem $a_n = \int_0^\infty t^n\sigma(t)\,dt$ is determined. As it is well known, this implies the convergence to $(L_1^H(\alpha))/(\alpha^2)$ of the Stieltjes type continued fraction associated with its divergent expansion, i.e., the Stieltjes summability, and, equivalently, the convergence to $L_1^H(\alpha)$ of any $[N,N+j]$, $j\ge 1$, sequence of Padé approximants to the divergent series (2.12).

3. BOREL SUMMABILITY AND CONVERGENCE OF THE RELATED APPROXIMATION METHODS

We have seen so far that the Stieltjes method fails, in the general case, to sum the divergent perturbation expansion to the exact solution. This is not the case for the Borel one, as we will now prove. From (2.6) we have

$$L_I(\alpha) = 2\alpha^2\int_0^\infty e^{-\sqrt{s'}}G^2 \times \sum_{n=1}^{\infty}\left(\frac{f((x'/y')^{1/2}n\pi)}{n^2\pi^2}\frac{1}{1-[(\alpha s'y')/(n^2\pi^2)]} + \frac{f((y'/x')^{1/2}n\pi)}{n^2\pi^2}\frac{1}{1+[(\alpha s'x')/(n^2\pi^2)]}\right). \tag{3.1}$$

We can write

$$L_I(\alpha) = 2\alpha^2\int_0^\infty e^{-\sqrt{s'}}F(\alpha s')\,ds', \tag{3.2}$$

where $F(\alpha)$ is given by the Stieltjes transform

$$F(\alpha) = \int_{-\infty}^{+\infty}\frac{d\rho(t)}{1+\alpha t}\,dt, \tag{3.3}$$

the function $\rho(t)$, bounded and increasing for $-\infty < t < \infty$, being of course defined as follows:

$$\rho(t) = G^2\sum_{n=1}^{\infty}\left[\frac{f((x'/y')^{1/2}n\pi)}{n^2\pi^2}\theta\left(t+\frac{y'}{n^2\pi^2}\right) + \frac{f((y'/x')^{1/2}n\pi)}{n^2\pi^2}\theta\left(t-\frac{x'}{n^2\pi^2}\right)\right]. \tag{3.4}$$

It is a simple matter to see that the Stieltjes transform (3.3) defines a meromorphic function in the whole complex α plane, having simple poles with positive residues at $\alpha = -(n^2\pi^2)/(x')$, $n=1,2,\cdots$; $\alpha = +(n^2\pi^2)/(y')$, $n=1,2,\ldots$, and that its convergent Taylor expansion around $\alpha=0$ is given by

$$F(\alpha) = 2\sum_{m=0}^{\infty}(-\alpha)^m\sum_{n=1}^{\infty}\frac{G^2}{n^2\pi^2}\left[f\left(\left(\frac{x'}{y'}\right)^{1/2}n\pi\right)\left(-\frac{y'}{n^2\pi^2}\right)^m + f\left(\left(\frac{y'}{x'}\right)^{1/2}n\pi\right)\left(\frac{x'}{n^2\pi^2}\right)^m\right]. \tag{3.5}$$

Defining, as usual, the Borel transform of order 2 of the formal perturbation expansion (2.9) of $(L_I(\alpha))/(2\alpha^2)$ through the power series

$$\frac{L_I^B(\alpha)}{2\alpha^2} = \sum_{m=0}^{\infty}\frac{A_m}{(2m+1)!}(-\alpha)^m, \tag{3.6}$$

formula (2.8) shows that the convergent expansions (3.5) and (3.6) coincide, so that

$$\frac{L_I^B(\alpha)}{2\alpha^2} \equiv F(\alpha). \tag{3.7}$$

The Borel sum of order 2 of the series (2.9) is defined by the integral $\int_0^\infty e^{-\sqrt{s'}}(L_I^B(\alpha s))/(2\alpha^2 s^2)ds$; then, through (3.7) and (3.2), since this last integral is uniformly and absolutely convergent in any compact having no intersection with the real axis, we can conclude:

Theorem 3.1: The divergent perturbation expansion (2.9) is Borel summable of order 2 to the exact solution $L_I(\alpha)$ in the whole α plane cut along the real axis.

Let us now prove the convergence to the solution of the approximation method proposed in Ref. 7, which consists in taking the Padé approximants on the Borel transform analytic at the origin.

In this procedure, thus, approximants to $(L_I(\alpha))/(2\alpha^2)$ are defined by

$$f_B^{[N,N+j]}(\alpha) = \int_0^\infty e^{-\sqrt{s}} F^{[N,N+j]}(\alpha s) ds, \quad (3.8)$$

where $F^{[N,N+j]}(\alpha)$ are the $[N, N+j]$ Padé approximants to the Borel transform $F(\alpha)$ defined by (3.3).

We have now the following:

Theorem 3.2:

$$\lim_{N\to\infty} f_B^{[N,N+j]}(\alpha) = L_I(\alpha), \quad j \geq 1 \quad (3.9)$$

uniformly in any compact of the α plane.

Proof [12]*:* Let us remark that we have to prove only the convergence of $f_B^{[N,N+j]}(\alpha)$ to $(L_I(\alpha))/(2\alpha^2)$; as it is well known, this implies the convergence of any sequence $f_B^{[N,N+j]}$, $j > -1$, and then also the validity of (3.9) for any $j \geq 1$, since the $f_B^{[N,N+j+2]}$ approximants to $(L_I(\alpha))/(2\alpha^2)$ are the $f_B^{[N,N+j]}$ ones to $L_I(\alpha)$. Now, since $F(\alpha)$ is meromorphic in the whole α plane having simple poles with positive residues along the real axis, it is known[13] that

$$\lim_{N\to\infty} F^{[N,N-1]}(\alpha) = F(\alpha) \quad (3.10)$$

uniformly in any compact containing none of the poles of $F(\alpha)$. Now for α in the cut plane, the boundedness of $|F(\alpha)|$ as $|\alpha| \to \infty$ as well as that of $|F^{[N,N-1]}(\alpha)|$ for any N are clearly sufficient to justify in (3.8) the interchange of the limit $N \to \infty$ with the integral, in spite of the nonuniformity of the convergence of $F^{[N,N-1]}(\alpha)$ to $F(\alpha)$ at infinity. We can then conclude

$$\lim_{N\to\infty} F_B^{[N,N-1]}(\alpha) = L_I(\alpha)/2\alpha^2,$$

uniformly in any compact of the α plane cut along the real axis.

Let us address ourselves now to the question of the convergence of the generalized Padé approximants, introduced in Ref. 6 and proposed in Ref. 14, in the framework of a superposition of the Stieltjes and Borel summation methods, as an approximation method for functions whose divergent Taylor expansion is of Stieltjes type, but not Stieltjes summable. We will now briefly recall the method of Ref. 14, to which the reader is referred for a detailed treatment.

A formal power series

$$\sum_{n=0}^{\infty} C_n z^n \quad (3.11)$$

is said to be $(S,B;m)$ summable if and only if its Borel transform of order m, defined as

$$\sum_{n=0}^{\infty} \frac{C_n}{(mn)!} z^n \quad (3.12)$$

is Stieltjes summable. Since, as it is well known, there exists in this case one and only one positive measure $q(x)$ in $[0,\infty)$, such that the Stieltjes sum of (3.12) can be written in the form

$$F_m(z) = \int_0^\infty \frac{dq(x)}{1+zx}, \quad (3.13)$$

the $(S,B;m)$ sum of (3.11) has the expression

$$f(z) = \int_0^\infty e^{-a} da \int_0^\infty \frac{dq(x)}{1+a^m zx} = \int_0^\infty e^{-a} F_m(za^m) da. \quad (3.14)$$

The generalized Padé approximants to the series (3.11) are now defined as follows:

$$f_m^{[N,N+j]}(z) = \int_0^\infty e^{-a} F_m^{[N,N+j]}(za^m) da, \quad (3.15)$$

where, as usual, the $[N, N+j]$ Padé approximants to $F_m(z)$ are indicated by $F_m^{[N,N+j]}(z)$.

We now have

Theorem 3.3: The divergent perturbation expansion $\sum_{m=0}^{\infty} A_m(-\alpha)^m$ is $(S,B;1)$ summable to $(L_I(\alpha))/(2\alpha)^2$; or, equivalently, any $[N, N+j], j \geq 1$, sequence of generalized Padé approximants converges to $L_I(\alpha)$, uniformly in any compact of the α plane cut along the real axis.

Proof: We have to show the Stieltjes summability of the Borel transform of order 1 of the expansion of $(L_I(\alpha))/(2\alpha^2)$, i.e., of the series

$$\sum_{m=0}^{\infty} \frac{Am}{m!}(-\alpha)^m = \sum_{m=0}^{\infty} \frac{(2m+1)!}{m!}(-\alpha)^m \sum_{n=1}^{\infty} \frac{G^2}{m^2\pi^2} \times \left[\frac{1}{x'} f\left(\left(\frac{y'}{x'}\right)^{1/2} n\pi\right)\left(\frac{\alpha x'}{n^2\pi^2}\right)^m + \frac{1}{y'} f\left(\left(\frac{x'}{y'}\right)^{1/2} n\pi\right)\left(-\frac{\alpha y'}{n^2\pi^2}\right)^m\right]. \quad (3.16)$$

It is easily seen that the above series is the divergent Taylor expansion of the function

$$\frac{L_I^{B1}(\alpha)}{2\alpha^2} = \int_{-\infty}^{+\infty} \frac{\chi(t) dt}{1+\alpha t}; \quad (3.17)$$

where

$$\chi(t) = \frac{2t}{y'} \sum_{n=1}^{\infty} \frac{f((y'/x')^{1/2} n\pi)}{n\pi} e^{-(n^2\pi^2/4y')} \theta(-t) + \frac{2t}{x'} \sum_{n=1}^{\infty} \frac{f((x'/y')^{1/2} n\pi)}{n\pi} e^{-(n^2\pi^2/4x')t} \theta(t) > 0, \quad -\infty < t < \infty. \quad (3.18)$$

The analyticity properties of (3.17) are, of course, the same as those of $L_I(\alpha)$.

Now, since one has trivially

$$\sum_{m=0}^{\infty} \left(\frac{A_m}{m!}\right)^{-1/m} = \infty,$$

the Hamburger moment problem for the coefficients $A_m/m!$ is determined by the Carleman criterion. The same Hamburger theorem employed in the proof of (2.3) ensures then the convergence to $(L^{B1}(\alpha))/(2\alpha^2)$ of the Stieltjes type continued fraction associated with the series (3.16), i.e., the Borel transform is Stieltjes summable. As we know, this means the convergence of the $[N, N]$ and $[N, N-1]$ Padé sequence, $j > 0$, is also convergent to $(L^{B1}(\alpha))/(2\alpha^2)$. Then, proceeding in exactly the same way as in Theorem 3.2, we can conclude that any sequence $[N, N+j]$ of generalized Padé approximants $f_1^{[N,N+j]}$, $j \geq 1$, converges to $L_I(\alpha)$, uniformly in any compact of the cut α plane.

---

*Fulbright Scholar. Supported in part by funds from the National Science Foundation.

† Address after July 1, 1972: Istituto Naz. di Fisica Nucleare, Sez. di Bologna, Via Imerio 46, 40126 Bologna, Italy.

[1] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[2] Standard textbook references on summation methods are: E. Borel, *Leçons sur les series divergentes* (Gauthier-Villars, Paris, 1928); H. S. Wall, *Continued Fractions* (Van Nostrand, New York, 1947); G. H. Hardy, *Divergent Series* (Oxford U.P., Oxford, 1948).
[3] For a recent review of results on the subject, see B. Simon in Proceedings of the 1972 Coral Gables Conference on Fundamental Interactions (Gordon and Breach, London), to be published.
[4] V. I. Ogievetsky, Dokl. Akad. Nauk SSSR **109**, 919 (1956).
[5] S. Graffi, N.Y.U. Technical Report 25/71 (to be published in J. Math. Phys.).
[6] G. A. Baker, Phys. Rev. **161**, 434 (1967).
[7] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[8] J. Shohat and J. D. Tamarkin, *The Problem of Moments* (Amer. Math. Soc., Providence, R.I., 1943).
[9] H. Hamburger, Math. Ann. **81**, 235 (1920).
[10] H. S. Wall, Trans. Am. Math. Soc. **33**, 511 (1931).
[11] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (New York and London, 1965).
[12] This proof is identical to the one given in Ref. 5 for the particular case of a pure magnetic or a pure electric field. It is presented here for completeness.
[13] H. S. Wall, Ref. 2, Chap. 18.
[14] S. Graffi, V. Grecchi, and G. Turchetti, Nuovo Cimento B **4**, 131 (1971).

Radiative transfer in adjacent half-spaces with specular reflection

Y. Yener and M. N. Özişik
Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina 27607
(Received 7 April 1972; Revised Manuscript Received 5 June 1972)

Radiative transfer in absorbing, emitting, isotropically scattering, gray adjacent half-spaces with specular reflection at the interface is solved rigorously by the application of normal-mode expansion technique. An alternative method based on the superposition of half-space problems is also presented for the solution of the problem. Once the expansion coefficients for the solutions are evaluated from the given relations, the physical quantities such as the angular distribution of radiation intensity and the net radiative heat flux anywhere in the medium can readily be determined.

I. INTRODUCTION

The problem of radiative transfer in adjacent media which absorb, emit, and scatter radiation is of interest in many applications. Davison[1] and Chandrasekhar[2] were among the earliest investigators who studied the angular distribution of radiation at the interface of two adjoining media. In the field of neutron transport theory, related problems have been investigated[3–6] for the cases involving a transparent interface. In the present study, radiative transfer in absorbing, emitting, isotropically scattering, gray, adjacent half spaces with specular reflection at the interface is solved rigorously using Case's[7] normal-mode expansion technique. An alternative method of treatment based on the superposition of half-space problems is developed for the solution of this problem. The present analysis have the advantage that, once the expansion coefficients are determined for a given set of parameters, the physical quantities such as the angular distribution of the radiation intensity and the net radiative heat flux anywhere in the medium are immediately determined.

In Sec. II we present a rigorous solution of the radiation problem in two adjacent half spaces for the nonconservative (i.e., $\omega_1 < 1, \omega_2 < 1$) case, while in Sec. III we solve the problem for a combination of a conservative and a nonconservative (i.e., $\omega_1 < 1$, $\omega_2 = 0$) media. Finally, in Sec. IV we present the method of superposition.

The two-region radiative transfer considered here may find applications in boundary layer heat transfer involving two different streams separated by a semitransparent barrier, in solidification and melting problems, and in two-region heat conduction when radiation effects are important in such problems.

II. NONCONSERVATIVE CASE ($\omega_1 < 1$, $\omega_2 < 1$)

The equations of radiative transfer for absorbing, emitting, isotropically scattering adjacent half spaces are given as

$$\mu \frac{\partial I_1(\tau,\mu)}{\partial \tau} + I_1(\tau,\mu) = S_1(\tau) + \frac{\omega_1}{2}\int_{-1}^{1} I_1(\tau,\mu')d\mu', \quad \tau < 0, \tag{1a}$$

$$\mu \frac{\partial I_2(\tau,\mu)}{\partial \tau} + I_2(\tau,\mu) = S_2(\tau) + \frac{\omega_2}{2}\int_{-1}^{1} I_2(\tau,\mu')d\mu', \quad \tau > 0. \tag{1b}$$

In writing the boundary conditions for this problem, consideration will be given to reflection of radiation at the interface. According to predictions by the classical electromagnetic theory, reflection and refraction at the interface vary with direction. However, because of the complexity of analysis, the reflectivity is generally taken independent of direction even for one region problems. For the more involved two-region problem considered here, it is reasonable to assume constant reflectivity and neglect the condensation effects. With this consideration we write the boundary conditions for the above problem as

$$I_1(0,\mu) = \rho_1 I_1(0,-\mu) + \Gamma_2 I_2(0,\mu), \quad \mu < 0, \tag{2a}$$

$$I_2(0,\mu) = \rho_2 I_2(0,-\mu) + \Gamma_1 I_1(0,\mu), \quad \mu > 0, \tag{2b}$$

and at infinity as

$$\lim_{|\tau|\to\infty} I_i(\tau,\mu) \to P_i(\tau,\mu), \quad i = 1 \text{ or } 2, \tag{2c}$$

where $I_i(\tau,\mu)$ is the intensity of radiation, τ is the optical variable, ω_i is the single scattering albedo, and μ is the cosine of the angle between the positive τ direction and the directed intensity. $S_i(\tau) = (1-\omega_i) \times n_i^2\sigma T_i^4(\tau)/\pi$ is the inhomogeneous source term, $P_i(\tau,\mu)$ is the corresponding particular solution, σ is the Stefan–Boltzmann constant, T is the temperature, and ρ_i and Γ_i, $i = 1, 2$, are the reflectivity and transmissivity of the interface. We note that $\Gamma_1 = (1-\rho_1)/n^2$ and $\Gamma_2 = (1-\rho_2)/n^2$, where n is the relative refractive index (i.e., n_1/n_2). When $n = 1$, both media have identical refractive indices, then reflectivity vanishes and the interface becomes transparent. When the two media have identical refractive indices but separated by a thin, semitransparent, reflecting layer, the above boundary conditions with reflection are still valid, but with $n = 1$.

We proceed to write the desired solutions of the equations of radiative transfer as a linear sum of the normal modes and a particular solution in the form[7]

$$I_1(\tau,\mu) = A(-\nu_0)\phi_1(-\nu_0,\mu)e^{\tau/\nu_0} + \int_0^1 A(-\eta)\phi_1(-\eta,\mu)e^{\tau/\eta}d\eta + P_1(\tau,\mu), \quad \tau < 0,\ \mu \in (-1,1), \tag{3a}$$

$$I_2(\tau,\mu) = A(\eta_0)\phi_2(\eta_0,\mu)e^{-\tau/\eta_0} + \int_0^1 A(\eta)\phi_2(\eta,\mu)e^{-\tau/\eta}d\eta + P_2(\tau,\mu), \quad \tau > 0,\ \mu \in (-1,1). \tag{3b}$$

In writing Eqs. (3) we have omitted those elementary solutions that diverge at infinity, thus the resulting solutions satisfy the requirement of Eqs. (2c). Here, $A(-\nu_0)$, $A(-\eta)$, $A(\eta_0)$, and $A(\eta)$ are the expansion coefficients which are to be determined by constraining these solutions to meet the boundary conditions given by Eqs. (2a) and (2b). The normal modes are defined as[7,8]

$$\phi_i(\xi,\mu)=\frac{\omega_i\xi}{2}\frac{1}{\xi-\mu},\quad \xi=\begin{cases}-\nu_0, & i=1\\ \eta_0, & i=2\end{cases},\tag{4a}$$

$$\phi_i(n,\mu)=\tfrac{1}{2}\ \omega_i\eta[P/(\eta-\mu)]+\lambda_i(\eta)\delta(\eta-\mu),$$
$$\eta\in(-1,1)\quad i=1 \text{ or } 2,\tag{4b}$$

$$\lambda_i(\eta)=1-\omega_i\eta\tanh^{-1}\eta,\quad i=1\text{ or }2,\tag{4c}$$

and the discrete eigenvalues (ν_0,η_0) are the zeros of the dispersion function

$$\Lambda_i(z)=1+\omega_i\frac{z}{2}\int_{-1}^{1}\frac{d\mu}{\mu-z}=1-\omega_i z\tanh^{-1}\left(\frac{1}{z}\right),$$
$$i=1\text{ or }2.\tag{4d}$$

Here P is a mnemonic symbol used to denote that the ensuing integral is evaluated in the Cauchy principal-value sense, and $\delta(x)$ is the Dirac delta function.

Before pursuing the analysis, we note that once the intensity of radiation is determined, the net radiative heat flux $q(\tau)$ is obtained from

$$q_i(\tau)=2\pi\int_{-1}^{1}I_i(\tau,\mu)\mu\,d\mu,\quad i=1\text{ or }2.\tag{5}$$

Analysis: The introduction of the solutions given by Eqs. (3) into Eqs. (2a) and (2b), respectively, yields

$$\begin{aligned}&M_1(\mu)+\rho_1[A(-\nu_0)\phi_1(-\nu_0,\mu)\\&\quad+\int_0^1 A(-\eta')\phi_1(-\eta',\mu)d\eta']\\&\quad+\Gamma_2\left(A(\eta_0)\phi_2(-\eta_0,\mu)+\int_0^1 A(\eta')\phi_2(-\eta',\mu)d\eta'\right)\\&=A(-\nu_0)\phi_1(\nu_0,\mu)+\int_0^1 A(-\eta')\phi_1(\eta'\mu)d\eta',\\&\mu>0,\end{aligned}\tag{6a}$$

$$\begin{aligned}&M_2(\mu)+\rho_2\left(A(\eta_0)\phi_2(-\eta_0,\mu)+\int_0^1 A(\eta')\phi_2(-\eta',\mu)d\eta'\right)\\&\quad+\Gamma_1\Big(A(-\nu_0)\phi_1(-\nu_0,\mu)\\&\quad+\int_0^1 A(-\eta')\phi_1(-\eta',\mu)d\eta'\Big)\\&=A(\eta_0)\phi_2(\eta_0,\mu)+\int_0^1 A(\eta')\phi_2(\eta',\mu)d\eta',\\&\mu>0,\end{aligned}\tag{6b}$$

where

$$M_1(\mu)\equiv\rho_1P_1(0,\mu)-P_1(0,-\mu)+\Gamma_2P_2(0,-\mu),\quad \mu>0,\tag{7a}$$

$$M_2(\mu)\equiv\rho_2P_2(0,-\mu)-P_2(0,\mu)+\Gamma_1P_1(0,\mu),\quad \mu>0,\tag{7b}$$

We note by the half-range completeness theorem[7] that an arbitrary Hölder function defined in the interval $\mu\in(0,1)$ can be expanded in terms of $\phi_1(\nu_0,\mu)$, $\phi_1(\eta,\mu)$ or $\phi_2(\eta_0,\mu)\phi_2(\eta,\mu)$, and the right-hand side of Eqs. (6) are such expansions. Equations (6a) and (6b) are two coupled singular integral equations; they may, however, be transformed to coupled regular Fredholm integral equations by making use of the orthogonality properties of the normal modes and of the results of various normalization integrals summarized in the Appendix.

We operate Eq. (6a) first by the operator $\int_0^1 \mu H_1(\mu)\times\phi_1(\nu_0,\mu)d\mu$ and then by $\int_0^1\mu H_1(\mu)\phi_1(\eta,\mu)d\mu$ and, respectively, obtain

$$\begin{aligned}&\left(\frac{1}{\nu_0}N_1(\nu_0)-\rho_1\frac{\phi_1(-\nu_0,\nu_0)}{H_1(\nu_0)}\right)A(-\nu_0)\\&\quad-\Gamma_2\frac{\phi_2(-\eta_0,\nu_0)}{H_1(\eta_0)}A(\eta_0)=G_1(\nu_0)\\&\quad+\int_0^1\left(\rho_1\frac{\phi_1(-\eta',\nu_0)}{H_1(\eta')}A(-\eta')\right.\\&\quad\left.+\Gamma_2\frac{\phi_2(-\eta',\nu_0)}{H_1(\eta')}A(\eta')\right)d\eta',\end{aligned}\tag{8a}$$

$$\begin{aligned}&\frac{1}{\eta}N_1(\eta)A(-\eta)=G_1(\eta)+\rho_1\frac{\phi_1(-\nu_0,\eta)}{H_1(\nu_0)}A(-\nu_0)\\&\quad+\Gamma_2\frac{\phi_2(-\eta_0,\eta)}{H_1(\eta_0)}A(\eta_0)\\&\quad+\int_0^1\left(\rho_1\frac{\phi_1(-\eta',\eta)}{H_1(\eta')}A(-\eta')\right.\\&\quad\left.+\Gamma_2\frac{\phi_2(-\eta',\eta)}{H_1(\eta')}A(\eta')\right)d\eta'.\end{aligned}\tag{8b}$$

Now we operate Eq. (6b) first by the operator $\int_0^1\mu H_2(\mu)\times\phi_2(\eta_0,\mu)d\mu$ and then by $\int_0^1\mu H_2(\mu)\phi_2(\eta,\mu)d\mu$ and, respectively, obtain

$$\begin{aligned}&\left(\frac{1}{\eta_0}N_2(\eta_0)-\rho_2\frac{\phi_2(-\eta_0,\eta_0)}{H_2(\eta_0)}\right)A(\eta_0)\\&\quad-\Gamma_1\frac{\phi_1(-\nu_0,\eta_0)}{H_2(\nu_0)}A(-\nu_0)\\&\quad=G_2(\eta_0)+\int_0^1\left(\Gamma_1\frac{\phi_1(-\eta',\eta)}{H_2(\eta')}A(-\eta')\right.\\&\quad\left.+\rho_2\frac{\phi_2(-\eta',\eta_0)}{H_2(\eta')}A(\eta')\right)d\eta',\end{aligned}\tag{9a}$$

$$\begin{aligned}&\frac{1}{\eta}N_2(\eta)A(\eta)=G_2(\eta)+\Gamma_1\frac{\phi_1(-\nu_0,\eta)}{H_2(\nu_0)}A(-\nu_0)\\&\quad+\rho_2\frac{\phi_2(-\eta_0,\eta)}{H_2(\eta_0)}A(\eta_0)\\&\quad+\int_0^1\left(\Gamma_1\frac{\phi_1(-\eta',\eta)}{H_2(\eta')}A(-\eta')\right.\\&\quad\left.+\rho_2\frac{\phi_2(-\eta',\eta)}{H_2(\eta')}A(\eta')\right)d\eta'.\end{aligned}\tag{9b}$$

Here we have defined

$$G_i(\xi)\equiv\frac{1}{\xi}\int_0^1\mu H_i(\mu)\phi_i(\xi,\mu)M_i(\mu)d\mu,\quad i=1\text{ or }2.\tag{10a}$$

The normalization integrals are given by

$$N_i(\xi) = \frac{\omega_i \xi^2}{2} H_i(\xi) \frac{d\Lambda_i(z)}{dz}\Bigg|_{z=\xi}, \quad \xi = \begin{cases} \nu_0, & i=1 \\ \eta_0, & i=2 \end{cases}, \tag{10b}$$

$$N_i(\eta) = \eta H_i(\eta)\Lambda_i^+(\eta)\Lambda_i^-(\eta), \; i = 1 \text{ or } 2, \quad \eta \in (0,1), \tag{10c}$$

where

$$\Lambda_i^+(\eta)\Lambda_i^-(\eta) = \lambda_i^2(\eta) + [\tfrac{1}{2}(\omega_i\eta\pi)]^2 \equiv 1/g_i(\omega_i,\eta), \tag{10d}$$

and the function $g_i(\omega_i, \eta)$ is tabulated.[9] Further, the $H_i(z)$ function for isotropic scattering is obtained from the solution of the nonlinear integral equation[10]

$$\frac{1}{H_i(z)} = 1 + \frac{\omega_i z}{2}\int_0^1 H_i(\mu)\frac{d\mu}{z+\mu}, \quad i = 1 \text{ or } 2, \tag{10e}$$

and it can be shown that

$$\frac{1}{H_i(-\xi)} = 0, \quad \xi = \begin{cases} \nu_0, & i=1 \\ \eta_0, & i=2 \end{cases}, \tag{10f}$$

$$\int_0^1 H_i(\mu)\phi_i(\xi,\mu)d\mu = 1,$$
$$\xi \in (-1,1) \text{ or } \xi = \begin{cases} \pm\nu_0, & i=1 \\ \pm\eta_0, & i=2 \end{cases}. \tag{10g}$$

Equations (8) and (9) provide four relations for the determination of the four unknown expansion coefficients. These equations can be written more concisely in matrix notation as

$$(\mathbf{N}_0 - \mathbf{K}_0)\mathbf{A}_0 = \mathbf{G}_0 + \int_0^1 \mathbf{K}_0(\eta')A(\eta')d\eta', \tag{11a}$$

$$\mathbf{N}(\eta)\mathbf{A}(\eta) = \mathbf{G}(\eta) + \mathbf{K}(\eta)\mathbf{A}_0 + \int_0^1 \mathbf{K}(\eta,\eta')\mathbf{A}(\eta')d\eta', \tag{11b}$$

where we have defined

$$\mathbf{A}_0 \equiv \begin{bmatrix} A(-\nu_0) \\ A(\eta_0) \end{bmatrix}, \quad \mathbf{A}(\eta) \equiv \begin{bmatrix} A(-\eta) \\ A(\eta) \end{bmatrix}, \tag{12a}$$

$$\mathbf{G}_0 \equiv \begin{bmatrix} G_1(\nu_0) \\ G_2(\eta_0) \end{bmatrix}, \quad \mathbf{G}(\eta) \equiv \begin{bmatrix} G_1(\eta) \\ G_2(\eta) \end{bmatrix}, \tag{12b}$$

$$\mathbf{N}_0 \equiv \begin{bmatrix} \dfrac{N_1(\nu_0)}{\nu_0} & 0 \\ 0 & \dfrac{N_2(\eta_0)}{\eta_0} \end{bmatrix}, \quad \mathbf{N}(\eta) \equiv \frac{1}{\eta}\begin{bmatrix} N_1(\eta) & 0 \\ 0 & N_2(\eta) \end{bmatrix}, \tag{12c}$$

$$\mathbf{K}_0 \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\nu_0,\nu_0)}{H_1(\nu_0)} & \Gamma_2 \dfrac{\phi_2(-\eta_0,\nu_0)}{H_2(\eta_0)} \\ \Gamma_1 \dfrac{\phi_1(-\nu_0,\eta_0)}{H_2(\nu_0)} & \rho_2 \dfrac{\phi_2(-\eta_0,\eta_0)}{H_2(\eta_0)} \end{bmatrix}, \tag{12d}$$

$$\mathbf{K}_0(\eta) \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\eta,\nu_0)}{H_1(\eta)} & \Gamma_2 \dfrac{\phi_2(-\eta,\nu_0)}{H_1(\eta)} \\ \Gamma_1 \dfrac{\phi_1(-\eta,\eta_0)}{H_2(\eta)} & \rho_2 \dfrac{\phi_2(-\eta,\eta_0)}{H_2(\eta)} \end{bmatrix}, \tag{12e}$$

$$\mathbf{K}(\eta) \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\nu_0,\eta)}{H_1(\nu_0)} & \Gamma_2 \dfrac{\phi_2(-\eta_0,\eta)}{H_1(\eta_0)} \\ \Gamma_1 \dfrac{\phi_1(-\nu_0,\eta)}{H_1(\nu_0)} & \rho_2 \dfrac{\phi_2(-\eta_0,\eta)}{H_2(\eta_0)} \end{bmatrix}, \tag{12f}$$

$$\mathbf{K}(\eta,\eta') \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\eta',\eta)}{H_1(\eta')} & \Gamma_2 \dfrac{\phi_2(-\eta',\eta)}{H_1(\eta')} \\ \Gamma_1 \dfrac{\phi_1(-\eta',\eta)}{H_2(\eta')} & \rho_2 \dfrac{\phi_2(-\eta',\eta)}{H_2(\eta')} \end{bmatrix}. \tag{12g}$$

In the above equations, the components of the vectors $\mathbf{G}_0$ and $\mathbf{G}(\eta)$ involve the particular solutions $P_1(\tau,\mu)$ and $P_2(\tau,\mu)$ which depend on the type of the inhomogeneous source terms $S_i(\tau)$, $i = 1$ or 2, of the equations of radiative transfer equations (1a) and (1b). Once the type of the source term is specified, a particular solution of the equation of radiative transfer can be constructed.[11]

Equations (11) are regular, Fredholm type integral equations which can be solved numerically for the four unknown expansion coefficients. Once these expansion coefficients are determined, the intensity of radiation is evaluated from Eq. (3) and the net radiative heat flux according to Eq. (5), i.e.,

$$q_1(\tau) = 2\pi(1-\omega_1)\Big(-A(-\nu_0)\nu_0 e^{\tau/\nu_0} - \int_0^1 A(-\eta)\eta e^{\tau/\eta}d\eta + (1-\omega_1)^{-1}\int_{-1}^1 P_1(\tau,\mu)\mu\, d\mu\Big), \tag{13a}$$

$$q_2(\tau) = 2\pi(1-\omega_2)\Big(A(\eta_0)\eta_0 e^{-\tau/\eta_0} + \int_0^1 A(\eta)\eta e^{-\tau/\eta}d\eta + (1-\omega_2)^{-1}\int_{-1}^1 P_2(\tau,\mu)\mu\, d\mu\Big). \tag{13b}$$

Here we have used the relation

$$\int_{-1}^1 \mu\phi_i(\xi,\mu)d\mu = \xi(1-\omega_i),$$
$$\xi \in (-1,1) \text{ or } \xi = \begin{cases} -\nu_0 & i=1 \\ \eta_0 & i=2 \end{cases} \tag{14}$$

Analytical Approximations: Up to this point our analysis has been mathematically rigorous, and the degree of precision of the final solution for the radiation intensity and the net radiative heat flux depends, of course, on how accurately the four expansion coefficients are determined. However, analytical approximations can also be obtained from Eqs. (11).

The first-order solution is obtained by neglecting the continuum coefficients entirely [i.e., $\mathbf{A}(\eta) \equiv 0$]; then the discrete coefficients are obtained from Eqs. (11a) as

$$\mathbf{A}_0^{(1)} = [\mathbf{N}_0 - \mathbf{K}_0]^{-1}\mathbf{G}_0. \tag{15}$$

The second-order solution is found by neglecting the contribution from the kernel $\mathbf{K}(\eta,\eta')$ in Eq. (11b), and by using in that equation the first-order solution for $\mathbf{A}_0^{(1)}$. Then, $\mathbf{A}^{(2)}(\eta)$ is obtained from Eq. (11b) as

$$\mathbf{A}^{(2)}(\eta) = \mathbf{N}^{-1}(\eta)[\mathbf{G}(\eta) + \mathbf{K}(\eta)\mathbf{A}_0^{(1)}], \qquad (16a)$$

and substitution of $\mathbf{A}^{(2)}(\eta)$ into Eq. (11a) yields $\mathbf{A}_0^{(2)}$ as

$$\mathbf{A}_0^{(2)} = [\mathbf{N}_0 - \mathbf{K}_0]^{-1}\left(\mathbf{G}_0 + \int_0^1 \mathbf{K}_0(\eta')\mathbf{A}^{(2)}(\eta')d\eta'\right). \qquad (16b)$$

Special Case $\omega_1 = \omega_2 < 1$: For this special case, Eqs. (11) are simplified since $\nu_0 = \eta_0$, $H_1(\xi) = H_2(\xi) \equiv H(\xi)$, $\phi_1(\xi,\mu) = \phi_2(\xi,\mu) \equiv \phi(\xi,\mu)$, and $N_1(\xi) = N_2(\xi) \equiv N(\xi)$, where $\xi = \eta_0$ or $\eta \in (-1,1)$:

$$\left(\frac{1}{\eta_0}N(\eta_0)\mathbf{I} - \frac{\phi(-\eta_0,\eta_0)}{H(\eta_0)}\boldsymbol{\rho}\right)\mathbf{A}(\eta_0) = \mathbf{G}_0 + \boldsymbol{\rho}\int_0^1 \mathbf{A}(\eta')\frac{\phi(-\eta',\eta_0)}{H(\eta')}d\eta', \qquad (17a)$$

$$\frac{1}{\eta}N(\eta)\mathbf{A}(\eta) = \mathbf{G}(\eta) + \frac{\phi(-\eta_0,\eta)}{H(\eta_0)}\boldsymbol{\rho}\mathbf{A}(\eta_0) + \boldsymbol{\rho}\int_0^1 \mathbf{A}(\eta')\frac{\phi(-\eta',\eta)}{H(\eta')}d\eta'. \qquad (17b)$$

where

$$\boldsymbol{\rho} \equiv \begin{bmatrix}\rho_1 & \Gamma_2\\ \Gamma_1 & \rho_2\end{bmatrix} \qquad (17c)$$

and $\mathbf{I}$ is the unit matrix.

III. THE CASE $\omega_1 < 1$, $\omega_2 = 1$

We now consider a situation in which medium 1 is non-conservative ($\omega_1 < 1$) and medium 2 is conservative ($\omega_2 = 1$). The equations of radiative transfer are given as

$$\mu\frac{\partial I_1(\tau,\mu)}{\partial\tau} + I_1(\tau,\mu) = S_1(\tau) + \frac{\omega_1}{2}\int_{-1}^1 I_1(\tau,\mu')d\mu', \quad \tau < 0, \qquad (18a)$$

$$\mu\frac{\partial I_2(\tau,\mu)}{\partial\tau} + I_2(\tau,\mu) = \tfrac{1}{2}\int_{-1}^1 I_2(\tau,\mu')d\mu', \quad \tau > 0, \qquad (18b)$$

subject to the boundary conditions (2a), (2b), and (2c) with $P_2(\tau,\mu)$ being set equal to the normal mode which is allowed by the boundary condition at infinity.

The solutions satisfying the boundary conditions at infinity are given as

$$I_1(\tau,\mu) = A(-\nu_0)\phi_1(-\nu_0,\mu)e^{\tau/\nu_0} + \int_0^1 A(-\eta)\phi_1(-\eta,\mu)e^{\tau/\eta}d\eta + P_1(\tau,\mu), \quad \tau < 0, \mu \in (-1,1), \qquad (19a)$$

$$I_2(\tau,\mu) = A + \int_0^1 A(\eta)\phi^*(\eta,\mu)e^{-\tau/\eta}d\eta, \quad \tau > 0,\ \mu \in (-1,1), \qquad (19b)$$

where the normal modes $\phi_1(-\nu_0,\mu)$ and $\phi_1(-\eta,\mu)$ are defined by Eqs. (4a) and (4b), respectively, and $\phi^*(\eta,\mu)$ is defined as

$$\phi^*(\eta,\mu) = \tfrac{1}{2}\eta\,[P/(\eta-\mu)] + \lambda^*(\eta)\delta(\eta-\mu), \qquad (20a)$$

where

$$\lambda^*(\eta) = 1 - \eta\tanh^{-1}\eta. \qquad (20b)$$

When the solutions given by Eqs. (19) are introduced into the boundary conditions (2a) and (2b), two coupled singular integral equations are obtained:

$$S_1(\mu) + \rho_1\left(A(-\nu_0)\phi_1(-\nu_0,\mu) + \int_0^1 A(-\eta')\phi_1(-\eta',\mu)d\eta'\right) + \Gamma_2\left(A + \int_0^1 A(\eta')\phi^*(-\eta',\mu)d\eta'\right) = A(-\nu_0)\phi_1(\nu_0,\mu) + \int_0^1 A(-\eta')\phi_1(\eta',\mu)d\eta', \qquad (21a)$$

$$S_2(\mu) + \Gamma_1\left(A(-\nu_0)\phi_1(-\nu_0,\mu) + \int_0^1 A(-\eta')\phi_1(-\eta',\mu)d\eta'\right) + \rho_2\int_0^1 A(\eta')\phi^*(-\eta',\mu)d\eta' = (1-\rho_2)A + \int_0^1 A(\eta')\phi^*(\eta',\mu)d\eta', \qquad (21b)$$

where

$$S_1(\mu) = \rho_1 P_1(0,\mu) - P_1(0,\mu), \qquad (22a)$$

$$S_2(\mu) = \Gamma_1 P_1(0,\mu). \qquad (22b)$$

These two coupled singular integral equations may be transformed to regular Fredholm integral equations by first operating Eq. (21a) by $\int_0^1 \mu H_1(\mu)\phi_1(\nu_0,\mu)d\mu$ and $\int \mu H_1(\mu)\phi_1(\eta,\mu)d\mu$, and then operating Eq. (21b) by $\int_0^1 \mu H^*(\mu)d\mu$ and $\int_0^1 \mu H^*(\mu)\phi^*(\eta,\mu)d\mu$. The resulting equations for the expansion coefficients can be written in the matrix form as

$$(\mathbf{N}_0^* - \mathbf{J}_0)\mathbf{A}_0^* = \mathbf{F}_0 + \int_0^1 \mathbf{J}_0(\eta')\mathbf{A}(\eta')d\eta', \qquad (23a)$$

$$\mathbf{N}^*(\eta)\mathbf{A}(\eta) = \mathbf{F}(\eta) + \mathbf{J}(\eta)\mathbf{A}_0^* + \int_0^1 \mathbf{J}(\eta,\eta')\mathbf{A}(\eta')d\eta', \qquad (23b)$$

where we have defined

$$\mathbf{A}^* \equiv \begin{bmatrix}A_1(-\nu_0)\\ A\end{bmatrix}, \quad \mathbf{A}(\eta) \equiv \begin{bmatrix}A(-\eta)\\ A(\eta)\end{bmatrix}, \qquad (24a)$$

$$\mathbf{F}_0 \equiv \begin{bmatrix}F_1(\nu_0)\\ F_2\end{bmatrix}, \quad \mathbf{F}(\eta) \equiv \begin{bmatrix}F_1(\eta)\\ F_2(\eta)\end{bmatrix}, \qquad (24b)$$

$$F_1(\xi) \equiv \frac{1}{\xi}\int_0^1 \mu H_1(\mu)\phi(\xi,\mu)S_1(\mu)d\mu, \qquad (24c)$$

$$F_2(\eta) \equiv \frac{1}{\eta}\int_0^1 \mu H^*(\mu)\phi^*(\eta,\mu)S_2(\mu)d\mu, \qquad (24d)$$

$$F_2 \equiv \int_0^1 \mu H^*(\mu)S_2(\mu)d\mu, \qquad (24e)$$

$$N_0^* \equiv \begin{bmatrix}\frac{1}{\nu_0}N_1(\nu_0) & 0\\ 0 & \frac{2}{\sqrt{3}}\end{bmatrix}, \quad N^*(\eta) \equiv \frac{1}{\eta}\begin{bmatrix}N_1(\eta) & 0\\ 0 & N^*(\eta)\end{bmatrix}, \qquad (25a)$$

$$N_1(\nu_0) = \frac{\omega_1\nu_0^2}{2} H_1(\nu_0) \left.\frac{d\Lambda_1(z)}{dz}\right|_{z=\nu_0}, \tag{25b}$$

$$N_1(\eta) = \eta H_1(\eta)\Lambda_1^+(\eta)\Lambda_1^-(\eta), \tag{25c}$$

$$N_1^*(\eta) = \eta H^*(\eta)\Lambda^+(\eta)\Lambda^-(\eta), \tag{25d}$$

$$\Lambda^+(\eta)\Lambda^-(\eta) = \lambda^{*2}(\eta) + \left(\frac{\eta\pi}{2}\right)^2 \equiv \frac{1}{g(1,\eta)}, \tag{25e}$$

$$\mathbf{J}_0 \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\nu_0,\nu_0)}{H_1(\nu_0)} & \Gamma_2(1-\omega_1)^{1/2} \\ \Gamma_1 \dfrac{\omega_1\nu_0}{H^*(\nu_0)} & \rho_2 \dfrac{2}{\sqrt{3}} \end{bmatrix}, \tag{26a}$$

$$\mathbf{J}_0(\eta) \equiv \begin{bmatrix} \omega_1\rho_1 \dfrac{\phi^*(-\eta',\nu_0)}{H_1(\eta')} & \Gamma_2 \dfrac{\phi^*(-\eta',\nu_0)}{H_1(\eta')} \\ \omega_1\Gamma_1 \dfrac{\eta'}{H^*(\eta')} & \rho_2 \dfrac{\eta'}{H^*(\eta')} \end{bmatrix}, \tag{26b}$$

$$\mathbf{J}(\eta) \equiv \begin{bmatrix} \rho_1 \dfrac{\phi_1(-\nu_0,\eta)}{H_1(\nu_0)} & \Gamma_2(1-\omega_1)^{1/2} \\ \Gamma_1 \dfrac{\phi_1(-\nu_0,\eta)}{H^*(\nu_0)} & 0 \end{bmatrix}, \tag{26c}$$

$$\mathbf{J}(\eta,\eta') \equiv \phi^*(-\eta',\eta) \begin{bmatrix} \dfrac{\rho_1\omega_1}{H_1(\eta')} & \dfrac{\Gamma_2}{H_1(\eta')} \\ \dfrac{\Gamma_1\omega_1}{H^*(\eta')} & \dfrac{\rho_2}{H^*(\eta')} \end{bmatrix}, \tag{26d}$$

and $H^*(z)$ function is the solution of

$$\frac{1}{H^*(z)} = 1 + \frac{z}{2}\int_0^1 H^*(\mu)\,\frac{d\mu}{z+\mu}\,. \tag{26e}$$

The solution of Eqs. (23) yields the unknown expansion coefficients. Analytical approximations can be obtained as described previously.

IV. SOLUTION BY SUPERPOSITION OF HALF-SPACE PROBLEMS

In this section we present a method of solution of the adjacent half-spaces problem considered in Sec. II by the superposition of the solutions of single half-space problems. Although this method eventually yields the same set of relations for the expansion coefficients as those given by Eqs. (11), it provides better insight to the physical significance of the results given by Eqs. (11) as well as to the physical significance of various orders of approximations obtained from them.

The radiation problem defined by Eqs. (1) and (2) are written more compactly in the form

$$L_j I_j(\tau,\mu) = S_j\,, \tag{27a}$$

where we have defined the operator L_j as

$$L_j I_j \equiv \mu\frac{\partial I_j}{\partial\tau} + I_j - \frac{\omega_j}{2}\int_{-1}^1 I_j\,d\mu' \tag{27b}$$

with

$$j = \begin{cases} 1, & \tau<0 \\ 2, & \tau>0 \end{cases}, \tag{27c}$$

subject to the boundary conditions

$$I_1(0,\mu) = \rho_1 I_1(0,-\mu) + \Gamma_2 I_2(0,\mu), \quad \mu<0, \tag{28a}$$

$$I_2(0,\mu) = \rho_2 I_2(0,-\mu) + \Gamma_1 I_1(0,\mu), \quad \mu>0. \tag{28b}$$

We now represent the intensity $I_j(\tau,\mu)$ by the superposition of the intensities $I_{ji}(\tau,\mu)$ of single half-space problems in the form

$$I_j(\tau,\mu) = \sum_{i=1}^N I_{ji}(\tau,\mu) + \Psi_j(\tau,\mu), \tag{29}$$

where $I_{ji}(\tau,\mu)$ are the solutions of the following simple problems:

$$L_1 I_{11}(\tau,\mu) = S_1 \quad \text{and} \quad L_2 I_{21}(\tau,\mu) = S_2$$

$$L_j I_{ji}(\tau,\mu) = 0, \quad i = 2,\cdots,N; \quad j = \begin{cases} 1, & \tau<0 \\ 2, & \tau>0 \end{cases}, \tag{30}$$

subject to the boundary conditions

$$I_{ji}(0,\mu) = (1-\delta_{1i})\Big(\rho_j I_{j,i-1}(0,-\mu) + \sum_{k=1}^2 (1-\delta_{jk})\Gamma_k I_{k,i-1}(0,\mu)\Big), \quad j = \begin{cases} 1, & \mu<0 \\ 2, & \mu>0 \end{cases}, \tag{31a}$$

and

$$\lim_{|\tau|\to\infty} I_{ji}(\tau,\mu) \to \delta_{1i}P_j(\tau,\mu), \quad i = 1,2,\cdots,N, \quad j = \begin{cases} 1, & \tau<0 \\ 2, & \tau>0 \end{cases}. \tag{31b}$$

Here δ_{ji} is the Kronecker delta, and $P_i(\tau,\mu)$ is a particular solution of Eq. (30).

The function $\Psi_j(\tau,\mu)$ satisfies the following equation:

$$L_j\Psi_j(\tau,\mu) = 0, \quad j = \begin{cases} 1, & \tau<0 \\ 2, & \tau>0 \end{cases}, \tag{32}$$

subject to the boundary conditions

$$\Psi_j(0,\mu) = \rho_j[I_{jN}(0,-\mu) + \Psi_j(0,-\mu)] + \sum_{k=1}^2 (1-\delta_{jk})\Gamma_k[I_{kN}(0,\mu) + \psi_k(0,\mu)], \quad j = \begin{cases} 1, & \mu<0 \\ 2, & \mu>0 \end{cases}, \tag{33a}$$

and

$$\lim_{|\tau|\to\infty} \Psi_j(\tau,\mu) \to 0, \quad j = \begin{cases} 1, & \tau<0 \\ 2, & \tau>0 \end{cases}. \tag{33b}$$

Scrutiny of the radiation problem defined by Eqs. (30) and (31) reveals that the intensity function $I_{ji}(\tau,\mu)$ approaches zero as N becomes infinite. Then, in Eq. (33a) the forcing functions $I_{jN}(0,-\mu)$ and $I_{kN}(0,\mu)$ vanish, and the solution $\Psi_j(\tau,\mu)$ of Eq. (32) subject to boundary conditions Eqs. (33) become trivial for $N \to \infty$. Therefore, for $N \to \infty$, Eq. (29) becomes

$$I_j(\tau,\mu) = \sum_{i=1}^{\infty} I_{ji}(\tau,\mu), \quad j = \begin{cases} 1, & \tau < 0 \\ 2, & \tau > 0 \end{cases}. \tag{34}$$

We now proceed to construct the solutions of the single half-space problems defined by Eqs. (30) and (31). The solutions for $I_{ji}(\tau,\mu)$ which satisfy the boundary conditions at infinity as given by Eq. (31b) can be written as a linear sum of the normal modes and a particular solution in the form

$$I_{1i}(\tau,\mu) = A_i(-\nu_0)\phi_1(-\nu_0,\mu)e^{\tau/\nu_0} + \int_0^1 A_i(-\eta)\phi_1(-\eta,\mu)e^{\tau/\eta}d\eta + \delta_{1i}P_1(\tau,\mu), \tag{35a}$$

$$I_{2i}(\tau,\mu) = A_i(\eta_0)\phi_2(\eta_0,\mu)e^{-\tau/\eta_0} + \int_0^1 A_i(\eta)\phi_2(\eta,\mu)e^{-\tau/\eta}d\eta + \delta_{1i}P_2(\tau,\mu). \tag{35b}$$

Substitution of Eqs. (35) into Eq. (34) yields the radiation intensity for the two adjacent half-spaces problem as

$$I_1(\tau,\mu) = \left(\sum_{i=1}^{\infty} A_i(-\nu_0)\right)\phi_1(-\nu_0,\mu)e^{\tau/\nu_0} + \int_0^1\left(\sum_{i=1}^{\infty} A_i(-\eta)\right) \times \phi_1(-\eta,\mu)e^{\tau/\eta}d\eta + P_1(\tau,\mu),$$
$$\tau < 0,\ \mu \in (-1,1), \tag{36a}$$

$$I_2(\tau,\mu) = \left(\sum_{i=1}^{\infty} A_i(\eta_0)\right)\phi_2(\eta_0,\mu)e^{-\tau/\eta_0} + \int_0^1\left(\sum_{i=1}^{\infty} A_i(\eta)\right) \times \phi_2(\eta,\mu)e^{-\tau/\eta}d\eta + P_2(\tau,\mu),$$
$$\tau > 0,\ \mu \in (-1,1). \tag{36b}$$

The normal modes appearing in these equations have been defined previously, and the expansions coefficients $A_i(-\nu_0)$, $A_i(-\nu)$, $A_i(\eta_0)$, and $A_i(\eta)$ are the half-space problem expansion coefficients associated with the solution given by Eqs. (35). These expansion coefficients are evaluated by constraining the solutions given by Eqs. (35) to meet the boundary conditions (31a) and then solving the resulting singular integral equation by utilizing the orthogonality property of the normal modes and various normalization integrals in a similar manner described previously. In this case, explicit relations could be obtained for these expansion coefficients; omitting the details we present below the resulting expressions for $A_i(-\nu_0)$, $A_i(-\eta)$, $A_i(\eta_0)$, and $A_i(\eta)$:

$$A_1(-\xi) = \frac{1}{N_1(\xi)}\int_0^1 \mu H_1(\mu)\phi_1(\xi,\mu)F_1(\mu)d\mu, \quad \xi = \nu_0, \eta, \tag{37a}$$

$$A_1(\xi) = \frac{1}{N_2(\xi)}\int_0^1 \mu H_2(\mu)\phi_2(\xi,\mu)F_2(\mu)d\mu, \quad \xi = \eta_0, \eta, \tag{37b}$$

where

$$F_1(\mu) = -P_1(0,-\mu), \tag{38a}$$

$$F_2(\mu) = -P_2(0,\mu), \tag{38b}$$

$$\mathbf{N}(\nu_0)\mathbf{A}_i = \mathbf{K}_0\mathbf{A}_{i-1} + \int_0^1 \mathbf{K}_0(\eta')\mathbf{A}_{i-1}(\eta')d\eta' + \delta_{i2}\mathbf{S}_0, \quad i = 2,3,\ldots, \tag{39a}$$

$$\mathbf{N}(\eta)\mathbf{A}_i(\eta) = \mathbf{K}(\eta)\mathbf{A}_{i-1} + \int_0^{+1} \mathbf{K}(\eta,\eta')\mathbf{A}_{i-1}(\eta')d\eta' + \delta_{i2}\mathbf{S}(\eta), \quad i = 2,3,\ldots, \tag{39b}$$

where

$$\mathbf{A}_i = \begin{bmatrix} A_i(-\nu_0) \\ A_i(\eta_0) \end{bmatrix}, \quad \mathbf{A}_i(\eta) = \begin{bmatrix} A_i(-\eta) \\ A_i(\eta) \end{bmatrix}, \tag{40a}$$

$$\delta_{i2} = \begin{bmatrix} \delta_{i2} & 0 \\ 0 & \delta_{i2} \end{bmatrix}, \tag{40b}$$

$$\mathbf{S}_0 = \begin{bmatrix} S_1(\nu_0) \\ S_2(\eta_0 \end{bmatrix}, \quad \mathbf{S}(\eta) = \begin{bmatrix} S_1(\eta) \\ S_2(\eta) \end{bmatrix}, \tag{40c}$$

where

$$S_i(\xi) = \frac{1}{\xi}\int_0^1 \mu H_i(\mu)\phi_i(\xi,\mu)R_i(\mu)d\mu, \quad i = 1,2 \tag{40d}$$

and

$$R_1(\mu) = \rho_1 P_1(0,\mu) + \Gamma_2 P_2(0,-\mu), \tag{40e}$$

$$R_2(\mu) = \rho_2 P_2(0,-\mu) + \Gamma_1 P_1(0,\mu). \tag{40f}$$

Other quantities are the same as defined in Eqs. (12).

A comparison of the solutions given by Eqs. (3) and (36) implies that if these two solutions are identical, we should have

$$A(\xi) = \sum_{i=1}^{\infty} A_i(\xi), \quad \xi = -\nu_0, +\eta_0, \pm\eta \tag{41}$$

Indeed, the substitution of the coefficients $A_i(\xi)$ given by Eqs. (37) and (39) into Eq. (41) has shown that the results are identical. In fact, the results as given by Eqs. (37) and (39) characterize the solution of the Fredholm integral equations (11) for the expansion coefficients by the method of successive iterations and provide an explicit relation for various orders of approximations.

ACKNOWLEDGMENTS

The authors wish to acknowledge discussions on this subject with Dr. C. E. Siewert. This work was supported in part by the National Science Foundation through Grant GK11935.

APPENDIX: HALF-RANGE COMPLETENESS AND ORTHOGONALITY THEOREMS

Here we state the half-range completeness theorem presented initially by Case[7,8] and the half-range orthogonality theorem. In addition, a summary of the necessary normalization integrals is given.

Theorem I: The eigenfunction $\phi(\eta_0,\mu)$ and $\phi(\eta,\mu)$, $\eta \in (0,1)$ are complete on the half range in the sense that an arbitrary Hölder function $\psi(\mu)$ defined for $\mu \in (0,1)$ can be expanded in the form

$$\psi(\mu) = A(\eta_0)\phi(\eta_0,\mu) + \int_0^{+1} A(\eta)\phi(\eta,\mu)d\eta, \quad \mu \in (0,1). \quad \text{(A1)}$$

Theorem II: The eigenfunctions $\phi(\eta_0,\mu)$ and $\phi(\eta,\mu)$, $\eta \in (0,1)$ are orthogonal with respect to the weight function $\mu H(\mu)$ on the interval $0 \le \mu \le 1$, i.e.,

$$\int_0^1 \mu H(\mu)\phi(\xi,\mu)\phi(\xi',\mu)d\mu = 0, \quad \xi \neq \xi', \ \xi,\xi' = \eta_0 \text{ or } \in (0,1). \quad \text{(A2)}$$

Normalization Integrals:

$$\int_0^1 \mu H_i(\mu)\phi_i(\eta,\mu)\phi_i(\eta',\mu)d\mu = \eta H_i(\eta)\Lambda_i^+(\eta)\Lambda_i^-(\eta)\delta(\eta-\eta'), \quad \text{(A3)}$$

$$\int_0^1 \mu H_i(\mu)\phi_i^2(\xi,\mu)d\mu = \frac{\omega\xi^2}{2} H_i(\xi) \left.\frac{d\Lambda_i(z)}{dz}\right|_{z=\xi}, \quad \text{(A4)}$$

where

$$\xi = \begin{cases} \nu_0, & i=1 \\ \eta_0, & i=2 \end{cases}, \quad \eta \text{ and } \eta' \in (0,1),$$

$$\int_0^1 \mu H_i(\mu)\phi_i(\eta,\mu)\phi_j(-\xi,\mu)d\mu = \eta \frac{\phi_j(-\xi,\eta)}{H_i(\xi)}, \quad \text{(A5)}$$

where

$$\eta = \begin{cases} \nu_0, & i=1 \\ \eta_0, & i=2 \end{cases}, \quad \text{or} \quad \eta \in (0,1),$$

$$\xi = \begin{cases} \nu_0, & j=1 \\ \eta_0, & j=2 \end{cases}, \quad \text{or} \quad \xi = \eta' \in (0,1).$$

Normalization integrals (A3)–(A5) have been derived by making use of the properties of the eigenfunctions and the properties of H function given by Eqs. (10).

[1] B. Davison, Can. J. Res. A 28, 303 (1950).
[2] S. Chandrasekhar, Can. J. Phys. 29, 14 (1951).
[3] A. Korn, Nukleonika 9, 237 (1967).
[4] M. R. Mendelson and G. C. Summerfield, J. Math. Phys. 5, 668 (1964).
[5] G. R. Bond and G. E. Siewert, Nucl. Sci. Eng. 35, 277 (1969).
[6] N. J. McCormick, Nucl. Sci. Eng. 37, 243 (1969).
[7] K. M. Case, Ann. Phys. (N.Y.) 9, 1 (1960).
[8] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[9] K. M. Case, F. de Hoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (U.S. Government Printing Office, Washington, D.C., 1953), Vol. 1.
[10] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[11] M. N. Özişik and C. E. Siewert, Nucl. Sci. Eng. 40, 491 (1970).