JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 13, NUMBER 10 OCTOBER 1972

Kinetic Theory of Scattering by a Plasma Cylinder

V. H. Weston
Division of Mathematical Sciences, Purdue University, Lafayette, Indiana 47907
(Received 17 May 1971)

The kinetic theory of scattering by a circular homogeneous isotropic plasma cyclinder is treated for plane wave incidence parallel to the axis of the cylinder. The relativistic form of the Vlasov equation is inverted, subject to the specular boundary condition, expressing the electronic distribution function in terms of the electric field intensity. After inverting Maxwell's equations and eliminating the distribution function, a set of Fredholm integral equations of the second kind are obtained for the angular Fourier components of the electric field. Since for low temperature the Neumann series converges, the low temperature solution is easily obtained. The first order temperature corrections are thus derived for the reflection coefficients associated with the Fourier components.

A significant amount of the analysis has been performed on plasma scattering problems based upon the linearized coupled Maxwell–Vlasov or Boltzman equations. However, these problems have been limited to slab or half-space geometrics. The mathematical methods employed have been based either on singular integral equations,[1–3] Fourier transforms[4–6] or series.[7] As such, these techniques were applicable because of the planar geometry involved.

Here, nonplanar geometry is considered, where the kinetic theory of plane wave scattering by a plasma circular cylinder is treated. The plasma will be taken to be homogeneous and isotropic. The linearized problem will be considered, with frequency sufficiently high such that the ion motions may be neglected.

The results in this paper will be mainly concerned with the case of the incident polarization being parallel to the axis of the cylinder. Although, with this simplification only transverse waves are produced in the plasma, it is important to investigate this case first, before the more interesting and more difficult case of the incident polarization perpendicular to axis is considered, wherein both transverse and longitudinal waves are induced in the plasma.

As a first step, the Boltzmann equation will be inverted, subject to the specular boundary condition. Because of the mathematical procedure used, the relativistic form of the equation is employed, and a collision term (corresponding to complex frequency) is included. The basic reasons for the use of the relativistic form (even for low temperature) and a collision term is to insure convergence. The nonrelativistic form could be employed; but the domain of the velocity space will have to be compact. The relativistic form insures this since it restricts the magnitude of the velocity to be less than the speed of light. The technique employed here to invert the Boltzmann equation is very general and is carried out initially for arbitrary cylindrical geometries with smooth boundaries.

The distribution function is then given an explicit form for the case of polarization parallel to the axis of a circular cylinder, from which, the current is expressed in terms of the electric field.

Maxwell's equations are inverted expressing E in terms of the current. The process employed, involves a Green's function closely associated with the corresponding cold plasma or dielectric problem. Eliminating the current from the two expressions, a set of one-dimensional Fredholm integral equations, of the second kind are obtained for the Fourier components (in the angular variable) of the electric field in the plasma. For the low temperature region, the integral equations may be solved by the Neumann series.

Asymptotic results for low temperature are considered, and the asymptotic form of the first iterate to the integral equations are computed, yielding the temperature correction to the electric field on the surface of the cylinder, from which the reflection coefficients may be computed.

The fundamental equations are Maxwell's equations [time dependence $\exp(-i\omega t)$ suppressed]

$$\nabla \times \mathbf{E} = i\omega\mu_0\mathbf{H}, \tag{1}$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0\mathbf{E} + \mathbf{j}, \tag{2}$$

where $\mathbf{j}$, the current due to the collective motion of the plasma, is related to the electronic perturbed distribution function f, by

$$\mathbf{j} = -ec\int \beta \mathbf{u} f d\mathbf{u}. \tag{3}$$

In the above equation, $-e$ is the electronic charge, $\mathbf{u}$ is the reduced velocity, and β is given by

$$\beta = (1 + u^2)^{-1/2}. \tag{4}$$

In combination with the above equations we have the relativistic Boltzmann equation (given by Clemmow and Willson)[8]

$$(-i\omega + \nu)f + \beta c\mathbf{u}\cdot\nabla f - (e/mc)\mathbf{E}\cdot\nabla_u f_0 = 0, \tag{5}$$

where the unperturbed distribution function f_0 is given by

$$f_0 = nF_0,$$
$$F_0 = \frac{\exp[-\lambda(1+u^2)^{1/2}]}{[(4\pi/\lambda)K_2(\lambda)]}, \qquad \lambda = \frac{mc^2}{kT}. \tag{6}$$

The geometry of the problem under consideration will be specified in terms of cylindrical polar coordinates (r, θ, z). The plasma will be contained in the circular cylinder $r \le a$, which is imbedded in free space, thus f and $\mathbf{j}$ will vanish for $r > a$.

Radiation will be incident normally to the plasma cylinder, producing a scattered field in the domain $r > a$, and an induced field in the plasma cylinder. The quantities $\mathbf{E}$, $\mathbf{H}$, and f will thus be independent of the z variable, and the problem reduces to two dimensions as far as the spatial variables are concerned.

INTEGRATION OF THE BOLTZMANN EQUATION

The first step in the analysis will be to invert Eq. (5) for $r < a$ expressing f in terms of $\mathbf{E}$ such that the distribution function f satisfies the specular boundary condition at $r = a$. This will be done for arbitrary polarization of $\mathbf{E}$. An explicit expression for f for the case where $\mathbf{E}$ has only the z component (corresponding to the polarization of the incident radiation being parallel to the cylinder), will then be presented.

Set

$$\nabla_u f_0 = -\mathbf{u} f_0', \tag{7}$$

$$g = neF_0'/u_r \beta mc^2, \tag{8}$$

$$\alpha = (\omega + i\nu)/\beta c u_r, \tag{9}$$

where the velocity $\mathbf{u}$ is expressed in cylinder polar coordinates (u_r, ϕ_u, u_z) in velocity space, with $\mathbf{u}_r = \mathbf{u} - u_z \hat{i}_z$. Equation (5) becomes

$$-i\alpha f + \hat{\mathbf{u}}_r \cdot \nabla f = -\mathbf{u}\cdot\mathbf{E} g, \tag{10}$$

which can be directly integrated by transforming the equation into a first order ordinary differential equation through the transformation of the variables (x, y) to (ξ, η) such that the ξ axis lies along the direction of the vector $\hat{\mathbf{u}}_r$. The result is

$$f(\mathbf{x}, \mathbf{u}) = e^{i\alpha l} f(\mathbf{x}_0, \mathbf{u}) - g\int_0^l e^{i\alpha t}\mathbf{u}\cdot\mathbf{E}(\mathbf{x} - \hat{\mathbf{u}}_r t)dt, \tag{11}$$

where

$$\mathbf{x} = \mathbf{x}_0 + l\hat{\mathbf{u}}_r. \tag{12}$$

For a fixed $\mathbf{x}$ and $\mathbf{u}$, the point $\mathbf{x}_0$ lies on the line passing through $\mathbf{x}$ in the direction of the vector $-\mathbf{u}_r$. The point $\mathbf{x}_0$ will be specified by requiring it to lie on the cylinder $r = a$, such that the direction of the vector $\mathbf{x} - \mathbf{x}_0$ is given by $\mathbf{u}_r$, in which case l is positive. Essentially $\mathbf{x}_0$ is found by tracing back from the point $\mathbf{x}$, along the velocity ray $\hat{\mathbf{u}}_r$ to the surface. Thus the unknown function $f(\mathbf{x}_0, \mathbf{u})$ is the velocity distribution at the point $\mathbf{x}_0$ on the surface. This quantity will be determined through successive applications of the specular boundary condition, which at an arbitrary point $\mathbf{a}$ on the surface is given by

$$f(\mathbf{a}, \mathbf{u}) = f(\mathbf{a}, \mathbf{u}^*),$$
$$\mathbf{u}^* = \mathbf{u} - 2(\mathbf{n}\cdot\mathbf{u})\mathbf{n},$$

where $\mathbf{n}$ is the local unit outward normal to the surface. The technique to be given here corresponds to tracing back the reflected rays undergoing multiple reflections and is applicable to smooth cylindrical geometries of arbitrary cross section.

As a first step to determine $f(\mathbf{x}_0, \mathbf{u})$, the specular boundary condition will be applied at the point $\mathbf{x}_0$,

$$f(\mathbf{x}_0, \mathbf{u}) = f(\mathbf{x}_0, \mathbf{u}^1),$$
$$\mathbf{u}^1 = \mathbf{u} - 2(\mathbf{n}_0\cdot\mathbf{u})\mathbf{n}_0. \tag{13}$$

Then tracing back the rays (in the direction $-\mathbf{u}^1$) from $\mathbf{x}_0$, to the point $\mathbf{x}^1$ on the surface, the following expression is obtained from Eq. (11) with appropriate substitution:

$$f(\mathbf{x}_0, \mathbf{u}) = f(\mathbf{x}_0, \mathbf{u}^1) = e^{i\alpha l_1} f(\mathbf{x}_1, \mathbf{u}^1) - g\int_0^{l_1} e^{i\alpha t}\mathbf{u}^1\cdot E(\mathbf{x}_0 - \mathbf{u}_r^1 t)\,dt, \tag{14}$$

where l_1 is the distance between points $\mathbf{x}_0$ and $\mathbf{x}_1$.

Similarly apply the specular boundary condition at the point $\mathbf{x}_1$, as follows:

$$f(\mathbf{x}_1, \mathbf{u}^1) = f(\mathbf{x}, \mathbf{u}^2),$$
$$\mathbf{u}^2 = \mathbf{u}^1 - 2(\mathbf{n}\cdot\mathbf{u}^1)\mathbf{u}^1,$$

where $\mathbf{n}$ is the local unit outward normal at $\mathbf{x}_1$. $\mathbf{u}^1$ is the reflected velocity vector corresponding to the incident vector $\mathbf{u}^2$ at $\mathbf{x}_1$. Repeating the process of tracing back along the ray $(-\mathbf{u}^2)$ to the point $\mathbf{x}_2$, one obtains

$$f(\mathbf{x}_1, \mathbf{u}^1) = f(\mathbf{x}_1, \mathbf{u}^2) = e^{i\alpha l_2} f(\mathbf{x}_2, \mathbf{u}^2) - g\int_0^{l_2} e^{i\alpha t}\mathbf{u}^2\cdot\mathbf{E}(\mathbf{x}_1 - \hat{\mathbf{u}}_r^2 t)dt,$$

where $l_2 = |\mathbf{x}_1 - \mathbf{x}_2|$.

The whole process is repeated for n reflections. Thus one obtains the following set of equations:

$$f(\mathbf{x}_{n-1}, \mathbf{u}^n) = e^{i\alpha l_n} f(\mathbf{x}_n, \mathbf{u}^n) - g\int_0^{l_n} e^{i\alpha t}\,\mathbf{u}^n\cdot\mathbf{E} \times (\mathbf{x}_{n-1} - \hat{\mathbf{u}}_r^n t)\,dt, \tag{15}$$

$$f(\mathbf{x}_n, \mathbf{u}^n) = f(\mathbf{x}_n, \mathbf{u}^{n+1}). \tag{16}$$

Combining the resulting equations, we obtain

$$f(\mathbf{x}_0, \mathbf{u}) = f(\mathbf{x}_n, \mathbf{u}^n)e^{i\alpha L_n} - g\sum_{m=1}^{n} e^{i\alpha L_{m-1}} \times \int_0^{l_m} e^{i\alpha t}\mathbf{u}^m\cdot\mathbf{E}(\mathbf{x}_{m-1} - \hat{\mathbf{u}}_r^m t)dt, \tag{17}$$

where

$$L_m = l_1 + l_2 + + l_m, \quad m \neq 0, \quad L_0 = 0, \quad m = 0 \tag{18}$$

Since

$$\text{real } \alpha = -(\nu/c)[(1 + u_z^2 + u_r^2)/u_r^2]^{1/2},$$

it is seen that if $\nu > 0$, real $\alpha < 0$, over the complete domain of velocity variables (u_r, u_z). If the nonrelativistic form of the Boltzmann equation was employed, real α would vanish, unless a cutoff in the velocity space $\mathbf{v}$ was required (i.e., integration restricted to a compact domain in velocity space). The relativistic form automatically ensures the cutoff with $|\mathbf{v}| < c$. Thus, for $\nu > 0$, it follows that $n \to \infty$,

$$f(\mathbf{x}_0, \mathbf{u}) = -g \sum_{m=1}^{\infty} e^{i\alpha L_{m-1}} \int_0^{l_m} e^{i\alpha t}\, \mathbf{u}^m \cdot \mathbf{E} \times (\mathbf{x}_{m-1} - \hat{\mathbf{u}}_r^m t)\, dt. \tag{19}$$

The summation can be expressed in terms of an integral along the multiply reflected rays. Set $\mathbf{s}$ as the vector directed back along the rays, with magnitude corresponding to the arc length along the rays from $\mathbf{x}_0$. The above expression has the form

$$f(\mathbf{x}_0, \mathbf{u}) = -g u_z \int_0^{\infty} e^{i\alpha s} E_z(\mathbf{x}_0 + \mathbf{s})\, ds \tag{20}$$

for polarization parallel to the cylinder, and the form

$$f(\mathbf{x}_0, \mathbf{u}) = +g \int_0^{\infty} e^{i\alpha s} \mathbf{E}(\mathbf{x}_0 + \mathbf{s}) \cdot d\mathbf{s} \tag{21}$$

for polarization perpendicular to the cylinder. This result holds for cylinders of arbitrary smooth cross section.

Returning to the case of polarization parallel to the axis of a circular cylinder, Eq. (19) can be expressed in a more explicit form. First, for a circular cylinder, the segments leave equal lengths, i.e.,

$$l_1 = l_2 = l_3 = \cdots = L,$$
$$L_m = (m-1)L.$$

Because of symmetry, the integral on the segment from $\mathbf{x}_{m-1}$ to $\mathbf{x}_m$, given by

$$\int_0^L e^{i\alpha t} E_z(r'(t), \theta(t, m))\, dt$$

(where polar coordinates are employed), can be placed in the form

$$\int_0^L e^{i\alpha t} E_z(r'(t), 2(m-1)\gamma + \theta(t))\, dt,$$

where 2γ is the angle subtended at the origin by the line segment, and $\theta(t)$ corresponds to the angular position of the point in the segment $(\mathbf{x}_0, \mathbf{x}_1)$. Employing a Fourier Series representation for E_z, the resulting expression can be summed over the index m to give a closed form result. Changing the variable of integration t to r', the final form for $f(\mathbf{x}_0, \mathbf{u})$ is given as follows:

$$f(\mathbf{x}_0, \mathbf{u}) = -\frac{ig}{4\pi} \sum_{n=-\infty}^{\infty} R_n(\gamma) e^{in\theta_0} [\sin(\alpha a \sin\gamma - n\gamma)]^{-1}, \tag{22}$$

where

$$R_n(\gamma) = \pm 2u_z \iint_A \frac{\cos\psi e^{-in\theta'}}{(r'^2 - a^2\cos^2\gamma)^{1/2}} \times E_z(r', \theta') H(r' - a|\cos\gamma|)\, dA', \tag{23}$$

where the plus or minus sign is taken according as $\sin\gamma$ is positive or negative. The function H is the Heaviside step function. The angle γ is given by

$$\gamma = \phi_u - \theta_0 - \tfrac{1}{2}\pi, \tag{24}$$

where θ_0 is the angular cordinate of $\mathbf{x}_0$. In expression (23) ψ is given by

$$\psi = \pm\, \alpha(r'^2 - a^2\cos^2\gamma)^{1/2} - n \arccos[(a/r')\cos\gamma], \tag{25}$$

the appropriate sign being taken according whether $\sin\gamma$ is positive or negative. It can be shown that the above series is uniformly convergent in θ and γ.

The results can be summarized as follows: For polarization parallel to the axis of the cylinder, the distribution function has the form

$$f(\mathbf{x}, \mathbf{u}) = \frac{-ig}{4\pi} e^{i\alpha l} \sum_{n=-\infty}^{\infty} \frac{R_n(\gamma) e^{in\theta_0}}{\sin(\alpha a \sin\gamma - n\gamma)} + g \int_0^{-l} e^{-i\alpha t} u_z E_z(\mathbf{x} + \hat{\mathbf{u}}_r t)\, dt, \tag{26}$$

where

$$\gamma = \phi_u - \theta_0 - \tfrac{1}{2}\pi, \tag{27}$$

and θ_0 is the angular coordinate of $\mathbf{x}_0$. To complete the expression we need to determine l and θ_0 in terms of the coordinates (r, θ) of $\mathbf{x}$ and the angular coordinate ϕ_u of $\mathbf{u}$.

It can be shown from geometry that

$$\sin(\phi_u - \theta_0) = (r/a)\sin(\phi_u - \theta), \tag{28}$$

$$\cos(\phi_u - \theta_0) = -[1 - (r^2/a^2)\sin^2(\phi_u - \theta)]^{1/2}, \tag{29}$$

$$l = r\cos(\phi_u - \theta) + [a^2 - r^2\sin^2(\phi_u - \theta)]^{1/2}. \tag{30}$$

EXPRESSION FOR THE CURRENT

The current will be evaluated in terms of E for the case where the polarization is parallel to the axis. In this case, we have

$$j_z = -ec \int \beta u_z f\, d\mathbf{u},$$

where f is given by Eq. (26). The current will be decomposed into the parts

$$j_z = \sum_{n=-\infty}^{\infty} j_n + j'. \tag{31}$$

The component j' is given by

$$j' = (-e^2 n/mc) \int (u_z^2/u_r) F_0' \int_0^l e^{-i\alpha t} E_z(\mathbf{x} + \hat{\mathbf{u}}_r t)\, dt du\, du. \tag{32}$$

Integrating with respect to the ϕ_u variable first, we find that with the substitutions $\phi_u + \pi = \psi$, $t = -R$, $x' = x + \cos\psi R$, and $y' = y + \sin\psi R$,

$$\int_0^{2\pi} d\phi_u \int_0^{-l} e^{-i\alpha t} E_z(\mathbf{x} + \hat{\mathbf{u}}_r t)\, dt = -\int_A (e^{i\alpha R}/R) E_z(\mathbf{x}')\, dA', \tag{33}$$

where $R = |\mathbf{x} - \mathbf{x}'|$ is the distance between the observation point $\mathbf{x}$ and the point $\mathbf{x}'$ of integration. The integration is taken over the disc $0 \leq r' \leq a$. The component j' is thus given by

$$j'(\mathbf{x}) = (e^2 n/mc)\int_A M(R)E_z(\mathbf{x}')dA', \tag{34}$$

where

$$M(R) = 2\int_0^\infty\int_0^\infty u_z^2 F_0'(e^{-i\alpha R}/R)du_z du_r \tag{35}$$

and

$$R = |\mathbf{x} - \mathbf{x}'). \tag{36}$$

The remaining components j_n, given by the relation

$$j_n = \frac{ie^2 n}{4\pi mc}\int \frac{u_z F_0' R_n(\gamma)e^{in\theta_0 + i\alpha l}d\mathbf{u}}{u_r \sin(\alpha a \sin\gamma - n\gamma)} \tag{37}$$

can be reduced to a more symmetric form. The parameters θ_0, l, γ are functions of ϕ_u and θ, and $R_n(\gamma)$ is given by Eq. (23). As a first step, the variable of integration ϕ_u will be replaced by t:

$$\begin{aligned} r\sin(\phi_u - \theta) &= t, \\ r\cos(\phi_u - \theta) &= \pm(r^2 - t^2)^{1/2}, \end{aligned} \tag{38}$$

where the plus and minus signs occur for the domains $\frac{1}{2}\pi + \theta \le \phi_u \le \frac{1}{2}\pi + \theta$ and $\frac{1}{2}\pi + \theta \le \phi_u \le \frac{3}{2}\pi + \theta$, respectively. The range of the variable t is from $-r$ to r. From Eq. (27) and (29), it follows that

$$\begin{aligned} a\sin\gamma &= -a\cos(\phi_u - \theta_0) = (a^2 - t^2)^{1/2}, \\ a\cos\gamma &= a\sin(\phi_u - \theta_0) = t, \end{aligned} \tag{39}$$

and $l = \pm(r^2 - t^2)^{1/2} + (a^2 - t^2)^{1/2}$,

where the appropriate sign chosen depends upon the original domain of the ϕ_u variable as indicated above.

From Eq. (23), we have

$$R_n(\gamma) = 2u_z\iint_A[\cos\psi(r')e^{-in\theta'}/(r'^2 - t^2)^{1/2}] \times E_z(r', \theta')H(r' - |t|)dA' \tag{40}$$

with

$$\psi(r') = \alpha(r'^2 - t^2)^{1/2} - n\arccos(t/r'). \tag{41}$$

If we employ Eq. (41) to define $\psi(r')$ as a function of r', it follows that

$$\psi(a) = \alpha a\sin\gamma - n\gamma, \tag{42}$$

$$\begin{aligned} n\theta_0 + \alpha l &= n\theta - n\gamma + n(\phi_u - \theta - \tfrac{1}{2}\pi) + \alpha l \\ &= n\theta + \psi(a) \pm \psi(r), \end{aligned} \tag{43}$$

where the sign is chosen according to the domain of r, it follows that

$$\int_0^{2\pi}\frac{R_n(\gamma)e^{in\theta_0 + i\alpha l}}{\sin(\alpha a\sin\gamma - n\gamma)}d\phi_u = 2e^{in\theta}\int_{-r}^{r}\frac{R_n(\gamma)e^{i\psi(a)}}{(r^2 - t^2)^{1/2}} \times \frac{\cos\psi(r)}{\sin\psi(a)}dt,$$

and combining (37), (40), and the above results, we have

$$j_n = (ie^2 n/\pi mc)\int_A E_z(r', \theta')e^{in(\theta - \theta')}M^n(r, r')dA', \tag{44}$$

where

$$M^n(r, r') = \int_0^\infty du_r\int_0^\infty du_z u_z^2 F_0' \times \int_{-t_0}^{t_0}\frac{\cos\psi(r)\cos\psi(r')e^{i\psi(a)}}{\sqrt{(r^2 - t^2)^{1/2}}\sqrt{(r'^2 - t^2)^{1/2}}\sin\psi(a)}dt, \tag{45}$$

where $t_0 = \min(r, r')$.

It can be shown that $M^n(r, r')$ has a logarithmic type singularity when r approaches r', provided that neither r or r' equals a. If $r = a$, then $M^n(a, r')$ has a singularity of the type $(a^2 - r'^2)^{1/2}$.

INTEGRAL EQUATION

The results of the preceding section where the current is expressed in terms of the electric intensity, may now be combined with Maxwell's equations to yield an integral equation for E. Maxwell's equations for polarization parallel to the axis of the cylinder reduce to

$$(\nabla^2 + k^2)E_z = -i\omega\mu_0 j_z, \tag{46}$$

where j_z vanishes for $r > a$. If E^i is the incident field, the above may be integrated directly to yield

$$E_z(\mathbf{x}_0) = E_z^i(\mathbf{x}_0) - \tfrac{1}{4}\omega\mu o\int_A H_0^{(1)}(kR_0)j_z(\mathbf{x})dA, \tag{47}$$

where $R_0 = |\mathbf{x} - \mathbf{x}_0|$ and $H_0^{(1)}(kR_0)$ is the Hankel function of order zero. Combining this with the following result derived from the previous section

$$j_z(\mathbf{x}) = (e^2 n/mc)\int_A [M(R) + (i/\pi)\sum e^{in(\theta - \theta')} \times M^{(n)}(r, r')]E_z(\mathbf{x}')dA', \tag{48}$$

we obtain a set of coupled integral equations, from which either E_z or j_z may be eliminated. Elimination of j_z, yields an integral equation for E_z, whose kernel is expandable in terms of a Fourier series in the θ variable, thus reducing the problem to a set of integral equations of one variable. However the resulting equations appear not to be amenable to solution or asymptotic approximation. Instead of this approach, we will use a modified technique, working directly with the Fourier components of E_z. Setting

$$E_z(r, \theta) = \sum_{n=-\infty}^{\infty} e^{in\theta}E^n(r), \tag{49}$$

Eq. (47) reduces to

$$L(r_0, k)E^n = -(i\omega\mu_0/2\pi)\int_0^{2\pi} e^{-in\theta_0}j_z(\mathbf{x}_0)r_0 d\theta_0, \tag{50}$$

where the operator $L(r_0, k)$ is given by

$$L(r, k) = \frac{d}{dr}\left(r\frac{d}{dr}\right) + \left(k^2 r - \frac{n^2}{r}\right). \tag{51}$$

An incident field of the form of plane wave excitation will be taken

$$E^i(\mathbf{x}_0) = e^{ikx_0} = \sum_{n=-\infty}^{\infty} J_n(kr_0)e^{in\theta_0}. \tag{52}$$

By requiring that the tangential electric and magnetic field (i.e., partial derivative of E with respect to r) be continuous across $r = a$, one obtains the following nonhomogeneous boundary condition for $E^n(r)$, associated with the above incident field,

$$H_n^{(1)\prime}(ka)E^n(a) - H_n^{(1)}(ka)E^{n\prime}(a) = 2i/\pi a, \tag{53}$$

where the prime indicates differentiation with respect to the variable r.

The system given by Eqs. (50) and (53) will be inverted after first modifying Eq. (50) as follows:

$$L(r_0, k_1)E^n = (k_1^2 - k^2)r_0 E^n(r_0) - (i\omega\mu_0/2\pi) \times \int_0^{2\pi} e^{-in\theta_0} j_z(\mathbf{x}_0) r_0 d\theta_0, \quad (54)$$

where

$$k_1^2 = k^2 \epsilon(\lambda), \quad (55)$$

$$\epsilon(\lambda) = 1 - \frac{\omega_p^2 \lambda^2}{\omega(\omega + i\nu)} \left(\int_\lambda^\infty \tau^{-2} K_2(\tau) \frac{d\tau}{K_2(\lambda)} \right). \quad (56)$$

In the above relation ω_p is the plasma frequency, and K_2 is the modified Bessel function. When $\lambda \to \infty$, it can be shown that

$$\epsilon(\lambda) \sim 1 - [\omega_p^2/\omega(\omega + i\nu)][1 - (5/2\lambda) + \cdots]. \quad (57)$$

Thus in the limit of zero temperature ($\lambda = \infty$), $\epsilon(\lambda)$ is the relative dielectric constant of the plasma medium. Hence inversion of Eq. (54) with respect to operator $L(r_0, k_1)$ will produce an integral equation which represents a perturbation about the cold temperature limit.

The following Green's function related to the operator $L(r_0, k_1)$ and satisfying the homogeneous boundary condition corresponding to Eq. (53) will be employed:

$$G(r, r_0) = \tfrac{1}{2} i\pi J_n(k_1 r_<) Z_n(k_1 r_>), \quad Z_n(k_1 r) = A_n J_n(k_1 r) + H_n^{(1)}(k_1 r) \quad (58)$$

$$-A_n = \frac{H_n^{(1)}(k_1 a) H_n^{(1)\prime}(ka) - H_n^{(1)\prime}(k_1 a) H_n^{(1)}(ka)}{J_n(k_1 a) H_n^{(1)\prime}(ka) - J_n'(k_1 a) H_n^{(1)}(ka)}.$$

In the above, the notation $r_<$ and $r_>$ is used to indicate $r_< = \min(r, r_0)$ and $r_> = \max(r, r_0)$. Equation (54) may now be inverted to yield

$$E^n(r) = (2i/\pi a C_n) J_n(k_1 r) - \int_0^a G(r, r_0) R(r_0) dr_0, \quad (59)$$

$$C_n = J_n(k_1 a) H_n^{(1)\prime}(ka) - J_n'(k_1 a) H_n^{(1)}(ka) \quad (60)$$

$$R(r_0) = (k_1^2 - k^2) r_0 E^n(r_0) - (i\omega\mu_0/2\pi) \times \int_0^{2\pi} e^{-in\theta_0} j_z(\mathbf{x}_0) r_0 d\theta_0. \quad (61)$$

From Eqs. (48) and (61) the integral term on the right-hand side of Eq. (59) may be placed in the form

$$\int_0^a G(r, r_0) R(r_0) dr_0 = \int_0^a r' E^n(r') [N_1(r, r') + N_2(r, r')] dr', \quad (62)$$

where

$$N_1(r, r') = (2\omega\omega_p^2/c^3) \int_0^a M^n(r_0, r') \times G(r, r_0) r_0 dr_0, \quad (63)$$

$$N_2(r, r') = (k_1^2 - k^2) G(r, r') - (i\omega\omega_p^2/c^3) \times \int_0^a \int_0^{2\pi} G(r, r_0) M(R) e^{-in\zeta} r_0 d\zeta dr_0 \quad (64)$$

with $R^2 = r^2 + r'^2 - 2rr'\cos\zeta$.

The Fredholm integral equation of the second kind is thus obtained:

$$E^n(r) = (2i/\pi a C_n) J_n(k_1 r) - \int_0^a r' [N_1(r, r') + N_2(r, r')] E^n(r') dr'. \quad (65)$$

It can be shown that the kernel is bounded. Briefly, this follows first of all from the requirement that, to invert the differential equations, restrictions are placed upon the parameters k, etc., such that the solution of the associated homogeneous equation with operator $L(r, k_1)$ and the homogeneous boundary conditions is unique, in which case the Green's function exists, and is bounded as well as being continuous (i.e., the constant C_n does not vanish). Secondly, the functions $M^n(r_0, r')$ and $M(R)$ have integrable singularities in their respective domains of integration in the expressions given by Eqs. (63) and (64).

We shall not be concerned here with obtaining estimates of the eigenvalues or eigenfunctions of the integral equation, which would correspond to the natural transverse plasma modes in a cylinder. For present purposes, the solution to the integral equation obtained by the Neumann series, which is valid if the norm of the integral operator is less than unity, will be considered. Instead of making estimates on the norm, use will be made of the fact (as shown in the next section) that the kernel vanishes as $\lambda \to \infty$. Hence in the low temperature region, the iterative solution is valid. Low temperatures corrections at least, to the scattered or total field can be obtained by the iterative procedure. The asymptotic behavior of the first iterate, given by

$$E^n(r) = (2i/\pi a C_n)\{J_n(k_1 r) - \int_0^a r' [N_1 + N_2] \times J_n(k_1 r') dr'\} \quad (66)$$

will be obtained in the next section.

Once the solutions have been found, the reflection coefficients associated with the scattered wave may be computed. If the total field outside the cylinder is given by

$$E(r, \theta) = \sum_{n=-\infty}^{\infty} e^{in\theta} [J_n(kr) + R_n H_n^{(1)}(kr)], \quad (67)$$

the reflection coefficients R_n may be obtained from the solution for $E^n(r)$ at $r = a$ as follows:

$$R_n = [E^n(a) - J_n(ka)]/H_n^{(1)}(ka). \quad (68)$$

ASYMPTOTIC SOLUTION

To determine the asymptotic nature of the kernel of the integral equation and the corresponding iterated solution in the limit of zero temperature, i.e., $\lambda \to \infty$, we will consider integrals of the form

$$M_m(R) = \int_0^\infty du_r \int_{-\infty}^\infty du_z F_0' u_z^2 \alpha^{-m} e^{i\alpha R}. \quad (69)$$

The asymptotic behavior for $m = 0$ is given in Appendix A. In a similar manner it can be shown that when

$$\lambda^{-1/2} \ll |[\omega + i\nu)/c\lambda]R| \ll 1,$$

the above integral has the asymptotic form

$$(1/6\pi)\lambda^{1/2 - m/2}(\omega + i\nu/c)^{-m} t^{m/3} \times \exp(-e^{-i\pi/3} t^{2/3} - im\pi/6),$$

where $t = (\omega + i\nu)\lambda^{1/2} R/c$.

This implies that the dominant contribution to integrals containing the product of the function M_m and a

slowly varying function (with respect to λ) will arise from the neighborhood of the point $R = 0$. Because of the factor $\lambda^{-m/2}$, the integrals containing M_m in the integrand will be essentially of higher order in λ than those that contain M_{m-1}, M_{m-2}, etc. To make use of this, we will integrate the integrals containing M_m in the integrand repeatedly by parts to produce higher order terms. The actual process of integration by parts will be performed in a different manner for the components $N_1(r,r')$ and $N_2(r,r')$ of the kernel.

The process of interation by parts for $N_2(r,r')$ will be achieved through the following identity:

$$\int_A (e^{i\alpha R}/R)\phi(\mathbf{x}_0)\,dA = (2i\pi/\alpha)\phi(\mathbf{x}') - (i/\alpha) \times \int_S (e^{i\alpha R}/R)\phi(\mathbf{x}_0)\hat{\mathbf{R}}\cdot\mathbf{n}\,ds + (i/\alpha)\int_A e^{i\alpha R}/R\hat{\mathbf{R}}\cdot\nabla_0\phi\,dA$$

for a function ϕ which is continuous in the domain A with boundary S. In the above, $\mathbf{x}_0$ is the variable of integrable, $\mathbf{x}'$ is a fixed point, $\hat{\mathbf{R}} = (\mathbf{x}_0 - \mathbf{x}')/R$, and $\mathbf{n}$ is the unit outward normal to the boundary S. By applying the above repeatedly one obtains

$$\int_A \frac{e^{i\alpha R}}{R}\,\phi(\mathbf{x}_0)\,dA \sim \frac{2i\pi}{\alpha}\,\phi(\mathbf{x}') - \frac{i\pi}{\alpha^3}\,\nabla^2\phi(\mathbf{x}') - \frac{i}{\alpha}\int_S \frac{e^{i\alpha R}}{R}\,\mathbf{n}\cdot\hat{\mathbf{R}}\left(\phi - \frac{1}{i\alpha}\,\frac{\partial\phi}{\partial R}\right)ds, \quad (70)$$

where higher-order terms are neglected. The above result will be applied to the integral involved in the expression for $N(r,r')$ given by Eq. (64), namely to

$$(-i\omega\omega_p^2/c^3)\int G(r,r_0)M(R)e^{-in\zeta}\,dA,$$

where the domain of integration with respect to the variables (r_0,ζ) is the disc $0 \le r_0 \le a$. By setting $\phi = G(r,r_0)e^{-in\zeta}$ and recalling expression (35) given for $M(R)$, application of (70) will yield the form

$$N_2(r,r') = N_2^S + N_2^d + N_2^0,$$

where N_2^S represents a boundary ($r_0 = a$) type of contribution, N_2^0 represents a contribution from the neighborhood of the point $R = 0$, and N_2^d represents a contribution from the discontinuity in the derivatives of G (arising from the boundary $r_0 = r$). It can be shown that the leading terms of the appropriate expressions are given by

$$N_2^S(r,r') = \frac{-\omega\omega_p^2}{c^3}\,a\left(G(r,a)B_1(a,r') + i\,\frac{\partial G(r,a)}{\partial a}\,B_2(a,r')\right), \quad (71)$$

$$N_2^d(r,r') = \frac{-i\omega\omega_p^2}{c^3}\,B_2(r,r'), \quad (72)$$

where

$$B_j(r,r') = \int_0^{2\pi} d\zeta\,\frac{e^{-in\zeta}}{R}\left(\frac{r - r'\cos\zeta}{R}\right)^j M_j(R) \quad (73)$$

with

$$R^2 = r^2 + r'^2 - 2rr'\cos\zeta.$$

The remaining term has a component which cancels out the first term in the right-hand side of expression (64) for $N_2(r,r')$ and, thus, is given by

$$N_2^0(r,r') \sim (\omega\omega_p^2/c^3)\pi k_1^2 M_3(0)G(r,r'). \quad (74)$$

With the comments made initially about the asymptotic behavior of $M_1(R)$ and $M_2(R)$, etc., and from the asymptotic relation

$$M_3(0) \sim [(\omega + i\nu)/c]^{-3}(1/\pi\lambda), \quad (75)$$

it is seen that the component N_2 of the kernel tends to zero as $\lambda \to \infty$.

Before investigating the asymptotic behavior of N_1, the asymptotic behavior of the first iterate of the solution of the integral equation for N_2 component of the kernel will be obtained for $r = a$. We will need to consider integrals of the form

$$\int_0^a N_2(a,r')J_n(k_1r')r'dr'. \quad (76)$$

First it can be shown from the Wronskian relation for the Bessel functions that the combination $N_2^S + N_2^d$ at $r = a$, reduces to

$$N_2^S(a,r') + N_2^d(a,r') = -(\omega\omega_p^2/c^3)\tfrac{1}{2}\pi i a Z_n(k_1a)\{B_1(a,r')J_n(k_1a) + iJ_n'(k_1a)B_2(a,r').$$

Inserting this expression into (75), it is seen that the following integrals

$$\int_0^a B_j(a,r')J_n(k_1r')r'dr', \qquad j = 1,2,$$

have to be asymptotically evaluated. The technique for doing so is similar in manner to the technique to obtain the asymptotic behavior of $N_2(r,r')$. However, in this case Eq. (69) has to be modified since the point given by $R = 0$ lies on the circle $r' = a$. The details will not be given here; but it can be shown that

$$\int_0^a B_1(a,r')J_n(k_1r')r'dr' \sim 2iM_2(0)J_n(k_1a) + \tfrac{1}{2}\pi M_3(0)J_n'(k_1a),$$

$$\int_0^a B_2(a,r')J_n(k_1r')r'dr' \sim \tfrac{1}{2}\pi M_3(0)J_n(k_1a);$$

hence, we have

$$\int_0^a (N_2^d + N_2^S)J_n(k_1r')r'dr' \sim (\omega\omega_p^2/c^3)\pi a Z_n(k_1a) \times [J_n(k_1a)]^2 M_2(0). \quad (77)$$

The remaining integral involving N_2^0 given by Eq. (74) is evaluated by direct integration and yields

$$\int_0^a N_2^0(a,r')J_n(k_1r')r'dr' \sim (\omega\omega_p^2/c^3)\tfrac{1}{4}i\pi^2 \times M_3(0)Z_n(k_1a)\{J_n^2(k_1a)[(k_1a)^2 - n^2] + a^2[J_n(k_1a)]^2\}. \quad (78)$$

We will now return to a brief discussion of the asymptotic behavior of $N_1(r,r')$ as $\lambda \to \infty$ and, in more detail, of the contribution of N_1 to the first iterate. Recall that N_1 is given by

$$N_1(r,r') = (2\omega\omega_p^2/c^3)\int_0^a M^n(r_0,r')G(r,r_0)r_0\,dr_0,$$

where M^n is given by expression (45). The integral representation for M^n contains the term $\exp(i\psi(a))/$

$\sin\psi(a)$ in the integrand, which for $\nu \neq 0$ and $|t| < a$, can be expanded in the convergent series

$$e^{i\psi(a)}[\sin\psi(a)]^{-1} = -2i \sum_{m=0}^{\infty} \exp[i(2m+2)\psi(a)].$$

With the expansion of the term $\cos\psi(r)\cos\psi(r')$ in the integrand into exponential components, the dominant contribution of the product is given by

$$e^{i\psi(a)}[\cos\psi(r)\cos\psi(r')/\sin\psi(a)] \sim \exp i[2\psi(a) - \psi(r) - \psi(r')],$$

since the exponent on the right-hand side contains the quantity

$$\alpha[2(a^2-t^2)^{1/2} - (r^2-t^2)^{1/2} - (r'^2-t^2)^{1/2}].$$

Referring back to the asymptotic analysis of integrals given by Eq. (69) and ignoring for the present the square root factors in the integral for M^n will be exponentially small except in the neighborhood of the boundary $r = a, r' = a$, and the dominant part of the integral arises from around $|t| = a$. Thus M^n is essentially a boundary type term. The asymptotic behavior of N_1 can be obtained through integration by parts with the variable r_0, such that factors $1/\alpha$ are introduced in the integrand. As is seen from the previous discussion, introduction of higher-order terms in α^{-1} produces higher-order terms in $\lambda^{-1/2}$.

The asymptotic behavior of the first iterate

$$\int_0^a N_1(a,r')J_n(k_1r')r'dr', \tag{79}$$

where

$$N_1(a,r') = (2\omega\omega_p^2/c^3)\tfrac{1}{2}i\pi Z_n(k_1a)\int_0^a M^n(r_0,r') \times J_n(k_1r')r'dr'$$

will be considered in more detail. Substitute in the integral representation for M^n and change the order of integration. Expression (79) then becomes

$$(\omega\omega_p^2/c^3)i\pi Z_n(k_1a)\int_0^\infty du_r \int_0^\infty du_z u_z^2 F_0' \times \int_{-\infty}^{\infty} dt\, e^{i\psi(a)}[P_n(t)]^2/\sin\psi(a), \tag{80}$$

where $P_n(t) = \int_{|t|}^a r\cos\psi(r)/(r^2-t^2)^{1/2}J_n(k_1r)dr.$ (81)

Next, the expression for $P_n(t)$ will be integrated by parts twice using the relations

$$\frac{r\cos\psi(r)}{(r^2-t^2)^{1/2}} = \frac{1}{\alpha}\frac{d\sin\psi(r)}{dr} + \frac{nt}{\alpha r}\frac{\cos\psi(r)}{(r^2-t^2)^{1/2}},$$

$$\sin\psi(r) = \frac{\cos\psi(r)}{\alpha(r^2-t^2)^{1/2}} + \frac{nt}{\alpha r^2}\sin\psi - \frac{1}{\alpha r}\frac{d}{dr} \times [(r^2-t^2)^{1/2}\cos\psi]$$

to obtain

$$P_n(t) \sim (1/\alpha)J_n(k_1a)\sin\psi(a) + (1/a\alpha^2)(a^2-t^2)^{1/2} \cos\psi(a)J_n'(k_1a) + (nt/a_2^2)J_n(k_1a)\sin\psi(a) + O(1/\alpha^3).$$

Squaring $P_n(t)$, and retaining teams in α up to α^{-3}, we have

$$\int_{-a}^{a} \frac{e^{i\psi(a)}}{\sin\psi(a)}[P_n(t)]^2 dt \sim \frac{iaJ_n^2(k_1a)}{\alpha^2} + \frac{a\pi}{2\alpha^3}J_n'(k_1a)J_n'(k_1a) + \frac{1}{\alpha^2}\int_{-a}^{a} dt\, e^{i2\psi(a)}f(t). \tag{82}$$

It can be shown that the last term yields a contribution the order of α^{-4} and hence neglected. Finally, by inserting expression (82) into (80), the asymptotic behavior of the first iterate (79) is given by

$$(\omega\omega_p^2/c^3)\pi aZ_n(k_1a)J_n(k_1a)[-M_2(0)J_n(k_1a) + \tfrac{1}{2}i\pi M_3(0)J_n'(k_1a)]. \tag{83}$$

By combining this with Eqs. (77) and (78), the following is obtained:

$$\int_0^a [N_1(a,r') + N_2(a,r')]J_n(k_1r')r'dr' = \tfrac{1}{4}i\pi^2(\omega\omega_p^2/c^3)M_3(0)Z_n(k_1a)Q(a),$$

$$Q(a) = [aJ_n'(k_1a) + J_n(k_1a)]^2 + J_n^2(k_1a)[(k_1a)^2 - n^2 - 1]. \tag{84}$$

From Eq. (66), the first-order temperature corrections to the nth Fourier component of the field on the field on the surface is given by

$$E^n(a) = \frac{2i}{\pi aC_n}\left(J_n(k_1a) - \frac{i\pi\omega\omega_p^2}{\lambda 4(\omega+i\nu)^3}Z_n(k_1a)Q(a)\right), \tag{85}$$

where the asymptotic form of $M_3(0)$ given by Eq. (75) is employed. The reflection coefficients can now be computed from (85) and (68).

It can be shown that expression (85) is equivalent to

$$E^n(a) = [2iJ_n(k_2a)/\pi aC_n(k_2a)] + O(\delta^2), \tag{86}$$

where

$$k_2 = k_1(1+\delta),\ \delta = -\omega_p^2\omega/2(\omega+i\nu)^3\lambda,$$

and $C_n(k_2a)$ is obtained from Eq. (60) with k_1 replaced by k_2. This is shown by expanding the above expression out in a power series in δ. Thus as the first temperature correction is concerned, the plasma behaves like a dielectric with relative dielectric constant $\epsilon(\lambda)(1+\delta)^2$, where $\epsilon(\lambda)$ is given by Eq. (56).

The reflection coefficients can now be computed from (86) and (68).

FORMAL ASYMPTOTIC EXPANSION

For polarization parallel to the axis of the cylinder, the technique employed in the previous section on the asymptotic solution to the integral equation, suggests a more formal but straightforward asymptotic approach. Namely the solution of the Boltzmann equation given by Eqs. (11) and (19), can yield a formal asymptotic series in inverse powers of α, by integrating the integrals by parts a sufficient number of times. The asymptotic properties of such a power series when integrated with respect to the velocity variables is made use of in the preceding section.

By integration by parts, we obtain

$$f(\mathbf{x},\mathbf{u}) \sim (g/i\alpha)\sum_{n=0}(i\alpha)^{-n}\{(\hat{\mathbf{u}}_r\cdot\nabla)^n\mathbf{E}(\mathbf{x}) + e^{i\alpha l}[(\hat{\mathbf{u}}_r^1\cdot\nabla)^n - (\hat{\mathbf{u}}_r\cdot\nabla)^n]\mathbf{E}_z(\mathbf{x}_0) + e^{i\alpha l+i\alpha l_1}[(\hat{\mathbf{u}}_r^2\cdot\nabla)^n - (\hat{\mathbf{u}}_r^1\cdot\nabla)^n]\mathbf{E}_z(\mathbf{x}_1) + \cdots\}.$$

For smooth convex shapes such as the circular cylinder, the contribution from the point $\mathbf{x}$ need only be considered when $\mathbf{x}$ is sufficiently far from the boundary. When $\mathbf{x}$ is close to the boundary, we need to consider the contribution due to the point $\mathbf{x}_0$ on the boundary, corresponding to the point of the first reflected velocity ray. The contribution due to the other points $\mathbf{x}_1$, etc., could be neglected. A simplification occurs by noting that

$$\mathbf{u}_r^1 = \mathbf{u}_r - 2(\mathbf{u}_r\cdot\mathbf{n})\mathbf{n},$$

where $\mathbf{n}$ is the unit outward normal to the surface at $\mathbf{x}_0$. In this case

$$(\hat{\mathbf{u}}_r^1\cdot\nabla) - (\hat{\mathbf{u}}_r\cdot\nabla) = -2(\hat{\mathbf{u}}_r\cdot\mathbf{n})\frac{\partial}{\partial n}.$$

Such a formal asymptotic expansion could be used at least for polarization parallel to the axis, for a nonhomogeneous plasma (but uniform with respect to the z variable). However, the case of the other polarization should be investigated first since longitudinal waves will have to be taken into account, and such expansions as above will have to be modified. This problem will be more difficult to handle since it is a singular perturbation problem.

CONCLUSION

It has been shown, employing kinetic theory, that the problem of scattering by a homogeneous circular plasma cylinder for polarization parallel to the axis, is reduced to a set of one-dimensional Fredholm integral equations of the second kind, for which the Neumann series converges for low temperature. The first-order temperature correction to the field on the surface is computed, from which the reflection coefficients may be obtained. The analysis could be extended to examine further the asymptotic approximations to the kernel and obtain estimates of the eigenvalues. However, of more importance would be to extend the analysis to the case of polarization perpendicular to the axis, in which case both longitudinal and transverse waves are included in the plasma. This problem should be reducible in a similar manner to two coupled integral equations.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under Grant GP-22583.

APPENDIX A

The asymptotic form of $M(R)$ will be obtained for $\lambda \to \infty$. From Eqs. (6), (7), and (35) we have

$$2\pi K_2(\lambda)RM(R) = \lambda^2\int_0^\infty\int_0^\infty \exp[i\alpha R - \lambda(1+u^2)^{1/2}] \times u_z^2(1+u^2)^{-1/2}\,du_r\,du_z. \quad \text{(A1)}$$

Integrate by parts with respect to u_z, and employ the resulting expression for the modified Hankel function

$$(1+u_r^2)^{1/2}K_1(q(1+u_r^2)^{1/2}) = \int_0^\infty \exp[iq(1+u^2)^{1/2}]\,du_z$$

to reduce (A1) to

$$2\pi K_2(\lambda)\,RM(R) = \lambda^2\int_0^\infty (1+u_r^2)^{1/2} \times K_1[q(1+u_r^2)^{1/2}]/q\;du_r,$$

where

$$q = \lambda - i[(\omega+i\nu)/cu_r]R. \quad \text{(A2)}$$

For $\lambda \gg 1$, the above expression becomes

$$RM(R) \sim \frac{\lambda^{5/2}e^\lambda}{2\pi}\int_0^\infty \frac{(1+u^2)^{1/2}}{q3/2} \times \exp[-q(1+u^2)^{1/2}]\,du. \quad \text{(A3)}$$

Saddle-point techniques may be employed to evaluate the integral when $|\omega+i\nu|R \gg c\lambda^{1/2}$. The saddle point occurs at u_0 lying in the fourth quadrant, where

$$u_0^3 = -i[(\omega+i\nu)/c\lambda]R;$$

when $|u_0| \ll 1$, the saddle-point approximation to (A3) becomes

$$RM(R) \sim \lambda^{1/2}(6\pi)^{-1/2}\exp(-e^{-i\pi/3}t^{2/3}), \quad \text{(A4)}$$

where $t = (\omega+i\nu)\lambda^{1/2}R/c$.

1 B. U. Felderhof, Physica **29**, 662 (1963).
2 J. J. Bowman and V. H. Weston, Phys. Fluids **11**, 601 (1968).
3 V. H. Weston, Phys. Fluids **10**, 632 (1967).
4 E. C. Taylor, Phys. Fluids **6**, 1305 (1963).
5 E. C. Taylor, Radio Sci. **69**, 735 (1965).
6 V. D. Shafranov, Zh. Eksp. Teor. Fiz. **34**, 1475 (1958) [Sov. Phys. JETP 7, 1019 (1958)].
7 C. T. Swift, thesis (William and Mary, 1969).
8 P. C. Clemmow and A. J. Willson, Proc. Roy. Soc. (London) **A237**, 117 (1956).

Lagrangian Density for Perfect Fluids in General Relativity

John R. Ray
Clemson University, Physics Department, Clemson, South Carolina 29631
(Received 28 April 1972)

A previously discussed variational principle for a perfect fluid in general relativity was restricted to irrotational, isentropic motions of the fluid. It is proven that these restrictions can be dropped, and the original variational principle can be generalized to general motions of the perfect fluid. The form of the basic Lagrangian density is unchanged by these generalizations. An Eulerian fluid description is used throughout. As a by-product of our variational principle, the 4-velocity is required to have the generalized Clebsch form.

1. INTRODUCTION

In general relativity one often uses a perfect fluid energy–momentum tensor to represent the sources of the gravitational field. The derivation of the equations of motion from a variational principle is complicated in this case by the constraint equations satisfied by the fluid variables. Several variational formulations have been discussed in the literature.[1–4] In a presentation we gave some time ago[3] we constructed a Lagrangian and formulated a variational principle for isentropic (constant entropy) motion of a perfect fluid in general relativity. In the present paper we show that our original variational principle describes only irrotational, isentropic motions of a perfect fluid and we generalize the variational principle so that it describes general motions of a perfect fluid. Our generalization does not change the form of the basic Lagrangian density, but adds further constraint equations which must be satisfied

2. THE ORIGINAL VARIATIONAL PRINCIPLE

Let ρ be the rest density, U^i the Eulerian 4-velocity, ϵ the rest, specific internal energy, and s the rest, specific entropy associated with the fluid.[5] In general, a perfect fluid has two thermodynamic degrees of freedom which we choose to be ρ and s. In the isentropic case, s is a constant, and all thermodynamic variables can be written in terms of ρ. The basic Lagrangian density given in Ref. 3 is

$$L_1 = c_1(-g)^{1/2}R - c^{-1}(-g)^{1/2}\rho[c^2 + \epsilon(\rho)], \qquad c_1 = c^3/16\pi k, \tag{2.1}$$

where c is the speed of light and k is the Newtonian gravitational constant. The first term $c_1(-g)^{1/2}R$ is the Lagrangian density for the vacuum gravitational field, and $-c^{-1}(-g)^{1/2}\rho(c^2+\epsilon)$ is the Lagrangian density for the fluid. Setting

$$F = -\rho(c^2 + \epsilon), \tag{2.2}$$

we have

$$L_1 = c_1(-g)^{1/2}R + c^{-1}(-g)^{1/2}F. \tag{2.3}$$

The fluid variables satisfy the following constraint equations:

$$(\rho U^i)_{;i} = 0, \tag{2.4a}$$

$$g_{ik}U^iU^k + c^2 = 0, \tag{2.4b}$$

where the semicolon denotes covariant differentiation. Equation (2.4a) expresses conservation of mass, and Eq. (2.4b) is the usual kinematical constraint on the 4-velocity. The Lagrangian density we must consider is[3]

$$L_2 = L_1 + (-g)^{1/2}\lambda_1(g_{ik}U^iU^k + c^2) + (-g)^{1/2}\lambda_2(\rho U^i)_{;i}. \tag{2.5}$$

Variation of the action associated with L_2 with respect to variables $g_{ik}, \rho, U^i, \lambda_1, \lambda_2$ yields the equation, of motion

$$\frac{\partial L_2}{\partial g_{ik}} - \left(\frac{\partial L_2}{\partial g_{ik,r}}\right)_{,r} + \left(\frac{\partial L_2}{\partial g_{ik,r,s}}\right)_{,s,r} = 0, \tag{2.6a}$$

$$\frac{\partial L_2}{\partial \rho} - \left(\frac{\partial L_2}{\partial \rho_{,i}}\right)_{,i} = 0, \tag{2.6b}$$

$$\frac{\partial L_2}{\partial U^i} - \left(\frac{\partial L_2}{\partial U^i{}_{,k}}\right)_{,k} = 0, \tag{2.6c}$$

$$\frac{\partial L_2}{\partial \lambda_1} = 0 \Rightarrow g_{ik}U^iU^k + c^2 = 0, \tag{2.6d}$$

$$\frac{\partial L_2}{\partial \lambda_2} = 0 \Rightarrow (\rho U^i)_{;i} = 0. \tag{2.6e}$$

Solving Eqs. (2.6b)–(2.6e) yield the following equations for the multipliers λ_1, λ_2:

$$\lambda_{2,i}U^i = c^{-1}F', \tag{2.7a}$$

$$\lambda_1 = -(2c^3)^{-1}\rho F', \tag{2.7b}$$

$$F' = \frac{dF(\rho)}{d\rho}.$$

Via these equations for the multipliers, Einstein's equations, Eqs. (2.6a), become

$$G^{ik} = (8\pi k/c^4)\, T^{ik}, \tag{2.8}$$

where

$$T^{ik} = -c^{-2}\rho F'U^iU^k + g^{ik}(F - \rho F') \tag{2.9}$$

is the energy-momentum tensor for the fluid. The combined first and second law of thermodynamics in the local rest frame of the fluid is[6]

$$Tds = d\epsilon + Pd(1/\rho), \tag{2.10}$$

where P is the pressure and T the absolute temperature. For isentropic flow,

$$\frac{d\epsilon}{d\rho} = P/\rho^2 \tag{2.11}$$

Combining Eqs. (2.2), (2.9), and (2.11) yields

$$T^{ik} = [\rho(1 + \epsilon/c^2 + P/\rho c^2)U^iU^k + g^{ik}P], \tag{2.12}$$

which is the energy–momentum tensor for a perfect fluid. Writing out Eqs. (2.6c), using Eqs. (2.7a), (2.7b), gives for the 4-velocity

$$U_i = -(F')^{-1}c^3\lambda_{2,i}. \tag{2.13}$$

Calculating the acceleration

$$\frac{DU_i}{d\tau} = U_{i;k}U^k$$

of the fluid using Eqs. (2.13) together with Eqs. (2.2), (2.4b), (2.7a), and (2.11) yields the equations of motion for the fluid, that is, the relativistic Euler equations[6,7]

$$\rho\left(1 + \frac{\epsilon}{c^2} + \frac{P}{\rho c^2}\right)\frac{DU_i}{d\tau} = -P_{,i} - c^{-2}U_i U^k P_{,k}. \tag{2.14}$$

These equations can also be derived using the Bianchi identities

$$T^{ik}{}_{;k} = 0$$

together with Eqs. (2.4a)(2.4b), and (2.11).[6,7]

3. ISENTROPIC ROTATIONAL FLOW

The angular velocity of rotation of the fluid is given by[8]

$$\omega^i = (2)^{-1}\eta^{ikrs}U_k U_{r,s} \tag{3.1}$$

If we calculate the angular velocity of the fluid using Eqs. (2.13) for U_i we find a surprising result

$$\omega^i = 0. \tag{3.2}$$

That is, the variational principle discussed in the previous section describes only isentropic, irrotational flow. This is the so-called Lin difficulty which appears even in Newtonian fluid mechanics.[9-11] We shall use the same technique (trick!) as Lin to generalize our variational principle so that it describes rotational as well as irrotational, isentropic motions. We assign to each particle a number $X(x)$ and require that this number not change as we move with the fluid particle

$$\frac{DX}{d\tau} = X_{,i}U^i = 0. \tag{3.3}$$

Equation (3.3) implies the conservation of particle identity. The Lagrangian density we must now consider is

$$L_3 = L_2 + (-g)^{1/2}\lambda_3 X_{,i}U^i. \tag{3.4}$$

The variations are carried out as before, and we obtain Eqs. (2.6) with L_3 as the Lagrangian density plus two new equations

$$\frac{\partial L_3}{\partial X} - \left(\frac{\partial L_3}{\partial X_{,i}}\right)_{,i} = 0, \tag{3.5a}$$

$$\frac{\partial L_3}{\partial \lambda_3} = 0 \Rightarrow X_{,i}U^i = 0. \tag{3.5b}$$

By using L_3 in place of L_2 in Eqs. (2.6), the only change comes in the equation for U^i, where instead of Eq. (2.13) we obtain

$$U_i = -c^3(F')^{-1}\lambda_{2,i} + c^3(\rho F')^{-1}\lambda_3 X_{,i}. \tag{3.6}$$

Therefore U_i no longer suffers the Lin difficulty, that is, ω^i is not necessarily zero. Thus both rotational and irrotational motions are included in the extremals of the variational principle. The multiplier equations, Eqs. (2.7a), and (2.7b), are the same as before as are Einstein's equations, Eqs. (2.8), (2.9) and (2.12). The relativistic Euler equations (2.14) can again be obtained either by differentiating U_i or from the Bianchi identities. Thus using the Lagrangian density L_3 we can described all isentropic motions of a perfect fluid.

4. GENERAL FLUID MOTION

As we remarked earlier a perfect fluid has, in general, two thermodynamic degrees of freedom which we choose to be ρ and s. In the general motion of a perfect fluid, s is not constant throughout the fluid, but is constant for a given fluid particle.[6] This means there is no heat exchange between different parts of the fluid. This condition implies that s not change as we move with a given fluid particle

$$s_{,i}U^i = 0. \tag{4.1}$$

The Lagrangian density we must now consider is

$$L = L_3 + (-g)^{1/2}\lambda_4 s_{,i}U^i. \tag{4.2}$$

The equations of motion in this case are Eqs. (2.6) and (3.5) using L as the Lagrangian density plus the two additional equations

$$\frac{\partial L}{\partial s} - \left(\frac{\partial L}{\partial s_{,i}}\right)_{,i} = 0, \tag{4.3a}$$

$$\frac{\partial L}{\partial \lambda_4} = 0 \Rightarrow s_{,i}U^i = 0. \tag{4.3b}$$

In writing out the equations of motion in this case, we must remember that the fluid has two thermodynamic degrees of freedom ρ, s. Therefore, instead of Eq. (2.2), we have

$$F(\rho, s) = -\rho[c^2 + \epsilon(\rho, s)]. \tag{4.4}$$

Also we must use the combined first and second law of thermodynamics as given in Eq. (2.10). Writing out the equations of motions just as before yields solutions analogous to Eqs. (2.7a), and (2.7b)

$$\lambda_{2,i}U^i = c^{-1}F', \tag{4.5a}$$

$$\lambda_1 = -(2c^3)^{-1}\rho F', \tag{4.5b}$$

where now

$$F' = \left(\frac{\partial F}{\partial \rho}\right)_s.$$

For Einstein's equations we obtain

$$G^{ik} = (8\pi k/c^4)\ T^{ik} \tag{4.6}$$

with the energy–momentum tensor

$$T^{ik} = -c^{-2}\rho F' U^i U^k + g^{ik}(F - \rho F'). \tag{4.7}$$

Using Eq. (2.10), we obtain

$$\left(\frac{\partial \epsilon}{\partial \rho}\right)_s = \frac{P}{\rho^2}, \tag{4.8}$$

which when combined with Eqs. (4.4) and (4.7) yields

$$T^{ik} = [\rho(1 + \epsilon/c^2 + P/\rho c^2)U^i U^k + g^{ik}P], \qquad (4.9)$$

which is again the energy–momentum tensor for the perfect fluid. The velocity equation yields in this case

$$U_i = -c^3(F')^{-1}\lambda_{2,i} + c^3(\rho F')^{-1}\lambda_3 X_{,i} + c^3(\rho F')^{-1}\lambda_4 s_{,i}. \qquad (4.10)$$

The Euler equations can again either be obtained by taking the derivative of U_i or by using the Bianchi identities. For isentropic motions $s_{,i} = 0$, we again see that U_i does not suffer Lin difficulty.

5. CONCLUSIONS

We have shown that the variational principle given in Ref. 3 generalizes to include the general motions of a perfect fluid. The basic Lagrangian density L_1 is the same as in Ref. 3:

$$L_1 = c_1(-g)^{1/2}R - c^{-1}(-g)^{1/2}\rho(c^2 + \epsilon),$$

however, we must introduce two new contraint equations which the fluid variables must satisfy. These constraint equations are

$X_{,i}U^i = 0$, conservation of particle identity,

$s_{,i}U^i = 0$, conservation of entropy for fluid particles.

The first equation removes the Lin difficulty from the variational principle, that is, for isentropic fluid motions the extremals of our variational principle describe both rotational and irrotational fluid motions. The second equation removes the requirement of isentropic motions. The Lagrangian density L we must consider is that of Ref. 3 plus the two new constraints

$$L = L_1 + (-g)^{1/2}(\lambda_1(g_{ik}U^iU^k + c^2) + \lambda_2(\rho U^i)_{;i} + \lambda_3 X_{,i}U^i + \lambda_4 s_{,i}U^i).$$

When we vary the action associated with L with respect to the variables $g_{ik}, \rho, U^i, X, s, \lambda_1, \lambda_2, \lambda_3, \lambda_4$, we obtain the Einstein field equations for perfect fluid sources plus the equations of motion of the fluid. The Eulerian fluid description is used throughout our variational principle. In our formulation it follows from the variational principle that U_i has the generalized Clebsch form given in Eqs. (4.10). In the other variational principles that use the Eulerian fluid description, the generalized Clebsch form for U_i is either built directly into the Lagrangian density[2] or simply assumed.[4] The fact that we derive the generalized Clebsch form for U_i allows us to interpret some of the terms occurring in it as Lagrange multipliers associated with conservation laws. For example, in Eqs. (4.10), λ_2 is associated with conservation of mass, λ_3 is associated with conservation of particle identity, and λ_4 is associated with conservation of entropy. The other quantities occurring in Eqs. (4.10) are variables associated with the fluid. This is the same type of result one obtains in the nonrelativistic case.[9–11]

We feel that the variational principle presented in this paper is the most natural variational principle for a perfect fluid in general relativity. Our variational principle is a relativistic generalization of the usual nonrelativistic variational principle for perfect fluids. Besides the Einstein field equations and the relativistic Euler equations, our variational principle implies that the 4-velocity can be represented in the generalized Clebsch form.

ACKNOWLEDGMENT

The author thanks Dr. Bernard Schutz, Jr. for an interesting letter.

---

1 A. Taub, Phys. Rev. **94**, 1468 (1954).
2 K. Tam and J. O'Hanlon, Nuovo Cimento **62B**, 351 (1966).
3 J. Ray, Acta Phys. Polon. **30**, 481 (1966). Also J. Ray, thesis (Ohio University, 1964) (available from University Microfilms, Ann Arbor, Michigan).
4 B. Schutz, Jr., Phys. Rev. D **2**, 2762 (1970), and references contained therein.
5 By rest quantity we shall mean that the quantity is measured in the local rest frame of the fluid. By specific quantity we shall mean that the quantity is per unit mass. Thus, for example, ϵ, the rest, specific internal energy, is the internal energy per unit mass in the locat rest frame of the fluid. Our signature convention is (+++ −) and we define the Ricci tensor by $R_{ik} = R^l{}_{ilk}$.
6 L. Landau and E. Lifshitz, *Fluid Mechanics* (Addison–Wesley, Reading, Mass., 1959).
7 A. Lichnerowitz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
8 J. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1964), p. 172.
9 C. Lin, in *Liquid Helium*, Proceedings of the International School of Physics Enrico Fermi, course 21, edited by G. Careri (Academic, New York, 1963).
10 J. Serrin, in *Handbuch der Physik* (Springer-Verlag, Berlin, 1959), Vol. 8/1.
11 G. Uhlenbeck, Lectures at the University of Colorado, summer 1968, unpublished.

Kramers-Kronig Relations and Sum Rules*

Ryogo Kubo
Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11790
and
Department of Physics, University of Tokyo, Bunkyoku, Tokyo

and

Munetake Ichimura
Department of Physics, State University of New York, Stony Brook, New York 11790
(Received 27 April 1972)

The concept of moments, which are integrals of positive or negative integral powers ω^n weighted by real or imaginary parts of admittance functions, is here generalized so as to be applied to a wide category of admittance functions, including Lorentzian functions. The generalized moments are related to the derivatives or integrals of sum rules in a general sense. This analysis is based on differentiation and integration–mapping of admittance functions and the associated Kramers–Kronig relations. Some model calculations are also shown.

1. INTRODUCTION

An admittance function $X(\omega)$ is defined by

$$X(\omega) = X'(\omega) - iX''(\omega) = \int_0^\infty dt e^{-i\omega t}\phi(t), \qquad (1.1)$$

where $\phi(t)$ is a real response function of the system to a pulsive perturbation. If the initial value of ϕ is bounded, i.e.,

$$\phi(+0) < \infty,$$

then

$$X(\infty) = \lim_{\omega\to\infty} X(\omega) = 0 \qquad (1.2)$$

may be assumed by Abel's theorem. Further, if the static response is finite, namely if

$$X(0) = \int_0^\infty dt\, \phi(t) < \infty, \qquad (1.3)$$

[which may be more generally the limiting value of $X(\omega)$ as $\omega \to +0$], then the function $X(\omega)$ is analytic on the lower half-plane of ω, so that we have the equation

$$X(\omega) + \frac{1}{\pi i}\int_{-\infty}^{\infty} \frac{X(\nu)}{\nu - \omega}\, d\nu = 0, \qquad (1.4)$$

by applying Cauchy's theorem to an infinite semicircle in the lower half-plane of ν closed by a straight line along the real axis with a small semicircle around the point $\nu = \omega$. The integral means, thus, its principal value. Equation (1.4) gives the Kramers–Kronig relations[1]

$$X'(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X''(\nu)}{\nu - \omega}\, d\nu \qquad (1.5a)$$

and

$$X''(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X'(\nu)}{\nu - \omega}\, d\nu. \qquad (1.5b)$$

The real part $X'(\omega)$ and the imaginary part $X''(\omega)$ are the cosine- and sine-Fourier integrals

$$X'(\omega) = \int_0^\infty dt\, \cos\omega t \cdot \phi(t), \qquad (1.6a)$$

$$X''(\omega) = \int_0^\infty dt\, \sin\omega t \cdot \phi(t), \qquad (1.6b)$$

and so they are even and odd in ω, respectively.

Equations (1.5) lead to the sum rules

$$X'(0) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X''(\omega)}{\omega}\, d\omega \qquad (1.7a)$$

and

$$[\omega X'']_\infty \equiv \lim_{\omega\to\infty} \omega X''(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} X'(\omega)\, d\omega, \qquad (1.7b)$$

of which the f-sum rule and the thermodynamic sum rule are familiar special cases.[2-4] If the expansion

$$\frac{1}{\nu - \omega} \simeq \sum_{n=0}^{\infty} \frac{\omega^n}{\nu^{n+1}} = -\sum_{n=0}^{\infty} \frac{\nu^n}{\omega^{n+1}}$$

is inserted into Eqn. (1.4), it is transformed into

$$X(\omega) \sim -\sum \frac{\omega^n}{\pi i}\int_{-\infty}^{\infty} \frac{X(\nu)}{\nu^{n+1}}\, d\nu \qquad (1.8a)$$

or into

$$X(\omega) \sim -\sum \frac{1}{\omega^{n+1}}\frac{1}{\pi i}\int_{-\infty}^{\infty} X(\nu)\nu^n d\nu, \qquad (1.8b)$$

which are in fact merely formal. If these expansions were convergent, they would give a set of sum rules for positive and negative moments defined with respect to the real or the imaginary part of the admittance function $X(\omega)$ which should be equated to the expansion coefficients of $X(\omega)$ in powers of ω or $1/\omega$. Generally speaking, such moments in the form as suggested by Eqs. (1.8) may not exist at all, and the expansions (1.8) may not make sense or may not be convergent.

If the response function $\phi(t)$ is analytic in t at $t = 0$, the inversion of Eq. (1.6a) or Eq. (1.6b) gives the sum rules[3]

$$\phi^{(2n)}(0) = \frac{1}{\pi}\int_{-\infty}^{\infty} X'(\omega)\,\omega^{2n}\, d\omega \qquad (1.9)$$

or

$$\phi^{(2n+1)}(0) = \frac{1}{\pi}\int_{-\infty}^{\infty} X''(\omega)\,\omega^{2n+1}\, d\omega, \qquad (1.10)$$

depending on whether $\phi(t)$ is even or odd in t. If $\phi(t)$ is not differentiable at $t = 0$, beyond a certain order, integrals of higher-order moments will not be convergent.

The present work has been motivated by asking ourselves the question how the sum rules of the form (1.7) and those given by Eq. (1.9) or (1.10) are related. This question will be answered by introducing a reasonable generalization of the concept of moments for an admittance function which may or may not possess moments in the ordinary sense. Thus the generalized moments can be applied to derive a set of sum rules in relation to the moment expansions of admittance functions from a somewhat unified point of view. We do not aim here at a very great generality, but are satisfied at this stage to find such a generalization of basic concepts which will hopefully be useful to obtain a better understanding of nature of sum rules appearing in a great variety in many problems of physics and also for some practical procedures of analyzing admittance functions obtained by various methods of spectroscopy.

2. DIFFERENTIATION AND INTEGRATION OF RESPONSE AND ADMITTANCE FUNCTIONS

Physically, a response function $\phi(t)$ is expected to behave in a moderate way. It is analytic for real t's with possible exceptions at $t = 0$ or $t = \infty$. At $t = 0$, the microscopic law of dynamics should prevail as a matter of principle so that any response should be *analytic* at the origin.[3] In practice, however, this region of *dynamic coherence* may be limited to the immediate vicinity of the origin. If that is the case, the response function may look there nonanalytic. The simplest example,

$$\phi(t) = e^{-\gamma|t|}, \quad \gamma > 0, \tag{2.1}$$

is most familiar as a typical form of response functions. It is inconsistent with dynamics for $t \sim 0$, but is adopted as an idealization or as an approximation to be justified under such a circumstance where the microscopic characteristic time is far shorter than the macroscopic time scale. In some cases, a response function is approximated by a function diverging at $t = 0$; but we shall exclude such a function in the following treatment and assume that

$$\phi(0) \equiv \lim_{t \to +0} \phi(t) < \infty \tag{2.2}$$

namely that $\phi(0)$ is bounded. At $t = \infty$, the irreversibility reveals itself in a macroscopic system to result in the decay of a response. Thus we assume that

$$\phi(\infty) \equiv \lim_{t \to \infty} \phi(t) = 0. \tag{2.3}$$

If, further, the static response should be finite, we may expect that the condition (1.3) holds, or more precisely

$$X(0) = \lim_{\omega \to +0} \int_0^\infty e^{-i\omega t}\, \phi(t)\, dt < \infty. \tag{2.4}$$

For a given response function $\phi(t)$ and the corresponding admittance function $X(\omega)$, we note the following basic transformations;[2]

$$X(\omega) = \int_0^\infty dt\, e^{-i\omega t}\, \phi(t)$$

$$= \frac{\phi(0)}{i\omega} + \frac{1}{i\omega} \int_0^\infty dt\, e^{-i\omega t}\, \dot{\phi}(t) \tag{2.5}$$

$$= \Phi(0) - i\omega \int_0^\infty dt\, e^{-i\omega t}\, \Phi(t), \tag{2.6}$$

where

$$\Phi(t) = \int_t^\infty \phi(t')\, dt' \tag{2.7}$$

is the corresponding relaxation function. These are due to assumptions (2.2)–(2.4).

The transformation (2.5) defines a mapping of the admittance function $X(\omega)$ to

$$\mathfrak{D}X(\omega) = \int_0^\infty dt\, e^{-i\omega t}\, \dot{\phi}(t) = i\omega X(\omega) - [i\omega X]_\infty, \tag{2.8}$$

where we have

$$[i\omega X]_\infty \equiv \lim_{\omega \to \infty} i\omega X(\omega) = \phi(0), \tag{2.9}$$

$$[\omega X']_\infty = 0, \quad [\omega X'']_\infty = \phi(0). \tag{2.10}$$

This is due to Abel's theorem. This mapping may be called *differentiation mapping*. Similarly, Eq. (2.6) defines the mapping

$$X(\omega) \to \mathfrak{D}^{-1}X(\omega) = -\int_0^\infty dt\, e^{-i\omega t}\, \Phi(t) = \frac{X(\omega) - X(0)}{i\omega}, \tag{2.11}$$

where we have

$$X(0) = X'(0) = \int_0^\infty dt\ \phi(t), \quad X''(0) = 0. \tag{2.12}$$

This may be called *integration mapping*. Obviously, differentiation and integration are mutually inverse; namely

$$\mathfrak{D}\mathfrak{D}^{-1} = \mathfrak{D}^{-1}\mathfrak{D} = 1. \tag{2.13}$$

Differentiation or integration may be repeated as long as the derived response function

$$\phi_n(t) \equiv \frac{d^n}{dt^n}\, \phi(t) \tag{2.14}$$

or

$$\phi_{-n}(t) = (-)^n \int_t^\infty dt_1 \int_{t_1}^\infty dt_2 \cdots \int_{t_{n-1}}^\infty dt_n \phi(t_n) \tag{2.15}$$

exists and satisfies conditions (2.2)–(2.4). Thus, to a given admittance function $X(\omega)$, there will be associated a set of mapped admittance functions

$$\{X_l(\omega)\}, \quad l = -m, -m+1, \ldots, n \tag{2.16}$$

and the corresponding response functions $\phi_l(t)$. We then have

$$X_l(\omega) = \int_0^\infty dt \cdot e^{-i\omega t}\, \phi_l(t) \tag{2.17}$$

for each member of the set. This set is finite if further differentiation or integration leads to undesirable divergences. We may say then that the response function $\phi(t)$ or the admittance function $X(\omega)$ is *regular* to that order. If differentiation or integration can be repeated indefinitely, we have an infinite number of mapped functions. Such an admittance function $X(\omega)$ may be said to be infinitely regular. An obvious example of this is afforded by $\phi(t)$, Eq. (2.1), for which we have

$$\phi_n(t) = (-\gamma)^n e^{-\gamma|t|}, \tag{2.18a}$$

$$X_n(\omega) = \frac{(-\gamma)^n}{\gamma + i\omega} \tag{2.18b}$$

for positive and negative integers n. This can be generalized to a poly-disperse system, namely to

$$\phi(t) = \sum_j \phi_j \exp(-\gamma_j|t|), \tag{2.19}$$

for which we have

$$\phi_n(t) = \sum_j \phi_j(-\gamma_j)^n \exp(-\gamma_j|t|), \tag{2.20a}$$

$$X_n(t) = \sum_j \phi_j[(-\gamma_j)^n/(\gamma_j + i\omega)]. \tag{2.20b}$$

Note, however, that the weights ϕ_j in Eq. (2.19) are not necessarily positive nor real. Such an example is given by the admittance function

$$X(\omega) = [i\omega + \gamma/(i\omega\tau + 1)]. \tag{2.21}$$

(This may be regarded as a modified Lorentzian function satisfying the conditions $\lambda_1 = 0$ and $\lambda_2 =$ finite.)

If the number of relaxation modes is finite, the corresponding admittance is infinitely regular. If there

exist an infinite number of relaxation modes or a continuous spectrum of relaxation frequency γ, the order of regularity depends on convergence of its moments defined with respect to the weights ϕ_j.

It is emphasized here, in order to avoid misunderstanding, that the function (2.19) represents only a special family of response functions. The whole analysis made here is applicable to much wider classes of regular response functions.

3. MOMENT EXPANSIONS

The transformation (2.5) can be repeated to yield

$$X(\omega) = \frac{\phi(0)}{i\omega} + \frac{\phi_1(0)}{(i\omega)^2} + \cdots + \frac{\phi_n(0)}{(i\omega)^{n+1}} + \frac{1}{(i\omega)^{n+1}} \int_0^\infty dt\, e^{-i\omega t}\, \phi_{n+1}(t) \quad (3.1)$$

as far as $\phi_n(0)$ remains finite. Similarly the transformation (2.6) gives

$$X(\omega) = -\phi_{-1}(0) - \phi_{-2}(0)\, i\omega - \cdots - \phi_{-n-1}(0)\,(i\omega)^n + (i\omega)^{n+1} \int_0^\infty dt\, e^{-i\omega t}\, \phi_{-n-1}(t). \quad (3.2)$$

These series are terminated with residual terms corresponding to the order of regularity of $X(\omega)$. If $X(\omega)$ is infinitely regular, the expansion (3.1) or (3.2) may be pushed to infinite order. However, the resulting series may not be convergent, but merely semi-convergent or asymptotic. Keeping this observation in mind, we write formal expansions of $X(\omega)$ in the following way:

$$X(\omega) \sim (1/i) \sum_{n \geq 0} (\lambda_n/\omega^{n+1}) \quad (3.3a)$$

$$X(\omega) \sim i \sum_{n \geq 0} \lambda_{-n-1}\, \omega^n. \quad (3.3b)$$

This defines the moment-expansions of $X(\omega)$; the expression (3.3a) is for *positive* moments and (3.3b) for *negative* moments. Depending on the order of regularity of $X(\omega)$, the positive or the negative moment-expansion is terminated at a finite order or extended to infinite order. The moments λ_n are related to the response function ϕ by

$$\lambda_n = i^{-n}\, \phi_n(0), \quad n \lesseqgtr 0 \quad (3.4)$$

or

$$\lambda_n = i^{-n} \left[\left(\frac{d}{dt}\right)^n \phi\right]_0, \quad (3.5a)$$

$$\lambda_{-n} = (-i)^n \int_0^\infty dt_1 \int_{t_1}^\infty dt_2 \cdots \int_{t_{n-1}}^\infty dt_n \phi(t_n). \quad (3.5b)$$

The negative moments can also be expressed as

$$\lambda_{-n-1} = \frac{(-i)^{n+1}}{n!} \int_0^\infty t^n \phi(t)\, dt \quad (3.6)$$

if the integral is convergent.

For the example of a monodisperse system, we have by Eq. (2.18)

$$X(\omega) = 1/(\gamma + i\omega) \quad (3.7)$$

and

$$\lambda_n = i^n \gamma^n, \quad n \gtreqless 0. \quad (3.8)$$

If the system is polydisperse as represented by Eq. (2.19), the moments are given by

$$\lambda_n = i^n \sum \phi_j \gamma_j^n = i^n \langle \gamma^n \rangle\, \phi(0), \quad (3.9)$$

that is the moment-average of relaxation frequencies. This may diverge depending on the nature of the spectrum of relaxation frequencies.

If the response function $\phi(t)$ is *analytic* at $t = 0$, it can be divided into the even and the odd parts. The positive moment-expansion (3.3a) contains only odd powers of ω^{-1} if ϕ is even in t and only even powers of ω^{-1} if ϕ is odd. This implies that the expansion (3.3) will then be only *asymptotic*. Simple examples of this are

$$\phi(t) = e^{-ct^2}, \quad \phi(t) = te^{-ct^2}. \quad (3.10)$$

Differentiation or integration mapping of an admittance function is a very simple mapping for the moment-expansions. Definition (2.8) means for the positive moment-expansion (3.3a) that the first ($n = 0$) term be removed, the powers of ω^{-1} be lowered by one, and the whole be multiplied by i; namely,

$$\mathfrak{D} X(\omega) = X_1 \sim \sum_{k \geq 0} (\lambda_{k+1}/\omega^{k+1}).$$

Similarly, integration (2.11) means for the negative moment-expansion (3.3b) to remove the first term ($n = 0$), to lower the powers of ω by one, and to divide the whole by i; namely,

$$\mathfrak{D}^{-1} X(\omega) = X_{-1} \sim \sum_{k \geq 0} \lambda_{-k-2}\, \omega^k.$$

More generally we have

$$X_n(\omega) \equiv \mathfrak{D}^n X(\omega) \sim i^{n-1} \sum_{k \geq 0} (\lambda_{n+k}/\omega^{k+1}) = (i\omega)^n \left(X(\omega) - \frac{1}{i} \sum_{k=0}^{n-1} \frac{\lambda_k}{\omega^{k+1}}\right) \quad (3.11a)$$

and

$$X_{-n}(\omega) \equiv \mathfrak{D}^{-n} X(\omega) \sim i^{-n+1} \sum_{k \geq 0} \lambda_{-n-k}\, \omega^k = \frac{1}{(i\omega)^n} \left(X(\omega) - i \sum_{k=0}^{n=1} \lambda_{-k-1}\, \omega^k\right). \quad (3.11b)$$

4. KRAMERS–KRONIG RELATIONS AND SUM RULES

To the order as it is regular, an addmittance function $X(\omega)$ is associated with its differential or integration mappings (2.16). Each member of these admittance functions satisfies the Kramers–Kronig relations

$$X_n'(\omega) = \frac{1}{\pi} \int_{-\infty}^\infty \frac{X_n''(\nu)}{\nu - \omega}\, d\nu, \quad (4.1a)$$

$$X_n''(\omega) = -\frac{1}{\pi} \int_{-\infty}^\infty \frac{X_n'(\nu)}{\nu - \omega}\, d\nu, \quad (4.1b)$$

where X_n' and $-X_n''$ are the real and the imaginary parts of X_n, respectively.

These relations, (4.1), are in themselves worth noting though they are obvious: Because the use of such relations for $n \neq 0$, instead of that for the original admittance, can be more advantageous for a detailed analysis. For example, if $X(\omega)$ is the conductivity function of electrons, we know that

$$[i\omega X]_\infty = [\omega X'']_\infty = ne^2/m \quad (4.2)$$

in the customary notations. Thus the mapping $X(\omega) \to X_1(\omega)$ subtracts away the high frequency part of the admittance. If $X(\omega)$ is a susceptibility function, the static part is subtracted by integration. This may be repeated to the extent that we have good knowledge of moments, leaving the unknown part of the admittance in favor of the Kramers–Kronig relations.

The Kramers–Kronig relations (4.1) provide us with a set of sum rules for the generalized moments, as we shall see in the following. This will clarify the reason why λ_n introduced in (3.3) are named as the generalized moments. As we have noticed in the Introduction, Eqs. (4.1) gives the equalities

$$X_n'(0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X_n''(\omega)}{\omega}\, d\omega. \tag{4.3a}$$

$$[\omega X_n'']_{\infty} = \frac{1}{\pi} \int_{-\infty}^{\infty} X_n'(\omega)\, d\omega. \tag{4.3b}$$

When X_n' and X_n'' are expressed in terms of the original admittance function $X(\omega)$, there will be obtained sum rules for $X(\omega)$. We note first the following equalities;

$$X_n''(0) = 0, \quad [\omega X_n']_{\infty} = 0, \tag{4.4}$$

$$-(X_{n+1}''/\omega)_0 = X_n'(0) = -[\omega X_{n-1}'']_{\infty} \tag{4.5}$$

The first of (4.4) is obvious from (2.12) applied to $\phi_n(t)$, and the second from (2.10) applied to X_n. Equations (2.8) and (2.11) applied to X_n give (4.5). Now we find

$$i^n \lambda_n = [\omega X_n'']_{\infty} = -X_{n+1}'(0) \tag{4.6}$$

by (3.4), (2.10), (2.17), and (4.5). Thus we write Eq. (4.3b) as

$$\lambda_n = \frac{i^{-n}}{\pi} \int_{-\infty}^{\infty} X_n'(\omega)\, d\omega, \quad n \gtrless 0, \tag{4.7}$$

where the explicit forms of the integrand is found from (3.11) as follows: for positive moments,

$$X_{2n}'(\omega) = (-)^n \omega^{2n} \left(X'(\omega) + \sum_{k=0}^{n-1} \frac{i\lambda_{2k+1}}{\omega^{2k+2}} \right), \tag{4.8a}$$

$$X_{2n+1}'(\omega) = (-)^n \omega^{2n+1} \left(X''(\omega) - \sum_{k=0}^{n} \frac{\lambda_{2k}}{\omega^{2k+1}} \right) \tag{4.8b}$$

and, for negative moments,

$$X_{-2n-1}'(\omega) = (-)^n \omega^{-2n-1} \left(-X''(\omega) - \sum_{k=0}^{n-1} \lambda_{-2k-2}\, \omega^{2k+1} \right), \tag{4.9a}$$

$$X_{-2n}'(\omega) = (-)^n \omega^{-2n} \left(X'(\omega) - \sum_{k=0}^{n-1} i\lambda_{-2k-1}\, \omega^{2k} \right). \tag{4.9b}$$

Thus we see that the right-hand side of expression (4.7) deserves the name of moments. The weight functions are not necessarily the original admittance functions, but are modified suitably so as to avoid obvious divergences and to yield higher moments, positive or negative, in a successive manner. Equation (4.7) shows that the even moments are *real* and odd moments are *imaginary*. This is rather a matter of convenience, but is useful to secure some elegance in the formal expressions. It is, of course, easy to derive (3.5) and (3.6) from Eq. (4.7).

By Eqs. (3.5) and (3.6), Eq. (4.7) may be written as

$$\left[\left(\frac{d}{dt} \right)^n \phi \right]_0 = \frac{1}{\pi} \int_{-\infty}^{\infty} X_n'(\omega)\, d\omega \tag{4.10a}$$

for positive moments, and

$$\frac{1}{(n-1)!} \int_0^{\infty} t^{n-1} \phi(t)\, dt = \int_0^{\infty} dt_1 \int_{t_1}^{\infty} dt_2 \cdots \int_{t_{n-1}}^{\infty} dt_n \phi(t_n) = \frac{1}{\pi} \int_{-\infty}^{\infty} X_{-n}'(\omega)\, d\omega \tag{4.10b}$$

for negative moments. These should properly be called sum rules.

If the function $\phi(t)$ is analytic at $t = 0$ and is even in t, all positive odd moments vanish identically. Then Eq. (4.7) recovers the usual form

$$\lambda_{2n} = (-)^n \phi^{(2n)}(0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \omega^{2n} X'(\omega)\, d\omega. \tag{4.11}$$

Vanishing of the odd moment-integrals can also be easily proved. If the function $\phi(t)$ is odd, all positive even moments vanish.

The equality (4.10b) represents a great variety of definite integrals to be defined by suitable choices of the function $\phi(t)$.

The expression (4.10b) can be identified with the moments of relaxation-time distribution, namely

$$\langle \tau^n \rangle = \frac{(-)^n}{(n-1)!} \int_0^{\infty} t^{n-1} \phi(t)\, dt \Big/ \int_0^{\infty} \phi(t)\, dt = (-)^n \int_{\infty}^{\infty} X_{-n}'(\omega)\, d\omega \Big/ \int_{-\infty}^{\infty} X'(\omega)\, d\omega, \quad n \geq 1. \tag{4.12}$$

This is equal to

$$\langle \tau^n \rangle = \sum \gamma_j^{-n} \phi_j \Big/ \sum \phi_j \tag{4.13}$$

for a polydisperse system (for which the positive λ_n are essentially the moments of relaxation frequencies), but more generally it is a formal definition of such moments because the decay of $\phi(t)$ needs not be a superposition of exponential decay.

5. MODEL CALCULATIONS

A puzzling problem one may encounter in using the Kramers–Kronig relations or the sum rules will be that the integration range of frequency is practically limited by some finite bounds. Experimental data may not be available in wider ranges of frequency because of technical difficulties or of unavoidable noises. It is, therefore, very desirable to develop a method to estimate or to correct errors caused by such limitations. Leaving this general task to the future, we discuss here some examples of application of the foregoing analysis to a few typical models which illustrate some important features of our problem.

For this purpose we consider an oscillator the frequency of which is randomly modulated.[5-6] It follows the equation of motion

$$\dot{x} = i\Omega(t)\, x, \tag{5.1}$$

where x is the coordinate of the oscillator and $\Omega(t)$ is the modulated frequency, its average being taken as the zero of the frequency scale. If the modulation is described by a stationary Gaussian process, the relaxation function $\phi(t)$ of the oscillator is given by

$$\phi(t) = \exp\left(-\int_0^t (t-\tau)\psi(\tau)\, d\tau \right), \tag{5.2}$$

where $\psi(t)$ is the correlation function of the frequency modulation, namely

$$\psi(t) = \langle \Omega(t_0)\,\Omega(t_0 + t)\rangle\,. \tag{5.3}$$

The admittance function $X(\omega)$ is then defined by

$$\begin{aligned} X(\omega) &\equiv X'(\omega) - iX''(\omega) \\ &= \int_0^\infty e^{-i\omega t}\,\phi(t)\,dt, \end{aligned} \tag{5.4}$$

of which the real part $X'(\omega)$ represents the absorption of energy when the oscillator is subject to a periodic driving force and the imaginary part represents the dispersion.

If the modulation $\Omega(t)$ changes in time very slowly, the absorption curve $X'(\omega)$ is Gaussian, reflecting the Gaussian assumption for $\Omega(t)$. If the modulation is very fast, the absorption curve is motional-narrowed to a Lorentzian form. This corresponds to the change of the relaxation function $\phi(t)$ from a Gaussian decay

$$\phi(t) \sim \exp[-\tfrac{1}{2}\,\psi(0)\,t^2] \tag{5.5}$$

in a slow-modulation case to a simple exponential decay

$$\phi(t) \sim \exp\left(-\,t\int_0^\infty \psi(\tau)\,d\tau\right) \tag{5.6}$$

in a fast-modulation case. Details of the line shape and the associated moments reflect the nature of modulation in a more delicate way. In order to see this, we choose here two typical examples for the decay $\psi(\tau)$ of correlation of the frequency modulation; namely,

$$\text{(a)}\quad \psi(\tau) = \psi(0)\,\exp(-\,\alpha\,|\,\tau\,|), \tag{5.7a}$$

and

$$\text{(b)}\quad \psi(\tau) = \psi(0)\,\exp(-\,\alpha^2\tau^2). \tag{5.7b}$$

The frequency and the time can be scaled by the width of the unnarrowed Gaussian line, which is thus given by

$$X'(\omega) = (2\pi)^{-1/2}\,\exp(-\,\omega^2/2). \tag{5.8}$$

Correspondingly, we assume in Eqs. (5.7) that

$$\psi(0) = 1. \tag{5.9}$$

For convenience we introduce here the modified moments λ'_n, which are real and are defined by

$$\lambda'_n = (-\,i)^n\,\lambda_n. \tag{5.10}$$

Explicitly they are expressed in terms of the response function or the admittance function as follows:

$$\begin{aligned}
&\lambda'_0 = \lambda_0 = \phi(0) = \frac{2}{\pi}\int_0^\infty X'(\omega)\,d\omega = 1,\\
&\lambda'_1 = -\,i\lambda_1 = -\,\dot\phi(0) = -\,\frac{2}{\pi}\int_0^\infty (\omega X'' - \lambda'_0)\,d\omega,\\
&\lambda'_2 = -\,\lambda_2 = \ddot\phi(0) = -\,\frac{2}{\pi}\int_0^\infty (\omega^2 X' - \lambda'_1)\,d\omega,\\
&\lambda'_3 = i\lambda_3 = -\,\dddot\phi(0) = \frac{2}{\pi}\int_0^\infty (\omega^3 X'' - \omega^2\lambda'_0 + \lambda'_2)\,d\omega,\\
&\lambda'_4 = \lambda_4 = \phi^{(4)}(0) = \frac{2}{\pi}\int_0^\infty (\omega^4 X' - \omega^2\lambda'_1 + \lambda'_3)\,d\omega, \text{ etc.},
\end{aligned} \tag{5.11}$$

and

$$\begin{aligned}
&\lambda'_{-1} = i\lambda_{-1} = \int_0^\infty \phi(t)\,dt = \frac{2}{\pi}\int_0^\infty \frac{X''(\omega)}{\omega}\,d\omega,\\
&\lambda'_{-2} = -\,\lambda_{-2} = \int_0^\infty t\phi(t)\,dt = -\,\frac{2}{\pi}\int_0^\infty \frac{1}{\omega^2}\,(X' - \lambda'_{-1})\,d\omega,\\
&\lambda'_{-3} = -\,i\lambda_{-3} = \frac{1}{2}\int_0^\infty t^2\phi(t)\,dt\\
&\qquad = -\,\frac{2}{\pi}\int_0^\infty \frac{1}{\omega^3}(X'' - \lambda'_{-2})\,d\omega,\\
&\lambda'_{-4} = \lambda_{-4} = \frac{1}{6}\int_0^\infty t^3\phi(t)\,dt\\
&\qquad = \frac{2}{\pi}\int_0^\infty \frac{1}{\omega^4}\,(X' - \lambda'_{-1} + \lambda'_{-3}\omega^2)\,d\omega, \quad \text{etc.}
\end{aligned} \tag{5.12}$$

Exact values of positive moments are found from expansions of the relaxation function $\phi(t)$ at $t = 0$; namely, for the example (5.7a),

$$\phi(t) = 1 - \tfrac{1}{2}\,t^2 + (1/3!)\,\alpha\,|\,t\,|^3 + (1/4!)(3 - \alpha^2)t^4 + \cdots \tag{5.13a}$$

and, for the example (5.7b),

$$\phi(t) = 1 - \tfrac{1}{2}t^2 + (1/4!)\,(3 + 2\alpha^2)\,t^4 + \cdots, \tag{5.13b}$$

which give

$$\begin{aligned}
&\lambda_0 = 1, \quad \lambda'_1 = 0, \quad \lambda'_2 = -\,1,\\
&\lambda'_3 = -\,\alpha \quad \text{for (a)}, \quad \lambda'_3 = 0 \quad \text{for (b)},\\
&\lambda'_4 = 3 - \alpha^2 \quad \text{for (a)}, \quad \lambda'_4 = 3 + 2\alpha^2 \quad \text{for (b)}.
\end{aligned} \tag{5.14}$$

The moments defined by Eqs. (5.11) and (5.12) are written as

$$\lambda'_n = \frac{2}{\pi}\int_0^\infty \omega^n\,F_n(\omega)\,d\omega.$$

By integrating to an upper bound M we define the *incomplete moments*

$$\lambda'_n(M) = \frac{2}{\pi}\int_0^M \omega^n\,F_n(\omega)\,d\omega.$$

In order to see how these converge to λ'_n, we computed $\lambda'_n(M)$ for $-\,4 \le n \le 4$ to $M = 10$ for several values of α ranging from the ideal Gaussian limit to a strongly narrowed limit. Numerical integration becomes more and more difficult for higher positive moments and for larger values of α because contributions from far wings more heavily predominate. The negative moments show better convergence even for larger values of α.

It is possible to correct the incomplete moments $\lambda'_n(M)$, Eq. (5.15), to better approximations. We define the error for λ_n by

$$\Delta\lambda_n(M) = \lambda_n - \lambda_n(M), \tag{5.15}$$

which can be written as

$$\Delta\lambda_n(M) = -\,\frac{2i}{\pi}\,\frac{\lambda_{n+1}}{M} + \frac{2}{\pi}\,(-\,i)^{n+2}\int_M^\infty \frac{X'_{n+2}(\omega)}{\omega^2}\,d\omega. \tag{5.16}$$

For an even positive moment, this is shown as follows:

$$\Delta\lambda_{2n} = \frac{2}{\pi}\int_M^\infty \omega^{2n}\left(X'(\omega) + \sum_{k=0}^{n-1}\frac{i\lambda_{2k+1}}{\omega^{2k+2}}\right)d\omega$$

$$= \frac{2}{\pi} \int_M^\infty \omega^{2n} \left(\frac{(-)^{n+1}}{\omega^{2n+2}} X'_{2n+2}(\omega) - \sum_{k=0}^{n} \frac{i\lambda_{2k+1}}{\omega^{2k+2}} + \sum_{k=0}^{n-1} \frac{i\lambda_{2k+1}}{\omega^{2k+2}} \right) d\omega$$

$$= -\frac{2}{\pi} \frac{i\lambda_{2n+1}}{M} + \frac{2}{\pi} \int_M^\infty \frac{(-)^{n+1}}{\omega^2} X'_{2n+2}(\omega)\, d\omega.$$

Derivation of Eq. (5.16) is done for other cases in much the same way. Therefore, the errors are bounded by

$$|\Delta\lambda_n(M)| \le \frac{2}{\pi} \frac{1}{M} \{ |\lambda_{n+1}| + \max_{|\omega|>M} |X'_{n+2}(\omega)| \}. \qquad (5.17)$$

The second term in the bracket on the right-hand side of the above inequality may be ignored when M is sufficiently large, so that the first term on the right-hand side of Eq. (5.16) may be used as a correction to the incomplete moment. Thus, for a modified incomplete moment of the nth order we may take

$$\lambda''_n(M) = \lambda'_n(M) + \delta\lambda'_n(M) \qquad (5.18)$$

as an improved approximation to λ'_n, where

$$\delta\lambda'_n(M) = (2/\pi)\lambda'_{n+1}/M. \qquad (5.19)$$

In order to save space we discuss here only the results of computation for $\alpha = 0$ and 2, although computation has been made also for $\alpha = 0.5$, 1, and 5. The line shape is Gaussian for $\alpha = 0$ whereas it is considerably narrowed for $\alpha = 2$. Figures 1 and 2 show the incomplete moments $\lambda'_n(M)$ and the corrected moments $\lambda''_n(M)$ as functions of M. Positive even moments λ_0, λ'_2, and λ'_4 converge fast. Corrections $\delta\lambda'_n$ are zero to these moments. Positive odd moments, on the other hand, show much slower convergence. However, the corrected moments λ''_1, converge much faster to the exact limit values which are zero for both. The third moment $\lambda''_3(M)$ attains a minimum at $M \sim 6$ and then starts to deviate again. This is due to errors in numerical integration which uses finite meshes in integrating over time. As shown in Eq. (5.11) the third moment is defined with respect to the function

$$X'_3(\omega) = -\omega^3 X''(\omega) + \omega^2 + 1.$$

The asymptotic expansion (3.3a) is so good that the terms on the right-hand side cancel out almost completely. Thus the computed values of $X'_3(\omega)$ after this cancellation are hardly above the error bound of computation. When ω becomes larger, say, than six, then the values of λ''_3 beyond there are subject to computational errors. This situation is somewhat similar to what is encountered in obtaining spectroscopic data from experiments.

Negative moments are found to be nicely convergent. As is shown in Fig. 2, the incomplete moments increase with the upper bound of integration monotonously. When corrected, the convergence becomes surprisingly improved. The corrected moments $\lambda''_{-n}(M)$ attain their limit values at the values of M as small as three or four. A nice feature about this is that the correction λ'_{-n} is determined by λ'_{-n+1}, so that the correction can be made sucessively.

The examples given by (5.7a) and (5.7b) with a large α are instructive as models of non-Gaussian spectra. The spectral line shape is close to Lorentzian with a strong peak at the center and with long tails which are above the unnarrowed Gaussian tails at far wings. These extended tails make the convergence of positive moments poorer than that in the Gaussian case. Furthermore, the positive moments are determined by the behavior of $\phi(t)$ at $t = 0$. This means that they delicately depend on the shape of the spectrum at far wings. A subtle difference almost indistinguishable in the spectral shape reveals itself in the values of moments as should be expected by Eqs. (5.13) which give different moments for different models (a) and (b).

Figure 3 shows the positive moments for the model (a). The zeroth moment converges rather fast, but the second moment converges rather slowly. The convergence of the first moment $\lambda'_1(M)$ is greatly improved if it is corrected to $\lambda''_1(M)$ using the exact value of the second moment λ'_2. The third and fourth moments $\lambda'_3(M)$ and $\lambda'_4(M)$ converge only slowly. The convergence is not much improved by correcting

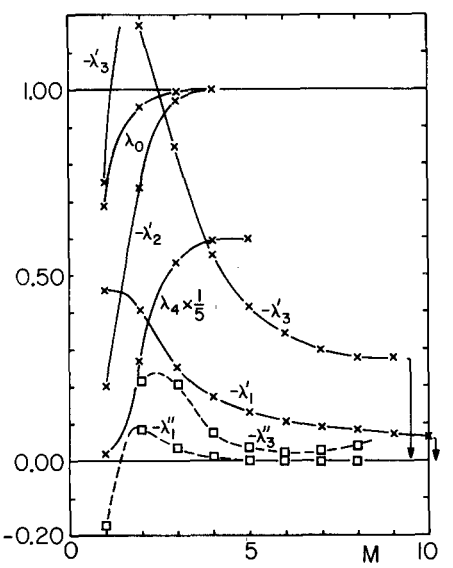


FIG. 1. Incomplete positive moments and corresponding corrected moments for a Gaussian spectrum ($\alpha = 0$).

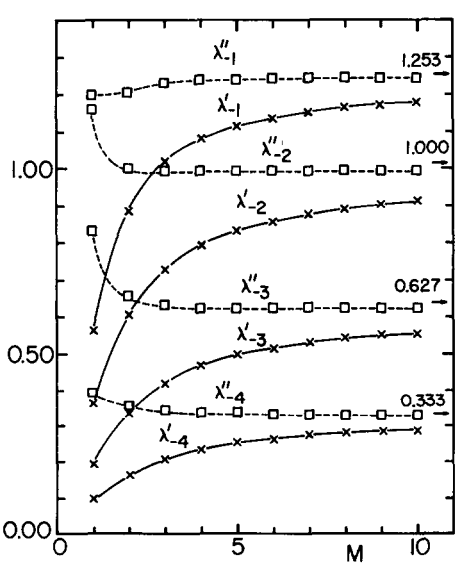


FIG. 2. Incomplete negative moments and corresponding corrected moments for a Gaussian spectrum ($\alpha = 0$). The numbers on the right are exact limits.

them to $\lambda''_3(M)$ and $\lambda''_4(M)$, respectively, using the exact values of λ'_4 and λ'_5. The numbers given on the right in this figure indicate the exact limit values to which each curve of moments should approach as M tends to infinity. The poor convergence of these higher moments is due to the contributions from extended tails of spectra. The accuracy of computation becomes poor at far wings, which makes the convergence further worse and untrustable. Figure 4 shows the corresponding positive moments for the model (b). The zeroth, first and second moments behave almost in the same way as in case (a). The third moment, however, must vanish, and the fourth moment is very large, as is indicated by Eqs. (5.13). Thus the third incomplete moment $\lambda'_3(M)$ and the corrected moment $\lambda''_3(M)$ behave quite differently from the corresponding moments in Fig. 3.

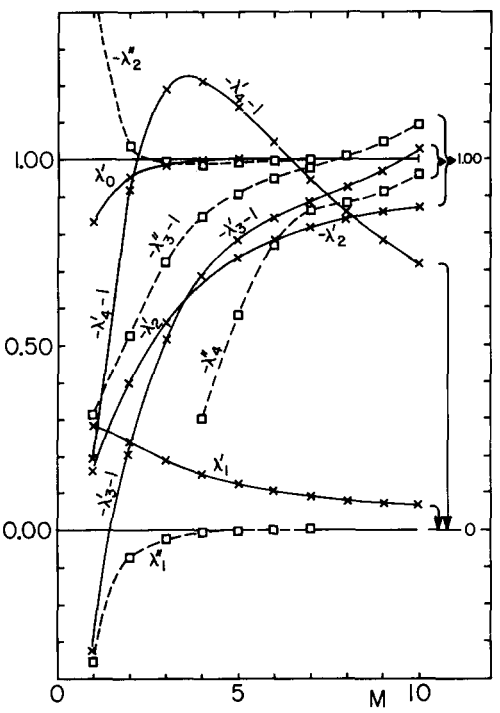


FIG. 3. Incomplete positive moments and corresponding corrected moments for model (a) with $\alpha = 2$. The exact limit of the third and the fourth moments λ'_3 and λ'_4 are -2 and -1, respectively, and that of the fifth moment λ'_5, which is used to obtain corrected moments, is 12.

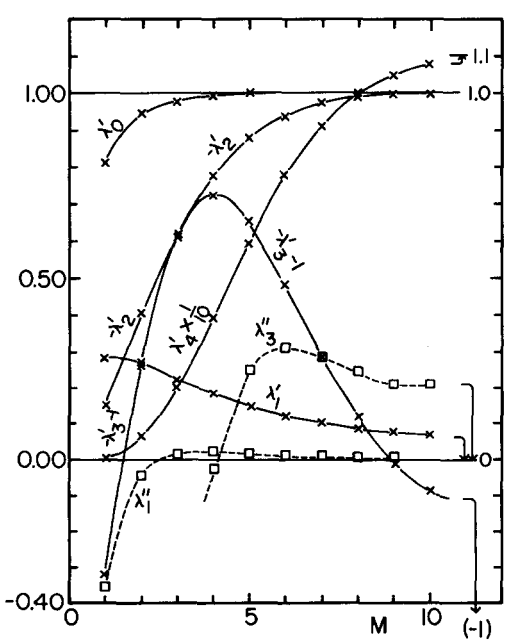


FIG. 4. Incomplete positive moments and corresponding corrected moments for model (b) with $\alpha = 2$. The exact limit of the third, the fourth, and the fifth moments λ'_3, λ'_4, and λ'_5 are 0, 11, and 0, respectively.

Contrary to positive moments, the behavior of negative moments for these models is qualitatively much the same as in the Gaussian case. The higher-order negative moments are much larger in magnitude because the relaxation decay of the function $\phi(t)$ is much slower. The convergence is always very good if the incomplete moments are corrected by Eq. (5.23). Figures 5 and 6 show this convergence. The numbers

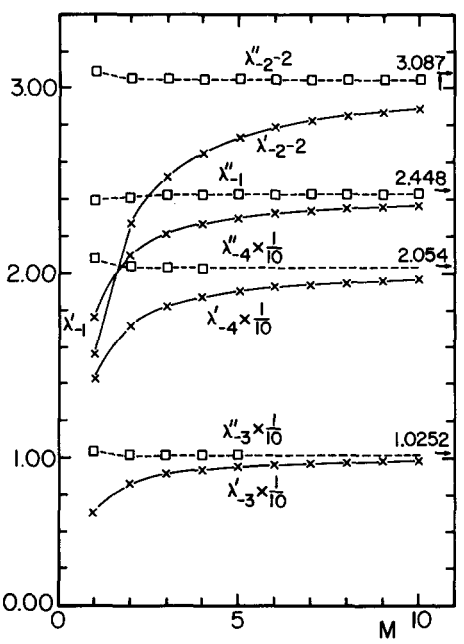


FIG. 5. Incomplete negative moments and corresponding corrected moments for model (a) with $\alpha = 2$. The exact values are $\lambda'_{-1} = 2.448$, $\lambda'_{-2} = 5.087$, $\lambda'_{-3} = 10.252$, $\lambda'_{-4} = 20.54$.

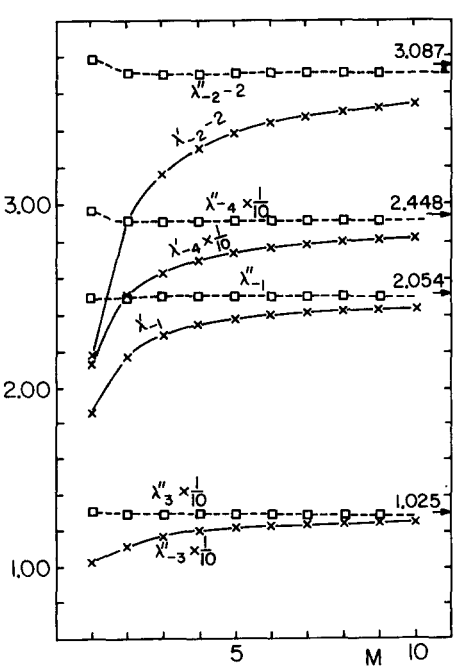


FIG. 6. Incomplete negative moments and corresponding corrected moments for model (b) with $\alpha = 2$. The exact values are $\lambda'_{-1} = 2.530$, $\lambda'_{-2} = 5.766$, $\lambda'_{-3} = 13.02$, $\lambda'_{-4} = 29.39$.

on the right are the limit values of each moment computed by Eq. (3.6).

The numerical analysis here made gives some insight to the problem one may encounter in practical evaluation of moments from spectra obtained by experiments. Higher positive moments are difficult to evaluate not only because the contribution at wings are dominated, but also because they delicately depend on the spectral shapes. This means also that they provide some valuable information if they could be evaluated. On the other hand, the negative moments will, in general, be more easily evaluated. They may be used as measures of the distribution of relaxation times.

6. CONCLUDING REMARKS

We have shown in this work that the definition of moments can be generalized so as to be applied to admittance functions of a rather wide class. The positive moments are defined with respect to the asymptotic expansion of an admittance function at large frequencies, whereas the negative moments are defined with respect to the power series expansion at small frequencies. The generalized moments are expressed in terms of moment integrals defined by the real or the imaginary part of the admittance function which should be suitably modified to secure the desired convergence. They are also given by the derivatives of the response function at $t = +0$ or by certain time integrals defined in terms of the response function. This fact is a very general statement of sum rules which contains most of known sum rules.

Application of this analysis may not be limited to admittance functions. For example, if one write a frequency dependent mobility function $\mu(\omega)$ as

$$\mu(\omega) = [i\omega + \gamma(\omega)]^{-1},$$

the frequency-dependent friction $\gamma(\omega)$ is also related to a function $\phi(t)$, which is in fact the correlation function of a random force.[7,8] In this sense it is possible to define moments for the function $\gamma(\omega)$ instead of doing it for $\mu(\omega)$. In the spirit of the continued fraction expansion introduced by Mori,[9] this process can even be carried on further. The moments introduced at one stage are related in some way to those defined at another stage. Their convergence properties may be different; but they must have some connections. There are a number of such questions, which have, however, not been dealt with in this work.

It should be also noted that the Kramers–Kronig relation can be applied to derivatives of an admittance function with respect to a parameter involved in it, and so the sum rules can also be applied to such derivatives. This is obvious, but may be a useful remark.

The numerical examples treated in Sec. 5 are only a sort of theoretical model, but they illustrate some features of the problems one may have in applying the analysis to a set of spectroscopic data.

ACKNOWLEDGMENTS

The greater part of this study was done when one of the authors (R. K) was visiting the Institute for Theoretical Physics, the State University of New York at Stony Brook. It is a great pleasure for him to express his deep appreciation to hospitality extended to him by Professor C. N. Yang.

---

* Work supported in part by the NSF Grant No. GP-32998X, the Scientific Research Fund of the Japanese Ministry of Education, and the U. S. Atomic Energy Commission No. AT(11-1)-3001.

1 H. A. Kramers, Atti. Congr. Fis. **11**, 22 (1927); R. Kronig, J. Opt. Soc. Amer. **12**, 547 (1926); see also B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U. P., London, 1959), p. 167 for a mathematical account.
2 R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957).
3 R. Kubo, in *Lectures in Theoretical Physics*, edited by W. Brittin (Interscience, New York, 1959), p. 120.
4 L. Kadanoff and P. C. Martin, Ann. Phys (N.Y.) **24**, 419 (1963).
5 R. Kubo, in *Fluctuation, Relaxation and Resonance in Magnetic Systems*, edited by D. ter Haar (Oliver and Boyd, Edinburgh, 1962), p. 23.
6 R. Kubo, J. Math. Phys. **4**, 174 (1962).
7 H. Mori, Progr. Theoret. Phys. **33**, 127 (1965).
8 R. Kubo, Rept. Progr. Phys. **29**, Part 1, 255 (1965).
9 H. Mori, Progr. Theoret. Phys. **34**, 399 (1965).

Steady-State Solutions in the Two-Group Theory of Neutron Diffusion

E. E. Burniston
Mathematics Department, North Carolina State University, Raleigh, North Carolina 27607

and

T. W. Mullikin
Mathematics Department, Purdue University, Lafayette, Indiana 47907

and

C. E. Siewert
Nuclear Engineering Department, North Carolina State University, Raleigh, North Carolina 27607

(Received 31 March 1972)

Functional analysis arguments are used to prove the existence of a unique solution to the integral form of the two-group neutron-transport equation for subcritical half-spaces. The analytic properties of the solutions are discussed and used to prove that the partial indices of canonical solutions of the matrix Riemann problem, basic to **H**-matrix or half-range completeness considerations, are nonnegative.

I. INTRODUCTION

There has been considerable interest[1–7] in recent years in the multigroup version of the neutron-transport equation, basically because a great deal of the fine structure of such energy-dependent processes as scattering, absorption, and fission can be maintained in this model without actually requiring solutions to the more general energy-dependent form of the trans-

on the right are the limit values of each moment computed by Eq. (3.6).

The numerical analysis here made gives some insight to the problem one may encounter in practical evaluation of moments from spectra obtained by experiments. Higher positive moments are difficult to evaluate not only because the contribution at wings are dominated, but also because they delicately depend on the spectral shapes. This means also that they provide some valuable information if they could be evaluated. On the other hand, the negative moments will, in general, be more easily evaluated. They may be used as measures of the distribution of relaxation times.

6. CONCLUDING REMARKS

We have shown in this work that the definition of moments can be generalized so as to be applied to admittance functions of a rather wide class. The positive moments are defined with respect to the asymptotic expansion of an admittance function at large frequencies, whereas the negative moments are defined with respect to the power series expansion at small frequencies. The generalized moments are expressed in terms of moment integrals defined by the real or the imaginary part of the admittance function which should be suitably modified to secure the desired convergence. They are also given by the derivatives of the response function at $t = +0$ or by certain time integrals defined in terms of the response function. This fact is a very general statement of sum rules which contains most of known sum rules.

Application of this analysis may not be limited to admittance functions. For example, if one write a frequency dependent mobility function $\mu(\omega)$ as

$$\mu(\omega) = [i\omega + \gamma(\omega)]^{-1},$$

the frequency-dependent friction $\gamma(\omega)$ is also related to a function $\phi(t)$, which is in fact the correlation function of a random force.[7,8] In this sense it is possible to define moments for the function $\gamma(\omega)$ instead of doing it for $\mu(\omega)$. In the spirit of the continued fraction expansion introduced by Mori,[9] this process can even be carried on further. The moments introduced at one stage are related in some way to those defined at another stage. Their convergence properties may be different; but they must have some connections. There are a number of such questions, which have, however, not been dealt with in this work.

It should be also noted that the Kramers–Kronig relation can be applied to derivatives of an admittance function with respect to a parameter involved in it, and so the sum rules can also be applied to such derivatives. This is obvious, but may be a useful remark.

The numerical examples treated in Sec. 5 are only a sort of theoretical model, but they illustrate some features of the problems one may have in applying the analysis to a set of spectroscopic data.

ACKNOWLEDGMENTS

The greater part of this study was done when one of the authors (R. K) was visiting the Institute for Theoretical Physics, the State University of New York at Stony Brook. It is a great pleasure for him to express his deep appreciation to hospitality extended to him by Professor C. N. Yang.

---

* Work supported in part by the NSF Grant No. GP-32998X, the Scientific Research Fund of the Japanese Ministry of Education, and the U. S. Atomic Energy Commission No. AT(11-1)-3001.

1 H. A. Kramers, Atti. Congr. Fis. **11**, 22 (1927); R. Kronig, J. Opt. Soc. Amer. **12**, 547 (1926); see also B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U. P., London, 1959), p. 167 for a mathematical account.
2 R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957).
3 R. Kubo, in *Lectures in Theoretical Physics*, edited by W. Brittin (Interscience, New York, 1959), p. 120.
4 L. Kadanoff and P. C. Martin, Ann. Phys (N.Y.) **24**, 419 (1963).
5 R. Kubo, in *Fluctuation, Relaxation and Resonance in Magnetic Systems*, edited by D. ter Haar (Oliver and Boyd, Edinburgh, 1962), p. 23.
6 R. Kubo, J. Math. Phys. **4**, 174 (1962).
7 H. Mori, Progr. Theoret. Phys. **33**, 127 (1965).
8 R. Kubo, Rept. Progr. Phys. **29**, Part 1, 255 (1965).
9 H. Mori, Progr. Theoret. Phys. **34**, 399 (1965).

Steady-State Solutions in the Two-Group Theory of Neutron Diffusion

E. E. Burniston
Mathematics Department, North Carolina State University, Raleigh, North Carolina 27607

and

T. W. Mullikin
Mathematics Department, Purdue University, Lafayette, Indiana 47907

and

C. E. Siewert
Nuclear Engineering Department, North Carolina State University, Raleigh, North Carolina 27607

(Received 31 March 1972)

Functional analysis arguments are used to prove the existence of a unique solution to the integral form of the two-group neutron-transport equation for subcritical half-spaces. The analytic properties of the solutions are discussed and used to prove that the partial indices of canonical solutions of the matrix Riemann problem, basic to **H**-matrix or half-range completeness considerations, are nonnegative.

I. INTRODUCTION

There has been considerable interest[1–7] in recent years in the multigroup version of the neutron-transport equation, basically because a great deal of the fine structure of such energy-dependent processes as scattering, absorption, and fission can be maintained in this model without actually requiring solutions to the more general energy-dependent form of the trans-

port equation. In fact, the multigroup model has proved adequate for so many reactor calculations that multigroup diffusion theory is perhaps the most widely used method in reactor design analysis. Since efficient multigroup diffusion codes often make use of transport theory, for example, to define improved boundary conditions, we develop here the fundamental analysis required to place the two-group model on a basis equally as firm as that provided by Case[8] for the one-speed theory.

It seems that even the very basic subcriticality conditions for infinite media have not been resolved definitively for the multigroup model, and hence in Sec. II, we first seek the general conditions required to ensure the existence of a unique solution to the half-space albedo problem, based on the two-group model. These same conditions can, naturally, be shown to be the infinite-media subcriticality conditions.

Our principal goal here is to point out that half-space problems in two-group transport theory can be reduced to a convenient computational form and to provide the appropriate existence and uniqueness theorems required to ensure that any computational results can be interpreted and used with confidence. We shall rely rather heavily on a previous paper,[7] hereafter referred to as SBK, in which the required analysis was given for all cases but one. In addition to resolving the one elusive case not included in SBK, we are confident that the functional analysis arguments developed in Sec. II, and used in Sec. III to establish definitively, in the manner of Goh'berg and Krein,[9] the very important proof that the partial indices of a canonical solution to a basic matrix Riemann problem are nonnegative, will prove very useful in the analysis of the more general models.

We consider the homogeneous, steady-state neutron-transport equation written in a convenient form as

$$\mu \frac{\partial}{\partial x}\Psi(x,\mu) + \Sigma\Psi(x,\mu) = \mathbf{C}\int_{-1}^{1}\Psi(x,\mu')d\mu', \tag{1}$$

where the two elements of $\Psi(x,\mu)$ are the angular fluxes in each of the two energy groups, and $\mathbf{C}$, with nonnegative elements, is the group-transfer matrix. By choosing to measure distances in terms of the optical variable x, defined in terms of σ_2 (the smaller of the two total cross sections σ_1 and σ_2), we can write

$$\Sigma = \begin{vmatrix} \sigma & 0 \\ 0 & 1 \end{vmatrix}, \qquad \sigma = \frac{\sigma_1}{\sigma_2} > 1. \tag{2}$$

Since the solutions to typical half-space problems based on Eq. (1) can be expressed in terms of the solution to the albedo problem, we seek a bounded (at infinity) *matrix* solution of Eq. (1) subject to the boundary condition

$$\Psi(\mu_1,\mu_2;0,\mu) = \begin{vmatrix} \delta(\mu-\mu_1) & 0 \\ 0 & \delta(\mu-\mu_2) \end{vmatrix}, \qquad \mu_1,\mu_2,\mu \in (0,1). \tag{3}$$

We can now enter

$$\Psi(\mu_1,\mu_2;x,\mu) = \hat{\Psi}(\mu_1,\mu_2;x,\mu) + \begin{vmatrix} \delta(\mu-\mu_1)e^{-\sigma x/\mu} & 0 \\ 0 & \delta(\mu-\mu_2)e^{-x/\mu} \end{vmatrix} \tag{4}$$

into Eq. (1) to obtain

$$\mu\frac{\partial}{\partial x}\hat{\Psi}(\mu_1,\mu_2;x,\mu) + \Sigma\hat{\Psi}(\mu_1,\mu_2;x,\mu) = \mathbf{C}\mathbf{F}(\mu_1,\mu_2;x), \tag{5}$$

where

$$\mathbf{F}(\mu_1,\mu_2;x) = \int_{-1}^{1}\hat{\Psi}(\mu_1,\mu_2;x,\mu)d\mu + \begin{vmatrix} e^{-\sigma x/\mu_1} & 0 \\ 0 & e^{-x/\mu_2} \end{vmatrix}; \tag{6}$$

from Eq. (3) we note that

$$\hat{\Psi}(\mu_1,\mu_2;0,\mu) = \mathbf{0}, \qquad \mu \in (0,1). \tag{7}$$

Equation (5) can be solved at once to yield

$$\hat{\Psi}(\mu_1,\mu_2;x,\mu) = \frac{1}{\mu}\int_0^x \begin{vmatrix} e^{-\sigma(x-x')/\mu} & 0 \\ 0 & e^{-(x-x')/\mu} \end{vmatrix} \times \mathbf{C}\mathbf{F}(\mu_1,\mu_2;x')dx', \quad \mu > 0, \tag{8a}$$

and

$$\hat{\Psi}(\mu_1,\mu_2;x,\mu) = -\frac{1}{\mu}\int_x^\infty \begin{vmatrix} e^{\sigma(x'-x)/\mu} & 0 \\ 0 & e^{(x'-x)/\mu} \end{vmatrix} \times \mathbf{C}\mathbf{F}(\mu_1,\mu_2;x')dx', \quad \mu < 0, \tag{8b}$$

which can be entered into Eq. (6) to establish the integral equation

$$\mathbf{F}(\mu_1,\mu_2;x) = \int_0^\infty \begin{vmatrix} E_1(\sigma|x-x'|) & 0 \\ 0 & E_1(|x-x'|) \end{vmatrix} \times \mathbf{C}\mathbf{F}(\mu_1,\mu_2;x')dx' + \mathbf{Q}(\mu_1,\mu_2;x), \tag{9}$$

where

$$\mathbf{Q}(\mu_1,\mu_2;x) = \begin{vmatrix} e^{-\sigma x/\mu_1} & 0 \\ 0 & e^{-x/\mu_2} \end{vmatrix}. \tag{10}$$

Here $E_1(x)$ is the standard exponential integral:

$$E_1(x) = \int_0^1 e^{-x/\nu}\frac{d\nu}{\nu}. \tag{11}$$

We now wish to argue that Eq. (9) admits a unique solution for all subcritical media. For the sake of notational convenience, we prefer to write Eq. (9) as

$$\mathbf{F}(\mu_1,\mu_2;x) = \mathbf{L}\mathbf{F}(\mu_1,\mu_2;x) + \mathbf{Q}(\mu_1,\mu_2;x), \tag{12}$$

where $\mathbf{L}$ denotes the integral operator.

II. EXISTENCE THEOREM

By investigating Eq. (12), the linear integral equation for $\mathbf{F}(\mu_1,\mu_2;x)$, we find in this section a condition sufficient to ensure the existence of a solution to the singular H-matrix equation discussed in SBK. To establish this condition, we consider

$$\mathbf{f} = \mathbf{L}\mathbf{f} + \mathbf{q} \tag{13}$$

in the function space $\mathfrak{L}_1$ of *vector* functions with norm

$$\|\mathbf{f}\| = \max_{i=1,2}\left\{\int_0^\infty |f_i(x)|dx\right\}, \tag{14}$$

where $f_i, i = 1$ and 2 denote the two elements of $\mathbf{f}$. Note that we take Eq. (13) to be a *vector* version of the *matrix* equation given by Eq. (12).

Theorem 1: If ρ denotes the dominant eigenvalue of the nonnegative matrix $\Sigma^{-1}\mathbf{C}$ and if $2\rho < 1$, then the equation

$$\mathbf{f} = \mathbf{L}\mathbf{f} + \mathbf{q}, \tag{15}$$

with $\mathbf{q}$ in $\mathcal{L}_1$, has a unique solution $\mathbf{f}$ in $\mathcal{L}_1$ given by the series

$$\mathbf{f} = \sum_{n=0}^{\infty} \mathbf{L}^n\mathbf{q}. \tag{16}$$

To prove the theorem, we note that the series given by Eq. (16) converges in $\mathcal{L}_1$ provided the spectral radius $\|\mathbf{L}\|_{sp}$ of $\mathbf{L}$, which can be computed from[10]

$$\|\mathbf{L}\|_{sp} = \lim_{n\to\infty} \|\mathbf{L}^n\|^{1/n}, \tag{17}$$

satisfies $\|\mathbf{L}\|_{sp} < 1$. Since the kernels in the matrix integral operator $\mathbf{L}$ are nonnegative, it follows that

$$|\mathbf{L}^n\mathbf{q}| \leq \mathbf{L}^n|\mathbf{q}|, \tag{18}$$

where by $|\mathbf{q}|$ we mean the vector

$$|\mathbf{q}| = \begin{vmatrix} |q_1| \\ |q_2| \end{vmatrix}. \tag{19}$$

It can be shown that

$$\int_0^\infty \mathbf{L}|\mathbf{q}|(x)dx \leq 2\Sigma^{-1}\mathbf{C}\int_0^\infty |\mathbf{q}(x)|dx, \tag{20}$$

and thus it follows that

$$\int_0^\infty \mathbf{L}^n|\mathbf{q}|(x)dx \leq (2\Sigma^{-1}\mathbf{C})^n\int_0^\infty |\mathbf{q}(x)|dx \tag{21}$$

and, hence, that

$$\|\mathbf{L}^n\mathbf{q}\| \leq 2^n\|\mathbf{q}\|[\text{max row sum of } (\Sigma^{-1}\mathbf{C})^n]. \tag{22}$$

For a nonnegative $m \times m$ matrix, the maximum of the row sums is the ∞-norm induced on the matrix when multiplying vectors $\mathbf{v}$ with norm

$$\|\mathbf{v}\|_\infty = \max_{i=1,2,\ldots,m} \{|v_i|\}, \tag{23}$$

and thus we can write

$$\|\mathbf{L}\|_{sp} \leq 2\lim_{n\to\infty}\|(\Sigma^{-1}\mathbf{C})^n\|_\infty^{1/n} = 2\|\Sigma^{-1}\mathbf{C}\|_{sp}. \tag{24}$$

Since the spectral radius of a finite-dimensional matrix is the maximum of the absolute values of the eigenvalues of the matrix and since $\Sigma^{-1}\mathbf{C}$ is nonnegative, it follows[11] that the spectral radius coincides with the dominant root ρ. The condition $2\rho < 1$, thus, guarantees that

$$\|\mathbf{L}\|_{sp} \leq 2\rho < 1, \tag{25}$$

which proves that the series given by Eq. (16) converges in $\mathcal{L}_1$. The proof of Theorem 1 is therefore established.

It is immediately apparent that our Theorem 1 is valid for the N-group version of Eq. (12). For the two-group case, we find that the condition $2\rho < 1$ can be written explicitly in the form

$$c_{11} + \sigma c_{22} + [(c_{11} + \sigma c_{22})^2 - 4\sigma C]^{1/2} < \sigma, \tag{26}$$

which is equivalent to the two conditions

$$1 - (1/\sigma)c_{11} - c_{22} > 0 \tag{27a}$$

and

$$1 - (2/\sigma)c_{11} - 2c_{22} + (4/\sigma)C > 0. \tag{27b}$$

Here the elements of $\mathbf{C}$ are denoted by c_{ij} and $C = \det\mathbf{C}$. Henceforth, we shall assume that the inequalities (27) are satisfied.

We now wish to show that $\mathbf{F}(u, v; x)$, the *matrix* solution of Eq. (12) corresponding to the inhomogeneous term $\mathbf{Q}(u, v; x)$, defines a function which satisfies the singular $\mathbf{H}$-matrix equation discussed in SBK. We thus consider

$$\mathbf{F}(u, v; x) = \mathbf{L}\mathbf{F}(u, v; x) + \mathbf{Q}(u, v; x), \tag{28}$$

with

$$\mathbf{Q}(u, v; x) = \begin{vmatrix} e^{-\sigma x/u} & 0 \\ 0 & e^{-x/v} \end{vmatrix}, \quad \mathrm{Re}u > 0, \ \mathrm{Re}v > 0. \tag{29}$$

In regard to Eq. (28), we now wish to establish

Theorem 2: For each value of $x \in [0, \infty)$, the function $\mathbf{F}(u, v; x)$ is analytic in u and v for $\mathrm{Re}u > 0$ and $\mathrm{Re}v > 0$.

To prove the theorem, we first note that we can write, subject to the conditions given by inequalities (27),

$$\mathbf{F}(u, v; x) = \sum_{n=0}^{\infty} \mathbf{F}_n(u, v; x), \tag{30}$$

where for fixed x, the function

$$\mathbf{F}_n(u, v; x) = \mathbf{L}^n\mathbf{Q}(u, v; x) \tag{31}$$

can clearly be seen to be analytic in u and v, for $\mathrm{Re}u > 0$ and $\mathrm{Re}v > 0$, since the first and second columns of $\mathbf{F}_n(u, v; x)$ are independent of v and u, respectively. Now if $|\mathbf{F}|$ denotes the matrix formed by replacing the elements F_{ij} of $\mathbf{F}$ by $|F_{ij}|$, then with the obvious interpretation of inequality between matrices, we can write

$$|\mathbf{F}_n| \leq \mathbf{L}^n|\mathbf{Q}| \leq \mathbf{L}^n\mathbf{I}, \quad \mathrm{Re}u > 0 \text{ and } \mathrm{Re}v > 0, \tag{32}$$

since $|\mathbf{Q}| \leq \mathbf{I}$. It now follows from the definition of $\mathbf{L}$ that

$$|\mathbf{F}_n(u, v; x)| \leq (2\Sigma^{-1}\mathbf{C})^n, \quad x \in [0, \infty) \tag{33}$$

for all $\mathrm{Re}u > 0$ and $\mathrm{Re}v > 0$.

Each entry in the matrix $\mathbf{F}(u, v; x)$ is analytic in u and v for $\mathrm{Re}u > 0$ and $\mathrm{Re}v > 0$, since it is the uniform limit of analytic functions, which follows by use of the norm

$$\|\mathbf{F}\| = \max_{i,j} \{|F_{ij}|\} \tag{34}$$

and the estimate

$$\left\|\mathbf{F}(u, v; x) - \sum_{n=0}^{N} \mathbf{F}_n(u, v; x)\right\| \leq \sum_{\alpha=N+1}^{\infty} \|(2\Sigma^{-1}\mathbf{C})^\alpha\|. \tag{35}$$

The right-hand side of Eq. (35) is independent of u and v and goes to zero as $N \to \infty$, since inequalities (27) require that

$$\lim_{n\to\infty} \|(2\Sigma^{-1}\mathbf{C})^n\|^{1/n} < 1. \tag{36}$$

Having shown that $\mathbf{F}(u, v; x)$ is analytic in u and v for $\mathrm{Re}u > 0$ and $\mathrm{Re}v > 0$, we now note

Theorem 3: For $\mathrm{Re}u > 0$ ($u \notin [0, 1]$) and $\mathrm{Re}v > 0$ ($v \notin [0, 1]$), the first and second columns of $\mathbf{F}(u, v; x)$ can be represented by

$$\mathbf{F}_1(u; x) = \mathbf{Q}_1(u; x) - \mathbf{F}\left(u, \frac{u}{\sigma}; x\right) u\int_{-1}^{1} \boldsymbol{\Theta}(\mu)\mathbf{C}\frac{d\mu}{\sigma\mu - u} \times \begin{vmatrix} 1 \\ 0 \end{vmatrix} + u\int_0^1 \mathbf{F}(\sigma\mu, \mu; x)\boldsymbol{\Theta}(\mu)\mathbf{C}\frac{d\mu}{\sigma\mu - u} \begin{vmatrix} 1 \\ 0 \end{vmatrix} \tag{37a}$$

and

$$\mathbf{F}_2(v; x) = \mathbf{Q}_2(v; x) - \mathbf{F}(\sigma v, v; x)v\int_{-1}^{1} \boldsymbol{\Theta}(\mu)\mathbf{C}\frac{d\mu}{\mu - v} \begin{vmatrix} 0 \\ 1 \end{vmatrix} + v\int_0^1 \mathbf{F}(\sigma\mu, \mu; x)\boldsymbol{\Theta}(\mu)\mathbf{C}\frac{d\mu}{\mu - v} \begin{vmatrix} 0 \\ 1 \end{vmatrix}. \tag{37b}$$

Here

$$\boldsymbol{\Theta}(\mu) = \begin{vmatrix} \theta(\mu) & 0 \\ 0 & 1 \end{vmatrix}, \tag{38}$$

with $\theta(\mu) = 1$, $\mu \in (-1/\sigma, 1/\sigma)$, $\theta(\mu) = 0$, $\mu \notin (-1/\sigma, 1/\sigma)$,

$$\mathbf{Q}_1(u, x) = e^{-\sigma x/u} \begin{vmatrix} 1 \\ 0 \end{vmatrix} \quad \text{and} \quad \mathbf{Q}_2(v; x) = e^{-x/v} \begin{vmatrix} 0 \\ 1 \end{vmatrix}. \tag{39}$$

Note that the variable x enters Eqs. (37) only as a parameter.

To prove Theorem 3, we first operate on Eq. (28) to obtain

$$\mathbf{LF} = \mathbf{L}(\mathbf{LF}) + \mathbf{LQ}. \tag{40}$$

Some elementary analysis can now be used to deduce that

$$\mathbf{LQ}(u, v; x) = u\left[-\mathbf{Q}\left(u, \frac{u}{\sigma}; x\right)\int_{-1}^{1}\psi(\mu)\frac{d\mu}{\sigma\mu - u} + \int_0^1 \mathbf{Q}(\sigma\mu, \mu; x)\psi(\mu)\frac{d\mu}{\sigma\mu - u}\right]\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} + v\left(-\mathbf{Q}(\sigma v, v; x)\int_{-1}^{1}\psi(\mu)\frac{d\mu}{\mu - v} + \int_0^1 \mathbf{Q}(\sigma\mu, \mu; x)\,\psi(\mu)\,\frac{d\mu}{\mu - v}\right)\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}, \tag{41}$$

where

$$\boldsymbol{\psi}(\mu) = \theta(\mu)\mathbf{C}. \tag{42}$$

Now since

$$\mathbf{LF} = (\mathbf{I} - \mathbf{L})^{-1}\mathbf{LQ} \quad \text{and} \quad \mathbf{F} = (\mathbf{I} - \mathbf{L})^{-1}\mathbf{Q}, \tag{43}$$

we conclude that $\mathbf{LF}$ can be expressed in terms of $\mathbf{F}$ and subsequently used in $\mathbf{F} = \mathbf{LF} + \mathbf{Q}$ to give Eqs. (37).

If we now define

$$\mathbf{F}(\sigma z, z; 0) = \tilde{\mathbf{H}}(z), \qquad \mathrm{Re}z > 0, \tag{44}$$

then clearly $\mathbf{H}(z)$ will be analytic in the half plane $\mathrm{Re}z > 0$, and from Eqs. (37) it follows that

$$\tilde{\mathbf{H}}(z)\boldsymbol{\Lambda}(z) = \mathbf{I} + z\int_0^1 \tilde{\mathbf{H}}(\mu)\boldsymbol{\psi}(\mu)\frac{d\mu}{\mu - z}, \qquad \mathrm{Re}z > 0, \tag{45}$$

where

$$\boldsymbol{\Lambda}(z) = \mathbf{I} + z\int_{-1}^{1}\boldsymbol{\psi}(\mu)\frac{d\mu}{\mu - z}. \tag{46}$$

Equation (45) clearly relates $\mathbf{H}(z)$ in the right-half plane to $\mathbf{H}(\mu)$, $\mu \in [0, 1]$. Since, by Theorem 1, the existence of a unique $\mathbf{F}(\sigma z, z; x)$ has been established, and since Theorem 2 ensures that $\mathbf{F}(\sigma z, z; x)$ is analytic for $\mathrm{Re}z > 0$, it follows from Eq. (44) that there is at least one $\mathbf{H}$ matrix, analytic for $\mathrm{Re}z > 0$, which satisfies Eq. (45). It is also clear that there is at least one $\mathbf{H}$ matrix, analytic for $\mathrm{Re}z > 0$, which satisfies the N-group version of Eq. (45) when the conditions of Theorem 1 are satisfied.

III. THE RIEMANN PROBLEM AND PARTIAL INDICES

Equation (45) as established in the previous section can now be used to derive the system of singular integral equations discussed in SBK. Since $\mathbf{H}(z)$ is analytic for $\mathrm{Re}z > 0$, we deduce, upon invoking the Plemelj formulas[12] and Eq. (45), that

$$\tilde{\mathbf{H}}(\mu)\boldsymbol{\Lambda}^{\pm}(\mu) = \mathbf{I} + \mu P\int_0^1 \tilde{\mathbf{H}}(\nu)\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - \mu} \pm \pi i\mu\tilde{\mathbf{H}}(\mu)\boldsymbol{\psi}(\mu), \qquad \mu \in (0, 1), \tag{47}$$

where the $+(-)$ superscript denotes the limiting value as $z \to \mu$ in the upper (lower) half-plane. We can now average the two Eqs. (47) to find

$$\tilde{\mathbf{H}}(\mu)\boldsymbol{\lambda}(\mu) = \mathbf{I} + \mu P\int_0^1 \tilde{\mathbf{H}}(\nu)\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - \mu}, \qquad \mu \in (0, 1), \tag{48}$$

where

$$\boldsymbol{\lambda}(\mu) = \mathbf{I} + \mu P\int_{-1}^{1}\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - \mu}. \tag{49}$$

Since the integral term in Eq. (48) does not involve all of the elements of $\tilde{\mathbf{H}}(\mu)$ on the interval $0 < \mu < 1$, we prefer to replace Eq. (48) by the equivalent system considered in SBK:

$$\tilde{\mathbf{H}}(\mu)\boldsymbol{\lambda}(\mu) = \mathbf{I} + \mu P\int_0^1 \tilde{\mathbf{H}}(\nu)\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - \mu}, \qquad \mu \in \left(0, \frac{1}{\sigma}\right), \tag{50}$$

and

$$\tilde{\mathbf{H}}(\mu)\boldsymbol{\lambda}(\mu)\mathbf{M}^{(2)}(\mu) = \left(\mathbf{I} + \mu P\int_0^1 \tilde{\mathbf{H}}(\nu)\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - \mu}\right)\mathbf{M}^{(2)}(\mu), \qquad \mu \in \left(\frac{1}{\sigma}, 1\right), \tag{51}$$

where

$$\mathbf{M}^{(2)}(\mu) \propto \begin{vmatrix} -\Lambda_{12}(\mu) \\ \Lambda_{11}(\mu) \end{vmatrix}. \tag{52}$$

If, as in SBK, we now introduce the sectionally analytic matrix

$$\mathbf{N}(z) = \frac{1}{2\pi i}\int_0^1 \tilde{\mathbf{H}}(\nu)\boldsymbol{\psi}(\nu)\frac{d\nu}{\nu - z}, \tag{53}$$

then the singular integral equations given by Eqs. (50) and (51) can be reduced to the equivalent inhomogeneous Riemann problem

$$\tilde{\mathbf{N}}^{+}(\mu) = \mathbf{G}(\mu)\tilde{\mathbf{N}}^{-}(\mu) + \tilde{\boldsymbol{\psi}}(\mu)[\tilde{\boldsymbol{\Lambda}}^{-}(\mu)]^{-1}, \qquad \mu \in (0, 1), \tag{54}$$

where

$$\mathbf{G}(\mu) = \tilde{\boldsymbol{\Lambda}}^{+}(\mu)[\tilde{\boldsymbol{\Lambda}}^{-}(\mu)]^{-1}. \tag{55}$$

Except for proof that the partial indices of a canonical

solution of the Riemann problem with the homogeneous boundary condition

$$\mathbf{\Phi}^{+}(\mu) = \mathbf{G}(\mu)\mathbf{\Phi}^{-}(\mu), \qquad \mu \in (0, 1), \tag{56}$$

are nonnegative when both of the conditions $\det\mathbf{C} < 0$ *and* $\det\mathbf{\Lambda}(\infty) > 0$ apply, the analysis reported in SBK establishes the existence of a unique solution to the system of equations given by Eqs. (50) and (51) and the linear constraint

$$\left(\mathbf{I} + \nu_i \int_0^1 \tilde{\mathbf{H}}(\nu)\psi(\nu)\frac{d\nu}{\nu - \nu_i}\right)\mathbf{M}(\nu_i) = \mathbf{0}, \qquad i = 1, 2, \ldots, \kappa, \tag{57}$$

where $\nu_i, i = 1, 2, \ldots, \kappa$, are the zeros, with positive real (imaginary) part, of

$$\Lambda(z) = \det\mathbf{\Lambda}(z), \tag{58}$$

and

$$\mathbf{\Lambda}(\nu_i)\mathbf{M}(\nu_i) = \mathbf{0}, \qquad i = 1, 2, \ldots, \kappa. \tag{59}$$

In a similar vein, the half-range completeness theorem basic to the elementary solutions of Eq. (1) has been proved in SBK, except for the one elusive case $\det\mathbf{C} < 0$ *and* $\Lambda(\infty) > 0$. That completeness theorem also follows at once if a proof that the partial indices of $\mathbf{\Phi}(z)$ are nonnegative can be established.

In the manner of Goh'berg and Krein,[9] we now can show at once that the results of Sec. II guarantee that the partial indices of $\mathbf{\Phi}(z)$ are nonnegative for all choices of the basic parameters which satisfy inequalities (27). Note[2] that the conditions given by inequalities (27) ensure that all ν_i must be real, which includes the case $\det\mathbf{C} < 0$ *and* $\det\mathbf{\Lambda}(\infty) > 0$.

The general solution (of finite degree at infinity) to the Riemann problem defined by Eq. (54) can be written as

$$\tilde{\mathbf{N}}(z) = \frac{1}{2\pi i}\mathbf{\Phi}(z)\left(\int_0^1 \mathbf{K}(\nu)\frac{d\nu}{\nu - z} + \mathbf{P}(z)\right), \tag{60}$$

where

$$\mathbf{K}(\nu) = [\mathbf{\Phi}^{+}(\nu)]^{-1}\tilde{\psi}(\nu)[\tilde{\mathbf{\Lambda}}^{-}(\nu)]^{-1}, \tag{61}$$

$\mathbf{P}(z)$ is a matrix of polynomials, and $\mathbf{\Phi}(z)$ is a canonical solution of the Riemann problem defined by Eq. (56). Without loss of generality, we consider $\mathbf{\Phi}(z)$ to be of ordered normal form at infinity so that

$$\lim_{|z|\to\infty} \mathbf{\Phi}(z)\begin{vmatrix} z^{\kappa_1} & 0 \\ 0 & z^{\kappa_2} \end{vmatrix} = \mathbf{A}, \tag{62}$$

where $\mathbf{A}$ is a constant nonsingular matrix, and κ_1 and $\kappa_2 \geq \kappa_1$ are the partial indices.

We note from Eq. (53) that $z\mathbf{N}(z)$ must be bounded as $|z|$ tends to infinity, and thus the proposition that $\kappa_1 \leq -1$ yields, from Eq. (60), the requirement that

$$\begin{vmatrix} 1 \\ 0 \end{vmatrix}^{T} \int_0^1 \mathbf{K}(\nu)d\nu = \mathbf{0}. \tag{63}$$

Cauchy's integral theorem can now be used to represent $\mathbf{\Phi}^{-1}(z)$, which subsequently can be evaluated at the origin to yield

$$\begin{vmatrix} 1 \\ 0 \end{vmatrix}^{T} \int_0^1 \mathbf{K}(\nu)d\nu = -\begin{vmatrix} 1 \\ 0 \end{vmatrix}^{T} \mathbf{\Phi}^{-1}(0). \tag{64}$$

It thus follows that Eqs. (63) and (64) imply that $\mathbf{\Phi}(0)$ is singular, which, of course, contradicts the notion of $\mathbf{\Phi}(z)$ being a canonical solution. We conclude, therefore, that if Eqs. (50) and (51) admit a solution, then the partial indices of $\mathbf{\Phi}(z)$, the canonical solution of the Riemann problem defined by Eq. (56), cannot be negative; since the analysis of Sec. II, when inequalities (27) are satisfied, does establish the existence of a solution to Eqs. (50) and (51), we conclude that the partial indices must be nonnegative, when inequalities (27) are satisfied. Again, we note that inequalities (27) include the one case not resolved definitely in SBK. It is also apparent that the crucial proof that the partial indices for the matrix Riemann problem required in the half-range completeness theorem for the N-group problem can be taken as established, for these cases when Theorem 1 applies.

Finally, we should like to mention that Pahor and Suhadolc[13] have established the existence of a unique solution to Eq. (15); their proof, however, is based on conditions more restrictive than those of our Theorem 1.

ACKNOWLEDGMENT

This work was supported in part by Joint Advisory Group Grant AFOSR-69-1779.

1 A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 181 (1966).
2 C. E. Siewert and P. S. Shieh, J. Nucl. Energy **21**, 383 (1967).
3 T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. **33**, 297 (1968).
4 S. Pahor and J. K. Shultis, J. Nucl. Energy **23**, 477 (1969).
5 D. R. Metcalf and P. F. Zweifel, Nucl. Sci. Eng. **33**, 318 (1968).
6 E. E. Burniston, C. E. Siewert, P. Silvennoinen, and P. F. Zweifel, Nucl. Sci. Eng. **45**, 331 (1971).
7 C. E. Siewert, E. E. Burniston, and J. T. Kriese, "Two-Group Neutron-Transport Theory: Existence and Uniqueness of the H-Matrix," J. Nucl. Energy (in press).
8 K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
9 I. C. Goh'berg and M. G. Krein, Usp. Mat. Nauk. (N.S.)**13**, 3 (1958) [Amer. Math. Soc. Transl. **14**, Ser. II. 217 (1960)].
10 F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1965).
11 S. R. Gantmacher, *Applications of the Theory of Matrices* (Interscience, New York, 1959).
12 N. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
13 S. Pahor and A. Suhadolc (submitted for publication).

A Perturbation Method for Two Synchronously Tuned, Coupled, Autonomous, Nonlinear Oscillators

William Streifer
Department of Electrical Engineering, University of Rochester, Rochester, New York 14627

and

K. Bradley Paxton
KAD Research Laboratory, Eastman Kodak Company, Rochester, New York 14650

(Received 11 May 1972)

A perturbation procedure is developed for two synchronously tuned, coupled, autonomous, nonlinear oscillators. The procedure results in ordinary nonlinear differential equations for the slowly varying amplitude envelopes and the slowly varying phase difference of the two oscillators. A method of obtaining initial values is included as are two examples for coupled van der Pol and linear oscillators.

INTRODUCTION

In this note we develop a perturbation method for studying two weakly coupled, synchronously tuned, autonomous, nonlinear oscillators. The nonlinearities are also assumed small. The method is quite similar to those developed by Krylov, Bogoliubov, and Mitropolsky[1] among others; but their methods are not applicable to the particular case in which the oscillators are synchronously tuned. The method results in differential equations for the slowly varying envelopes and phases of the oscillations.

Perturbation series for the initial values of the envelopes and phases are obtained, and the method is illustrated by application to coupled linear and van der Pol oscillators.

PERTURBATION SOLUTIONS

The coupled differential equations are

$$\ddot{x} + x = \epsilon f(x, \dot{x}, y, \dot{y}, \epsilon) \tag{1a}$$

and

$$\ddot{y} + y = \epsilon g(x, \dot{x}, y, \dot{y}, \epsilon), \tag{1b}$$

where f and g are assumed to be analytic functions of $x, \dot{x}, y, \dot{y}$, and ϵ, expressible in the form

$$f = \sum_{j=0}^{\infty} f_j(x, \dot{x}, y, \dot{y})\epsilon^j \tag{2a}$$

and

$$g = \sum_{j=0}^{\infty} g_j(x, \dot{x}, y, \dot{y})\epsilon^j, \tag{2b}$$

with f_0 and g_0 not identically zero. Equations (1a) and (1b) have unique solutions for specified initial values of $x, \dot{x}, y$, and $\dot{y}$.

We seek solutions in the form

$$x = a\cos\psi + \epsilon x_1(a, b, \psi, \phi) + \epsilon^2 x_2(a, b, \psi, \phi) + \cdots \tag{3a}$$

and

$$y = b\cos\phi + \epsilon y_1(a, b, \psi, \phi) + \epsilon^2 y_2(a, b, \psi, \phi) + \cdots, \tag{3b}$$

where a, b, ψ, and ϕ satisfy the differential equations

$$\dot{a} = \epsilon A_1(a, b, \theta) + \epsilon^2 A_2(a, b, \theta) + \cdots, \tag{3c}$$

$$\dot{b} = \epsilon B_1(a, b, \theta) + \epsilon^2 B_2(a, b, \theta) + \cdots, \tag{3d}$$

$$\dot{\psi} = 1 + \epsilon C_1(a, b, \theta) + \epsilon^2 C_2(a, b, \theta) + \cdots, \tag{3e}$$

$$\dot{\phi} = 1 + \epsilon D_1(a, b, \theta) + \epsilon^2 D_2(a, b, \theta) + \cdots, \tag{3f}$$

and

$$\theta = \psi - \phi. \tag{3g}$$

Thus

$$\dot{\theta} = \epsilon(C_1 - D_1) + \epsilon^2(C_2 - D_2) + \cdots, \tag{3h}$$

and so θ is a slowly varying function of t. Furthermore, x_j and y_j are all periodic functions of ψ and ϕ, but do not contain the fundamental, while A_j, B_j, C_j, and D_j are all periodic functions of θ. The functions x_j, y_j, A_j, B_j, C_j and D_j, are determined iteratively; first for $j = 1$, then $j = 2$, and so on by the method described below.

For $\epsilon = 0$ the solutions are

$$x^{(0)} = a\cos\psi \tag{4a}$$

and

$$y^{(0)} = b\cos\phi, \tag{4b}$$

where $\dot{a}^{(0)} = 0$, $\dot{b}^{(0)} = 0, \dot{\psi}^{(0)} = 1$, $\dot{\phi}^{(0)} = 1$, and the superscript has been added to denote the $\epsilon = 0$ solution. These results are now employed to determine x_1, A_1, and C_1 and later y_1, B_1, and D_1. The procedure for calculating x_1, A_1, and C_1 is to substitute $x^{(0)}$ and $y^{(0)}$ in the right side of (1a) and retain only first powers of ϵ,

$$\epsilon f(x, \dot{x}, y, \dot{y}, \epsilon) = \epsilon f_0(x^{(0)}, \dot{x}^{(0)}, y^{(0)}, \dot{y}^{(0)}), \tag{5}$$

where

$$\dot{x}^{(0)} = -a\sin\psi + O(\epsilon) \tag{6a}$$

and

$$\dot{y}^{(0)} = -b\sin\phi + O(\epsilon). \tag{6b}$$

The right side of Eq. (5) is expanded in an exponential Fourier series with known coefficients, viz.,

$$\epsilon f_0(x^{(0)}, \dot{x}^{(0)}, y^{(0)}, \dot{y}^{(0)}) = \epsilon \sum_{n,m} h_{nm}(a, b)e^{i(n\psi + m\phi)}. \tag{7}$$

It is convenient to assume that this series contains a finite number of terms, which is all one would retain in practice.

To evaluate the left side of (1a), we differentiate

$$x = a\cos\psi + \epsilon x_1(a, b, \psi, \phi) \tag{8}$$

and obtain

$$\dot{x} = -\dot{\psi}\, a\sin\psi + \dot{a}\cos\psi + \epsilon(\dot{a}x_{1a} + \dot{b}x_{1b} + \dot{\psi}x_{1\psi} + \dot{\phi}x_{1\phi}), \tag{9}$$

where the subscripts denote partial derivatives, e.g., $x_{1a} = \partial x_1/\partial a$. Upon substituting Eqs. (3c) to (3f) in Eq. (9), that equation becomes

$$\dot{x} = -a\sin\psi - \epsilon a C_1(a, b, \theta)\sin\psi + \epsilon A_1(a, b, \theta)\cos\psi + \epsilon x_{1\psi} + \epsilon x_{1\phi} + O(\epsilon^2). \tag{10}$$

The above calculation is repeated to obtain

$$\ddot{x} = -a\cos\psi - \epsilon 2aC_1\cos\psi - \epsilon 2A_1\sin\psi + \epsilon(x_{1\psi\psi} + 2x_{1\psi\phi} + x_{1\phi\phi}) + O(\epsilon^2), \quad (11)$$

where the fact that $\dot{a}, \dot{b}$, and $\dot{\theta}$ are all $O(\epsilon)$ have been employed. When this result and the expression for x are substituted in Eq. (1a) and first powers of ϵ are equated, we obtain a partial differential equation for x_1, which also contains the unknowns $A_1(a,b,\theta)$ and $C_1(a,b,\theta)$, viz.,

$$x_{1\psi\psi} + 2x_{1\psi\phi} + x_{1\phi\phi} - 2A_1\sin\psi - 2aC_1\cos\psi + x_1 = \sum_{n,m} h_{nm}e^{i(n\psi+m\phi)}. \quad (12)$$

To solve Eq. (12) for x_1, A_1, and C_1 we separate it into resonant and nonresonant parts. The resonant part contains the A_1 and C_1 terms which are equated to the Fourier series terms with $n + m = \pm 1$,

$$-2A_1(a,b,\theta)\sin\psi - 2aC_1(a,b,\theta)\cos\psi = \sum_{\substack{n,m \\ n+m=\pm1}} h_{nm}(a,b)e^{i(n\psi+m\phi)}; \quad (13)$$

whereas the nonresonant part contains the Fourier terms with $n + m \neq \pm 1$,

$$x_{1\psi\psi} + 2x_{1\psi\phi} + x_{1\phi\phi} + x_1 = \sum_{\substack{n,m \\ n+m\neq\pm1}} h_{nm}(a,b)e^{i(n\psi+m\phi)}. \quad (14)$$

The resonant equation (13) is solved by substituting $\phi = \psi - \theta$ in the exponent to obtain

$$e^{i(n\psi+m\phi)} = e^{i[(n+m)\psi - m\theta]} = e^{-im\theta}(\cos\psi \pm i\sin\psi).$$

Thus Eq. (13) becomes

$$-2A_1(a,b,\theta)\sin\psi - 2aC_1(a,b,\theta)\cos\psi = \sum_{\substack{n,m \\ n+m=\pm1}} h_{nm}(a,b)e^{-im\theta}(\cos\psi \pm i\sin\psi) \quad (15)$$

and so by equating $\sin\psi$ and $\cos\psi$ terms,

$$A_1 = -\frac{i}{2}\sum_{\substack{n,m \\ n+m=\pm1}} (n+m)h_{nm}(a,b)e^{-im\theta}, \quad (16a)$$

and

$$C_1 = -\frac{1}{2a}\sum_{\substack{n,m \\ n+m=\pm1}} h_{nm}(a,b)e^{-im\theta}. \quad (16b)$$

To solve the nonresonant equation (14), we write $x_1(a,b,\psi,\phi)$ as a Fourier series:

$$x_1(a,b,\psi,\phi) = \sum_{\substack{n,m \\ n+m=\pm1}} \xi^{(1)}_{nm}(a,b)e^{i(n\psi+m\phi)},$$

and substitute in Eq. (14), which thereupon becomes

$$\sum_{\substack{n,m \\ n+m\neq\pm1}} [1-(n+m)^2]\xi^{(1)}_{nm}(a,b)e^{i(n\psi+m\phi)} = \sum_{\substack{n,m \\ n+m\neq\pm1}} h_{nm}(a,b)e^{i(n\psi+m\phi)}.$$

Thus the coefficients are given by

$$\xi^{(1)}_{nm}(a,b) = \{h_{nm}(a,b)/[1-(n+m)^2]\}, \quad n+m \neq \pm 1. \quad (17)$$

Note that Eq. (17) is defined for all n and m, with $n + m \neq \pm 1$, and all other terms are included in the determination of A_1 and C_1. Resonant terms act to drive the fundamental amplitude and phase, whereas the nonresonant terms produce harmonics.

The above calculation must be repeated with Eq. (1b) to determine y_1, B_1, and D_1. In place of (13) and (14), we obtain

$$-2B_1\sin\phi - 2bD_1\cos\phi = \sum_{\substack{n,m \\ n+m=\pm1}} k_{nm}(a,b)e^{i(n\psi+m\phi)} \quad (18a)$$

and

$$y_{1\psi\psi} + 2y_{1\psi\phi} + y_{1\phi\phi} + y_1 = \sum_{\substack{n,m \\ n+m\neq\pm1}} k_{nm}(a,b)e^{i(n\psi+m\phi)}, \quad (18b)$$

where $k_{nm}(a,b)$ are the Fourier coefficients of $g_0(x^{(0)}, \dot{x}^{(0)}, y^{(0)}, \dot{y}^{(0)})$. Then after substituting $\psi = \phi + \theta$ in (18a) and performing the same manipulations as above, we obtain

$$B_1 = -\frac{i}{2}\sum_{\substack{n,m \\ n+m=\pm1}} (n+m)k_{nm}(a,b)e^{in\theta}, \quad (19a)$$

and

$$D_1 = -\frac{1}{2b}\sum_{\substack{n,m \\ n+m=\pm1}} k_{nm}(a,b)e^{in\theta}. \quad (19b)$$

Finally the series representation for y_1,

$$y_1(a,b,\psi,\phi) = \sum_{\substack{n,m \\ n+m\neq\pm1}} \eta^{(1)}_{nm}(a,b)e^{i(n\psi+m\phi)}, \quad (20)$$

has coefficients given by

$$\eta^{(1)}_{nm}(a,b) = \{k_{nm}(a,b)/[1-(n+m)^2]\}, \quad n+m \neq \pm 1. \quad (21)$$

The calculation of x_2, A_2, C_2 and y_2, B_2, D_2 is carried out in the same way as the first-order terms.

INITIAL VALUES

In many cases it is desirable to express the initial values $a^{(i)}, b^{(i)}, \psi^{(i)}$, and $\phi^{(i)}$ in terms of given initial values of $x_i, \dot{x}_i, y_i$, and $\dot{y}_i$. Such expressions can be derived without integrating the equations for $\dot{a}, \dot{b}, \dot{\psi}$, and $\dot{\phi}$, i.e., using A_j, B_j, C_j, and D_j. Indeed, this is essential since the equations for $\dot{a}$, etc., may only be integrable numerically, and thus the initial values must be known in advance. The procedure is not complicated albeit laborious; our discussion is limited to a brief outline.

The initial values x_i and $\dot{x}_i$ are given by Eqs. (3a) and (9) evaluated at $a^{(i)}, b^{(i)}, \psi^{(i)}$, and $\phi^{(i)}$; y_i and $\dot{y}_i$ are given by corresponding equations. Then upon employing Eqs. (3c)–(3f), combining the equations for x_i and $\dot{x}_i$, and rearranging terms, we obtain

$$a^{(i)}e^{i\psi^{(i)}} = x_i - i\dot{x}_i + \sum_{j=1}^{\infty}\epsilon^j W_j(a^{(i)}, b^{(i)}, \psi^{(i)}, \phi^{(i)}), \quad (22)$$

where W_j contains x_j and all terms in the equation for $\dot{x}_i$ multiplied by ϵ^j. Perturbation series are now assumed for $a^{(i)}$ and $\psi^{(i)}$,

$$a^{(i)} = \sum_{j=0}^{\infty}\epsilon^j a_j(x_i, \dot{x}_i, y_i, \dot{y}_i) \quad (23a)$$

and

$$\psi^{(i)} = \sum_{j=0}^{\infty} \epsilon^j \psi_j(x_i, \dot{x}_i, y_i, \dot{y}_i), \tag{23b}$$

and zero order terms ($\epsilon^0 = 1$) are collected in Eq. (22). The results are

$$a_0 = (x_i^2 + \dot{x}_i^2)^{1/2} \tag{24a}$$

and

$$\psi_0 = -\tan^{-1}(\dot{x}_i/x_i), \tag{24b}$$

and similar equations hold for b_0 and ϕ_0.

To obtain expressions for the higher-order terms a_j and ψ_j, we expand $a^{(i)}e^{i\psi^{(i)}}$ in a Taylor series about a_0, ψ_0, viz.,

$$\begin{aligned} a^{(i)}e^{i\psi^{(i)}} &= a_0 e^{i\psi_0} + (a^{(i)} - a_0)e^{i\psi_0} + ia^{(i)}e^{i\psi_0}(\psi^{(i)} - \psi_0) \\ &\quad + e^{i\psi_0}\sum_{l=2}^{\infty}\left(\frac{i^l}{l!}a_0(\psi^{(i)} - \psi_0)^l \right. \\ &\quad \left. + \frac{i^{l-1}}{(l-1)!}(\psi^{(i)} - \psi_0)^{l-1}(a^{(i)} - a_0)\right). \end{aligned} \tag{25}$$

The perturbation series (23a) and (23b) are substituted in Eq. (25), and the resultant expression is combined with Eq. (22). By equating all terms multiplying ϵ^j for each j, we find

$$a_j + ia_0\psi_j = G_j(a_0, \cdots a_{j-1}; b_0, \cdots b_{j-1}; \psi_0, \cdots \psi_{j-1}; \phi_0, \cdots \phi_{j-1}) \tag{26}$$

so that

$$a_j = \mathrm{Re}(G_j) \quad \text{and} \quad \psi_j = (1/a_0)\,\mathrm{Im}(G_j). \tag{27}$$

Thus expressions for $(a_1, b_1, \psi_1, \phi_1)$, $(a_2, b_2, \psi_2, \phi_2)$ are successively determined as functions of $(a_0, b_0, \psi_0, \phi_0)$, which in turn are known functions of $x_i, \dot{x}_i, y_i, \dot{y}_i$ [cf. Eq. (24a) and (24b)].

EXAMPLES AND DISCUSSION

Consider two coupled van der Pol oscillators described by the equations

$$\ddot{x} + x = \epsilon[c_0\dot{x}(1 - x^2) + c_1 y] \tag{28a}$$

and

$$\ddot{y} + y = \epsilon[c_2\dot{y}(1 - y^2) + c_3 x + c_4 y], \tag{28b}$$

where c_1 and c_3 are coupling constants and c_4 is proportional to the detuning. The results of carrying out the perturbation calculation are

$$\dot{a} = \epsilon[\tfrac{1}{2}c_0 a(1 - \tfrac{1}{4}a^2) - \tfrac{1}{2}c_1 b\,\sin\theta], \tag{29a}$$

$$\dot{b} = \epsilon[\tfrac{1}{2}c_2 b(1 - \tfrac{1}{4}b^2) + \tfrac{1}{2}c_3 a\,\sin\theta], \tag{29b}$$

$$\dot{\psi} = 1 - \epsilon(c_1 b/2a)\cos\theta, \tag{29c}$$

$$\dot{\phi} = 1 - \tfrac{1}{2}\epsilon c_4 - \epsilon(c_3 a/2b)\cos\theta, \tag{29d}$$

so that

$$\dot{\theta} = \epsilon\{\tfrac{1}{2}[-(c_1 b/a) + (c_3 a/b)]\cos\theta + \tfrac{1}{2}c_4\}. \tag{29e}$$

It does not appear possible to analytically integrate Eqs. (29a), (29b), and (29e), and this is generally the case. However, the equations for $\dot{a}, \dot{b}$, and $\dot{\theta}$ are substantially easier to integrate numerically than the original equations, since the former describe slowly varying amplitudes and frequencies. Furthermore, the equations for $\dot{a}, b$, and $\dot{\theta}$ yield insight into the system behavior. Consider, for example, setting $\dot{\theta} = 0$ in Eq. (29e), which corresponds to synchronization. Then equilibrium values of a and b can be determined. Furthermore, occasionally a conservation condition relating a and b can be obtained as, for example,[2] if f and g in Eqs. (1a) and (1b) are given by $f = xw(r^2)$ and $g = yw(r^2)$, where $r^2 = x^2 + y^2$. The three equations relating a, b, and θ, then reduce to two equations in a and θ alone, which are amenable to study in the phase plane.

As a second example consider the linear coupled equations

$$\ddot{x} + x = \epsilon^2 c_1 y \tag{30a}$$

and

$$\ddot{y} + y = \epsilon c_4 y + \epsilon^2 c_3 x. \tag{30b}$$

The solutions obtained by linear analysis are

$$x = a_0\{\cos(t) + (\epsilon b_0 c_1/a_0 c_4)\cos[(1 - \tfrac{1}{2}\epsilon c_4)t]\} + O(\epsilon^2), \tag{31a}$$

and

$$y = b_0\{\cos[(1 - \tfrac{1}{2}\epsilon c_4)t] - (\epsilon a_0 c_3/b_0 c_4)\cos(t)\} + O(\epsilon^2), \tag{31b}$$

where a_0 and b_0 are constants.

The results of our analysis, which follow from Eqs. (29a)–(29c), are

$$\dot{a} = -\tfrac{1}{2}\epsilon^2 c_1 b\,\sin\theta, \tag{32a}$$

$$\dot{b} = \tfrac{1}{2}\epsilon^2 c_3 a\,\sin\theta, \tag{32b}$$

$$\dot{\psi} = 1 - (\epsilon^2 c_1 b/2a)\cos\theta, \tag{32c}$$

$$\dot{\phi} = 1 - \tfrac{1}{2}\epsilon c_4 - (\epsilon^2 c_3 a/2b)\cos\theta, \tag{32d}$$

$$\dot{\theta} = \tfrac{1}{2}\epsilon c_4 + \tfrac{1}{2}\epsilon^2[(c_3 a/b) - (c_1 b/a)]\cos\theta. \tag{32e}$$

To integrate these equations, we assume

$$a = a_0[1 + O(\epsilon)] \quad \text{and} \quad b = b_0[1 + O(\epsilon)]$$

so that

$$\dot{\theta} = \tfrac{1}{2}\epsilon c_4 + \tfrac{1}{2}\epsilon^2\{[c_3(a_0/b_0) - c_1(b_0/a_0)]\}[1 + O(\epsilon)]\cos\theta, \tag{33}$$

and the integrals are[3]

$$\cos\theta = \frac{A\cos(\gamma t) - B}{A - B\cos(\gamma t)}, \tag{34a}$$

and

$$\sin\theta = (A/\gamma)[1 + (B/A)\cos\theta]\sin\gamma t, \tag{34b}$$

where

$$A = \tfrac{1}{2}\epsilon c_4, \tag{34c}$$

$$B = -\tfrac{1}{2}\epsilon^2[(c_1 b_0)/a_0) - (c_3 a_0/b_0)], \tag{34d}$$

and

$$\gamma = (A^2 - B^2)^{1/2}. \tag{34e}$$

After some manipulation we obtain

$$\cos\theta = \cos(At) - (B/2A)[1 - \cos(2At)] + O(\epsilon^2) \tag{35a}$$

and

$$\sin\theta = \sin(At) + (B/2A)\sin(2At) + O(\epsilon^2), \quad (35b)$$

and so

$$a = a_0[1 + (\epsilon c_1/c_4)(b_0/a_0)\cos(At)] + O(\epsilon^2), \quad (36a)$$

$$b = b_0[1 - (\epsilon c_3/c_4)(a_0/b_0)\cos(At)] + O(\epsilon^2), \quad (36b)$$

$$\psi = t - (\epsilon c_1/c_4)(b_0/a_0)\sin(At) + O(\epsilon^2), \quad (36c)$$

$$\phi = (1 - \tfrac{1}{2}\epsilon c_4)t - (\epsilon c_3/c_4)(a_0/b_0)\sin(At) + O(\epsilon^2). \quad (36d)$$

Upon substitution in $x = a\cos\psi$ and $y = b\cos\phi$, the results agree identically with Eqs. (31a) and (31b).

[1] N. N. Bogoliubov and Y. A. Mitropolysky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (Gordon and Breach, New York, 1961).

[2] K. B. Paxton and W. Streifer, Appl. Opt. **10**, 1164 (1971).

[3] G. P. Bois, *Tables of Indefinite Integrals* (Dover, New York, 1961), p. 121.

Symmetry in Einstein-Maxwell Space-Time

George Debney

Virginia Polytechnic Institute and State University, Department of Mathematics, Blacksburg, Virginia 24061

(Received 4 April 1972)

By using the complex null tetrad as basis for the tangent space, a Killing vector field ("symmetry") is introduced into the system of Einstein's equations with Maxwell's equations. The two bivectors $F_{\mu\nu}$ and $K_{\mu;\nu}$ (the associated Killing bivector) are assumed to have a principal null direction in common. Killing's equations, Maxwell's equations, and Einstein's equations are then written down for the case where this special direction is also a principal null geodesic for the Weyl conformal tensor. A certain analog of the Goldberg–Sachs theorem is proved. The static cases, plus a sizeable class of the static algebraically special cases are examined, to wit: where the special direction is also shear-free. In particular, all such algebraically special spaces must be Petrov Type D as a result of a coupling of the principal null directions for $F_{\mu\nu}$. This algebraically special metric is derived as an example of the static classes and is a static generalization of the Reissner–Nordström metric.

1. INTRODUCTION AND PRELIMINARIES

This work is concerned with properties of spaces admitting the vacuum Einstein–Maxwell equations plus a Killing vector field. The formalism used towards these ends is that of the complex null tetrad.

Section 2 develops some of the relationships between the electromagnetic field in general relativity and the existing geometry of the principal null directions for that field. This is mostly a synthesis of known results scattered through the literature.

Section 3 derives a result in general relativity with electromagnetic theory which says that given certain conditions, including a Killing vector field, one has a theorem of the sort "geodesic and shear-free implies algebraically special." Certainly, such a result is next to trivial when the electromagnetic field is null; but the new results are in the context of a nonnull electromagnetic field.

Section 4 takes up the subject of lightlike Killing vector fields (for any space–time) in a separate context. Some of the cumbersome cases in more general Killing vector studies are eliminated by the separate results possible here.

Section 5 derives the equations necessary to study first any vacuum Einstein–Maxwell space–time with a Killing vector field whose Killing bivector and electromagnetic bivector possess a principal null direction in common. This is specialized to the study of hypersurface orthogonal (in particular, static) Killing vector fields. The primary equations include algebraically general as well as algebraically special cases. The field equations are included in their general form.

Section 6 derives as an example a static generalization of the Reissner–Nordström solution for a point charge. This solution has two arbitrary constants: one real and associated with mass and one complex and associated with a static electromagnetic field. It can be made to admit as many as four Killing vector fields and, in that event, be the Reissner–Nordström solution. More general axially symmetric cases are possible, however, by a less restrictive choice of the function $p(\zeta, \bar{\zeta})$. With any choice, however, the electromagnetic field remains unaltered for the whole class.

Reference 1, Secs. 2–4, contains most of the necessary details surrounding the mathematical development of this section. In the following work, quite extensive use is made of the notion of a complex null tetrad $\{\mathbf{e}_a | a = 1, 2, 3, 4\}$ and its dual $\{\epsilon^a | a = 1, 2, 3, 4\}$ ($\mathbf{e}_1$ and $\mathbf{e}_2$ are complex conjugates of one another while $\mathbf{e}_3$ and $\mathbf{e}_4$ are real; all four vectors are null in the sense that their squares, $\mathbf{e}_a \cdot \mathbf{e}_a$, are zero). If Greek indices refer to components of an object with respect to a coordinate system $\{x^\mu\}$ and Latin indices refer to components of an object with respect to a general basis (tetrad), we have $\mathbf{e}_a = e_a{}^\mu \partial_\mu$ and $\epsilon^a = \epsilon^a{}_\mu dx^\mu$, where $\epsilon^a(\mathbf{e}_b) = \delta^a_b$; i.e., $\epsilon^a{}_\mu e_b{}^\mu = \delta^a{}_b$ and $e_a{}^\mu \epsilon^a{}_\nu = \delta^\mu{}_\nu$. The choice of signature for the metric tensor $g = g_{\mu\nu} dx^\mu dx^\nu = g_{ab}\epsilon^a\epsilon^b [\epsilon^a\epsilon^b \equiv \tfrac{1}{2}(\epsilon^a \otimes \epsilon^b + \epsilon^b \otimes \epsilon^a)]$ is made to be (+++−). Locally over the C^∞ Lorentz manifold, the tetrad components $g_{ab} = g(\mathbf{e}_a, \mathbf{e}_b)$ of the metric may be given the form $g_{ab} = \tfrac{1}{2}(\delta^1{}_a\delta^2{}_b + \delta^2{}_a\delta^1{}_b) + \tfrac{1}{2}(\delta^3{}_a\delta^4{}_b + \delta^4{}_a\delta^3{}_b)$ so that $g = 2\epsilon^1\epsilon^2 + 2\epsilon^3\epsilon^4$.

The set of transformations on the tetrad $\mathbf{e}_a \to \mathbf{e}_{a'}$ preserving the above form for g_{ab} (i.e., for which $g_{a'b'} = g_{ab}$) is called the set of Lorentz transformations. The proper orthochronous subgroup of these is given by

$$\begin{bmatrix} \exp(iB)\mathbf{e}_{1'} \\ \exp(-iB)\mathbf{e}_{2'} \\ \exp(A)\mathbf{e}_{3'} \\ \exp(-A)\mathbf{e}_{4'} \end{bmatrix} = |1 - \alpha\beta|^{-1}$$

$$\sin\theta = \sin(At) + (B/2A)\sin(2At) + O(\epsilon^2), \quad (35b)$$

and so

$$a = a_0[1 + (\epsilon c_1/c_4)(b_0/a_0)\cos(At)] + O(\epsilon^2), \quad (36a)$$

$$b = b_0[1 - (\epsilon c_3/c_4)(a_0/b_0)\cos(At)] + O(\epsilon^2), \quad (36b)$$

$$\psi = t - (\epsilon c_1/c_4)(b_0/a_0)\sin(At) + O(\epsilon^2), \quad (36c)$$

$$\phi = (1 - \tfrac{1}{2}\epsilon c_4)t - (\epsilon c_3/c_4)(a_0/b_0)\sin(At) + O(\epsilon^2). \quad (36d)$$

Upon substitution in $x = a\cos\psi$ and $y = b\cos\phi$, the results agree identically with Eqs. (31a) and (31b).

[1] N. N. Bogoliubov and Y. A. Mitropolysky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (Gordon and Breach, New York, 1961).

[2] K. B. Paxton and W. Streifer, Appl. Opt. **10**, 1164 (1971).

[3] G. P. Bois, *Tables of Indefinite Integrals* (Dover, New York, 1961), p. 121.

Symmetry in Einstein-Maxwell Space-Time

George Debney

Virginia Polytechnic Institute and State University, Department of Mathematics, Blacksburg, Virginia 24061

(Received 4 April 1972)

By using the complex null tetrad as basis for the tangent space, a Killing vector field ("symmetry") is introduced into the system of Einstein's equations with Maxwell's equations. The two bivectors $F_{\mu\nu}$ and $K_{\mu;\nu}$ (the associated Killing bivector) are assumed to have a principal null direction in common. Killing's equations, Maxwell's equations, and Einstein's equations are then written down for the case where this special direction is also a principal null geodesic for the Weyl conformal tensor. A certain analog of the Goldberg–Sachs theorem is proved. The static cases, plus a sizeable class of the static algebraically special cases are examined, to wit: where the special direction is also shear-free. In particular, all such algebraically special spaces must be Petrov Type D as a result of a coupling of the principal null directions for $F_{\mu\nu}$. This algebraically special metric is derived as an example of the static classes and is a static generalization of the Reissner–Nordström metric.

1. INTRODUCTION AND PRELIMINARIES

This work is concerned with properties of spaces admitting the vacuum Einstein–Maxwell equations plus a Killing vector field. The formalism used towards these ends is that of the complex null tetrad.

Section 2 develops some of the relationships between the electromagnetic field in general relativity and the existing geometry of the principal null directions for that field. This is mostly a synthesis of known results scattered through the literature.

Section 3 derives a result in general relativity with electromagnetic theory which says that given certain conditions, including a Killing vector field, one has a theorem of the sort "geodesic and shear-free implies algebraically special." Certainly, such a result is next to trivial when the electromagnetic field is null; but the new results are in the context of a nonnull electromagnetic field.

Section 4 takes up the subject of lightlike Killing vector fields (for any space–time) in a separate context. Some of the cumbersome cases in more general Killing vector studies are eliminated by the separate results possible here.

Section 5 derives the equations necessary to study first any vacuum Einstein–Maxwell space–time with a Killing vector field whose Killing bivector and electromagnetic bivector possess a principal null direction in common. This is specialized to the study of hypersurface orthogonal (in particular, static) Killing vector fields. The primary equations include algebraically general as well as algebraically special cases. The field equations are included in their general form.

Section 6 derives as an example a static generalization of the Reissner–Nordström solution for a point charge. This solution has two arbitrary constants: one real and associated with mass and one complex and associated with a static electromagnetic field. It can be made to admit as many as four Killing vector fields and, in that event, be the Reissner–Nordström solution. More general axially symmetric cases are possible, however, by a less restrictive choice of the function $p(\zeta, \bar{\zeta})$. With any choice, however, the electromagnetic field remains unaltered for the whole class.

Reference 1, Secs. 2–4, contains most of the necessary details surrounding the mathematical development of this section. In the following work, quite extensive use is made of the notion of a complex null tetrad $\{\mathbf{e}_a | a = 1, 2, 3, 4\}$ and its dual $\{\epsilon^a | a = 1, 2, 3, 4\}$ ($\mathbf{e}_1$ and $\mathbf{e}_2$ are complex conjugates of one another while $\mathbf{e}_3$ and $\mathbf{e}_4$ are real; all four vectors are null in the sense that their squares, $\mathbf{e}_a \cdot \mathbf{e}_a$, are zero). If Greek indices refer to components of an object with respect to a coordinate system $\{x^\mu\}$ and Latin indices refer to components of an object with respect to a general basis (tetrad), we have $\mathbf{e}_a = e_a{}^\mu \partial_\mu$ and $\epsilon^a = \epsilon^a{}_\mu dx^\mu$, where $\epsilon^a(\mathbf{e}_b) = \delta^a_b$; i.e., $\epsilon^a{}_\mu e_b{}^\mu = \delta^a{}_b$ and $e_a{}^\mu \epsilon^a{}_\nu = \delta^\mu{}_\nu$. The choice of signature for the metric tensor $g = g_{\mu\nu} dx^\mu dx^\nu = g_{ab}\epsilon^a\epsilon^b [\epsilon^a\epsilon^b \equiv \tfrac{1}{2}(\epsilon^a \otimes \epsilon^b + \epsilon^b \otimes \epsilon^a)]$ is made to be (+++−). Locally over the C^∞ Lorentz manifold, the tetrad components $g_{ab} = g(\mathbf{e}_a, \mathbf{e}_b)$ of the metric may be given the form $g_{ab} = \tfrac{1}{2}(\delta^1{}_a\delta^2{}_b + \delta^2{}_a\delta^1{}_b) + \tfrac{1}{2}(\delta^3{}_a\delta^4{}_b + \delta^4{}_a\delta^3{}_b)$ so that $g = 2\epsilon^1\epsilon^2 + 2\epsilon^3\epsilon^4$.

The set of transformations on the tetrad $\mathbf{e}_a \to \mathbf{e}_{a'}$ preserving the above form for g_{ab} (i.e., for which $g_{a'b'} = g_{ab}$) is called the set of Lorentz transformations. The proper orthochronous subgroup of these is given by

$$\begin{bmatrix} \exp(iB)\mathbf{e}_{1'} \\ \exp(-iB)\mathbf{e}_{2'} \\ \exp(A)\mathbf{e}_{3'} \\ \exp(-A)\mathbf{e}_{4'} \end{bmatrix} = |1 - \alpha\beta|^{-1}$$

$$\times \begin{pmatrix} 1 & \bar{\alpha}\beta & -\bar{\alpha} & \beta \\ \alpha\bar{\beta} & 1 & -\alpha & \bar{\beta} \\ -\bar{\beta} & -\beta & 1 & -\beta\bar{\beta} \\ \alpha & \bar{\alpha} & -\alpha\bar{\alpha} & 1 \end{pmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \mathbf{e}_4 \end{bmatrix}, \quad (1.1)$$

where A, B, α, β are parameters; $\alpha\beta \neq 1$; A and B are real; α and β are complex. (A bar above a symbol denotes complex conjugation.)

As a basis for the space of bivectors (2-forms) over this manifold, take the set $\{\epsilon^A | A = \text{I, II, III, IV, V, VI}\}$ or its dual $\{e_A\}$ introduced in Ref. 1; i.e., the bivector $B_{\mu\nu}dx^\mu \wedge dx^\nu = B_{ab}\epsilon^a \wedge \epsilon^b = B_A\epsilon^A$, where $B_{\mu\nu} = -B_{\nu\mu}$. The adjoint ${}^*\hat{B} \equiv B^*_{\mu\nu}dx^\mu \wedge dx^\nu = B^*_A\epsilon^A$ is another bivector formed as follows:

$$B^*_{\mu\nu} \equiv \tfrac{1}{2}\eta_{\mu\nu\rho\sigma}B^{\rho\sigma}, \quad (1.2a)$$

or

$$B^*_{ab} = \tfrac{1}{2}\eta_{abcd}B^{cd}, \quad (1.2b)$$

where $\eta_{\mu\nu\rho\sigma} = [\det(g_{\mu\nu})]^{1/2} \cdot \epsilon_{\mu\nu\rho\sigma}$ with $\epsilon_{\mu\nu\rho\sigma}$ the completely skew-symmetric Levi-Civita permutation symbol. It turns out that $\eta_{1234} = i = \sqrt{-1}$ where the 1, 2, 3, 4 are *tetrad* indices as used in (1.2b). (Unless otherwise stated, all number indices will refer to the complex null tetrad.) A bivector B is *null* if and only if $B_AB^A = 0 = B^*_AB^A$; otherwise it is said to be *nonnull*.

The following canonical forms are possible through the transformations (1.1) using choices of α and β only:

$$B \text{ null: } B = 2B_{\text{III}}\epsilon^3 \wedge \epsilon^1 + 2B_{\text{VI}}\epsilon^3 \wedge \epsilon^2 \quad (\text{or } = 2B_{\text{I}}\epsilon^4 \wedge \epsilon^2 + 2B_{\text{IV}}\epsilon^4 \wedge \epsilon^1), \quad (1.3a)$$

$$B \text{ nonnull: } B = 2B_{\text{II}}(\epsilon^1 \wedge \epsilon^2 + \epsilon^3 \wedge \epsilon^4) + 2B_{\text{V}}(\epsilon^2 \wedge \epsilon^1 + \epsilon^3 \wedge \epsilon^4), \quad (1.3b)$$

where $B_{\text{III}} = B_{31}$ and $B_{\text{II}} = \frac{1}{2}(B_{12} + B_{34})$. The transformation freedom left on (1.1) after transforming a null bivector to (1.3a) is

$$\begin{aligned} \mathbf{e}_{1'} &= \exp(-i\text{B}) \cdot [\mathbf{e}_1 + \beta\mathbf{e}_4], \\ \mathbf{e}_{2'} &= \exp(i\text{B}) \cdot [\mathbf{e}_2 + \bar{\beta}\mathbf{e}_4], \\ \mathbf{e}_{3'} &= \exp(-\text{A}) \cdot [\mathbf{e}_3 - \bar{\beta}\mathbf{e}_1 - \beta\mathbf{e}_2 - \beta\bar{\beta}\mathbf{e}_4], \\ \mathbf{e}_{4'} &= \exp(\text{A}) \cdot [\mathbf{e}_4], \end{aligned} \quad (1.4)$$

the so-called *null rotations about* $\mathbf{e}_4$ (which is then a principal null direction for B). The transformation freedom left on (1.1) after the nonnull bivector is put into the form (1.3b) is

$$\begin{aligned} \mathbf{e}_{1'} &= \exp(-i\text{B})\mathbf{e}_1, \quad \mathbf{e}_{2'} = \exp(i\text{B})\mathbf{e}_2, \\ \mathbf{e}_{3'} &= \exp(-\text{A})\mathbf{e}_3, \quad \mathbf{e}_{4'} = \exp(\text{A})\mathbf{e}_4; \end{aligned} \quad (1.5)$$

i.e., a simple *scaling* is all that can be allowed. Both $\mathbf{e}_3$ and $\mathbf{e}_4$ are now principal null directions for B.) It turns out that in the latter case, $B_{\text{II}'} = B_{\text{II}}$ under (1.5) so that no further simplification of B_{II} (such as making it real, or pure imaginary, or constant) is possible. This is in contrast to the null case where $B_{\text{III}'} = \exp(\text{A} - i\text{B})B_{\text{III}}$ under (1.4).

If we define

$$\mathfrak{G}^{(-)} \equiv B + i^*B, \quad (1.6)$$

then

$$\mathfrak{G}^{(-)} = 2B_{\text{I}}\epsilon^{\text{I}} + 2B_{\text{II}}\epsilon^{\text{II}} + 2B_{\text{III}}\epsilon^{\text{III}} = \mathfrak{G}^{(-)}{}_A\epsilon^A. \quad (1.7)$$

The following product is an invariant with respect to both coordinate and tetrad transformations:

$$\mathcal{K} \equiv \tfrac{1}{2}\mathfrak{G}^{(-)}{}_AB^A = 2(B_{\text{I}}B_{\text{III}} - B^2_{\text{II}}). \quad (1.8)$$

The existence of canonical forms for B and the invariance of (1.8) lets us obtain

$$B \text{ null} \Rightarrow \mathcal{K} = 0; B \text{ nonnull} \Rightarrow \mathcal{K} = -2B^2_{\text{II}}. \quad (1.9)$$

The equations for a source-free Maxwell field, whose field tensor is the bivector B are written classically as

$$B_{[\mu\nu;\sigma]} = 0 = B^{\mu\nu}{}_{;\nu}. \quad (1.10)$$

Equations (1.10) have a particularly simple form if one uses (1.6) and exterior differentiation. The following is equivalent to (1.10):

$$d\mathfrak{G}^{(-)} = 0. \quad (1.11)$$

Hence, the source-free Maxwell equations may be written in the form of (1.11).

2. EINSTEIN-MAXWELL THEORY: THE ELECTROMAGNETIC BIVECTOR IN GENERAL RELATIVITY

Let $F_{\mu\nu}$ be a source-free Maxwell field defined locally over the manifold; i.e., $F = F_{\mu\nu}dx^\mu \wedge dx^\nu = F_{ab}\epsilon^a \wedge \epsilon^b$ is an electromagnetic bivector which satisfies $d\mathfrak{F}^{(-)} = 0$. To couple this with relativity theory, the equations ($R_{\mu\nu}$ is the Ricci tensor and $T_{\mu\nu}$ is the Maxwell energy–momentum tensor)

$$R_{\mu\nu} = -8\pi T_{\mu\nu}, \quad (2.1)$$

where $T_{\mu\nu} = (4\pi)^{-1}(F_{\mu\sigma}F_\nu{}^\sigma - \frac{1}{4}g_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma})$ (2.2)

must be satisfied in addition to the Maxwell equations.

If the connection coefficients $\Gamma_{abc} = -\Gamma_{bac}$ are defined through the first structure equations ($\Gamma_{abc} \equiv g_{am}\Gamma^m{}_{bc}$)

$$d\epsilon^a = \Gamma^a{}_{bc}\epsilon^b \wedge \epsilon^c, \quad (2.3)$$

the null vector field $\mathbf{e}_4$ has the following interpretation of its connection coefficients Γ_{4ab}:

Γ_{424} = geodesy: $\mathbf{e}_4$ is geodesic $\Leftrightarrow \Gamma_{424} = 0$,

Γ_{422} = shear: $\mathbf{e}_4$ is shearfree $\Leftrightarrow \Gamma_{422} = 0$,

Γ_{423} = rotation,

Γ_{421} = complex expansion: $\text{Re}(\Gamma_{421})$ = expansion, $\text{Im}(\Gamma_{421})$ = twist.

The physical interpretations of these quantities have been given the most attention for the case where $\Gamma_{424} = 0$—the geodesic null congruence $\mathbf{e}_4$. (See, for example, the work of Sachs.[2])

If F is a *null* bivector and if $\mathbf{e}_4$ is a principal null direction for F, then it can be shown that $d\mathfrak{F}^{(-)} = 0$ implies

$$\Gamma_{424} = \Gamma_{422} = 0; \quad (2.4)$$

i.e., $\mathbf{e}_4$ is then geodesic and shear-free. This is the Mariot–Robinson theorem, and its proof is straightforward using F in its canonical form (1.3a).

The Goldberg–Sachs theorem[3] and additional lemmas in Ref. 3 were generalized by Robinson and Schild.[4] In terms of the notation used in this paper, Ref. 3 contains the following theorems:

(1) *A vacuum space–time* ($R_{\mu\nu} = 0$) *is algebraically special* ($C^{(5)} = C^{(4)} = 0$) *if and only if there exists a geodesic and shear-free lightlike vector field* ($\mathbf{e}_4$). *Such a vector field is a degenerate principal null direction for the Riemann* (Weyl conformal) *curvature tensor.*

(2) *If with the tetrad* $\{\mathbf{e}_a\}$, $\mathbf{e}_4$ *is geodesic and shear-free and the Ricci tensor has the form* $R_{42} = R_{41} = R_{44} = R_{11} = R_{22} = 0$, *then* $C^{(5)} = C^{(4)} = 0$; *i.e., the space is algebraically special.* (This is not redundant to the theorem above whenever the field equations are more general than $R_{\mu\nu} = 0$.)

A corollary of the latter has application to electromagnetic fields and may be stated as follows:

Theorem 1: If an electromagnetic field F in Einstein–Maxwell theory has a principal null direction $\mathbf{e}_4$ which is geodesic and shear-free, then the space is algebraically special. Furthermore $C^{(5)} = C^{(4)} = 0$ ($\mathbf{e}_4$ *is a degenerate principal null direction for the Weyl tensor.*)

Proof: Using the canonical form for a null or nonnull F, a straightforward calculation shows that the tensor T_{ab} has sufficient zero components to apply the second theorem above. QED

A null field F *always* has its principal null direction to be geodesic and shearfree. (A nonnull field generally has no such properties.) Hence we see that *a null electromagnetic field in Einstein–Maxwell theory is algebraically special,* as shown by Sachs.

One theorem of Robinson and Schild is given (where the tetrad is chosen so that F_{II} is the only nonzero component of $\mathfrak{F}^{(-)}$) by the following. *If* $C^{(5)} = C^{(4)} = 0$ ($\mathbf{e}_4$ *a degenerate principal null direction for the Weyl tensor), then* $\mathbf{e}_4$ *is geodesic and shear-free if and only if the* "weak field equations"

$$P^{(-)}{}_{242} = 0 = P^{(-)}{}_{442}, \tag{2.5}$$

where $P_{abc} \equiv -R_{a[b;c]} + \frac{1}{6} g_{a[b} R_{,c]}$ *and* $P^{(-)} \equiv P + i^*P$, *are satisfied.* In the case of a nonnull electromagnetic field, Eqs. (2.5) become

$$|F_{II}|^2 \Gamma_{422} = 0 = |F_{II}|^2 \Gamma_{424}.$$

Hence, no new information can be gained in the nonnull electromagnetic case since "geodesic and shear-free $\Longleftrightarrow$ geodesic and shear-free" is all that emerges. There is no direct analog, for electromagnetic fields, of the Goldberg–Sachs theorem; but certain familiar additional constraints can give rise to a certain kind of analog. Introduction of a Killing vector field into the system provides one such constraint.

Finally the Maxwell equations, $d\mathfrak{F}^{(-)} = 0$, may be written for a nonnull electromagnetic field F in canonical form as

$$F_{II,1} = -2F_{II}\Gamma_{314}, \quad F_{II,2} = -2F_{II}\Gamma_{423},$$
$$F_{II,3} = 2F_{II}\Gamma_{312}, \quad F_{II,4} = 2F_{II}\Gamma_{421}. \tag{2.6}$$

Equations (2.6) are simple by virtue of both $\mathbf{e}_3$ and $\mathbf{e}_4$ being principal null directions for F, making $F_I = F_{III} = 0$.

3. INTRODUCTION OF A SYMMETRY INTO THE EINSTEIN–MAXWELL SYSTEM

It is sufficient to study the *nonnull* electromagnetic field F in this section since the main purpose here is to try to say something about algebraically special spaces and their relation to the geodesic and shear-free property; certainly much has already been said about the *null* field F in the literature.

Let $\mathbf{K} \equiv K^a \mathbf{e}_a$ be a (contravariant) Killing vector field defined locally over the manifold, and let $K \equiv K_a \epsilon^a$ be its dual (covariant) field. Then $(\pm)\, dK$ is called the *Killing bivector*[1] (KBV) and it has components $K_{a;b}$. (Recall $K_{a;b} + K_{b;a} = 0$ is a form of Killing's equations.) Just as with any bivector, dK may be classified according to whether it is null or nonnull and may be put into a canonical form by Lorentz transformations on the tetrad. If we define $B \equiv dK$, then $d\mathfrak{B}^{(-)} = 0$ if $R_{\mu\nu} = 0$; however, $R_{\mu\nu} = -8\pi T_{\mu\nu}$ for the electromagnetic field and only $dB = 0$ holds true in that case. Killing's equations with a first set of integrability conditions may be written as follows:

$$K_{a;b} = -K_{b;a}, \tag{3.1}$$

$$K_{a;bc} = R_{abcm} K^m, \tag{3.2}$$

$$K_{a;}{}^{b}{}_{b} = R_{am} K^m. \tag{3.3}$$

All three are quite tedious in their most general form, but become more manageable if $\mathbf{e}_3$ or $\mathbf{e}_4$ is a principal null direction for dK. Consider the following lemma.

Lemma 1: Let F be a nonnull electromagnetic field satisfying an Einstein–Maxwell system. Let $\mathbf{K}$ be a Killing vector field admitted by the system for which $\mathbf{e}_4$ is a principal null direction for both dK and F. If the space–time is algebraically special, with $\mathbf{e}_4$ a degenerate principal null direction for the Weyl tensor, then $\mathbf{e}_4$ is geodesic and shear-free; i.e.,

$$C^{(5)} = C^{(4)} = 0 \Rightarrow \mathbf{e}_4 \text{ geodesic and shear-free.}$$

Proof: Since $\mathbf{e}_4$ is a principal null direction for dK, it follows that $K_{1;4} = K_{2;4} = K_{4;1} = K_{4;2} = 0$; i.e., $K_I = 0 = \overline{K}_I$. Hence two equations from (3.2) are of the form

$$4K_{II}\Gamma_{422} = -C^{(4)}K^1 - C^{(5)}K^4, \tag{3.4a}$$

$$4K_{II}\Gamma_{424} = C^{(5)}K^2 - C^{(4)}K^3, \tag{3.4b}$$

whenever dK is a nonnull KBV. So in this case $C^{(5)} = C^{(4)} = 0$ implies $\Gamma_{422} = \Gamma_{424} = 0$; i.e., $\mathbf{e}_4$ is geodesic and shear-free. In the other case where dK is a null KBV, the Eqs. (3.2) include

$$C^{(4)}K^2 - C^{(3)}K^3 = 0 = C^{(3)}K^1 + C^{(4)}K^4. \tag{3.5}$$

Then if $C^{(5)} = C^{(4)} = 0$, $C^{(3)}K^1 = C^{(3)}K^3 = 0$. This

last condition then splits into two cases: $C^{(3)} \neq 0$ or $C^{(3)} = 0$.

Suppose $C^{(3)} \neq 0$. Then $K^1 = K^2 = K^3 = 0$. Hence $\mathbf{K} = K^4\mathbf{e}_4$ and $K^4 \to 1$ by a transformation, making $\mathbf{K} = \mathbf{e}_4$. By virtue of being a lightlike Killing vector it is clear by covariant differentiation that $\mathbf{e}_4$ is geodesic and shear-free.

Suppose $C^{(3)} = 0$. At this point we refer to the work of Robinson and Schild.[4] If $C^{(5)} = C^{(4)} = C^{(3)} = 0$, then "degeneracy $d_{(2)}$" of Ref. 4 is satisfied: "Theorem I" of that reference has the result that "$d_{(2)}$" implies Eq. (2.5) above in the present paper. Hence, for the nonnull electromagnetic field, (2.5) gives us that $\Gamma_{424} = \Gamma_{422} = 0$. QED

(Notice that in the last paragraph no essential use was made of a Killing vector field; hence we have the lemma: *If F is nonnull and $C^{(5)} = C^{(4)} = C^{(3)} = 0$, then $\mathbf{e}_4$ is geodesic and shear-free in an Einstein–Maxwell space-time.*)

The lemma above and Theorem 1 go together to form the following theorem.

Theorem 2: Let F be a nonnull electromagnetic bivector satisfying an Einstein–Maxwell system. Let $\mathbf{K}$ be a Killing vector field admitted by the system for which the tetrad vector $\mathbf{e}_4$ is a principal null direction for both dK and F. Then the space-time is algebraically special, with $\mathbf{e}_4$ a degenerate principal null direction for the Weyl tensor, if and only if $\mathbf{e}_4$ is geodesic and shearfree; i.e., $C^{(5)} = C^{(4)} = 0$ if and only if $\mathbf{e}_4$ is geodesic and shear-free.

At this stage the "charged-Kerr" solution (see for example Ref. 5) with its timelike Killing vector and nonnull electromagnetic field goes as an example consistent with this theorem. Another consistent example is in the work mentioned by Mas.[6]

4. LIGHTLIKE KILLING VECTOR FIELDS IN ANY SPACE-TIME

A digression on lightlike Killing vector fields appears in order at this point. If it is possible to conclude something useful in general here no matter what the space-time, exclusion of these special cases in subsequent discussions is then made possible.

Let $\mathbf{K} = K^4\mathbf{e}_4$ be a Killing vector field, without loss of generality. A Lorentz transformation on the complex null tetrad merely of the form (1.5) with $B = 0$ may transform $K^4 \to 1$. Then if $\mathbf{K} = K^a\mathbf{e}_a$, $K^a = \delta^a{}_4$ or $K_a = g_{a4}$. Consequently

$$K_{a;b} = -\Gamma_{4ab} \tag{4.1}$$

are the components of the KBV. Since $K_{1;1} = K_{2;2} = K_{3;3} = 0$,

$$\Gamma_{411} = \Gamma_{422} = \Gamma_{343} = 0. \tag{4.2}$$

Also notice that $K_{4;2} = -\Gamma_{442} \equiv 0$ and that $K_{2;4} = -\Gamma_{424}$; hence

$$\Gamma_{414} = \Gamma_{424} = 0. \tag{4.3}$$

Equations (4.2) and (4.3) imply that $\mathbf{e}_4$ is *geodesic* and *shear-free*. Moreover $K_{4;2} = 0 \Rightarrow K_{\mathrm{I}} = 0$, in bivector notation. Hence $\mathbf{e}_4$ is a principal null direction for its own KBV.

More relationships from (4.1) and Killing's equations give the following:

$$\Gamma_{421} + \Gamma_{412} = 0 = \Gamma_{344} = \Gamma_{342} - \Gamma_{423}. \tag{4.4}$$

Since one usually defines the *complex expansion* of e_4, $z \equiv \Gamma_{421}$, then $z + \bar{z} = 0$. If $K_{\mathrm{II}} \neq 0$, $\mathbf{e}_{3'}$ can be made to coincide with the other principal null direction for dK, giving $K_{\mathrm{III}} \equiv K_{3;1} = 0$. If dK is a null KBV then $K_{\mathrm{II}} = 0$ and $K_{\mathrm{III}} \neq 0$, given that $K_{\mathrm{I}} = 0$.

Consider the expression $K_{[a;b]}K^{a;b}$. This has the explicit form $-\frac{1}{2}(\Gamma_{421} - \Gamma_{412})^2 = -\frac{1}{2}(z - \bar{z})^2$ whenever $K_{\mathrm{II}} \neq 0$. So, for this case,

$$K_{[a;b]}K^{a;b} = -\tfrac{1}{2}(2i)^2[\mathrm{Im}(z)]^2 = 2[\mathrm{Im}(z)]^2. \tag{4.5}$$

Therefore

$$\mathrm{Im}(z) = [\tfrac{1}{2}K_{a;b}K^{a;b}]^{1/2} \tag{4.6}$$

for $K_{\mathrm{II}} \neq 0$. If $K_{\mathrm{II}} = 0$ then $K_{1;2} = 0 \Rightarrow \Gamma_{412} = 0$. Therefore $z = 0$ for dK being a null KBV (by definition $K_{a;b}K^{a;b} = 0$ here).

Consider the expression $K^a{}_{;a}$. This has the explicit form $K^1{}_{;1} + K^2{}_{;2} + K^3{}_{;3} + K^4{}_{;4} = 0$, since $\mathbf{K}$ is a Killing vector. Killing's equations give $K_{1;2} + K_{2;1} = 0 = K_{3;4} + K_{4;3}$. Hence

$$2\,\mathrm{Re}(z) \equiv \Gamma_{412} + \Gamma_{421} = 0 = \Gamma_{344}, \tag{4.7}$$

when $K_{\mathrm{II}} \neq 0$. To conform to other authors' notation when dealing with lightlike congruences, we have thus far established that the complex expansion z of $\mathbf{e}_4$ has the general form

$$z = \tfrac{1}{2}K^a{}_{;a} + i[\tfrac{1}{2}K_{[a;b]}K^{a;b}]^{1/2}. \tag{4.8}$$

In all cases where $\mathbf{e}_4$ is proportional to a Killing vector $\mathbf{K}$, $K^a{}_{;a} = 0$.

Finally consider the integrability condition (3.3). Since $\mathbf{K}$ is lightlike,

$$0 = (K_aK^a)_{;b} = 2K_{a;b}K^a,$$
$$0 = K_{a;b}K^{a;b} + K_{a;b}{}^bK^a = K_{a;b}K^{a;b} + R_{ab}K^aK^b. \tag{4.9}$$

By virtue of $K^a = \delta^a{}_4$, Eq. (4.9) becomes

$$0 = K_{a;b}K^{a;b} + R_{44}; \tag{4.10}$$

hence $K_{a;b}K^{a;b} = 0 \iff R_{44} = 0$. This gives rise to the following theorem. (One shows $K^*_{a;b}K^{a;b} = 0$ here by noting that $z = 0$ implies $\mathbf{K}$ is hypersurface orthogonal.)

Theorem 3: Let $\mathcal{K}^\mu$ be a lightlike Killing vector congruence in a space-time. Then $R_{\mu\nu}\mathcal{K}^\mu\mathcal{K}^\nu = 0$ if and only if the complex expansion of $\mathcal{K}^\mu$, $z \equiv 0$. The KBV $\mathcal{K}_{\mu;\nu}$ is then null.

An application of the theorem above (in addition to the obvious cases: $R_{\mu\nu} = 0$ and $R_{\mu\nu} = \lambda g_{\mu\nu}$) is where $\mathcal{K}^\mu$ is also a principal null direction for an electromagnetic field. Then it is well known that $T_{\mu\nu}\mathcal{K}^\nu = \alpha\mathcal{K}_\mu$, so that $T_{\mu\nu}\mathcal{K}^\mu\mathcal{K}^\nu = 0$, where $T_{\mu\nu}$ is the electromagnetic energy-momentum tensor. And hence the Einstein–Maxwell equations are sufficient to give $z = 0$.

5. THE HYPERSURFACE ORTHOGONAL KILLING VECTOR (WITH NONNULL KBV) TOGETHER WITH A NONNULL ELECTROMAGNETIC FIELD

In this section we confine ourselves to the Einstein-Maxwell case admitting a Killing vector field which is hypersurface orthogonal. Further assumptions are that the Killing bivector and the electromagnetic field are both nonnull. One additional condition is added: *that the* KBV *and the electromagnetic field both have a principal null direction* (here $\mathbf{e}_4$) *in common which is also geodesic.*

If a space-time admits a *timelike* hypersurface orthogonal Killing vector field, it is said to be *static*. In this case the orthogonal hypersurfaces are spacelike and evolve with time and can be more easily visualized from a physical point of view; examples are the Schwarzschild and certain Weyl/Levi-Civita solutions (for vacuum space-times) or the Reissner-Nordström solution (for the Einstein-Maxwell case). The hypersurface orthogonal Killing vector field of this section is not necessarily timelike but could be spacelike just as well. These instances have practically the same mathematical properties, however, and timelike is assumed at an intermediate stage. (The lightlike case is excluded in view of the fact that $z = 0$ for this case, as shown in Sec. 4.)

Using the assumptions at the beginning of this section we reduce the electromagnetic field to its canonical components and do the same for the nonnull KBV:

$$F \leftrightarrow (0, F_{\mathrm{II}}, F_{\mathrm{III}}), \tag{5.1}$$

$$dK \leftrightarrow (0, K_{\mathrm{II}}, 0), \tag{5.2}$$

where $F_{\mathrm{II}} \equiv \frac{1}{2}(F_{12} + F_{34})$, $F_{\mathrm{III}} = F_{31}$, $K_{\mathrm{II}} = \frac{1}{2}(K_{1;2} + K_{3;4})$. Then Eqs. (3.2) and (3.3) yield the following general set of integrability conditions on $\mathbf{K}(= K^a\mathbf{e}_a)$:

$$\begin{aligned}
2K_{\mathrm{I};1} &= 4K_{\mathrm{II}}\Gamma_{421} = C^{(4)}K^2 - C^{(3)}K^3,\\
2K_{\mathrm{I};2} &= 4K_{\mathrm{II}}\Gamma_{422} = -C^{(4)}K^1 - C^{(5)}K^4,\\
2K_{\mathrm{I};3} &= 4K_{\mathrm{II}}\Gamma_{423} = C^{(3)}K^1 + C^{(4)}K^4,\\
2K_{\mathrm{I};4} &= 4K_{\mathrm{II}}\Gamma_{424} = C^{(5)}K^2 - C^{(4)}K^3;
\end{aligned} \tag{5.3}$$

$$\begin{aligned}
2K_{\mathrm{II};1} &= 2K_{\mathrm{II},1} = (C^{(3)} + R_{12})K^2 - C^{(2)}K^3,\\
2K_{\mathrm{II};2} &= 2K_{\mathrm{II},2} = -(C^{(3)} + R_{12})K^1\\
&\quad - R_{32}K^2 - C^{(4)}K^4,\\
2K_{\mathrm{II};3} &= 2K_{\mathrm{II},3} = C^{(2)}K^1 + R_{32}K^3\\
&\quad + (C^{(3)} + R_{34})K^4,\\
2K_{\mathrm{II};4} &= 2K_{\mathrm{II},4} = C^{(4)}K^2 - (C^{(3)} + R_{34})K^3;
\end{aligned} \tag{5.4}$$

$$\begin{aligned}
2K_{\mathrm{III};1} &= -4K_{\mathrm{II}}\Gamma_{311} = (C^{(2)} - R_{13})K^2 - C^{(1)}K^3,\\
2K_{\mathrm{III};2} &= -4K_{\mathrm{II}}\Gamma_{312} = (R_{13} - C^{(2)})K^1\\
&\quad + R_{33}K^3 - C^{(3)}K^4,\\
2K_{\mathrm{III};3} &= -4K_{\mathrm{II}}\Gamma_{313} = C^{(1)}K^1 - R_{33}K^2\\
&\quad + (C^{(2)} + R_{13})K^4,\\
2K_{\mathrm{III};4} &= -4K_{\mathrm{II}}\Gamma_{314} = C^{(3)}K^2 - (C^{(2)} + R_{13})K^3.
\end{aligned} \tag{5.5}$$

These are the direct generalization (with corrected numerical factors) of those in Ref. 1. A set of "Maxwell equations" for the KBV, quite similar to (2.6), can be obtained from (5.3)-(5.5): These are

$$K_{\mathrm{II},1} - \tfrac{1}{2}R_{12}K^2 + \tfrac{1}{2}R_{13}K^3 = -2K_{\mathrm{II}}\Gamma_{314}, \tag{5.6a}$$

$$K_{\mathrm{II},2} + \tfrac{1}{2}R_{12}K^1 + \tfrac{1}{2}R_{23}K^3 = -2K_{\mathrm{II}}\Gamma_{423}, \tag{5.6b}$$

$$K_{\mathrm{II},3} - \tfrac{1}{2}R_{13}K^1 - \tfrac{1}{2}(R_{23} + R_{33})K^3 - \tfrac{1}{2}R_{34}K^4 = 2K_{\mathrm{II}}\Gamma_{312}, \tag{5.6c}$$

$$K_{\mathrm{II},4} + \tfrac{1}{2}R_{34}K^3 = 2K_{\mathrm{II}}\Gamma_{421}. \tag{5.6d}$$

For the electromagnetic field (5.1), the Maxwell equations $d\mathfrak{F}^{(-)} = 0$ yield

$$0 = F_{\mathrm{II},1} - 2F_{\mathrm{II}}\Gamma_{314} - F_{\mathrm{III},4} + F_{\mathrm{III}}(\Gamma_{421} - \Gamma_{124} - \Gamma_{344}), \tag{5.7a}$$

$$0 = F_{\mathrm{II},2} + 2F_{\mathrm{II}}\Gamma_{423} + F_{\mathrm{III}}\Gamma_{422}, \tag{5.7b}$$

$$0 = F_{\mathrm{II},3} - 2F_{\mathrm{II}}\Gamma_{312} + F_{\mathrm{III},2} + F_{\mathrm{III}}(\Gamma_{423} + \Gamma_{342} + \Gamma_{122}), \tag{5.7c}$$

$$0 = F_{\mathrm{II},4} - 2F_{\mathrm{II}}\Gamma_{421} + F_{\mathrm{III}}\Gamma_{424}. \tag{5.7d}$$

The condition that $\mathbf{e}_4$ be *geodesic* ($\Gamma_{424} = 0$) reduces the last of (5.3) to

$$C^{(5)}K^2 - C^{(4)}K^3 = 0. \tag{5.8}$$

The additional requirement that $\mathbf{K}$ be hypersurface orthogonal involves only two cases:

$$\mathrm{Im}(K_{\mathrm{II}}) = 0 \quad \text{and} \quad K^1 = K^2 = 0, \tag{5.9a}$$

$$\mathrm{Re}(K_{\mathrm{II}}) = 0 \quad \text{and} \quad K^3 = K^4 = 0 \tag{5.9b}$$

[see, for example, the derivation of (5.9a) and (5.9b) in Ref. 1]. Hence, these together imply (5.8) is satisfied by $C^{(5)}K^2 = C^{(4)}K^3 = 0$. The case (5.9a) is that for which $\mathbf{K}\cdot\mathbf{K} = 2K^3K^4$ and can be spacelike or timelike. On the other hand (5.9b) is that for which $\mathbf{K}\cdot\mathbf{K} = 2K^1K^2 > 0$ and can only be spacelike.

For the remainder of this section only (5.9a) is examined. This case includes all *static* electromagnetic examples of the spaces underlying this section. Then $\mathbf{K} = K^3\mathbf{e}_3 + K^4\mathbf{e}_4$ and $2K_{\mathrm{II}} = K_{3;4} \neq 0$.

The geodesic condition (5.8) implies $C^{(4)} = 0$. Then (5.3) gives $\Gamma_{423} = \Gamma_{413} = 0$. From (5.4), $K_{\mathrm{II},2} = 0 = K_{\mathrm{II},1}$. Hence (5.3) then implies $C^{(2)} = 0$. Since $K_{\mathrm{II},4}$ is real, (5.4) implies $C^{(3)} = \overline{C^{(3)}}$. Hence $\Gamma_{421} = \Gamma_{412}$ from (5.3); therefore, $\mathbf{e}_4$ is hypersurface orthogonal. (Algebraically special cases here fall into the Robinson-Trautman class.[7]) In addition (5.5) gives $\Gamma_{321} = \Gamma_{312}$ so that $\mathbf{e}_3$ is also hypersurface orthogonal. Since $K^1 = K_{\mathrm{II},2} = \Gamma_{423} = 0$, Eq. (5.6b) implies $R_{23} = 0$. But $T_{23} = \pi^{-1}\overline{F_{\mathrm{III}}}F_{\mathrm{II}}$. Hence, with F nonnull this dictates $F_{\mathrm{III}} = 0$ for consistency; and so $\mathbf{e}_3$ *is also a principal null direction for* F. Since $T_{33} = \pi^{-1}F_{\mathrm{III}}\overline{F_{\mathrm{III}}}$, $R_{33} = 0$. The energy-momentum tensor then has the general form

$$(T_{ab}) = (4\pi)^{-1}\begin{pmatrix} 0 & |F_{II}|^2 & 0 & 0 \\ |F_{II}|^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -|F_{II}|^2 \\ 0 & 0 & -|F_{II}|^2 & 0 \end{pmatrix}. \tag{5.10}$$

Equation (5.6a) gives $\Gamma_{314} = 0$ and (5.5) gives $\Gamma_{313} = 0$. In summary we have the following list of relations:

(i) Both $\mathbf{e}_3$ and $\mathbf{e}_4$ are geodesic and hypersurface orthogonal principal null directions for dK and F;
(ii) $C^{(4)} = C^{(2)} = 0$;
(iii) $K_{II,1} = K_{II,2} = 0 = \Gamma_{423} = \Gamma_{314}$;
(iv) $R_{34} = 2|F_{II}|^2 = -R_{12}$, with other $R_{ab} = 0$.

Maxwell equations (5.7a)–(5.7d) are now

$$F_{II,1} = F_{II,2} = 0, \tag{5.11a}$$

$$F_{II,3} = 2F_{II}\Gamma_{312}, \tag{5.11b}$$

$$F_{II,4} = 2F_{II}\Gamma_{421}. \tag{5.11c}$$

Killing's equations ($K_{a;b} + K_{b;a} = 0$) simplify to

$$\begin{aligned} K_{1;1} = 0 &\Rightarrow K^3\Gamma_{311} + K^4\Gamma_{411} = 0, \\ K_{3;3} = 0 &\Rightarrow K^4{}_{,3} + K^4\Gamma_{343} = 0, \\ K_{4;4} = 0 &\Rightarrow K^3{}_{,4} + K^3\Gamma_{344} = 0, \\ K_{1;2} = 0 &\Rightarrow K^3\Gamma_{312} + K^4\Gamma_{412} = 0, \\ K_{1;3} = K_{3;1} = 0 &\Rightarrow K^4{}_{,1} + K^4\Gamma_{341} = 0 = K^4\Gamma_{413}, \\ K_{2;4} = K_{4;2} = 0 &\Rightarrow K^3{}_{,2} - K^3\Gamma_{342} = 0 = K^4\Gamma_{324}, \\ K_{3;4} = 2K_{II} &\Rightarrow K^4{}_{,4} + K^4\Gamma_{344} = 2K_{II}, \\ K_{4;3} = -2K_{II} &\Rightarrow K^3{}_{,3} - K^3\Gamma_{343} = -2K_{II}. \end{aligned} \tag{5.12}$$

Equations (5.3)–(5.6) become

$$\begin{aligned} 2K_{II}\Gamma_{421} &= -\tfrac{1}{2}C^{(3)}K^3, \\ 2K_{II}\Gamma_{422} &= -\tfrac{1}{2}C^{(5)}K^4, \\ 2K_{II,3} &= [C^{(3)} + 2|F_{II}|^2]K^4, \\ 2K_{II,4} &= -[C^{(3)} + 2|F_{II}|^2]K^3, \\ 2K_{II}\Gamma_{311} &= \tfrac{1}{2}C^{(1)}K^3, \\ 2K_{II}\Gamma_{312} &= \tfrac{1}{2}C^{(3)}K^4. \end{aligned} \tag{5.13}$$

At this stage no transformations of the (allowed) type (1.5) have been utilized to simplify expressions. Furthermore the field equations, or "structure equations", for the gravitational field have not been imposed.

The following notation is introduced: $z \equiv \Gamma_{421}$, $x \equiv \Gamma_{312}$, $\sigma \equiv \Gamma_{422}$, and $\tau \equiv \Gamma_{311}$. (No similarity to names of Newman–Penrose spin coefficients is intended.) The structure equations may be written as

$$\begin{aligned} d(z\epsilon^1 + \sigma\epsilon^2) + 2(z\epsilon^1 + \sigma\epsilon^2) \wedge (\Gamma_{IIa}\epsilon^a) \\ = \tfrac{1}{2}[C^{(5)}\epsilon^4 \wedge \epsilon^2 + C^{(3)}\epsilon^3 \wedge \epsilon^1], \end{aligned} \tag{5.14}$$

$$\begin{aligned} d(\Gamma_{IIm}\epsilon^m) + (z\epsilon^1 + \sigma\epsilon^2) \wedge (\tau\epsilon^1 + x\epsilon^2) \\ = \tfrac{1}{2}[C^{(3)}(\epsilon^1 \wedge \epsilon^2 + \epsilon^3 \wedge \epsilon^4) \\ + R_{34}(\epsilon^2 \wedge \epsilon^1 + \epsilon^3 \wedge \epsilon^4) + R_{23}\epsilon^3 \wedge \epsilon^2], \end{aligned} \tag{5.15}$$

$$\begin{aligned} d(\tau\epsilon^1 + x\epsilon^2) + (2\Gamma_{IIm}\epsilon^m) \wedge (\tau\epsilon^1 + x\epsilon^2) \\ = \tfrac{1}{2}[C^{(3)}\epsilon^4 \wedge \epsilon^2 + C^{(1)}\epsilon^3 \wedge \epsilon^1 \\ + R_{13}(\epsilon^2 \wedge \epsilon^1 + \epsilon^3 \wedge \epsilon^4) - R_{33}\epsilon^3 \wedge \epsilon^2]. \end{aligned} \tag{5.16}$$

Equation (5.14) reduces to

$$\begin{aligned} -z_{,2} + z\Gamma_{342} + \sigma_{,1} - 2\sigma\Gamma_{121} - \sigma\Gamma_{341} &= 0, \\ z_{,3} - z(x + \Gamma_{343}) - \sigma\tau &= \tfrac{1}{2}C^{(3)}, \\ z_{,4} - z^2 - \sigma\bar{\sigma} - z\Gamma_{344} &= 0, \\ \sigma_{,4} - \sigma(2z + 2\Gamma_{124} - \Gamma_{344}) &= \tfrac{1}{2}C^{(5)}, \\ \sigma_{,3} - \sigma(x + 2\Gamma_{123} + \Gamma_{343}) - z\bar{\tau} &= 0. \end{aligned} \tag{5.17}$$

Equation (5.15) reduces to

$$\begin{aligned} -\Gamma_{II1,2} + \Gamma_{II2,1} - \Gamma_{II1}\Gamma_{122} - \Gamma_{II2}\Gamma_{121} \\ + zx - \sigma\tau = \tfrac{1}{2}[C^{(3)} - R_{34}], \\ -\Gamma_{II3,1} + \Gamma_{II1,3} + \Gamma_{II1}(\Gamma_{123} - x) \\ - \Gamma_{II2}\tau - \Gamma_{II3}\Gamma_{341} = 0, \\ -\Gamma_{II4,1} + \Gamma_{II1,4} + \Gamma_{II1}(\Gamma_{124} - z) \\ - \Gamma_{II2}\bar{\sigma} + \Gamma_{II4}\Gamma_{341} = 0, \\ -\Gamma_{II3,2} + \Gamma_{II2,3} - \Gamma_{II1}\bar{\tau} - \Gamma_{II2}(\Gamma_{123} + x) \\ - \Gamma_{II3}\Gamma_{342} = R_{23} = 0, \\ -\Gamma_{343,4} + \Gamma_{344,3} - 2\Gamma_{343}\Gamma_{344} = [C^{(3)} + R_{34}], \\ -\Gamma_{123,4} + \Gamma_{124,3} - \Gamma_{124}\Gamma_{343} - \Gamma_{123}\Gamma_{344} = 0, \end{aligned} \tag{5.18}$$

where $\Gamma_{IIa} \equiv \frac{1}{2}(\Gamma_{12a} + \Gamma_{34a})$. Finally, Eq. (5.16) reduces to

$$\begin{aligned} -\tau_{,2} + x_{,1} - \tau(\Gamma_{342} + 2\Gamma_{122}) + x\Gamma_{341} &= -\tfrac{1}{2}R_{13} = 0, \\ \tau_{,3} + \tau(\Gamma_{343} + 2\Gamma_{123} - 2x) &= \tfrac{1}{2}C^{(1)}, \\ \tau_{,4} + \tau(\Gamma_{344} + 2\Gamma_{124} - z) - \bar{\sigma}x &= 0, \\ x_{,3} + x(\Gamma_{343} - x) - \tau\bar{\tau} &= -\tfrac{1}{2}R_{33} = 0, \\ x_{,4} + x(\Gamma_{344} - z) - \sigma\tau &= \tfrac{1}{2}C^{(3)}. \end{aligned} \tag{5.19}$$

At this point one may transform K^3 (for example) to a constant with the function A of (1.5) so that $K^3{}_{,i} = 0$ in Killing's equations (5.12). In particular this provides for $\Gamma_{344} = \Gamma_{341} = 0$. Hence

$$\begin{aligned} &2K_{II} = K^3\Gamma_{343}, \quad 2K_{II} = K^4{}_{,4}, \quad K^4{}_{,3} = -K^4\Gamma_{343} \\ &K^4{}_{,1} = K^4{}_{,2} = 0, \quad \text{and} \quad xK^3 + zK^4 = 0. \end{aligned} \tag{5.20}$$

Furthermore Γ_{124} and Γ_{123} may be transformed to zero by the function B of (1.5). Consequently Eqs. (5.17)–(5.19) become

$$\begin{aligned} z_{,2} - \sigma_{,1} + 2\sigma\Gamma_{121} &= 0, \\ z_{,3} - z\Gamma_{343} &= x_{,4}, \\ z_{,4} - z^2 - \sigma\bar{\sigma} &= 0, \\ \sigma_{,3} - \sigma(x + \Gamma_{343}) - z\bar{\tau} &= 0, \\ \sigma_{,4} - 2\sigma z &= \tfrac{1}{2}C^{(5)}, \\ -\Gamma_{121,2} + \Gamma_{122,1} - 2\Gamma_{122}\Gamma_{121} \\ + 2zx - 2\sigma\tau &= C^{(3)} - R_{34}, \\ \Gamma_{121,3} + \Gamma_{343,1} - x\Gamma_{121} - \tau\Gamma_{122} &= 0, \end{aligned}$$

$$
\begin{aligned}
\Gamma_{121,4} - z\Gamma_{121} - \bar{\sigma}\Gamma_{122} &= 0, \\
\Gamma_{122,3} + \Gamma_{343,2} - x\Gamma_{122} - \bar{\tau}\Gamma_{121} &= 0, \\
-\Gamma_{343,4} &= C^{(3)} + R_{34}, \\
\Gamma_{343,1} = \Gamma_{343,2} &= 0, \\
\tau_{,2} - x_{,1} + 2\tau\Gamma_{122} &= 0, \\
\tau_{,3} + \tau(\Gamma_{343} - 2x) &= \tfrac{1}{2}C^{(1)}, \\
\tau_{,4} - \tau z - \bar{\sigma}x &= 0, \\
x_{,3} + x(\Gamma_{343} - x) - \tau\bar{\tau} &= 0, \\
x_{,4} - xz - \sigma\tau &= \tfrac{1}{2}C^{(3)} \\
&(\Rightarrow \sigma\tau \text{ is real}).
\end{aligned}
\tag{5.21}
$$

Consider the case where $z = 0$. Then $x = 0$, and vice-versa from (5.20). Furthermore $\sigma = \tau = 0$. Hence $C^{(i)} = 0$ for $i = 1, 2, 3, 4, 5$. The space is therefore *conformally flat*, with F_{II} (and therefore F) being constant. The metric of this space is included in the work of Cahen and Leroy[8] in their study of the most general conformally flat Einstein-Maxwell spaces with a nonnull electromagnetic field.

With $z \neq 0$ assumed, the results of this section are summarized in the following lemma.

Lemma 2: Let $\mathbf{K} = K^a\mathbf{e}_a$ be a Killing vector field in an Einstein-Maxwell space where $F = F_{ab}\epsilon^a \wedge \epsilon^b$ is the electromagnetic bivector. Furthermore, let

(i) $-dK = K_{a;b}\epsilon^a \wedge \epsilon^b$ be the nonnull KBV,
(ii) F be nonnull,
(iii) $\mathbf{e}_3$ be a principal null direction for dK,
(iv) $\mathbf{e}_4$ be a principal null direction for both dK and F,
(v) $\mathbf{e}_4$ be geodesic,
(vi) $\mathbf{K}$ be hypersurface orthogonal with $\mathrm{Im}(K_{II}) = 0$.

Then

(a) $\mathbf{e}_3$ is also geodesic and is a principal null direction for F;
(b) $\mathbf{e}_3$ is shear-free $\Leftrightarrow$ $\mathbf{e}_4$ is shear-free (from $K_{2;2} = 0$);
(c) $\mathbf{e}_3$ and $\mathbf{e}_4$ are hypersurface orthogonal;
(d) $\Gamma_{423} = \Gamma_{413} = \Gamma_{314} = \Gamma_{324} = 0$;
(e) $K_{II,1} = K_{II,2} = 0 = F_{II,1} = F_{II,2}$;
(f) $C^{(4)} = C^{(2)} = 0$;
(g) $\arg(F_{II}) = \text{const}$.

It should be noted that as long as $C^{(5)}C^{(1)} \neq 9C^{(3)2}$, this lemma applies to algebraically general cases.

6. ALGEBRAICALLY SPECIAL STATIC CASES

From an invariant viewpoint, the algebraically special cases from Sec. 5 above are those for which either (1) $\mathbf{e}_3$ or $\mathbf{e}_4$ is geodesic and shearfree or (2) $C^{(5)}C^{(1)} = 9C^{(3)2}$. Here we propose to examine the case (1): $\sigma = 0$ and $z \neq 0$.

By part (b) of the lemma above, $\mathbf{e}_3$ *and* $\mathbf{e}_4$ are geodesic and shear-free; hence $\sigma = \tau = 0$ and the space is Petrov Type D. The structure equations (5.21) yield

$$z_{,1} = z_{,2} = 0, \tag{6.1a}$$

$$z_{,3} = z\Gamma_{343} + xz + \tfrac{1}{2}C^{(3)}, \tag{6.1b}$$

$$z_{,4} = z^2, \tag{6.1c}$$

$$x_{,1} = x_{,2} = 0, \tag{6.1d}$$

$$x_{,3} = x^2 - x\Gamma_{343}, \tag{6.1e}$$

$$x_{,4} = xz + \tfrac{1}{2}C^{(3)}, \tag{6.1f}$$

$$\Gamma_{343,1} = \Gamma_{343,2} = 0, \tag{6.1g}$$

$$-\Gamma_{343,4} = C^{(3)} + 2|F_{II}|^2, \tag{6.1h}$$

$$-\Gamma_{121,2} + \Gamma_{122,1} - 2\Gamma_{122}\Gamma_{121} + 2zx = C^{(3)} - 2|F_{II}|^2, \tag{6.1i}$$

$$\Gamma_{121,3} = x\Gamma_{121}, \tag{6.1j}$$

$$\Gamma_{121,4} = z\Gamma_{121}. \tag{6.1k}$$

Consider first the relation from (5.12) stating that $K^3x + K^4z = 0$. Since K^3 has been transformed to a constant, it is possible to make that constanty unity, so that $-K^4z = x$. Define $\mathcal{K} \equiv -K^4$. Then

$$x = \mathcal{K}z, \tag{6.2}$$

where $\mathcal{K}$ is real. Furthermore Killing's equations (5.12) imply

$$\mathcal{K}_{,1} = 0 = \mathcal{K}_{,2},$$

$$-\mathcal{K}_{,4} = 2K_{II} = \Gamma_{343},$$

$$-\mathcal{K}_{,3} = \mathcal{K}\Gamma_{343}.$$

In particular note that $(\log\mathcal{K})_{,3} = \mathcal{K}_{,4}$ is implied. From (6.1e) we have

$$(\log x)_{,3} = x - \Gamma_{343} = x + \mathcal{K}_{,4}.$$

Furthermore

$$(\log x)_{,3} = (\log\mathcal{K}z)_{,3} = (\log\mathcal{K})_{,3} + (\log z)_{,3} = \mathcal{K}_{,4} + (\log z)_{,3}.$$

Consequently

$$(\log z)_{,3} = x \quad \Rightarrow \quad z_{,3} = xz. \tag{6.3}$$

In particular this implies

$$z\Gamma_{343} = -\tfrac{1}{2}C^{(3)} = 2zK_{II}. \tag{6.4}$$

Next consider Eqs. (5.11a)-(5.11c). These have the general solution

$$F_{II} = \alpha z^2, \tag{6.5}$$

where α is a complex constant. Therefore $|F_{II}|^2 = \alpha\bar{\alpha}z^4$. Differentiating (6.4) and using (6.1g) with (6.1h), one obtains

$$C^{(3)}{}_{,4} - 3zC^{(3)} = 4\alpha\bar{\alpha}z^5. \tag{6.6}$$

The general solution to (6.6) is

$$C^{(3)} = (2C + 4\alpha\bar{\alpha}z)z^3, \tag{6.7}$$

where $C_{,1} = C_{,2} = C_{,4} = 0$. Furthermore, differentiation of (6.7) yields

$$C^{(3)}{}_{,3} = C_{,3}z^3 + 3Cxz^3 + 16\alpha\bar{\alpha}xz^4.$$

The relation $\pounds_{\mathbf{K}} C^{(3)} = 0$ yields $K^3 C^{(3)}{}_{,3} + K^4 C^{(3)}{}_{,4} = 0$. Hence $C^{(3)}{}_{,3} = \mathcal{K}C^{(3)}{}_{,4} = \mathcal{K}z^4[6C + 16\alpha\bar{\alpha}z]$. Putting all this together results in $C_{,3} = 0$ since $z \neq 0$. Therefore C is a *constant* in the expression (6.7).

The expression for the function $\mathcal{K}$ introduced in (6.2) is now computed. Since $\mathcal{K}_{,4} = \Gamma_{343}$, Eq. (6.4) gives

$$\mathcal{K}_{,4} = Cz^2 + 2\alpha\bar{\alpha}z^3. \tag{6.8}$$

Equation (6.8) has the general solution (since $\mathcal{K}_{,1} = \mathcal{K}_{,2} = 0$)

$$\mathcal{K} = Cz + \alpha\bar{\alpha}z^2 + \mathcal{K}_0, \tag{6.9}$$

where $\mathcal{K}_{0,1} = \mathcal{K}_{0,2} = \mathcal{K}_{0,4} = 0$. Since $\mathcal{K}_{,3} = \mathcal{K}\mathcal{K}_{,4}$ we get, after detailed calculation, $\mathcal{K}_{0,3} = 0$. Hence $\mathcal{K}_0$ is *constant*.

The last field equation (6.1i) now becomes

$$-\Gamma_{121,2} + \Gamma_{122,1} - 2\Gamma_{121}\Gamma_{122} = -2\mathcal{K}_0 z^2. \tag{6.10}$$

It can be shown that $\mathcal{K}_0$ in (6.9) and (6.10) is an essential constant and is transformed away if and only if $\Gamma_{121} = 0$. Now it is clear that $d\Gamma_{42} \wedge \Gamma_{42} = 0$, so that there exist functions φ and ζ for which

$$z\epsilon^1 = \Gamma_{42} = e^{\varphi}d\zeta. \tag{6.11}$$

If we compute $d\epsilon^1 = \Gamma_{2ab}\epsilon^a \wedge \epsilon^b$, we obtain $\bar{\varphi}_{,1} = -\Gamma_{121}$ and $\varphi_{,2} = \Gamma_{122}$ with $\varphi_{,4} = \varphi_{,3} = 0$. Equation (6.10) then becomes

$$\varphi_{,21} + \bar{\varphi}_{,12} + 2\varphi_{,2}\bar{\varphi}_{,1} = -2\mathcal{K}_0 z^2.$$

Hence, if we define $\mathrm{Re}(\varphi) \equiv p$ and $\mathrm{Im}(\varphi) \equiv q$, then

$$\varphi_{,21} + \bar{\varphi}_{,12} = p_{,12} + p_{,21} - i(q_{,12} - q_{,21}).$$

Furthermore

$$2\varphi_{,2}\bar{\varphi}_{,1} = 2(p_{,1}p_{,2} + q_{,1}q_{,2}) - 2i(p_{,2}q_{,1} - p_{,1}q_{,2}).$$

Therefore (6.10) now takes the form

$$\begin{aligned} -2\mathcal{K}_0 z^2 &= p_{,12} + p_{,21} + 2(p_{,1}p_{,2} + q_{,1}q_{,2}) \\ &\quad - i(q_{,12} - q_{,21} + 2p_{,2}q_{,1} - 2p_{,1}q_{,2}). \end{aligned} \tag{6.12}$$

Commutation relations on $\mathbf{e}_1$ and $\mathbf{e}_2$ give

$$p_{,12} - p_{,21} = p_{,2}\bar{\varphi}_{,1} - p_{,1}\varphi_{,2} = -i(p_{,1}q_{,2} + p_{,2}q_{,1}),$$

$$q_{,12} - q_{,21} = q_{,2}\bar{\varphi}_{,1} - q_{,1}\varphi_{,2} = q_{,2}p_{,1} - q_{,1}p_{,2} - 2iq_{,1}q_{,2}.$$

Together with (6.12) these give

$$-\mathcal{K}_0 z^2 = p_{,12} + p_{,1}p_{,2} + i(p_{,1}q_{,2}).$$

At this stage it is possible, by letting $i\mathrm{B} = -iq$ in the transformation (1.5), to transform q to zero since $q_{,3} = q_{,4} = 0$. Equation (6.12) and other calculations simplify, giving $p_{,12} = p_{,21}$ and

$$p_{,12} + p_{,1}p_{,2} = -2\mathcal{K}_0 z^2. \tag{6.13}$$

In addition to the Bianchi identities all being identically satisfied at this point, the following fact is easy to show: $[\mathbf{K}, \mathbf{e}_a] = 0$ for $a = 1, 2, 3, 4$. This means, in particular, that the curve parameter along $\mathbf{K}$ could serve as a coordinate, if needed. We choose coordinates, in fact, in the following manner. Let $r \equiv -1/z$ and define the real coordinate σ through $\epsilon^3 = d\sigma$, since $d\epsilon^3 = \Gamma_{4ab}\epsilon^a \wedge \epsilon^b = 0$. Then $dr = \mathcal{K}\epsilon^3 + \epsilon^4$ and

$$\begin{aligned} \epsilon^1 &= z^{-1}e^p d\zeta = -re^p d\zeta, \\ \epsilon^2 &= z^{-1}e^p d\bar{\zeta} = -re^p d\bar{\zeta}, \\ \epsilon^3 &= d\sigma, \\ \epsilon^4 &= dr - \mathcal{K}d\sigma; \end{aligned} \tag{6.14}$$

i.e., ζ and $\bar{\zeta}$ form the other two (complex) coordinates and $p = p(\zeta, \bar{\zeta})$. The corresponding contravariant basis is given by

$$\begin{aligned} \mathbf{e}_1 &= -r^{-1}e^{-p}\partial_\zeta, \quad \mathbf{e}_2 = -r^{-1}e^{-p}\partial_{\bar{\zeta}}, \\ \mathbf{e}_3 &= \partial_\sigma - \mathcal{K}\partial_r, \quad \mathbf{e}_4 = \partial_r. \end{aligned} \tag{6.15}$$

The metric $g_{ab}\epsilon^a\epsilon^b$ becomes

$$ds^2 = 2r^2 e^{2p}d\zeta d\bar{\zeta} + 2drd\sigma - 2\mathcal{K}d\sigma^2, \tag{6.16}$$

where $\mathcal{K} = -(c/r) + (\alpha\bar{\alpha}/r^2) + \mathcal{K}_0$. The electromagnetic field bivector is

$$F = F_{\mathrm{II}}\epsilon^{\mathrm{II}} + \overline{F_{\mathrm{II}}}\epsilon^{\mathrm{V}} = 2F_{34}\epsilon^3 \wedge \epsilon^4 = (4\alpha/r^2)d\sigma \wedge dr. \tag{6.17}$$

Note here that all the spaces in the class (6.16) have the same electromagnetic field no matter what the function e^{2p}.

By virtue of Eq. (6.13) the function $p = p(\zeta, \bar{\zeta})$ above obeys

$$e^{-2p}p_{,\zeta\bar{\zeta}} = -\mathcal{K}_0. \tag{6.18}$$

This equation occurs frequently in studies of algebraically special cases and expresses the curvature $(-\mathcal{K}_0)$ of the two-dimensional space $d\tau^2 = e^{2p}d\zeta d\bar{\zeta}$. Equation (6.18) implies also that $p(\zeta, \bar{\zeta})$ satisfies $(p_{,\zeta})^2 - p_{,\zeta\zeta} = G(\zeta)$, where G is an analytic function; i.e.,

$$(e^{-p})_{,\zeta\zeta} = e^{-p(\zeta,\bar{\zeta})}G(\zeta). \tag{6.19}$$

A second Killing vector field $\mathbf{L} = \eta\partial_\zeta + \bar{\eta}\partial_{\bar{\zeta}}$ satisfying $[\mathbf{K}, \mathbf{L}] = 0$ as well as $dL \wedge L = 0$ is possible if the following coupling with p is satisfied:

$$(e^{2p}\eta)_\zeta + (e^{2p}\bar{\eta})_{\bar{\zeta}} = 0, \tag{6.20}$$

where η is an analytic function of ζ only. The vector $\mathbf{L}$ is spacelike and it can be shown that neither $\mathbf{e}_3$ nor $\mathbf{e}_4$ is a principal null direction of the Killing bivector dL. Hence one can deduce that these properties together with $\mathbf{K}$ define an *axially symmetric* space. More detailed analysis of examples fitting into this class have been given at various times in the literature (see for example Witten[9]; Robinson and Trautman[7]).

1 G. Debney, J. Math. Phys. **12**, 1088 (1971).
2 R. K. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).
3 J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. **22**, 13 (1962).
4 I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1963).
5 G. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
6 L. Mas, Compt. Rend., Sér. A **270**, 837 (1970).
7 I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962).
8 M. Cahen and J. Leroy, J. Math. Mech. **16**, 501 (1966).
9 L. Witten, in *Gravitation* (Wiley, New York, 1962), Chap. 9.

Form Factors for Any Spin and Charge Coupling

D. H. Tchrakian
School of Mathematical and Physical Sciences, University of Sussex, Falmer, Brighton, England

(Received 5 April 1971)

A prescription for decomposition into form factors of the matrix elements of totally symmetric tensor currents between $(s, 0)$ and $(s', 0)$ spinor states of the incoming and outgoing particles of spins s and s' is presented. The case of the electromagnetic current is studied in some detail, and a procedure to modify the above prescription resulting in the grouping of the form factors into Dirac-type and Pauli-type classes is given. Several examples are used for illustration.

1. INTRODUCTION

The importance of the decomposition of a three-point scattering function into form factors has been discussed by several authors,[1-7] who give various different methods of effecting these decompositions for processes with on-mass-shell particles (i.e., particles between whose states the matrix element of the *current* is taken) of arbitrary pins, and with off-mass-shell particles (i.e., bosons which couple to the current) of both special[1] ($j = 0, 1$) and general spins[2-8]. These different methods can be classified under two distinct categories, which both have their analogues in the decomposition of four-point function decompositions. One is the canonical way of decomposing the Lorentz-covariant spinor function[5,9] associated with the process into invariant functions which have only dynamical singularities in the invariants of the process. The methods given by Refs. (4-8) belong to this category. The other category, to which the methods of Refs. 1-3 belong, is the direct utilization of the physical scattering function of the process, in particular, the helicity amplitudes.[10] From now on we shall concern ourselves with the first category exclusively. One advantage that this covariant approach has is that constraint conditions, such as Lorentz conditions and gauge conditions, that are expressed in a manifestly covariant form, can be incorporated into this framework directly. In fact, in this article, we are concerned eventually with precisely these considerations.

In the first category, Ref. 4 utilizes spinor functions that are constructed from Rarita-Schwinger spinor states, which possess decompositions into a number of form factors that is larger than the number of couplings obtained for the corresponding physical process by use of the Clebsch-Gordan coefficients (CGC) of the Poincaré group and the Wigner-Eckart theorem, presented in Ref. 5. The number of these invariant functions can be reduced to the number of independent physical couplings by projection of the spinor states onto physical scattering states that results in "redundancy conditions," which are rather complicated in these cases. Moreover, when more than one such spinor state is involved, the use of other spinor identities[4] is necessary. This is an indirect method and involves unnecessary complication.

References 5 and 6 give the vertex functions related to the matrix elements of the current such that the on-mass-shell particles of spins s and s' are represented by spinor states transforming according to the irreducible representations (IR) $(s, 0)$ and $(s', 0)$, respectively, of the homogeneous Lorentz group (HLG), and the boson of spin j that couples to the current by the IR of HLG $(j, 0)$.

Reference 7 gives the vertex functions where the three particles are represented by spinor states transforming according to the IR's (A_1, B_1): $A_1 \otimes B_1 \supset s$, (A_2, B_2): $A_2 \otimes B_2 \supset s'$, and (A, B): $A \otimes B \supset j$ of HLG, respectively. The distinguishing feature of this method is that in this way it can be arranged to have vertex functions without explicit factors of four-momenta appearing. In Refs. 5 and 6 four-momenta do feature explicitly.

Reference 8 gives form factor expansions of matrix elements of totally symmetric tensor currents which transform according to the representations $(\frac{1}{2}, \frac{1}{2}) \times \cdots$ (e.g., the vector and tensor currents), taken between spinor states relating to the sequence of IR's $(s, 0)$, $(s - \frac{1}{2}, \frac{1}{2}) \cdots, (\frac{1}{2}, s - \frac{1}{2}), (0, s)$ of HLG, where s is the greater of the spins of these two particles. These are the Fierz-Bargmann-Wigner spinor states.

It is clear that for the on-mass-shell particles, by far the best spinor states are those belonging to the IR of $(s, 0)$ of HLG, since these give rise to no redundant components. For the current (off-mass-shell particle), however, it may be advantageous to use spinors of more complicated representations which lend themselves naturally to the constraints of some dynamical model for the interactions. Examples of this situation are the weak-interaction currents (with the formulation of PCAC, etc.) and the electromagnetic currents, which couple to the vector potential[11] $A_\mu(x)$ and which motivate the choices of construction made in this paper.

We give a prescription for decomposing the matrix element of a current that can couple to a particle of spin-j, belonging to the symmetric tensor representation $(\frac{1}{2}, \frac{1}{2})_1 \otimes (\frac{1}{2}, \frac{1}{2})_2 \otimes \cdots \otimes (\frac{1}{2}, \frac{1}{2})_j$ of HLG taken between the spinor states $\{p(s', 0)\alpha|$ and $|k(s, 0)\beta\}$ of the on-mass-shell particles whose free-particle states belong to the IR's $[m', s']$ and $[m, s]$ of the Poincaré group. In this way we incorporate the simplicity of representing the on-shell particles with the advantage of having a tensorial current. This method is in fact the direct generalization of the method used to decompose the matrix element be-

1 G. Debney, J. Math. Phys. **12**, 1088 (1971).
2 R. K. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).
3 J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. **22**, 13 (1962).
4 I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1963).
5 G. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
6 L. Mas, Compt. Rend., Sér. A **270**, 837 (1970).
7 I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962).
8 M. Cahen and J. Leroy, J. Math. Mech. **16**, 501 (1966).
9 L. Witten, in *Gravitation* (Wiley, New York, 1962), Chap. 9.

Form Factors for Any Spin and Charge Coupling

D. H. Tchrakian
School of Mathematical and Physical Sciences, University of Sussex, Falmer, Brighton, England

(Received 5 April 1971)

A prescription for decomposition into form factors of the matrix elements of totally symmetric tensor currents between $(s, 0)$ and $(s', 0)$ spinor states of the incoming and outgoing particles of spins s and s' is presented. The case of the electromagnetic current is studied in some detail, and a procedure to modify the above prescription resulting in the grouping of the form factors into Dirac-type and Pauli-type classes is given. Several examples are used for illustration.

1. INTRODUCTION

The importance of the decomposition of a three-point scattering function into form factors has been discussed by several authors,[1-7] who give various different methods of effecting these decompositions for processes with on-mass-shell particles (i.e., particles between whose states the matrix element of the *current* is taken) of arbitrary pins, and with off-mass-shell particles (i.e., bosons which couple to the current) of both special[1] ($j = 0, 1$) and general spins[2-8]. These different methods can be classified under two distinct categories, which both have their analogues in the decomposition of four-point function decompositions. One is the canonical way of decomposing the Lorentz-covariant spinor function[5,9] associated with the process into invariant functions which have only dynamical singularities in the invariants of the process. The methods given by Refs. (4-8) belong to this category. The other category, to which the methods of Refs. 1-3 belong, is the direct utilization of the physical scattering function of the process, in particular, the helicity amplitudes.[10] From now on we shall concern ourselves with the first category exclusively. One advantage that this covariant approach has is that constraint conditions, such as Lorentz conditions and gauge conditions, that are expressed in a manifestly covariant form, can be incorporated into this framework directly. In fact, in this article, we are concerned eventually with precisely these considerations.

In the first category, Ref. 4 utilizes spinor functions that are constructed from Rarita-Schwinger spinor states, which possess decompositions into a number of form factors that is larger than the number of couplings obtained for the corresponding physical process by use of the Clebsch-Gordan coefficients (CGC) of the Poincaré group and the Wigner-Eckart theorem, presented in Ref. 5. The number of these invariant functions can be reduced to the number of independent physical couplings by projection of the spinor states onto physical scattering states that results in "redundancy conditions," which are rather complicated in these cases. Moreover, when more than one such spinor state is involved, the use of other spinor identities[4] is necessary. This is an indirect method and involves unnecessary complication.

References 5 and 6 give the vertex functions related to the matrix elements of the current such that the on-mass-shell particles of spins s and s' are represented by spinor states transforming according to the irreducible representations (IR) $(s, 0)$ and $(s', 0)$, respectively, of the homogeneous Lorentz group (HLG), and the boson of spin j that couples to the current by the IR of HLG $(j, 0)$.

Reference 7 gives the vertex functions where the three particles are represented by spinor states transforming according to the IR's (A_1, B_1): $A_1 \otimes B_1 \supset s$, (A_2, B_2): $A_2 \otimes B_2 \supset s'$, and (A, B): $A \otimes B \supset j$ of HLG, respectively. The distinguishing feature of this method is that in this way it can be arranged to have vertex functions without explicit factors of four-momenta appearing. In Refs. 5 and 6 four-momenta do feature explicitly.

Reference 8 gives form factor expansions of matrix elements of totally symmetric tensor currents which transform according to the representations $(\frac{1}{2}, \frac{1}{2}) \times \cdots$ (e.g., the vector and tensor currents), taken between spinor states relating to the sequence of IR's $(s, 0)$, $(s - \frac{1}{2}, \frac{1}{2}) \cdots, (\frac{1}{2}, s - \frac{1}{2}), (0, s)$ of HLG, where s is the greater of the spins of these two particles. These are the Fierz-Bargmann-Wigner spinor states.

It is clear that for the on-mass-shell particles, by far the best spinor states are those belonging to the IR of $(s, 0)$ of HLG, since these give rise to no redundant components. For the current (off-mass-shell particle), however, it may be advantageous to use spinors of more complicated representations which lend themselves naturally to the constraints of some dynamical model for the interactions. Examples of this situation are the weak-interaction currents (with the formulation of PCAC, etc.) and the electromagnetic currents, which couple to the vector potential[11] $A_\mu(x)$ and which motivate the choices of construction made in this paper.

We give a prescription for decomposing the matrix element of a current that can couple to a particle of spin-j, belonging to the symmetric tensor representation $(\frac{1}{2}, \frac{1}{2})_1 \otimes (\frac{1}{2}, \frac{1}{2})_2 \otimes \cdots \otimes (\frac{1}{2}, \frac{1}{2})_j$ of HLG taken between the spinor states $\{p(s', 0)\alpha|$ and $|k(s, 0)\beta\}$ of the on-mass-shell particles whose free-particle states belong to the IR's $[m', s']$ and $[m, s]$ of the Poincaré group. In this way we incorporate the simplicity of representing the on-shell particles with the advantage of having a tensorial current. This method is in fact the direct generalization of the method used to decompose the matrix element be-

tween two spin-$\frac{1}{2}$ particles of the electromagnetic current[12]:

$$\{p(\tfrac{1}{2},0)\alpha\,|\,J^{\mu}(\tfrac{1}{2},\tfrac{1}{2})\,|\,k(\tfrac{1}{2},0)^{*}\dot{b}\} = F_D(t)(M_1^{(+)\mu})_{a\dot{b}} + F_P(t)(N^{(+)\mu})_{a\dot{b}}, \quad (1)$$

with

$$M_1^{(\pm)\mu} = \sigma^{\mu} \pm p\tilde{\sigma}^{\mu}k, \quad M_2^{(\pm)\mu} = (k+p)P^{\mu},$$

$$N^{(+)\mu} = [p(\bar{\sigma}^{\mu}\sigma^{\nu} - \bar{\sigma}^{\nu}\sigma^{\mu}) + (\sigma^{\mu}\bar{\sigma}^{\nu} - \sigma^{\nu}\bar{\sigma}^{\mu})k]q_{\nu}; \quad (2)$$

$$P = p + k, \quad q = p - k.$$

The notation

$$k^{i}_{a\dot{b}} = \frac{k^{i}_{\mu}\sigma^{\mu}_{a\dot{b}}}{m_i}, \quad \tilde{k}^{\dot{a}b}_{i} = \frac{k^{\mu}_{i}\tilde{\sigma}^{\dot{a}b}_{\mu}}{m_i}, \quad \sigma_{\mu} = (\sigma_0, \sigma), \quad \text{and } \tilde{\sigma}_{\mu} = (\sigma_0, -\sigma),$$

where σ are the Pauli spin matrices.

The basis $N^{\mu}_{a\dot{b}}$ can be related to $M_1^{(+)}$ and $M_2^{(+)}$ linearly, through use of the Dirac equation, and the superscripts $^{(\pm)}$ signify vector and axial vector character. The equivalence of this two-component basis to the usual four-component basis given in terms of Dirac γ matrices is given in Refs. 13 and 14. We choose to use the $(2s+1)$-component $(s,0)$ representations throughout and not the $2(2s+1)$-component $[(s,0)\oplus(0,s)]$, because the former has some technical advantages in general cases, especially when $s \neq s'$.

This prescription is presented in Sec. 2, and some examples are given. In Sec. 3, the special case of the electromagnetic current is discussed in some detail, with special emphasis on the physical circumstance that this current actually couples to a vector potential $A_{\mu}(x)$ and not the electromagnetic field.[11] As an example, the matrix element of the electromagnetic current between two spin-1 particles is decomposed into form factors that can accommodate the above physical properties of the interaction, to illustrate the fact that the method given in Sec. 2 is amenable to this kind of procedure. In fact the procedure followed in Sec. 3 can serve as a prescription for decomposing the matrix element of the electromagnetic current between the spinor states of a particle of arbitrary spin.

2. FORM FACTORS FOR jth-RANK TENSOR CURRENTS

The jth-rank tensor currents we shall consider are those that couple to spin-j particle fields belonging to the IR $(\frac{1}{2},\frac{1}{2})$ of HLG which are themselves jth-rank *symmetric* tensors, satisfying the generalized Lorentz conditions

$$\partial_{\mu_n} A^{\mu_1\cdots\mu_n\cdots\mu_j}(x) = 0; \quad n = 1,\cdots,j, \quad (3)$$

from which it follows, that all bases in the decomposition which contain q^{μ}, the four-momentum of the tensor particle, will not contribute to the physical matrix element and hence can be neglected, and that the current is also a *symmetric* tensor.

This symmetry of the current and the Lorentz condition (3) will be the *only* constraints we shall need in our construction below.

A. The Spin Matrices

In this subsection we shall give the definitions, according to Ref. 13, of the spin matrices which are the direct generalizations of the matrices σ^{μ} used to construct the covariant bases in (2). Unlike σ^{μ}, however, which transforms according to the IR $(\frac{1}{2},\frac{1}{2}) \sim (\frac{1}{2},0)\otimes(\frac{1}{2},0)^{*}$ of the HLG and is a 2×2 matrix, spin matrices in the general case, are not necessarily square, that is, they transform according to the IR (s',s) of HLG, where $s \neq s'$. In the special case where $s = s'$, such matrices have also been defined and used in Ref. 14 to construct projection operators for propagators, i.e., covariant basis for two-point function, arising from the use of local fields with HLG transformation $(s,0)$. Since the spin matrices are so well discussed, we shall limit ourselves here simply to defining them and then exposing the symmetry properties of a few special cases which we shall need in the examples, later in this section.

A spin matrix is called a spin-j matrix if it transforms according to the IR (s',s) of HLG, where $\max(s',s) = j$. Such a matrix can be constructed from $2j$ spin-$\frac{1}{2}$ matrices $\sigma_{\mu_1},\cdots,\sigma_{\mu_{2j}}$ with the aid of the CGC of $SU(2)$. Clearly for $j > 1$ there are several modes of making the couplings, and this must be specified for each such matrix, since the symmetry properties depend on it.

From the point of view of presentation, the clearest way of defining these matrices is to build the spin-1 matrices from spin-$\frac{1}{2}$ matrices first, and then define the spin-$\frac{3}{2}$ matrices in terms of the latter, thus prescribing the construction of a spin-j matrix by means of spin-$\frac{1}{2}$ and spin-$(j-\frac{1}{2})$ matrices, etc., always remembering to record the modes of coupling, except in cases where $s' = s = j$, where the latter is unique.

Thus we write the spin-1 matrices:

$$\rho^{\mu_1\mu_2}(s',s)_{\alpha\dot{\beta}} = \rho^{\mu_1}(\tfrac{1}{2},\tfrac{1}{2})_{a\dot{b}}\rho^{\mu_2}(\tfrac{1}{2},\tfrac{1}{2})_{a',\dot{b}'}[\tfrac{1}{2}\tfrac{1}{2}s']^{aa'}_{\alpha}[\tfrac{1}{2}\tfrac{1}{2}s]^{\dot{b}\dot{b}'}_{\dot{\beta}}, \quad (4)$$

where $[\frac{1}{2}\frac{1}{2}1]^{aa'}_{\alpha}$ is a CGC of $SU(2)$, $\rho^{\mu}(\frac{1}{2},\frac{1}{2}) = (1/\sqrt{2})\sigma^{\mu}$, and repeated indices are summed over. s' and s can take on the values 0 and 1. The verification that (3) has the transformation of an IR (s',s) of HLG is performed in Ref. 13 by use of the orthogonality relations of the CGC's.

We next write the spin-$\frac{3}{2}$ matrices by means of (3):

$$\rho^{\mu_1\mu_2\mu_3}[(s',s):(L'L)]_{A\dot{B}} = \rho^{\mu_1}(\tfrac{1}{2},\tfrac{1}{2})_{a\dot{b}}\rho^{\mu_2\mu_3}(L',L)_{\alpha\dot{\beta}}[\tfrac{1}{2}L's']^{a\alpha}_{A}[\tfrac{1}{2}Ls]^{\dot{b}\dot{\beta}}_{\dot{B}}, \quad (4')$$

where s' and s can take on the values $\frac{1}{2}$ and $\frac{3}{2}$. It is now clear how to construct a (j',j) matrix for any j.

The symmetries that these matrices may possess may be obtained in the general case, by using the symmetry properties of the constituent spin matrix, e.g., the symmetry of $\rho^{\mu_2\mu_3}(L',L)$ in (4') and by further using the properties of the CG and recoupling (or Racah) coefficients of $SU(2)$. In the basic case of $\rho^{\mu_1\mu_2}(s',s)$, the symmetry properties can be derived from the properties of the CGC's and some basic spinor identities.[13]

Thus, from the properties of CGC's under the interchange of two indices, the following useful property

of the spin-1 matrix follows:

$$\rho^{\mu_1\mu_2}(1,1) = \rho^{\mu_2\mu_1}(1,1), \tag{5}$$

while using[13] $\sigma^{\mu}_{a\dot{b}}\sigma_{\mu\dot{a}'b'} = 2C^{-1}_{aa'}C^{-1}_{\dot{b}\dot{b}'}$ we find that

$$g_{\mu_1\mu_2}\rho^{\mu_1\mu_2}(1,1) = 0. \tag{6}$$

Another useful property of spin-1 matrices follows from the identity $\sigma_\mu\tilde{\sigma}_\nu = g_{\mu\nu} + \frac{1}{2}i\epsilon_{\mu\nu\tau\lambda}\sigma^\tau\tilde{\sigma}^\lambda$:

$$\rho^{\mu_1\mu_2}(1,0) = \tfrac{1}{2}i\epsilon^{\mu_1\mu_2\nu_1\nu_2}\rho_{\nu_1\nu_2}(1,0), \tag{7}$$

from which it is also apparent that $\rho^{\mu\nu}(1,0)$ and $\rho^{\mu\nu}(0,1)$ are antisymmetric and traceless with respect to their space–time indices μ and ν.

Using (5)–(7) as well as properties of CG and recoupling coefficients we can find the symmetry properties of the next higher spin matrices, namely spin-$\frac{3}{2}$ matrices. In this way, for $s' = s = \frac{3}{2}$ and $L' = L$ in (4′), we find, using the appropriate recoupling coefficient, that

$$\rho^{\mu_1\mu_2\mu_3}(\tfrac{3}{2},\tfrac{3}{2}) = \rho^{\mu_2\mu_1\mu_3}(\tfrac{3}{2},\tfrac{3}{2}), \tag{8}$$

but according to (4), $\rho^{\mu_1\mu_2\mu_3}(\frac{3}{2},\frac{3}{2})$ is symmetric in the interchange of μ_2 and μ_3 and therefore it is symmetric in all its space–time indices. It follows from (6) and (8) that it is also traceless with respect to any two of its space–time indices. In fact symmetry and tracelessness is a well-known[14] property of all square spin matrices, i.e., $s = s' = j$, which can be seen by carrying out the above procedure to arbitrary j.

Let us now consider the case $s' = \frac{3}{2}s = \frac{1}{2}$ in (4). This can be achieved through both $L' = 1, L = 0$ and $L' = 1, L = 1$; and the two matrices are related as follows:

$$\begin{aligned}
&\tfrac{\sqrt{3}}{2}\rho^{\mu_1\mu_2\mu_3}[(\tfrac{3}{2},\tfrac{1}{2}):(1,0)] = \rho^{\mu_3\mu_2\mu_1}[(\tfrac{3}{2},\tfrac{1}{2}):(1,1)] \\
&\quad -\tfrac{1}{2}\rho^{\mu_1\mu_2\mu_3}[\tfrac{3}{2},\tfrac{1}{2}):(1,1)], \\
&\tfrac{\sqrt{3}}{2}\rho^{\mu_1\mu_2\mu_3}[(\tfrac{3}{2},\tfrac{1}{2}):(1,1)] = \rho^{\mu_3\mu_2\mu_1}[\tfrac{3}{2},\tfrac{1}{2}):(1,0) \\
&\quad +\rho^{\mu_1\mu_2\mu_3}[(\tfrac{3}{2},\tfrac{1}{2}):(1,0)]
\end{aligned} \tag{9}$$

by use of the appropriate recoupling coefficients. It is clear from (9) that one is free to choose between $(L', L) = (1,0)$ and $(1,1)$. As in our example, we shall choose to use $(L', L) = (1,0)$; we shall not need (9) from now on. Finally, we give the following useful symmetry property of $\rho^{\mu_1\mu_2\mu_3}[(\frac{3}{2},\frac{1}{2}):(1,0)]$, obtained like (8) was, but by use of (7) and its consequences:

$$\begin{aligned}
\rho^{\mu_1\overline{\mu_2\mu_3}}[(\tfrac{3}{2},\tfrac{1}{2}):(1,0)] &= \rho^{\mu_2\overline{\mu_1\mu_3}}[(\tfrac{3}{2},\tfrac{1}{2}):(1,0)] \\
&\quad + \rho^{\mu_3\overline{\mu_2\mu_1}}[(\tfrac{3}{2},\tfrac{1}{2}):(1,0)],
\end{aligned} \tag{10}$$

where the notation $\overline{\mu\nu}$ serves to remind us of the additional property that the interchange of μ and ν is antisymmetric. It is, of course, easy to find the relation analogous to (10) for $\rho[(\frac{3}{2},\frac{1}{2}):(1,1)]$.

B. The Form Factors

In this subsection we present our prescription to decompose the matrix element of a tensor current of order j between $(s',0)$ and $(s,0)^*$ spinor states corresponding to the on-mass-shell particles whose free-particle states are labeled according to the IR's $[m', s']$ and $[m, s]$ of the Poincaré group. The currents are tensors in the sense that they transform according to

$$U(\Lambda)J^{\mu_1\cdots\mu_j}(x)U^{-1}(\Lambda) = J^{\nu_1\cdots\nu_j}(\Lambda x)\Lambda^{\mu_1}_{\nu_1}\cdots\Lambda^{\mu_j}_{\nu_j}. \tag{11}$$

Furthermore, these tensor currents are symmetric in all their space–time indices, that is, they couple to spin-j tensor fields of rank j.

We adopt the following notation:

$$\{p(s',0)\alpha\,|J_{\mu_1\cdots\mu_j}(0)\,|k(s,0)\dot{\beta}\} = M_{\mu_1\cdots\mu_j}(P,q)_{\alpha\dot{\beta}}, \tag{12}$$

with the decomposition (suppressing tensor and spinor indices)

$$\begin{aligned}
M(P,q) = \sum_{i=1}^{m} F_i^{(+)}(q^2)M_i^{(+)}(P,q) \\
+ \sum_{i=m+1}^{n} F_i^{(-)}(q^2)M_i^{(-)}(P,q),
\end{aligned} \tag{13}$$

where $F_i^{(+)}(q^2)$ are the m form factors corresponding to the positive-parity signature couplings and $F_i^{(-)}(q^2)$ the $(n-m)$ form factors for the negative-parity signature couplings. The $M_i^{(\pm)}(P,q)$ are the covariant bases of both parity signatures, respectively, and have the same transformation properties under HLG as $M(P,q)$. Our object is now to construct the set of independent bases $M_i^{(\pm)}$ out of the two independent four-momenta P and q of the process which are equal in number to the number of physical couplings of the corresponding vertex. The only constraints to be used are those given by (3), in addition of course to those imposed by the symmetry properties of the spin matrices that are used in the construction.

Before we actually carry out this construction, we must indicate how to proceed, such that the bases we obtain should be endowed with definite-parity signature.

It follows from the transformation properties of the $(s,0)$ spinor state[5] that any basis in the decomposition of (12) will undergo the following transformation under space reflection:

$$M^i_{(\mu)}(P,q)_{\alpha\dot{\beta}} \xrightarrow{\pi} \overline{\Pi}^{(s')}(p)^{\dot{\alpha}'\alpha}M^i_{(\mu)}(P,q)_{\alpha\dot{\beta}}\overline{\Pi}^{(s)}(k)^{\dot{\beta}\beta'}, \tag{14}$$

where $\Pi^{(s)}(k) = \rho^{\mu_1\cdots\mu_{2s}}(s,s)k_{\mu_1}\cdots k_{\mu_{2s}}$ as in Ref. (14), and $\overline{\Pi}(k) \equiv \Pi(\bar{k})$, where $\bar{k} = (k_0, -\mathbf{k})$ the space-reflected four-momentum. Thus, by replacing all four-momenta K by their space-reflected counterparts $\bar{K}$ in the right-hand side of (14), the parity conjugate of each basis can be found, and by adding (subtracting) it to (from) the original basis the positive (negative) signature combination can be obtained. The space-reflection character of the positive- (negative-) signature combination is that of a j th-order tensor (axial tensor) characterized by the space–time indices $(\mu) = (\mu_1\cdots\mu_j)$.

All that need be done now is the construction of the bases $M^i_{(\mu)}(P,q)_{\alpha\dot{\beta}}$, out of the four-momenta P, q and the (s', s) spin matrices, where the on-mass-shell particle spins s' and s $(s' > s)$, according to the following prescription:

(a) $s' > j$: Contract r of the space-time indices of the spin matrix with all permutations of the four-momenta P and q, giving all independent rank-$(2s'-r)$ tensor covariants, having taken into account the symmetry properties of the spin matrix. Then multiply (in direct product) by r' factors of the four-vector P, where r' is such that $(2s'-r)+r'=j$, thus obtaining all independent rank-j tensor covariants. The covariants arising from the multiplication in direct product of the four-vector q have been omitted since their contributions vanish under the constraint of (3). Repeat the process from $r = 2s'-j, \cdots, m$ to $r = 2s'$.

(a') $s' < j$: In this case the above process should be performed $r = j - 2s', \cdots, 2s'$.

(b) The rank-j tensor bases obtained in (a) or (a') should then be symmetrized in all space-time indices $\mu_1 \cdots \mu_j$. The bases can then be obtained in their final form accommodated by the decomposition (13) by (c).

(c) Casting all bases $M^{i}_{(\mu)}(P,q)_{\alpha\dot\beta}$ into parity-definite$^{(\pm)}$ form (combinations) with the help of (14) and the discussion following it.

We shall now illustrate the above prescription by some examples, (i)–(iii), which cover all cases $s' \gtrless j$ and $s' \geqslant s$ and hopefully are also of some physical interest. The notation we use is the one of (1) and (2), in which the incoming (outgoing) on-mass-shell particle has mass m (m'), spin s (s') and four-momentum k (p), with $P = p + k$ and $q = p - k$. In addition, we shall in the following use the shorthand

$$\begin{aligned} &\rho^{\mu_1 \cdots \mu_{2s'}}(s',s) = (\mu_1 \cdots \mu_{2s'})^{(s'-s)}, \\ &x_1^{\mu_1} x_2^{\mu_2} \cdots x_r^{\mu_r} \rho_{\mu_1\mu_2\cdots\mu_r}{}^{\mu_{r+1}\cdots\mu_{2s'}}(s',s) \\ &\quad = (x_1 x_2 \cdots x_r \mu_{r+1} \cdots \mu_{2s'})^{(s',s)} \end{aligned} \tag{15}$$

where x_i is the four-momentum of one of the on-mass-shell particles, divided by its mass. In this notation $\Pi^{(s)}(p) = (p,p,\cdots,p)^{(s,s)}$.

(i) $j = 1$, $s' = s = \frac{1}{2}$: According to (a), we can construct from P and q, or more conveniently, from k and p and the $(\frac{1}{2},\frac{1}{2})$ spin matrices the following bases:

$$(\mu)^{(\frac{1}{2},\frac{1}{2})}, (k)^{(\frac{1}{2},\frac{1}{2})}P^\mu, (p)^{(\frac{1}{2},\frac{1}{2})}P^\mu,$$

and by the use of (c) we find the bases of both parity signatures $\sigma_\pi = \pm 1$:

$$\begin{aligned} M_1^{(\pm)\mu} &= (\mu) \pm (p)(\tilde\mu)(k), \\ M^{(\pm)\mu} &= [(k) \pm (p)]P^\mu. \end{aligned} \tag{16}$$

The $\sigma_\pi = +1$ members of (16) are in agreement with the same members in (2). As it happens, there are in fact two $\sigma_\pi = +1$ and two $\sigma_\pi = -1$ couplings for this vertex.

(ii) $j = 1$, $s' = 1$, $s = 0$: There are two $\sigma_\pi = -1$ and one $\sigma_\pi = +1$ couplings for this vertex.

According to (a) we find the following bases:

$$(\mu k)^{(1,0)}, \quad (\mu p)^{(1,0)}, \quad (kp)^{(1,0)}P^\mu,$$

and by applying (c) we get the following σ_π-definite combinations

$$\begin{aligned} M_1^{(-)\mu} &= (\mu p)^{(1,0)}, \\ M_2^{(-)\mu} &= (kp)^{(1,0)}P^\mu, \\ M^{(+)\mu} &= (\mu k)^{(1,0)} + \frac{k\cdot p}{mm'}(\mu p)^{(1,0)} + \frac{P^\mu}{m'}(kp)^{(1,0)} \\ &= (\mu k)^{(1,0)} + \frac{k\cdot p}{mm'}M_1^{(-)\mu} + \frac{1}{2m'}M_2^{(-)\mu}, \end{aligned} \tag{17}$$

where it is clear that we have had to use the condition (3) once more on the third member ($\sigma_\pi = +1$). The contractions implied in (14) have been carried through by use of the orthogonality conditions on the CGC.

(iii) $j = 1$, $s' = s = 1$: There are four $\sigma_\pi = +1$ and three $\sigma_\pi = -1$ couplings for this vertex:

(a) $(\mu k)^{(1,1)}$, $(\mu p)^{(1,1)}$, $(kk)^{(1,1)}P^\mu$, $(pp)^{(1,1)}P^\mu$, $(kp)^{(1,1)}P^\mu$.

(c) At this stage, we have used the following useful identity:

$$(pp)^{(1,1)}(\widetilde{\mu\nu})^{(1,1)}(kk)^{(1,1)} = (p\tilde\mu k, p\tilde\nu k)^{(1,1)}, \tag{18}$$

where $(\widetilde{\mu,\nu})$ is just the spin matrix constructed from the spin matrices $\tilde\sigma^\mu$ and $\tilde\sigma^\nu$, and the right-hand side of (18) means

$$(p\tilde\mu k, p\tilde\nu k)_{\alpha\dot\beta} = (p\tilde\sigma^\mu k)_{a\dot b}(p\tilde\sigma^\nu k)_{a'\dot b'}[\tfrac{1}{2}\tfrac{1}{2}1]^{aa'}{}_{\alpha}[\tfrac{1}{2}\tfrac{1}{2}1]^{\dot b\dot b'}_{\dot\beta}. \tag{19}$$

In fact (18) is the special case ($s = 1$) of the more general identity

$$\begin{aligned} &(pp\cdots p)^{(s,s)}(\widetilde{\mu_1\mu_2\cdots\mu_{2s}})^{(s,s)}(kk\cdots k)^{(s,s)} \\ &\qquad = (p\tilde\mu_1 k, p\tilde\mu_2 k, \cdots, p\tilde\mu_{2s} k)^{(s,s)}, \end{aligned} \tag{20}$$

with a similar notation for the right-hand side of (20) as that of (18) [cf. (19)].

Using (14) and (18), we obtain the following σ_π — definite combinations

$$\begin{aligned} M_1^{(\pm)\mu} &= (\mu k)^{(1,1)} \pm (p\tilde\mu k, p)^{(1,1)}, \\ M_2^{(\pm)\mu} &= (\mu p)^{(1,1)} \pm (p\tilde\mu k, k)^{(1,1)}, \\ M_3^{(\pm)\mu} &= [(kk)^{(1,1)} \pm (pp)^{(1,1)}]P^\mu, \\ M_4^{(\pm)\mu} &= (kp)^{(1,1)}P^\mu. \end{aligned} \tag{21}$$

(iv) $j = 1$, $s' = \frac{3}{2}$, $s = \frac{1}{2}$: There are three $\sigma_\pi = +1$ and three $\sigma_\pi = -1$ couplings for this vertex:

(a) $(k\overline{\mu k})^{(\frac{3}{2},\frac{1}{2})}$, $(k\overline{\mu p})^{(\frac{3}{2},\frac{1}{2})}$, $(p\overline{kp})^{(\frac{3}{2},\frac{1}{2})}P^\mu$, $(p\overline{\mu k})^{(\frac{3}{2},\frac{1}{2})}$, $(p\overline{\mu p})^{(\frac{3}{2},\frac{1}{2})}$, $(k\overline{kp})^{(\frac{3}{2},\frac{1}{2})}P^\mu$,

where the designation of the antisymmetry used in (10) is used here. In fact the above bases have been constructed subject to the restriction of this antisymmetry, as well as of (10), which eliminates the basis $(\mu\overline{kp})$.

(c) The operation of (14) is carried out by using the orthogonality conditions of the CGC in a very similar manner as in (ii) yielding the following definite σ_π contributions:

$$M_1^{(\pm)\mu} = [(k\overline{\mu p})^{(\frac{3}{2},\frac{1}{2})} \mp (p\overline{\mu p})^{(\frac{3}{2},\frac{1}{2})}],$$

$$M_2^{(\pm)\mu} = [(k\bar{kp})^{(\frac{3}{2},\frac{1}{2})} \mp (p\bar{kp})^{(\frac{3}{2},\frac{1}{2})}]P^\mu, \tag{22}$$

$$M_3^{(\pm)\mu} = [(k\bar{\mu k})^{(\frac{3}{2},\frac{1}{2})} + (p\bar{\mu k})^{(\frac{3}{2},\frac{1}{2})}] + \frac{k\cdot p}{mm'}M_1^{(\mp)\mu} + \frac{1}{2m'}M_2^{(\mp)\mu}.$$

(v) $j = 1$, $s' = s = \frac{3}{2}$: There are five $\sigma_\pi = +1$ and five $\sigma_\pi = -1$ couplings for this vertex:

(a)

$$(\mu kk)^{(\frac{3}{2},\frac{3}{2})},\quad (\mu kp)^{(\frac{3}{2},\frac{3}{2})},\quad (kkk)^{(\frac{3}{2},\frac{3}{2})}P^\mu,\quad (kpp)^{(\frac{3}{2},\frac{3}{2})}P^\mu,$$
$$(\mu pp)^{(\frac{3}{2},\frac{3}{2})},\quad (ppp)^{(\frac{3}{2},\frac{3}{2})}P^\mu,\quad (pkk)^{(\frac{3}{2},\frac{3}{2})}P^\mu,$$

which are all the independent bases, satisfying (3), and the conditions coming from the symmetry of square spin matrices in all their space–time indices.

(c) Using (20) for $s = \frac{3}{2}$ in the application of (14), we get the following definite σ_π bases:

$$\begin{aligned}
M_1^{(\pm)\mu} &= (\mu kk)^{(\frac{3}{2},\frac{3}{2})} \pm (p\tilde{\mu}k,p,p)^{(\frac{3}{2},\frac{3}{2})},\\
M_2^{(\pm)\mu} &= (\mu pp)^{(\frac{3}{2},\frac{3}{2})} \pm (p\tilde{\mu}k,k,k)^{(\frac{3}{2},\frac{3}{2})},\\
M_3^{(\pm)\mu} &= (\mu kp)^{(\frac{3}{2},\frac{3}{2})} \pm (p\tilde{\mu}k,p,k)^{(\frac{3}{2},\frac{3}{2})},\\
M_4^{(\pm)\mu} &= [(kkk)^{(\frac{3}{2},\frac{3}{2})} \pm (ppp)^{(\frac{3}{2},\frac{3}{2})}]P^\mu,\\
M_5^{(\pm)\mu} &= [(kpp)^{(\frac{3}{2},\frac{3}{2})} \pm (pkk)^{(\frac{3}{2},\frac{3}{2})}]P^\mu.
\end{aligned} \tag{23}$$

(vi) $j = 2$, $s' = s = \frac{1}{2}$ ($j > 2s$): There are two $\sigma_\pi = +1$ and two $\sigma_\pi = -1$ couplings for this vertex.

(a)

$$(\mu)^{(\frac{1}{2},\frac{1}{2})}P^\nu,\quad (\nu)^{(\frac{1}{2},\frac{1}{2})}P^\mu,\quad (k)^{(\frac{1}{2},\frac{1}{2})}P^\mu P^\nu,\quad (p)^{(\frac{1}{2},\frac{1}{2})}P^\mu P^\nu,$$

which all satisfy (3). We next apply symmetrization in the indices $\mu\nu$ as this current must couple to a spin-2 particle for which there is a second-rank symmetric tensor field.

(b) $(\mu)P^\nu + (\nu)P^\mu,\quad (k)P^\mu P^\nu,\quad (p)P^\mu P^\nu$

and (c) finally we obtain the σ_π-definite bases:

$$\begin{aligned}
M_1^{(\pm)\mu} &= [(\mu) \pm (p)(\tilde{\mu})(k)]P^\nu + [(\nu) \pm (p)(\tilde{\nu})(k)]P^\mu,\\
M_2^{(\pm)\mu} &= [(k) \pm (p)]P^\mu P^\nu.
\end{aligned} \tag{24}$$

(vii) $j = 2$, $s = s = 1$: There are five $\sigma_\pi = +1$ and four $\sigma_\pi = -1$ couplings for this vertex.

(a)–(b)

$$(\mu\nu)^{(1,1)},\ (\mu k)^{(1,1)}P^\nu + (\nu k)^{(1,1)}P^\mu,\ (kk)P^\mu P^\nu,\ (kp)P^\mu P^\nu,$$
$$(\mu p)^{(1,1)}P^\nu + (\nu p)^{(1,1)}P^\mu,\ (pp)P^\mu P^\nu,$$

(c) with the σ_π-definite bases

$$\begin{aligned}
M_1^{(\pm)} &= (\mu\nu)^{(1,1)} \pm (p\tilde{\mu}k,p\tilde{\nu}k)^{(1,1)},\\
M_2^{(\pm)} &= [(\mu k)^{(1,1)} \pm (p\tilde{\mu}k,p)^{(1,1)}]P^\nu + [(\nu k)^{(1,1)} \pm (p\tilde{\nu}k,p)^{(1,1)}]P^\mu,\\
M_3^{(\pm)} &= [(\mu p)^{(1,1)} \pm (p\tilde{\mu}k,k)^{(1,1)}]P^\nu + [(\nu p)^{(1,1)} \pm (p\tilde{\nu}k,k)^{(1,1)}]P^\mu,\\
M_4^{(+)} &= [(kk) \pm (pp)]P^\mu P^\nu,\\
M_5^{(+)} &= (kp)P^\mu P^\nu.
\end{aligned} \tag{25}$$

It should be mentioned that the form factors occurring in the decompositions of (i)–(vii) are not the couplings occurring in the Clebsch-Gordan decomposition of the Poincaré group for the respective processes, although they are equal in number to these. A particular case for the comparison of such coupling functions (of q^2) labeled by $s \otimes s'$ and the orbital angular momentum l, with the form factors given by the method of the present article, for the same process ($j = 1$, $s = s' = \frac{1}{2}$), is to be found in Ref. 15.

3. ELECTROMAGNETIC INTERACTION

This is the special case of $j = 1$. The method of decomposition described in Sec. 2 does not use anything but the covariance property (under HLG) of the spinor matrix element, and the assumptions that it is analytic in the components of the four-momenta and possesses a decomposition into invariant functions (form factors) that are holomorphic in the scalar variable of the process. These assumptions are expected to accommodate all types of interactions, but that does not mean that the ensuing decompositions would be suitable for describing any interaction, for it may be the case that some interaction satisfies the above assumptions only under certain conditions governed by certain constraints. A case in point is the electromagnetic interaction, where the mediating quantum is not the photon field $F^{\mu\nu}(x)$ but the vector potential $A^\mu(x)$, which unlike the field, does not transform according to any IR of HLG. However, its use, or in other words the use of the electromagnetic current coupling to it, does not violate the assumption of covariance made in Sec. 2, provided the current satisfies the following constraint derived in Ref. 11:

$$\partial_\mu J^\mu(x) = 0, \tag{26}$$

in which case the coupling $J^\mu(x)A_\mu(x)$ is guaranteed to be a Lorentz scalar. In a purely S-matrix theoretic framework, the condition (26) is replaced by

$$q_\mu M^\mu(P,q) = 0, \quad q^2 = 0. \tag{27}$$

The derivation of (27) is given in Ref. 16, and its main ingredients (as in the derivation of the noncovariance of $A_\mu(x)$ and hence (26)) are the properties of the representation function of the little group of the Poincaré group for massless particles. Equations (26) and (27) comprise the massless particle gauge conditions.

In order to be able to enforce (27) in a convenient framework, we have chosen to write the matrix elements of currents of Sec. 2 in tensorial form. We have, in particular, avoided currents with $(j,0)$ HLG transformation properties because these can couple only to $(j,0)$ photon fields (related to $F_{\mu\nu}$) and not to the vector potential A_μ.

The reason that we must insist on the coupling to the vector potential is that such an interaction, as opposed to one mediated by a (1, 0) photon field, does not vanish in the limit of the four-momentum of the phonon, q_μ, vanishing. This property is desirable because it accommodates the Coulomb interaction. It follows from this argument that our approach in Sec. 2 of using tensorial rather than $(j,0)$ currents is compulsory only if the particle it couples to is mass-

less and mediates an interaction which has a long-range effect in the limit $q_\mu \to 0$, and not if it is a *free* particle, massless or massive, in which case, currents transforming according to any IR (A, B) of HLG can be used, subject in the massless case to the condition (3.8) of Ref. 17. It also follows that our approach is specially useful only for $j = 1$ and 2 (photons and gravitons).

We now proceed to give a prescription for $j = 1$, which, although consistent with that given in Sec. 2, modifies the decomposition of the current to satisfy the additional features of photonic interactions described above.

Needless to say, the masses of the on-shell particles will be taken to be equal, $m = m'$, in the following, since otherwise we would be excluding the crucially important possibility of $q_\mu \to 0$.

Our prescription is a very simple one. In Sec. 2 we obtained decompositions with respect to bases which can be divided into two distinct classes: those with the space–time index carrying the vector character belonging to the spin matrix and those with the vector index appearing on P^μ. Both of these are capable of contracting with a vector *field* to give an invariant vertex function. In the event, however, of the current coupling to a vector potential $A_\mu(x)$ related to the photon field $F_{\mu\nu}(x)$ through

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x), \tag{28}$$

we must satisfy the requirements discussed above, that is, current conservation and the appearance of a Coulomb term at $q_\mu \to 0$. In quantum electrodynamics ($j = 1$, $s' = s = \frac{1}{2}$), which will be our guideline, this is assured by using the principle of minimal coupling, whence it follows that the charge coupling is associated with the Dirac form factor. The Pauli form factor, on the other hand, does not contribute in this limit because its basis involves a factor of the photon four-momentum q_ν contracted with an antisymmetric spin matrix $\sigma^{\mu\nu}$, which in other words means that the current in this case is really coupled to $F_{\mu\nu}(x)$, the photon field (28).

The central objective of our prescription will then be the division of the bases into two classes, which give rise to Dirac- and Pauli-type form factors, respectively, as distinct from the two classes of bases mentioned above. This can be done essentially by constructing generalized spin matrices $\Sigma^{\mu\nu}$ that are antisymmetric in the indices μ and ν, but before doing that, we shall review the electron–photon situation, which will serve as a guideline.

The bases of expansion in this case are given by the $\sigma_\pi = +1$ members of (16). It can be checked, remembering $q \cdot P = 0$ for $m' = m$, that the constraint (27) is satisfied in this case. It is quite clear, however, that this situation is not accommodated by the principle of minimal coupling, because in the limit $q_\mu \to 0$, both $F_1^{(+)}$ and $F_2^{(+)}$ contribute, while only the Dirac form factor $F_D(t)$, which is the coefficient of $M_1^{(+)}$, should contribute. This situation can be rectified by the well-known artifice of defining a new basis, which includes a factor of q_ν, contracted to the spin matrix, such that in this limit it does not contribute, and such that the spin matrix $\Sigma^{\mu\nu}(\frac{1}{2}, \frac{1}{2})$ is antisymmetric in μ and ν, so that when contracted with $A(x)$ it gives rise to a coupling with the photon field $F_{\mu\nu}(x)$. In other words, we obtain the Pauli form factor F_P. Now this basis is related to $M_1^{(+)}$ and $M_2^{(+)}$ by

$$N^{(+)} = [(p)\tilde{\Sigma}^{\mu\nu} + \Sigma^{\mu\nu}(k)]q_\nu = -4mM_1^{(+)} + 2M_2^{(+)},$$
$$\Sigma^{\mu\nu}_{(\frac{1}{2})} = (\sigma^\mu\tilde{\sigma}^\nu - \sigma^\nu\tilde{\sigma}^\mu). \tag{29}$$

which permits us to write $M_2^{(+)}$ in terms of $N^{(+)}$ and $M_1^{(+)}$, thus redefining the form factors, namely obtaining $F_D(t)$ and $F_P(t)$ [cf. Eq. (2)]. We must note that (29) has the particularly convenient feature that $M_2^{(+)}$, the basis we desire to replace, does not have a kinematic coefficient, which would have imposed a kinematic constraint (in particular, a zero) on the redefined form factors. This is an important point to keep in mind when extending this procedure to higher spins.

The new decomposition with respect to $M_1^{(+)}$ and $N^{(+)}$ also satisfies the condition (27); what is more, in the $q_\mu \to 0$ limit, $N^{(+)}$ vanishes and so only $F_D(t)$ contributes, as required. [Considering the $\sigma_\pi = -1$ (axial vector) members of (16), and then proceeding as above to replace $M_2^{(-)}$ in terms of $N^{(-)} = (p\tilde{\Sigma}^{\mu\nu} - \tilde{\Sigma}^{\mu\nu}k)_{q\nu}$, we would find that condition (27) would force the coefficient of $M_1^{(-)}$ to vanish, so that in the limit $|\mathbf{q}| \to 0$ there would be no long-range interaction for the axial vector current.[16]]

We now proceed to give the spin matrices $\Sigma^{\mu\nu}_{(s)}$ for the arbitrary s case ($s' = s$). $\Sigma^{\mu\nu}_{(s)}{}^{\beta}_{\alpha}$ is obtained from the following matrix:

$$\rho^{\mu\mu_2\cdots\mu_{2s}}(s, s)_{\alpha\dot{\gamma}}\tilde{\rho}^{\nu\nu_2\cdots\nu_{2s}}(s, s)^{\dot{\gamma}\beta} - \rho^{\nu\mu_2\cdots\mu_{2s}}(s, s)_{\alpha\dot{\gamma}} \times \tilde{\rho}^{\mu\nu_2\cdots\nu_{2s}}(s, s)^{\dot{\gamma}\beta} \tag{30}$$

by contracting (for $s > \frac{1}{2}$, $\Sigma^{\mu\nu}_{(s)}$ is not a number matrix, but is momentum dependent), all the indices $\mu_2\cdots\mu_{2s}$ and ν_{2s}, with any combination of the four-momenta P and q or more conveniently, with k and p. The $\sigma_\pi = +1$ bases then are given [cf. (14)]

$$N^{(+)\mu}_{(i)_{\alpha\dot{\beta}}} = \{\Pi^{(s)}(p)\tilde{\Sigma}^{\mu\nu}_{i(s)} + \Sigma^{\mu\nu}_{i(s)}\Pi^{(s)}(k)\}_{\alpha\dot{\beta}}q_\nu, \tag{31}$$

which satisfies all the requirements of a basis for a Pauli form factor. The label i in (31) denotes the particular combination of k's and p's contracting the matrices (30).

It is straightforward but tedious in general to express the $N_i^{(+)}$ in terms of the $M_i^{(+)}$ by using the orthogonality relations of the CGC's when, for example, contracting the indices $\dot{\gamma}$ in (30). Once the N_i are expressed in terms of M_i, we are in a position to carry out our prescription to its conclusion by replacing all those M_i which would give neither Dirac-type nor Pauli-type form factors by those M_i which give Dirac-type form factors and N_i which give Pauli-type form factors. In general, the number of N_i arising from (30) and (31) is greater than the number of M_i that have to be replaced.

So far we have given our prescription only for $s' = s$, which is in fact a necessary condition if we are to expect interactions (Coulomb) in the $q^\mu \to 0$ limit. In the above procedure, (30) is without content for $s' \neq s$. However, this does not mean that it is not possible in this case to construct vector bases N_i^μ which contain a contacted factor of q_ν and are antisymmetric in μ and ν. Rather, such bases are easier to construct

for $s' \neq s$ because one can *always* find at least two space–time indices in whose interchange the spin matrix $\rho^{\mu_1\cdots\mu_{2s'}}(s', s)$ is antisymmetric, thus bypassing (30). It follows that we would then have three types of bases: (a) The two antisymmetric indices of the spin matrix contracted with P and q, the vector character being carried by P^μ; this basis does not contribute at $q^\mu \to 0$. (b) One of the antisymmetric indices contracted with q_ν; this basis also vanishes as $q^\mu \to 0$. (c) One of the antisymmetric indices contracted to P_ν; this basis could contribute at $q^\mu \to 0$, but the enforcement of condition (27) eliminates the form factors associated with this type of basis. Therefore, for $s' \neq s$ there is no Coulomb interaction in the $q^\mu \to 0$ limit, and thus we restrict our attention only to $s' = s$.

We shall now illustrate our prescription with an example, where $j = 1$ $s' = s = 1$, which is the simplest case after $s' = s = \frac{1}{2}$. The bases for this matrix element are given by the $\sigma_\pi = +1$ members of (21). As discussed above, in a minimal-type coupling framework the bases $M_1^{(+)}$ and $M_2^{(+)}$ could give rise to Dirac-type form factors, whereas $M_3^{(+)}$ and $M_4^{(+)}$ would give rise to neither Dirac- nor Pauli-type form factors. It is the latter two, therefore, that have to be replaced by $N_1^{(+)}$ and $M_{1,2}^{(+)}$. The following are the $N_i^{(+)\mu}$ that we can construct for $s = 1$, where all $(1,1)$ covariants $(xy)^{(1,1)}$ are implicitly denoted as (xy):

$$
\begin{aligned}
N_1^{(+)\mu} &= \{(pp)[(\tilde{\mu}k)(\nu k) - (\tilde{\nu}k)(\mu k)] + [(\mu k)(\tilde{\nu}k) \\
&\quad - (\nu k)(\tilde{\mu}k)](kk)\}q_\nu, \\
N_2^{(+)\mu} &= \{(pp)[(\tilde{\mu}p)(\nu p) - (\tilde{\nu}p)(\mu p)] + [(\mu p)(\tilde{\nu}p) \\
&\quad - (\nu p)(\tilde{\mu}p)](kk)\}q_\nu, \\
N_3^{(+)\mu} &= \{(pp)[(\tilde{\mu}k)(\nu p) - (\tilde{\nu}k)(\mu p)] + [(\mu k)(\tilde{\nu}p) \\
&\quad - (\nu k)(\tilde{\mu}p)](kk)\}q_\nu, \\
N_4^{(+)\mu} &= \{(pp)[(\tilde{\mu}p)(\nu k) - (\tilde{\nu}p)(\mu k)] + [(\mu p)(\tilde{\nu}k) \\
&\quad - (\nu p)(\tilde{\mu}k)](kk)\}q_\nu.
\end{aligned} \tag{32}
$$

Clearly not all these bases are needed just to replace $M_3^{(+)\mu}$ and $M_4^{(+)\mu}$, but we have listed all that follows from (30) and (31).

Carrying out the contractions in the matrix multiplication of spin-1 matrices, for example, over $\dot{\gamma}$ in (30), by use of the orthogonality relations of the GCG's we obtain the following:

$$
\begin{aligned}
&(\mu_1\mu_2)^{(1,1)}_{\alpha\alpha'}(\widetilde{\nu_1\nu_2})^{\dot{\alpha}'\dot{\beta}'}_{(1,1)}(\lambda_1\lambda_2)^{(1,1)}_{\dot{\beta}'\dot{\beta}} = (\mu_1\tilde{\nu}_1\lambda_1, \mu_2\tilde{\nu}_2\gamma_2)^{(1,1)}_{\alpha\dot{\beta}} \\
&\quad + \tfrac{1}{2}g^{\nu_1\nu_2}(\mu_1\mu_2)^{(1,0)}_{\alpha}(\lambda_1\lambda_2)^{(1,0)*}_{\dot{\beta}} - \tfrac{1}{2}[\tfrac{1}{2}\tfrac{1}{2}1]^{\alpha\alpha'}_{\alpha}[\tfrac{1}{2}\tfrac{1}{2}1]^{\dot{b}\dot{b}'}_{\dot{\beta}} \\
&\quad \times [(\rho^{\mu_2}\tilde{\rho}^{\nu_2}\rho^{\nu_1}\tilde{\rho}^{\mu_1}C^{-1})_{aa'}(C^{-1}\tilde{\rho}^{\lambda_2}\rho^{\lambda_1})_{\dot{b}\dot{b}'} \\
&\quad + (\rho^{\mu_2}\tilde{\rho}^{\mu_1}C^{-1})_{aa'}(C^{-1}\tilde{\rho}^{\lambda_2}\rho^{\nu_2}\tilde{\rho}^{\nu_1}\rho^{\lambda_1})_{\dot{b}\dot{b}'}],
\end{aligned} \tag{33}
$$

where we have used the notations of (15) and (20). Since there are only two independent momenta in the problem, in every case the last two terms of (33) turn out to be products of $\rho(1,0)_\alpha$ and $\rho(1,0)^*_{\dot{\beta}}$ covariants.

Using (33) in each member of (32), and then using the identity

$$
\begin{aligned}
&\rho^{\mu\tau}(1,0)_\alpha\rho^{\nu\lambda}(1,0)^*_{\dot{\beta}} + \rho^{\nu\lambda}(1,0)_\alpha\rho^{\mu\tau}(1,0)^*_{\dot{\beta}} \\
&\quad = -2(g^{\mu\nu}\rho^{\tau\lambda}(1,1)_{\alpha\dot{\beta}} - g^{\mu\lambda}\rho^{\nu\tau}(1,1)_{\alpha\dot{\beta}} \\
&\quad - g^{\nu\tau}\rho^{\mu\lambda}(1,1)_{\alpha\dot{\beta}} + g^{\tau\lambda}\rho^{\mu\nu}(1,1)_{\alpha\dot{\beta}})
\end{aligned} \tag{34}
$$

and its conjugate given by the transformation (14), we get

$$
\begin{aligned}
N_1^{(+)} &= m(\tau - 2)M_1^{(+)} - mM_2^{(+)} + \tfrac{3}{2}M_3^{(+)} - M_4^{(+)} \\
&\quad - \tfrac{1}{2}\hat{M}_3^{(+)} + \hat{M}_4^{(+)}, \\
N_2^{(+)} &= -mM_1^{(+)} + m(\tau - 2)M_2^{(+)} + \tfrac{3}{2}M_3^{(+)} - M_4^{(+)} \\
&\quad + \tfrac{1}{2}\hat{M}_3^{(+)} - \hat{M}_4^{(+)}, \\
N_3^{(+)} &= m(2 - 3\tau)M_1^{(+)} + mM_2^{(+)} + (4\tau - 1)M_3^{(+)} - 4M_4^{(+)}, \\
N_4^{(+)} &= -m(M_1^{(+)} + M_2^{(+)}) + 2M_4^{(+)},
\end{aligned} \tag{35}
$$

where $\tau = k\cdot p/m^2$ and $\hat{M}_3^{(+)\mu}$ and $\hat{M}_4^{(+)\mu}$ are the counterparts of $M_3^{(+)\mu}$ and $M_4^{(+)\mu}$, with P^μ replaced by q^μ in both, and are in fact those bases that were excluded in our construction in Sec. 2, under the constraint of (3). Here also, they may be excluded by invoking (3). We are then, it seems, in a position to replace $M_3^{(+)\mu}$ and $M_4^{(+)\mu}$ in terms of $N_1^{(+)\mu}\cdots N_4^{(+)\mu}$ and $M_1^{(+)\mu}$ and $M_2^{(+)\mu}$ by use of the relations (35), being careful in the process not to introduce a kinematic constraint onto the newly defined form factors. For example, replacing $M_3^{(+)}$ in terms of $N_3^{(+)}$ would result in a kinematic zero in the coefficients of $M_1^{(+)}$, $M_2^{(+)}$, and $M_4^{(+)}$ at $4k\cdot p = m^2$. Even under this restriction, however, this procedure does not seem to be unique, since we do not need all members of (35) to carry it out. Moreover, what appears to be more dangerous is that one could also find a way of replacing $M_1^{(+)}$ and $M_2^{(+)}$ in terms of $N_i^{(+)}$. This clearly is impossible, since there are two relations between, namely, $(N_1^{(+)} + N_2^{(+)})$, N_3, and N_4, not involving $M_1^{(+)}$ and $M_2^{(+)}$.

Here we enforce the condition (27) onto the matrix element,[18] resulting in the constraint

$$
F_1^{(+)}(q^2) = F_2^{(+)}(q^2). \tag{36}
$$

Calling both form factors in (36) $F_D(q^2)$, and hence obtaining the basis $(M_1^{(+)} + M_2^{(+)})^\mu = M_D^\mu$, we can rewrite relations (35):

$$
\begin{aligned}
N_1^{(+)} + N_2^{(+)} &= m(\tau - 3)M_D + 3M_3^{(+)} - 2M_4^{(+)}, \\
N_3^{(+)} &= -m(3\tau - 2)M_D + (4\tau - 1)M_3^{(+)} - 4M_4^{(+)}, \\
N_4^{(+)} &= -mM_D + 2M_4^{(+)}.
\end{aligned} \tag{37}
$$

It is then obvious that however we use Eq. (37) to substitute for $M_3^{(+)}$ and $M_4^{(+)}$, it is subsequently impossible to make a substitution for M_D without bringing back $M_3^{(+)}$ and $M_4^{(+)}$ into the expansion of the matrix element. The subscript D on the form factor signifies that it is a Dirac-type form factor, that is, it contributes to the Coulomb interaction in the limit $|\mathbf{q}| \to 0$. The two other form factors in this case will be, of course, Pauli-type.

Finally we should remark that our prescription in Sec. 2 is concerned with a rank-j tensor current that

couples to a spin-j particle, which means that this particle is not really off its mass shell at all since it has definite spin. The reason for this was our desire to obtain couplings for a vertex function that are equal in number to the couplings obtained directly by the use of the CGC's of the Poincaré group and the Wigner-Eckert theorem as in Ref. 5. The on-mass-shell property can be reversed, however, by not giving the tensor current a definite spin, which can be done simply by relaxing the condition (3) which was included in our prescription. The result will be to have a greater number of bases. The additional bases will be those with HLG transformation character

$$(s', s) \otimes q^{\mu_1} q^{\mu_2} \cdots q^{\mu_n} P^{\mu_{n+1}} \cdots P^{\mu_{2j}}, \tag{38}$$

where n ranges from 1 to $2j$. Thus for example, for $j = 1$, q^μ will carry the tensor character of a basis corresponding to a spin-0 particle, the so-called scalar photon.

We shall now demonstrate, in the special case of the electromagnetic current, the connection between the above approach and our prescription. In this case, the constraint to be satisfied in both approaches is (27). It is precisely this condition that leads to the two approaches giving the same result, through forcing the form factors associated with all bases of the form (38) to vanish, *provided* that the constraints arising from time-reversal invariance are already satisfied. These constraints will be nontrivial since the on-mass-shell particles (incoming/outgoing) are identical in the case under consideration. This symmetry can easily be found by the interchange $k \rightleftharpoons p$ in the bases of decomposition in Sec. 2B. Thus, for example, if $m' = m$, then $F_1^{(\pm)} = F_2^{(\pm)}$ in (21) and $F_1^{(\pm)} = F_2^{(\pm)}$ in (23) and $F_2^{(\pm)} = F_3^{(\pm)}$ in (25) etc. Now applying (27), results, for all s, into a constraint on the matrix element involving as many (s, s)-covariant σ_π-definite bases as there are bases whose vector character can be carried by P^μ. There are, therefore, just as many constraints on form factors as there are form factors with bases $(s, s) \otimes q^\mu$. In the event of time-reversal constraints being implemented, the latter form factors vanish, giving the same result as that obtained by our prescription. We illustrate the above by examples (iii) and (v) of Sec. 2B. We proceed by supposing that in both cases the separation of the form factors into Pauli and Dirac types has been carried out.

(iii) Under the time-reversal constraint, the new bases are $(M_1^{(+)} + M_2^{(+)})^\mu$, $N_I^{(+)\mu}$, $N_{II}^{(+)\mu}$ the Pauli-type bases, and the two additional bases:

$$\begin{aligned} \hat{M}_3^{(+)\mu} &= [(kk)^{(1,1)} + (pp)^{(1,1)}]q^\mu, \\ \hat{M}_4^{(+)\mu} &= (kp)^{(1,1)} q^\mu. \end{aligned} \tag{39}$$

Applying (27) with $q^2 \neq 0$, we end up with the constraint equation

$$q^2 \hat{F}_3(q^2)[(kk)^{(1,1)} + (pp)^{(1,1)}] + q^2 \hat{F}_4(q^2)(kp)^{(1,1)} = 0, \tag{40}$$

whence it follows that $\hat{F}_3(q^2) = \hat{F}_4(q^2) = 0$.

(v) Under the time-reversal constraint, the new bases are $(M_1^{(+)} + M_2^{(+)})^\mu, M_3^{(+)\mu}, N_I^{(+)\mu}, N_{II}^{(+)\mu}$, and the two additional bases are

$$\begin{aligned} \hat{M}_4^{(+)\mu} &= [(kkk)^{(\frac{3}{2},\frac{3}{2})} + (ppp)^{(\frac{3}{2},\frac{3}{2})}]q^\mu, \\ \hat{M}_5^{(+)\mu} &= [(kpp)^{(\frac{3}{2},\frac{3}{2})} + (pkk)^{(\frac{3}{2},\frac{3}{2})}]q^\mu. \end{aligned} \tag{41}$$

Applying (27) with $q^2 \neq 0$ we end up with $\hat{F}_4(q^2) = \hat{F}_5(q^2) = 0$.

ACKNOWLEDGMENTS

This work was started in Westfield College (London), and I am very grateful to Professor E. Leader for several discussions. I would also like to thank Dr. N. Dombey for reading the manuscript.

1 L. Durand III, P. C. DeCelles, and R. B. Marr, Phys. Rev. **126**, 1888 (1962).
2 E. de Rafael, Ann. Inst. Henri Poincaré 5, 83 (1966).
3 T. L. Trueman, Phys. Rev. **182**, 1469 (1969).
4 M. Scadron, Phys. Rev. **165**, 1640 (1968).
5 H. Joos, Fortschr. Physik **10**, 65 (1962).
6 M. S. Marinov, Ann. Phys. (N.Y.) **49**, 357 (1968).
7 S. Weinberg, Phys. Rev. **181**, 1893 (1969).
8 D. N. Williams and Y.-P. Yao, Phys. Rev. D 1, 1380 (1970).
9 H. P. Stapp, Phys. Rev. **125**, 3129 (1962).
10 In the case of the three-point function a direct multipole decomposition (Ref. 1) of the current between the helicity states of the on-mass-shell particles is carried out, while for four-point functions the most familiar way [see G. Cohen-Tannoudji, A. Morel, and H. Navelet, Ann. Phys. (N.Y.) **46**, 239 (1968)] is the use of the analytic properties of the t-channel center-of-mass helicity amplitudes and the corresponding crossing matrix, to find such combinations of s-channel center-of-mass helicity amplitudes that have only dynamical singularities.
11 S. Weinberg, Phys. Rev. **138**, B988 (1965).
12 $F_1(t)$ and $F_2(t)$ are the usual form factors, except that the factor $K/2m$ has been absorbed in $F_2(t)$.
13 A. O. Barut, I. J. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).
14 S. Weinberg, Phys. Rev. **133**, B1318 (1964).
15 D. H. Tchrakian, Nuovo Cimento **65A**, 535 (1970).
16 S. Weinberg, Phys. Rev. **135**, B1049 (1964).
17 S. Weinberg, Phys. Rev. **134**, B882 (1964).
18 Enforcing (27) onto the matrix element of the axial-vector current whose decomposition is given by the $\sigma_\pi = -1$ members of (21) modified according to the above prescription, we find $F_1^{(-)} = -F_2^{(-)}$ which results into the vanishing of the interaction when $|\mathbf{q}| \to 0$.

Neumann Series Solution for the Atom-Rigid Rotor Collision*

T. J. O'Brien and T. T. Holloway
Department of Chemistry, Texas Tech University, Lubbock, Texas 79409
(Received 17 February 1972; Revised Manuscript Received 15 May 1972)

The solutions of the coupled differential equations arising in the quantum mechanical discussion of the collision of an atom with a rigid, rotating diatom are written as Neumann series, i.e., expanded in terms of spherical Bessel functions. The coefficients of these series are generated by a set of coupled recursion relations. The formalism is limited to potentials less singular than r^{-2} at the origin.

1. INTRODUCTION

The exact quantum mechanical description of the collision of a structureless particle with a rigid rotor has been reduced to solving an infinite set of coupled equations.[1] In the close-coupling approximation, this infinite system of equations is truncated to a finite set which is then solved numerically.[2] As the size of the truncated set is increased, the exact solution is approached, but the computational effort becomes enormous.

Recently Gersten[3] presented an analytic method for representing, as an expansion in spherical Bessel functions, the solution of the radial Schrödinger equation describing a particle scattered by a spherically symmetric potential. The coefficients in this expansion are generated by a recursion formula. Thus, in the region in which the series converges, the solution can be generated very rapidly in a convenient analytic form.

In this paper we generalize Gersten's method to obtain a solution of the close-coupled equations of rotational excitation. Again the solutions are represented as series of spherical Bessel functions and the coefficients are generated by a set of coupled recursion formulas. In Sec. 2 this more general approach is applied, in detail, to the simple case of spherically symmetric scattering. Without explicit use of a second expansion procedure, we obtain a result entirely equivalent to Gersten's but of a different form. The method is applied to the coupled equations in Sec. 3. Section 4 contains the algorithm for the determination of the phase shift, and Sec. 5 contains a discussion of the significance and usefulness of the method.

2. THE RADIAL EQUATION FOR SPHERICALLY SYMMETRIC SCATTERING

The basic formulas used in this derivation are slight modifications of those discussed by Watson[4]: If $f(z)$ is analytic inside and on the circle $|z| = R$, and if C denotes the contour formed by this circle, then $f(z)$ has the representation

$$\left(\frac{\pi}{2}\right)^{1/2} z^{l} f(z) = \sum_{n=0}^{\infty} a_n j_{l+n}(z), \tag{2.1}$$

where

$$a_n = (2\pi i)^{-1} \oint_C f(z) A_{n,\, l+1/2}(z)\, dz \tag{2.2}$$

and l is an arbitrary, nonnegative integer. Here $j_l(z)$ is the regular spherical Bessel function,[5] and

$$A_{n,m}(z) = \frac{2^{n+m}(n+m)}{z^{n+1}} \sum_{\nu=0}^{[n/2]} \frac{\Gamma(n+m-\nu) z^{2\nu}}{2^{2\nu}\nu!} \tag{2.3}$$

is a Gegenbauer polynomial.[6]

The radial Schrödinger equation for the spherically symmetric case is

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2}\right) \psi_l(r) = U(r)\psi_l(r), \tag{2.4}$$

subject to the boundary conditions $\psi_l(0) = 0$ and

$$\psi_l(r) \underset{r\to\infty}{\sim} A[\sin(kr - \tfrac{1}{2}l\pi) + \tan\eta_l \cos(kr - \tfrac{1}{2}l\pi)]. \tag{2.5}$$

For potentials less singular than r^{-2}, the origin is a regular singular point and $\psi_l(r) = O(r^{l+1})$. The representation

$$\psi_l(r) = kr \sum_{n=0}^{\infty} f_n^l\, j_{l+n}(kr) \tag{2.6}$$

satisfies this condition and hence the boundary condition at the origin.

Because $krj_{l+n}(kr)$ is a solution of the homogeneous ($U \equiv 0$) form of Eq. (2.4), with l replaced by $l+n$, inserting the expansion (2.6) into Eq. (2.4) yields the expression

$$\sum_{n=0}^{\infty} [(l+n)(l+n+1) - l(l+1)] f_n^l\, j_{l+n}(kr) = \sum_{m=0}^{\infty} f_m^l\, r^2 U(r) j_{l+m}(kr). \tag{2.7}$$

Up to this point the derivation is identical to that presented by Gersten. However, his method of obtaining a recursion relation for the f_n^l is not applicable to the more general case of coupled equations. We will rederive what is essentially the same result but using a method which can also be used in Sec. 3 of this paper.

We note that, by setting $kr = z$, Eq. (2.7) is of the form of Eq. (2.1) with

$$a_n = [(l+n)(l+n+1) - l(l+1)] f_n^l \tag{2.8}$$

and

$$f(z) = \left(\frac{2}{\pi}\right)^{1/2} k^{-2} \sum_{m=0}^{\infty} f_m^l\, z^{2-l} U\left(\frac{z}{k}\right) j_{l+m}(z). \tag{2.9}$$

But combining Eqs. (2.2) and (2.9) gives an alternative expression for a_n:

$$a_n = \left(\frac{2}{\pi}\right)^{1/2} k^{-2} \sum_{m=0}^{\infty} f_m^l \times \left[(2\pi i)^{-1} \oint_C z^{-l+2} U\left(\frac{z}{k}\right) j_{l+m}(z) A_{n,\, l+1/2}(z) dz\right]. \tag{2.10}$$

For potentials which can be expanded as

$$U(r) = \sum_{\mu=-1}^{\infty} u_\mu r^\mu, \tag{2.11}$$

the integral in brackets in Eq. (2.10) can be evaluated by the residue theorem (see Appendix). The significant result is that the integral vanishes for $m > n-1$. Thus the rhs of Eq. (2.10) is a finite sum. Using the explicit result for the integral when $m \le n-1$, Eq. (A3), and equating Eqs. (2.8) and (2.10) yields the previously derived[3] recursion relation for the f_n^l:

$$[(l+n)(l+n+1)-l(l+1)]f_n^l = (n+l+\tfrac{1}{2})\sum_{m=0}^{n-1} f_m^l C_{mn}^l, \quad (2.12)$$

where

$$C_{mn}^l = 2^{n-m}\sum_{\kappa=0}^{N}\frac{u_{n-m-2\kappa-2}}{2^{2\kappa}k^{n-m-2\kappa}} \times \sum_{\nu=0}^{\kappa}\frac{(-1)^{\kappa-\nu}\Gamma(n+l+\frac{1}{2}-\nu)}{\nu!(\kappa-\nu)!\,\Gamma(m+l+\kappa+\frac{3}{2}-\nu)} \quad (2.13)$$

and $N=[(n-m-1)/2]$, the largest integer contained in $(n-m-1)/2$. For an arbitrary choice of f_0^l, which corresponds to a normalization constant, these equations can be used to generate, recursively, the set of coefficients f_n^l.

3. THE COUPLED RADIAL EQUATIONS FOR ASYMMETRIC SCATTERING

The generalized method of the preceeding section, which we used to obtain a known result although in a somewhat different form, will be used here to reduce the set of coupled differential equations occurring in the theory of rotational excitation to a system of coupled recursion relations. These equations[1] we write in the form

$$\left(\frac{d^2}{dr^2}+k_j^2-\frac{l(l+1)}{r^2}\right)\psi_{jl}(r) = \sum_{j'l'}\langle jl;J|U(r)|j'l';J\rangle\psi_{j'l'}(r), \quad (3.1)$$

where

$$k_j^2 = 2\mu\hbar^{-2}E-\mu j(j+1)/I \quad (3.2)$$

j,l and J are the quantum numbers specifying, respectively, the angular momentum of the rigid rotor, the orbital angular momentum of the atom and the total angular momentum of the system, μ is the reduced mass of the atom-diatomic system, I is the moment of inertia of the rigid rotor, and E is the total center-of-mass energy. The boundary conditions are

$$\psi_{jl}(0)=0 \quad (3.3)$$

and

$$\psi_{jl}(r) \underset{r\to\infty}{\sim} \delta_{jj_0}\delta_{ll_0}\exp[-i(k_jr-\tfrac{1}{2}l\pi)] - (k_{j_0}/k_j)^{1/2}S^J(jl,j_0l_0)\exp[i(k_jr-\tfrac{1}{2}l\pi)]. \quad (3.4)$$

Because of Eq. (3.4), ψ_{jl} should also be labeled by J,j_0, and l_0. To reduce clutter, however, this extra notation has been suppressed.

The ψ's can be represented as

$$\psi_{jl}(r) = k_jr\sum_{n=0}^{\infty} f_n^{jl}\, j_{l+n}(k_jr). \quad (3.5)$$

This is a valid representation for the regular solution [Eq. (3.3)] in some region about the origin, provided the potential matrix elements are less singular than r^{-2}. Inserting Eq. (3.5) into Eq. (3.1) yields

$$\sum_{n=0}^{\infty}[(l+n)(l+n+1)-l(l+1)]f_n^{jl}j_{l+n}(k_jr) = \sum_{j'l'}\sum_{m=0}^{\infty}k_j,k_j^{-1}f_m^{j'l'}r^2\langle jl;J|U|j'l';J\rangle j_{l'+m}(k_j,r). \quad (3.6)$$

By setting $k_jr=z$, this equation is of the form of Eq. (2.1) with

$$a_n = [(l+n)(l+n+1)-l(l+1)]f_n^{jl} \quad (3.7)$$

and

$$f(z) = (2/\pi)^{1/2}z^{-l}\sum_{j'l'}\sum_{m=0}k_j,k_j^{-3}f_m^{j'l'} \times z^2\langle jl;J|U(z/k_j)|j'l';J\rangle j_{l'+m}(k_j,z/k_j). \quad (3.8)$$

Inserting Eq. (3.8) into the integral expression for a_n, Eq. (2.2), gives

$$a_n = (2/\pi)^{1/2}\sum_{j'l'}\sum_{m=0}^{\infty}k_j,k_j^{-3}f_m^{j'l'} \times\left[(2\pi i)^{-1}\oint_C z^{2-l}\left\langle jl;J\left|U\left(\frac{z}{k_j}\right)\right|j'l';J\right\rangle \times j_{l'+m}\left(\frac{k_j,z}{k_j}\right)A_{n,l+1/2}(z)dz\right]. \quad (3.9)$$

As before, using the expansion

$$\langle jl;J|U(r)|j'l';J\rangle = \sum_{\mu=-1}^{\infty}u_\mu(jl,j'l';J)r^\mu, \quad (3.10)$$

the contour integral in Eq. (3.9) can be evaluated (see Appendix). In this case the integral vanishes for $m>l-l'+n-1$. For $m\le l-l'+n-1$, an explicit result can be obtained, Eq. (A7). Thus the sum over m in Eq. (3.9) is finite and can be readily evaluated. Equating this expression for a_n to Eq. (3.7) yields the set of coupled recursion relations for the f_n^{jl}:

$$[(l+n)(l+n+1)-l(l+1)]f_n^{jl} = (n+l+\tfrac{1}{2})k_j^{-n-l-1}\sum_{j'l'}\sum_{m=0}^{l-l'+n-1}f_m^{j'l'}C_{mn}(jl,j'l';J), \quad (3.11)$$

where

$$C_{mn}(jl,j'l';J) = 2^{l-l'+n-m}k_{j'}^{l'+m+1} \times\sum_{\kappa=0}^{K}\left(\frac{k_{j'}}{2}\right)^{2\kappa}u_{l-l'+n-m-2\kappa-2}(jl,j'l';J) \times\sum_{\nu=0}^{M}\frac{(-1)^{\kappa-\nu}(k_j/k_{j'})^{2\nu}\Gamma(n+l+\frac{1}{2}-\nu)}{\nu!(\kappa-\nu)!\,\Gamma(m+l'+\kappa+\frac{3}{2}-\nu)}. \quad (3.12)$$

Here $K=[(l-l'+n-m-1)/2]$ and $M=\min\{\kappa,[n/2]\}$.

At first sight Eq. (3.11) does not appear to be a truly recursive relation for the f_n^{jl} of channel (jl). The maximum value of m attained on the rhs is $l-l'+n-1$, so that, for $l>l'$, the values of m can become greater than n, the index of the lhs of Eq. (3.11). Thus, for $l>l'$, $f_m^{j'l'}$ is required for $m>n$ in order to calculate f_n^{jl}. However $f_m^{j'l'}$ is a coefficient of the wavefunction of a different channel ($l'\neq l$). The coefficients in this $(j'l')$ channel are also generated by a recursion relationship obtainable from Eq. (3.11) by interchanging (jl) with $(j'l')$. In order to obtain $f_n^{j'l'}$, f_m^{jl} is needed only up to $m=l'-l+n-1$, which is less than n since $l>l'$. Thus the series for the $f_n^{j'l'}$ can be "run ahead" to larger values of n. On each recursion of the system, by always using Eq. (3.11) for the smallest value of l first, $f_m^{j'l'}$ will be available when needed.

It should also be noted that closed channels (k_j imaginary) present no formal difficulty aside from the requirement of complex arithmetic. Since the analysis was performed in the complex plane there is no requirement that k_j be real. In particular, the series expansion for $j_l(k_j r)$, which was used in the Appendix, is valid for complex values of the argument.

4. EVALUATION OF THE PHASE SHIFTS

From the representation (2.6) and the boundary condition (2.5), since $j_{l+n}(kr) \sim \sin(kr - \frac{1}{2}l\pi - \frac{1}{2}n\pi)/(kr)$,

$$\tan\eta_l = -\left(\sum_{n=0}^{\infty} (-1)^n f_{2n+1}^l\right)\cdot\left(\sum_{n=0}^{\infty} (-1)^n f_{2n}^l\right)^{-1}. \tag{4.1}$$

As it must be, this result is independent of the choice of f_0^l. However, Eq. (4.1) is only formal. In fact (see Sec. 5), the radius of convergence of Eq. (2.6) may be smaller than the range of the potential, in which case the method of this paper could not be used to determine the phase shift at all. However, even if the radius of convergence is large enough so that the wavefunction has assumed its asymptotic form [Eq. (2.5)] before the series diverges, Eq. (4.1) makes the further assumption that $j_{l+n}(kr)$ has assumed it asymptotic form for all values of n which contribute.

Writing Eq. (2.5) more precisely as

$$\psi_l(r) \underset{r>a}{\sim} Akr[j_l(kr) - \tan\eta_l\, y_l(kr)], \tag{4.2}$$

where a is the range of the potential and y_l is the spherical Bessel function of the second kind,[5] the phase shift is given by the more complicated expression

$$\tan\eta_l = \frac{kRj_l(kR) - D(d/dr)[krj_l(kr)]|_{r=R}}{kRy_l(kR) - D(d/dr)[kry_l(kr)]|_{r=R}}, \tag{4.3}$$

where

$$D = kR\sum_{n=0}^{\infty} f_n^l j_{l+n}(kR)\cdot\left(\sum_{n=0}^{\infty} f_n^l \frac{d}{dr}[krj_{l+n}(kr)]\Big|_{r=R}\right)^{-1}, \tag{4.4}$$

and R is some value of r outside the range of the potential but inside the radius of convergence of the series (2.6).

This same consideration applies to the coupled equations. As in any close-coupled calculation,[7] N linearly independent sets of solutions to the N equations must be combined to satisfy the boundary condition (3.4). These sets of solutions ψ_{lj}^i, $i = 1, \ldots, N$, are obtained by choosing N linearly independent sets of values for the f_0^{jl}. For large r, each function of these N sets has a form similar to Eq. (4.2) or (2.5):

$$\psi_{lj}^i(r) \underset{r\to\infty}{\sim} A_{lj}^i[\sin(k_j r - \tfrac{1}{2}l\pi) + \tan\eta_{lj}^i \cos(k_j r - \tfrac{1}{2}l\pi)]. \tag{4.5}$$

The ψ_{lj}^i must then be combined[7] in the usual way to satisfy Eq. (3.4) or, equivalently, A_{lj}^i and η_{lj}^i must be combined to form the $S^J(jl, j_0 l_0)$.

5. DISCUSSION

We have presented a method for generating analytically the solutions of the coupled, differential equations appearing in the rotational excitation problem, Eq. (3.1). The coefficients of an expansion of the solutions in terms of spherical Bessel functions can be generated recursively. Thus the close-coupled solution can be obtained rapidly and in a convenient form.

The method is of the nature of a power series expansion in terms of the radial variable r. However, the series is rearranged to display the spherical Bessel functions, a natural function for the problem. Because of this the convergence properties of the series should be improved. The rate and radius of convergence will determine the utility of the method. We are analyzing the convergence properties of the series for several model potentials. Unfortunately it does not seem that any simple conclusions can be drawn. We will present a detailed discussion of this complicated question in another paper.

From preliminary work, it appears that the method can also be extended to the vibrational excitation problem, both collinear and three dimensional. We plan to investigate this possibility further.

APPENDIX

According to the residue theorem,[8] the value of the contour integral in Eq. (2.10) is the sum of the residues of the integrand at its singularities within the contour. Since $z^{-l}j_{l+m}(z)$ is entire and, under the assumptions on the potential, $z^2U(z/k)$ is analytic in some region containing the origin, the only singularities are due to the pole of $A_{n,l+1/2}$ at $z = 0$. Thus by expanding the integrand in powers of z, the residue (and hence the integral) can be found as the coefficient of z^{-1}. This coefficient can be obtained by repeated application of Cauchy's formula for the product of two series.[9] Proceeding in this way, we obtain

$$\begin{aligned} I(z) &\equiv z^{-l+2}U(z/k)j_{l+m}(z)A_{n,l+1/2}(z) \\ &= \left(\frac{\pi}{2}\right)^{1/2}(n + l + \tfrac{1}{2})\,2^{n-m}z^{m-n+1} \\ &\quad\times\left(\sum_{\mu=0}^{\infty} u_{2\mu-1}\frac{z^{2\mu-1}}{k^{2\mu-1}} + \sum_{\mu=0}^{\infty} u_{2\mu}\frac{z^{2\mu}}{k^{2\mu}}\right) \\ &\quad\times\sum_{\nu=0}^{[n/2]}\frac{\Gamma(n + l + \frac{1}{2} - \nu)z^{2\nu}}{2^{2\nu}\nu!} \\ &\quad\times\sum_{\kappa=0}^{\infty}\frac{(-1)^\kappa z^{2\kappa}}{\kappa!\,2^{2\kappa}\Gamma(l + m + \kappa + \frac{3}{2})} \\ &= \left(\frac{\pi}{2}\right)^{1/2}(n + l + \tfrac{1}{2})\,2^{n-m}z^{m-n+1} \\ &\quad\times\left(\sum_{\mu=0}^{\infty} z^{2\mu}\sum_{\kappa=0}^{\mu} S_{2\mu,\kappa}^{lnm} + \sum_{\mu=0}^{\infty} z^{2\mu-1}\sum_{\kappa=0}^{\mu} S_{2\mu-1,\kappa}^{lnm}\right), \end{aligned} \tag{A1}$$

where

$$S_{\sigma,\kappa}^{lnm} = \frac{u_{\sigma-2\kappa}}{2^{2\kappa}k^{\sigma-2\kappa}}\sum_{\nu=0}^{M}\frac{(-1)^{\kappa-\nu}\Gamma(n + l + \frac{1}{2} - \nu)}{\nu!(\kappa - \nu)!\,\Gamma(l + m + \kappa + \frac{3}{2} - \nu)} \tag{A2}$$

and $M = \min(\kappa, [n/2])$.

Setting $m - n + 1 + 2\mu = -1$ or $m - n + 1 + 2\mu - 1 = -1$ yields the value of μ corresponding to the z^{-1}

term: $\mu = [(n - m - 1)/2] \equiv N$. Since $\mu \geq 0$, there is no residue for $m > n - 1$. Further, N satisfies the condition $N \leq [n/2]$ so that, in all cases, $M = \kappa$. Thus

$$(2\pi i)^{-1} \oint_C I(z)dz = (\pi/2)^{1/2}(n + l + \tfrac{1}{2})\, 2^{n-m} \sum_{\kappa=0}^{N} S^{lnm}_{n-m-2,\kappa} \quad \text{(A3)}$$

for $m \leq n - 1$ and vanishes identically otherwise.

The residue of the integrand in Eq. (3.9) can be evaluated in much the same way:

$$\begin{aligned} I(z) &\equiv z^{2-l}\langle jl;J|\,U(z/k_j)\,|j'l';J\rangle j_{l'+m}(k_{j'}z/k_j)A_{n,l+1/2}(z) \\ &= (\pi/2)^{1/2}\, 2^{l-l'+n-m}(n + l + \tfrac{1}{2})(k_{j'}/k_j)^{l'+m} \\ &\quad \times z^{l'-l+m-n+1}\Bigg(\sum_{\mu=0}^{\infty} z^{2\mu} \sum_{\kappa=0}^{\mu} S^{nm}_{2\mu,\kappa}\,(jl,j'l';J) \\ &\quad + \sum_{\mu=0}^{\infty} z^{2\mu-1} \sum_{\kappa=0}^{\mu} S^{nm}_{2\mu-1,\kappa}\,(jl,j'l';J)\Bigg), \end{aligned} \quad \text{(A4)}$$

where, setting $M = \min\{\kappa, [n/2]\}$,

$$S^{nm}_{\sigma,\kappa}\,(jl,j'l';J) = \frac{u_{\sigma-2\kappa}(jl,j'l';J)}{k_j^{\sigma-2\kappa}\,2^{2\kappa}} \times \sum_{\nu=0}^{M} \frac{(-1)^{\kappa-\nu}(k_{j'}/k_j)^{2\kappa-2\nu}\Gamma(n + l + \frac{1}{2} - \nu)}{(\kappa-\nu)!\,\nu!\,\Gamma(l' + m + \kappa + \frac{3}{2} - \nu)}. \quad \text{(A5)}$$

According to Eq. (A4), the value of μ corresponding to the z^{-1} term is given by $l' - l + m - n + 1 + 2\mu = -1$ or $l' - l + m - n + 1 + 2\mu - 1 = -1$, that is,

$$\mu = [(l - l' + n - m - 1)/2] \equiv K. \quad \text{(A6)}$$

The condition $\mu \geq 0$ implies that the residue vanishes for $m \geq l - l' + n$. Thus

$$\begin{aligned} (2\pi i)^{-1} &\oint_C I(z)dz \\ &= (\pi/2)^{1/2}\, 2^{l-l'+n-m}(n + l + \tfrac{1}{2})(k_{j'}/k_j)^{l'+m} \\ &\quad \times \sum_{\kappa=0}^{k} S^{nm}_{n-m+l-l'-2,\kappa}(jl,j'l';J) \end{aligned} \quad \text{(A7)}$$

for $m \leq l - l' + n - 1$ and vanishes identically otherwise. Insertion of this result in Eq. (3.9) gives the coupled recursion relation (3.11).

* Work supported in part by a Robert A. Welch Foundation grant.

1 A. Arthurs and A. Dalgarno, Proc. Roy. Soc. (London) **A256**, 540 (1960).

2 W. A. Lester, Jr., in *Methods in Computational Physics* (Academic, New York, 1971), Vol. 10, p. 211; B. R. Johnson and D. Secrest, J. Chem. Phys. **48**, 4682 (1968); W. N. Sams and D. J. Kouri, *ibid.* **53**, 496 (1970); R. G. Gordon, *ibid.* **51**, 14 (1967).

3 A. Gersten, J. Math. Phys. **12**, 924 (1971).

4 G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966), Chap. 16.

5 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Department of Commerce, Natl. Bur. Stds., Washington, D.C., 1964), p. 437.

6 Reference 4, p. 283. *N.B.*: this is a different polynomial from the orthogonal Gegenbauer polynomial found in most texts.

7 W. A. Lester, Jr. and R. B. Bernstein, J. Chem. Phys. **48**, 4896 (1968); K. Smith, Argonne National Laboratory Report ANL-6095 (1960).

8 R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill, New York, 1960), p. 155.

9 Reference 5, p. 15, Eq. 3.6.21.

A Geometrical Theory of the Electromagnetic Field and the Gauge Transformation

Antonio F. Rañada

Departamento de Física Teórica, Universidad de Madrid, Madrid 3, Spain

(Received 14 March 1972)

A modification of Weyl's theory of the electromagnetic field is presented, such that the usual gravitational Lagrangian becomes invariant under the corresponding definition of gauge transformation. As another advantage the problem of the weight of the tensors in the construction of Lagrangians is eliminated.

I. INTRODUCTION

This paper proposes a geometrical interpretation of the electromagnetic field, closely related to Weyl's theory,[1] though it uses a different geometrical structure. The model includes a new geometrical definition of gauge transformation, which makes gauge invariance a weaker requirement than in Weyl's theory. Some important advantages follow. For instance, the usual gravitational Lagrangian becomes gauge invariant and the problem of the weights of the tensors is eliminated.

The present model has some common points with a theory presented in 1958 by Sciama.[2] He needed, however, a complex space–time of the Einstein–Schrödinger type. It seems that this is not necessary, the present interpretation being, therefore, much simpler.

II. MIXED THEORY OF THE GRAVITATIONAL FIELD

In order to describe the affine properties of space-time, we consider the 4-vector fields X_j[3]:

$$X_j = h_j{}^{\mu}\frac{\partial}{\partial x^{\mu}}, \quad [X_j, X_k] = c^{i}{}_{jk}X_i \quad (1)$$

and the differential forms

$$w^i = h^i{}_{\mu}dx^{\mu}, \quad (2)$$

$$w^k{}_l = A^k{}_{l\mu}dx^{\mu}, \quad (3)$$

such that

$$w^i(X_j) = h^i{}_{\mu}h_j{}^{\mu} = \delta^i_j. \quad (4)$$

term: $\mu = [(n - m - 1)/2] \equiv N$. Since $\mu \geq 0$, there is no residue for $m > n - 1$. Further, N satisfies the condition $N \leq [n/2]$ so that, in all cases, $M = \kappa$. Thus

$$(2\pi i)^{-1} \oint_C I(z)dz = (\pi/2)^{1/2}(n + l + \tfrac{1}{2})\, 2^{n-m} \sum_{\kappa=0}^{N} S^{lnm}_{n-m-2,\kappa} \quad \text{(A3)}$$

for $m \leq n - 1$ and vanishes identically otherwise.

The residue of the integrand in Eq. (3.9) can be evaluated in much the same way:

$$\begin{aligned} I(z) &\equiv z^{2-l}\langle jl;J|\,U(z/k_j)\,|j'l';J\rangle j_{l'+m}(k_{j'}z/k_j)A_{n,l+1/2}(z) \\ &= (\pi/2)^{1/2}\, 2^{l-l'+n-m}(n + l + \tfrac{1}{2})(k_{j'}/k_j)^{l'+m} \\ &\quad \times z^{l'-l+m-n+1}\Bigg(\sum_{\mu=0}^{\infty} z^{2\mu} \sum_{\kappa=0}^{\mu} S^{nm}_{2\mu,\kappa}\,(jl,j'l';J) \\ &\quad + \sum_{\mu=0}^{\infty} z^{2\mu-1} \sum_{\kappa=0}^{\mu} S^{nm}_{2\mu-1,\kappa}\,(jl,j'l';J)\Bigg), \end{aligned} \quad \text{(A4)}$$

where, setting $M = \min\{\kappa, [n/2]\}$,

$$S^{nm}_{\sigma,\kappa}\,(jl,j'l';J) = \frac{u_{\sigma-2\kappa}(jl,j'l';J)}{k_j^{\sigma-2\kappa}\,2^{2\kappa}} \times \sum_{\nu=0}^{M} \frac{(-1)^{\kappa-\nu}(k_{j'}/k_j)^{2\kappa-2\nu}\Gamma(n + l + \frac{1}{2} - \nu)}{(\kappa-\nu)!\,\nu!\,\Gamma(l' + m + \kappa + \frac{3}{2} - \nu)}. \quad \text{(A5)}$$

According to Eq. (A4), the value of μ corresponding to the z^{-1} term is given by $l' - l + m - n + 1 + 2\mu = -1$ or $l' - l + m - n + 1 + 2\mu - 1 = -1$, that is,

$$\mu = [(l - l' + n - m - 1)/2] \equiv K. \quad \text{(A6)}$$

The condition $\mu \geq 0$ implies that the residue vanishes for $m \geq l - l' + n$. Thus

$$\begin{aligned} (2\pi i)^{-1} &\oint_C I(z)dz \\ &= (\pi/2)^{1/2}\, 2^{l-l'+n-m}(n + l + \tfrac{1}{2})(k_{j'}/k_j)^{l'+m} \\ &\quad \times \sum_{\kappa=0}^{k} S^{nm}_{n-m+l-l'-2,\kappa}(jl,j'l';J) \end{aligned} \quad \text{(A7)}$$

for $m \leq l - l' + n - 1$ and vanishes identically otherwise. Insertion of this result in Eq. (3.9) gives the coupled recursion relation (3.11).

* Work supported in part by a Robert A. Welch Foundation grant.

1 A. Arthurs and A. Dalgarno, Proc. Roy. Soc. (London) **A256**, 540 (1960).

2 W. A. Lester, Jr., in *Methods in Computational Physics* (Academic, New York, 1971), Vol. 10, p. 211; B. R. Johnson and D. Secrest, J. Chem. Phys. **48**, 4682 (1968); W. N. Sams and D. J. Kouri, *ibid.* **53**, 496 (1970); R. G. Gordon, *ibid.* **51**, 14 (1967).

3 A. Gersten, J. Math. Phys. **12**, 924 (1971).

4 G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966), Chap. 16.

5 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Department of Commerce, Natl. Bur. Stds., Washington, D.C., 1964), p. 437.

6 Reference 4, p. 283. *N.B.*: this is a different polynomial from the orthogonal Gegenbauer polynomial found in most texts.

7 W. A. Lester, Jr. and R. B. Bernstein, J. Chem. Phys. **48**, 4896 (1968); K. Smith, Argonne National Laboratory Report ANL-6095 (1960).

8 R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill, New York, 1960), p. 155.

9 Reference 5, p. 15, Eq. 3.6.21.

A Geometrical Theory of the Electromagnetic Field and the Gauge Transformation

Antonio F. Rañada

Departamento de Física Teórica, Universidad de Madrid, Madrid 3, Spain

(Received 14 March 1972)

A modification of Weyl's theory of the electromagnetic field is presented, such that the usual gravitational Lagrangian becomes invariant under the corresponding definition of gauge transformation. As another advantage the problem of the weight of the tensors in the construction of Lagrangians is eliminated.

I. INTRODUCTION

This paper proposes a geometrical interpretation of the electromagnetic field, closely related to Weyl's theory,[1] though it uses a different geometrical structure. The model includes a new geometrical definition of gauge transformation, which makes gauge invariance a weaker requirement than in Weyl's theory. Some important advantages follow. For instance, the usual gravitational Lagrangian becomes gauge invariant and the problem of the weights of the tensors is eliminated.

The present model has some common points with a theory presented in 1958 by Sciama.[2] He needed, however, a complex space–time of the Einstein–Schrödinger type. It seems that this is not necessary, the present interpretation being, therefore, much simpler.

II. MIXED THEORY OF THE GRAVITATIONAL FIELD

In order to describe the affine properties of space-time, we consider the 4-vector fields X_j[3]:

$$X_j = h_j{}^{\mu}\frac{\partial}{\partial x^{\mu}}, \quad [X_j, X_k] = c^{i}{}_{jk}X_i \quad (1)$$

and the differential forms

$$w^i = h^i{}_{\mu}dx^{\mu}, \quad (2)$$

$$w^k{}_l = A^k{}_{l\mu}dx^{\mu}, \quad (3)$$

such that

$$w^i(X_j) = h^i{}_{\mu}h_j{}^{\mu} = \delta^i_j. \quad (4)$$

The quantities $h_k{}^\mu, A^i{}_{j\mu}$ are very often called—components of the vierbein and of the local affine connection—respectively[4-8]. The latin indexes will alway refer to the basis X_J, while the Greek ones will be related to the basis $X_\mu = \partial/\partial x^\mu$.

The torsion tensor $C^i{}_{kl}$ and the curvature tensor $R^i{}_{ljk}$ are determined by means of the structural equations of Cartan:

$$dw^i = -w^i{}_p \wedge w^p - \tfrac{1}{2} C^i{}_{jk} w^j \wedge w^k, \tag{5a}$$

$$dw^i{}_l = -w^i{}_p \wedge w^p{}_l - \tfrac{1}{2} R^i{}_{ljk} w^j \wedge w^k, \tag{5b}$$

from which one can deduce

$$C^i{}_{jk} = (h_j{}^\mu h_k{}^\nu - h_k{}^\mu h_j{}^\nu)(h^i{}_{\mu,\nu} + A^i{}_{q\nu} h^q{}_\mu), \tag{6a}$$

$$R^i{}_{ljk} = h_j{}^\mu h_k{}^\nu (A^i{}_{l\mu,\nu} - A^i{}_{l\nu,\mu} - A^i{}_{q\mu} A^q{}_{l\nu} + A^i{}_{q\nu} A^q{}_{l\nu}). \tag{6b}$$

(Our torsion and curvature tensors differ by a minus sign from those of Ref. 3.)

The metric properties of space–time are described by means of a tensor field g, of type (0, 2). Let us choose the basis X_k such that

$$g(X_i X_j) = \eta_{ij}, \tag{7}$$

η_{ij} being the Minkowski metric tensor with diagonal elements $(1, -1, -1, -1)$. It is then convenient to raise and lower the latin indexes by means of η_{ij}. One has then

$$g_{\mu\nu} = g\left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right) = h_{i\mu} h^i{}_\nu, \qquad h^i{}_\mu h_i{}^\nu = \delta_\mu{}^\nu. \tag{8}$$

As it is always done in general relativity we assume that the forms $w^i{}_j$ verify the condition

$$w^{ij} + w^{ji} = 0, \tag{9}$$

which implies that the length of a vector is invariant under parallel displacement.

We will consider the quantities $h_k{}^\mu, A^i{}_{j\mu}$ as independent basic fields of the theory. Such a formulation is usually called mixed theory. Another possibility, the so-called metric theory, is based on the assumption that the space–time is Riemannian and considers only the $h_k{}^\mu$ fields as independent fields.

The simplest Lagrangian which can be constructed is then

$$\mathcal{L} = R\mathcal{H}, \tag{10}$$

where $R = R^{ij}{}_{ij}$ and $\mathcal{H} = (\det h_k{}^\mu)^{-1} = \sqrt{|\det g_{\mu\nu}|}$.

Independent variations of $h_k{}^\rho, A^i{}_{j\mu}$ give the field equations

$$R_{ik} = 0, \qquad C^i{}_{kl} = 0, \tag{11}$$

Where R_{ik} is the Ricci tensor. As we see, they are the same as in the purely metric theory.

III. INTRODUCTION OF THE ELECTROMAGNETIC FIELD

Now let us assume that when the electromagnetic field A_μ is present, condition (9) is no longer adequate. More precisely, w^{ij} picks up a symmetric part and takes the form

$$w^i{}_j = {}^{(0)}w^i{}_j + \delta^i{}_j \alpha, \quad \alpha = A\mu\, dx^\mu, \tag{12}$$

where ${}^{(0)}w^i{}_j = A^i{}_{j\mu} dx^\mu$ verifies (9). The length L of a vector is then changed under parallel displacement. One has

$$dL = L\alpha \tag{13}$$

just as in the Weyl model.

From (12) and (6) we get

$$\begin{aligned} R^i{}_{jkl} &= {}^{(0)}R^i{}_{jkl} + \delta^i{}_j F_{kl}, \\ R_{jl} &= {}^{(0)}R_{jl} + F_{jl}, \\ C^i{}_{kl} &= {}^{(0)}C^i{}_{kl} + \delta^i{}_k A_l - \delta^i{}_l A_k, \\ R &= {}^{(0)}R, \end{aligned} \tag{14}$$

where the left index (0) means that the corresponding quantities are constructed only in terms of ${}^{(0)}w^i{}_j$ and $A_l = h_l{}^\mu A_\mu$, $F_{kl} = h_k{}^\alpha h_l{}^\beta (A_{\alpha,\beta} - A_{\beta,\alpha})$.

We now take the Lagrangian

$$\mathcal{L} = (R - \tfrac{\kappa}{32} R^i{}_{ikl} R_j{}^{jkl})\mathcal{H}, \tag{15}$$

where κ is the gravitational constant. From (14) we get

$$\mathcal{L} = (R - \tfrac{\kappa}{2} F_{\mu\nu} F^{\mu\nu})\mathcal{H}, \tag{16}$$

so that $\mathcal{L}$ has the usual form. Note that (15) is the only Lagrangian which is quadratic in $R^i{}_{jkl}$ and reduces to (10) as A_μ vanishes.

Now we submit $h_k{}^\mu, A^i{}_{j\mu}, A_\mu$ to independent variations.

The calculations are very similar to those of Refs. 6 and 8, where we refer for more details. Variation with respect to $A^i{}_{j\mu}$ gives

$${}^{(0)}C^i{}_{jk} = 0, \tag{17}$$

which implies

$$C^i{}_{jk} = \delta^i{}_j A_k - \delta^i{}_k A_j. \tag{18}$$

As we see, A_μ induces a torsion. In spite of (13) our geometry is therefore different from Weyl's.

Variation with respect of $h_k{}^\mu$ and A_μ give

$$\begin{aligned} {}^{(0)}R^k{}_j - \tfrac{1}{2}\delta^k{}_j R &= -\kappa(F^{kl}F_{jl} + \tfrac{1}{4}\delta^k{}_j F_{nl}F^{nl}), \\ (F^{\mu\nu}\mathcal{H})_{,\nu} &= 0, \end{aligned} \tag{19}$$

which are the usual Einstein–Maxwell coupled equations, with symmetric Einstein and energy–momentum tensors. This is somewhat surprising since it is known[6,7] that these tensors become nonsymmetric in twisted spaces. This peculiar situation is due to the fact that the first term in (15) depends only on ${}^{(0)}w^i{}_j$ which, on the other hand, does not affect the second one.

Now it seems natural to define the gauge transformation as

$$w^i{}_j \to w^i{}_j + \delta^i{}_j d\phi, \tag{20}$$

where ϕ is an arbitrary function. In terms of our fundamental fields this takes the form

$$h_k{}^\mu \to h_k{}^\mu, \quad A^i{}_{j\mu} \to A^i{}_{j\mu}, \quad A_\mu \to A_\mu + \phi,_\mu . \tag{21}$$

Let us recall the corresponding definition in Weyl's theory:

$$g_{\alpha\beta} \to \phi g_{\alpha\beta}, \quad A_\alpha \to A_\alpha + \tfrac{1}{2}(\log\phi)_{,\alpha} . \tag{22}$$

It is clear that (5b) and, consequently, the curvature tensor, is invariant under (20), as well as the metric tensor $g_{\alpha\beta}$. This means that the Lagrangians (10) and (16) are gauge invariant. As it is well known this is not the case in Weyl's theory. An important consequence is the elimination of the problem of the weights of the tensors in the construction of Lagrangians. More precisely one does not need to change the gravitational Lagrangian if electromagnetism is to be introduced. It is also easy to make invariant the Dirac or Klein–Gordon Lagrangians by the usual phase transformations

$$\psi \to e^{i\varphi}\psi$$

and the substitution $\partial_\mu \to \partial_\mu - iA_\mu$. This is in sharp contrast with the Weyl model. On the other hand, Eq. (5a) is not invariant under (20). This means that the torsion tensor is not invariant.

In Weyl's view, gauge invariance expresses the fact that a theory must not change under arbitrary point-dependent changes in the scale of length. Even in the absence of matter this requirement is too strong, because the gravitation constant links the units of length and energy. In our model, gauge invariance means independence on the way we compare the tangent spaces at two neighboring points, provided that the comparation device $(w^i{}_j)$ is changed according to (20). This is certainly a weaker condition.

It has been argued with respect to Weyl's theory that the variation of the scale of length under parallel displacement leads to contradictions [see for instance Chap. 13 in Ref. 9). The arguments are based on the following assumption: The mathematical parallel displacement of a vector is always the expression of the physical transport of a rod. This view is plausible, but difficult to apply to more general objects as spinors or tensors or to transport along a spacelike curve. We can, however, take the view that the parallel displacement gives only a rule useful to construct a covariant derivative. (See in this connection Ref. 10, Chap. VII.) In other words the scale of length is given by the metric tensor which we take as independent of the connection. If we transport the scale of length V^μ from P to P', we have to adjust this vector according to the value $g_{\mu\nu}(P')$ if it is to be interpreted as the unit of length at P'.

IV. CONCLUSIONS

We have given a geometrical interpretation of A_μ which has some important advantages over Weyl's theory, specially the gauge invariance of the gravitational Lagrangian.

The usual Einstein–Maxwell coupled equation can be interpreted in two different ways:

(a) as equations involving purely geometrical quantities in a non-Riemannian space–time;

(b) as equations involving geometrical $(h_k{}^\mu, A^i{}_{j\mu})$ and radiation (A_μ) quantities in a Riemannian space–time.

We stress the fact that in both cases the equations are identical.

ACKNOWLEDGMENT

The author is very indebted to Professor M. Soler for discussions and criticism.

1 H. Weyl, *Space-Time-Matter* (Dover, New York, 1950).
2 D. W. Sciama, Nuovo Cimento 8, 417 (1958).
3 S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1969).
4 H. Weyl, Phys. Rev. 77, 699 (1950).
5 R. Utiyama, Phys. Rev. 101, 1597 (1956).
6 T. W. B. Kibble, J. Math. Phys. 2, 212 (1961).
7 D. W. Sciama, *Festschrift for Infeld* (Pergamon, New York, 1969).
8 A. F. Rañada and M. Soler, J. Math. Phys. 13, 671 (1972).
9 R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
10 A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., London, 1965).

Conditions for the Existence of the Generalized Wave Operators

D. B. Pearson
Department of Applied Mathematics, University of Hull, Hull, Yorkshire, England
(Received 26th July 1971)

If $E_0(G)$ is the spectral projection operator associated with the free Hamiltonian H_0, corresponding to a bounded measurable subset G of $\mathbb{R}$, and $E_1(G)$ is associated with the total Hamiltonian $H = H_0 + V$, where the operator $E_1(G)VE_0(G)$ is of trace class, it is proved that the element $g = E_0(G)f$ belongs to the domain of the generalized wave operators $\Omega_\pm$ if and only if

$$\lim_{t\to\pm\infty} \|(1 - E_1(G))e^{-iH_0t}g\| = 0.$$

A stronger version of this result is also proved, from the theory of time-dependent scattering, and is applicable to scattering systems for which families $\{G_1\}$, $\{G_2\}$ of measurable sets may be found such that $E_1(G_2)VE_0(G_1)$ is of trace class.

INTRODUCTION

For a scattering system described by free Hamiltonian H_0 and total Hamiltonian H, where H_0 and H are self-adjoint, the generalized wave operators $\Omega_\pm$ are defined by

$$\Omega_\pm f = \operatorname*{s-lim}_{t\to\mp\infty} e^{iH_0t}e^{-iHt}P_1 f, \tag{1}$$

where P_0 is the orthogonal projection onto the abso-

where ϕ is an arbitrary function. In terms of our fundamental fields this takes the form

$$h_k{}^\mu \to h_k{}^\mu, \quad A^i{}_{j\mu} \to A^i{}_{j\mu}, \quad A_\mu \to A_\mu + \phi,_\mu . \tag{21}$$

Let us recall the corresponding definition in Weyl's theory:

$$g_{\alpha\beta} \to \phi g_{\alpha\beta}, \quad A_\alpha \to A_\alpha + \tfrac{1}{2}(\log\phi)_{,\alpha} . \tag{22}$$

It is clear that (5b) and, consequently, the curvature tensor, is invariant under (20), as well as the metric tensor $g_{\alpha\beta}$. This means that the Lagrangians (10) and (16) are gauge invariant. As it is well known this is not the case in Weyl's theory. An important consequence is the elimination of the problem of the weights of the tensors in the construction of Lagrangians. More precisely one does not need to change the gravitational Lagrangian if electromagnetism is to be introduced. It is also easy to make invariant the Dirac or Klein–Gordon Lagrangians by the usual phase transformations

$$\psi \to e^{i\varphi}\psi$$

and the substitution $\partial_\mu \to \partial_\mu - iA_\mu$. This is in sharp contrast with the Weyl model. On the other hand, Eq. (5a) is not invariant under (20). This means that the torsion tensor is not invariant.

In Weyl's view, gauge invariance expresses the fact that a theory must not change under arbitrary point-dependent changes in the scale of length. Even in the absence of matter this requirement is too strong, because the gravitation constant links the units of length and energy. In our model, gauge invariance means independence on the way we compare the tangent spaces at two neighboring points, provided that the comparation device $(w^i{}_j)$ is changed according to (20). This is certainly a weaker condition.

It has been argued with respect to Weyl's theory that the variation of the scale of length under parallel displacement leads to contradictions [see for instance Chap. 13 in Ref. 9). The arguments are based on the following assumption: The mathematical parallel displacement of a vector is always the expression of the physical transport of a rod. This view is plausible, but difficult to apply to more general objects as spinors or tensors or to transport along a spacelike curve. We can, however, take the view that the parallel displacement gives only a rule useful to construct a covariant derivative. (See in this connection Ref. 10, Chap. VII.) In other words the scale of length is given by the metric tensor which we take as independent of the connection. If we transport the scale of length V^μ from P to P', we have to adjust this vector according to the value $g_{\mu\nu}(P')$ if it is to be interpreted as the unit of length at P'.

IV. CONCLUSIONS

We have given a geometrical interpretation of A_μ which has some important advantages over Weyl's theory, specially the gauge invariance of the gravitational Lagrangian.

The usual Einstein–Maxwell coupled equation can be interpreted in two different ways:

(a) as equations involving purely geometrical quantities in a non-Riemannian space–time;

(b) as equations involving geometrical $(h_k{}^\mu, A^i{}_{j\mu})$ and radiation (A_μ) quantities in a Riemannian space–time.

We stress the fact that in both cases the equations are identical.

ACKNOWLEDGMENT

The author is very indebted to Professor M. Soler for discussions and criticism.

1 H. Weyl, *Space-Time-Matter* (Dover, New York, 1950).
2 D. W. Sciama, Nuovo Cimento 8, 417 (1958).
3 S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1969).
4 H. Weyl, Phys. Rev. 77, 699 (1950).
5 R. Utiyama, Phys. Rev. 101, 1597 (1956).
6 T. W. B. Kibble, J. Math. Phys. 2, 212 (1961).
7 D. W. Sciama, *Festschrift for Infeld* (Pergamon, New York, 1969).
8 A. F. Rañada and M. Soler, J. Math. Phys. 13, 671 (1972).
9 R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
10 A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., London, 1965).

Conditions for the Existence of the Generalized Wave Operators

D. B. Pearson
Department of Applied Mathematics, University of Hull, Hull, Yorkshire, England
(Received 26th July 1971)

If $E_0(G)$ is the spectral projection operator associated with the free Hamiltonian H_0, corresponding to a bounded measurable subset G of $\mathbb{R}$, and $E_1(G)$ is associated with the total Hamiltonian $H = H_0 + V$, where the operator $E_1(G)VE_0(G)$ is of trace class, it is proved that the element $g = E_0(G)f$ belongs to the domain of the generalized wave operators $\Omega_\pm$ if and only if

$$\lim_{t\to\pm\infty} \|(1 - E_1(G))e^{-iH_0t}g\| = 0.$$

A stronger version of this result is also proved, from the theory of time-dependent scattering, and is applicable to scattering systems for which families $\{G_1\}$, $\{G_2\}$ of measurable sets may be found such that $E_1(G_2)VE_0(G_1)$ is of trace class.

INTRODUCTION

For a scattering system described by free Hamiltonian H_0 and total Hamiltonian H, where H_0 and H are self-adjoint, the generalized wave operators $\Omega_\pm$ are defined by

$$\Omega_\pm f = \operatorname*{s-lim}_{t\to\mp\infty} e^{iH_0t}e^{-iHt}P_1 f, \tag{1}$$

where P_0 is the orthogonal projection onto the abso-

lutely continuous subspace of H_0. One of the most important problems of scattering theory is to discover what conditions must be imposed on H and H_0 for the wave operators $\Omega_\pm$ to exist and prove that these conditions are satisfied for systems of interest in quantum mechanics. For single-channel systems, it is equally important to prove the existence of the limits

$$\Omega_\pm^* f = \text{s-}\lim_{t\to\mp\infty} e^{iH_0 t} e^{-iHt} P_1 f, \tag{1'}$$

where P_1 is the orthogonal projection onto the absolutely continuous subspace of H. The conditions for $\Omega_\pm^*$ to exist as a strong limit may be derived from conditions for $\Omega_\pm$ to exist by interchanging the roles of the operators H_0 and H. If, for all elements f, g, the strong limits of Eqs. (1) and (1') both exist, it follows that the restriction of the scattering operator $S = \Omega_-^*\Omega_+$ to the absolutely continuous subspace of H_0 is unitary. The unitarity of S has important physical consequences for the system, and may be regarded as the defining property of single-channel systems.

It is known[1,2] that if $H = H_0 + V$, then the wave operators exist provided V is of finite rank. This result has been considerably strengthened by successive authors, using the properties of trace-class operators.[3] Kato[1] and Rosenblum[4] showed that the wave operators exist provided V is of trace class, and Kato[5] has subsequently shown that it is sufficient that $\psi(H) - \psi(H_0)$ be of trace class, where ψ is any function satisfying a set number of rather weak conditions. For example, $\Omega_\pm$ exist provided $(z - H)^{-N} - (z - H_0)^{-N}$ is of trace class for some positive number N and some complex number z. [The result of Kuroda,[6] requiring that $|V|^{1/2}(H_0 + i)^{-1}$ be of Schmidt class and that $\|Vx\| \leq a\|H_0 x\| + b\|x\|, a < 1$, for all x in the domain of H_0, may be derived as a special case of the above result with $N = 1$. It is sufficient to prove that $(H_0 + i)^{-1}V(H_0 + i)^{-1}$ is of trace class.]

Now if $(z - H)^{-N} - (z - H_0)^{-N}$ is of trace class, it may be deduced that there exist bounded measurable subsets G of the real line such that the operator $E_0(G)VE_1(G)$ is of trace class, where $E_0(G)$ is the spectral projection operator corresponding to the subset G and is defined in terms of the resolution of the identity for H_0, and $E_1(G)$ is similarly defined in terms of the resolution of the identity for H. [The operator $E_0(G)VE_1(G)$, which is always bounded, may be defined in a natural way even when the domain of $V = H - H_0$ is not dense in the Hilbert space.] Moreover, the union of all such subsets is $\mathbb{R}\backslash\sigma$, where σ is a set of Lebesgue measure zero, and this remains true for a wide class of functions ψ, given the more general condition that $\psi(H) - \psi(H_0)$ be of trace class.

The condition that $E_0(G)VE_1(G)$ be of trace class is an extremely weak one, and does not of itself guarantee the existence of the wave operators. (If H_0 is the one-dimensional "position operator," then $E_0(G)VE_1(G)$ may be of trace class for *every* bounded subset G, but the wave operators do not exist.) In the present paper we prove that if $E_1(G)VE_0(G)$ is of trace class, then $E_0(G)f$ belongs to the domain of $\Omega_\pm$ provided that

$$\lim_{t\to\pm\infty} \|E_1(\overline{G}) e^{-iH_0 t} E_0(G) f\| = 0,$$

where $\overline{G}$ denotes the complement of G. This is in fact a special case of Theorem 3 of Sec. 3, which provides sufficient conditions for an element of the space to be in the domain of $\Omega_\pm$, given the existence of a number of subsets G_1, G_2 such that $E_1(G_2)VE_0(G_1)$ is of trace class. Theorem 3 is actually the strongest possible result, in the sense that if $E_1(G)VE_0(G)$ is of trace class for *every* bounded measurable subset G (as in the case for a wide class of scattering systems), then we have a necessary and sufficient condition for an element to be in the domain of $\Omega_\pm$. Theorem 3 is in that case a generalisation of a theorem of Birman,[7] who postulates the existence of subsets G_1, G_2 such that $E_1(\overline{G}_2)E_0(G_1)$ is compact; the existence of subsets having this property is not a necessary condition for the existence of $\Omega_\pm$. The application of Theorem 3 to proving the existence of the wave operators for a class of singular potentials will be considered in a subsequent paper, and provides examples of scattering systems for which weaker versions of Theorem 3 are not adequate.

In Sec. 1, we define,[8] for any bounded linear operator L, the family of linear operators $F_{ab}(L)$ and the corresponding singular integral operators $F_{0\infty}(L)$ and $F_{-\infty\infty}(L)$. In Theorem 1 we consider the case where L is of rank 1, and in that case construct a dense set of elements belonging to the domain of $F_{-\infty\infty}(L)$. Theorem 1 is a generalization of a result proved in Ref. 9.

In Sec. 2, we use the properties of the F operators to derive, under very general conditions, an identity for the evolution operator $e^{iHt}e^{-iH_0 t}$, using techniques which have been applied in Ref. 10 to the case where V is of rank 1. The main result of Sec. 2 is summarized in Theorem 2.

In Sec. 3, we make the assumption that subsets G_1, G_2 exist such that $E_1(G_2)VE_0(G_1)$ is of trace class. The argument is analogous to that of Ref. 9, and in Theorem 3 and its corollaries we give conditions for an element f to be in the domain of $\Omega_\pm$.

1. THE DOMAIN OF $F_{-\infty\,\infty}(L)$ WHERE L IS OF RANK ONE

For any bounded linear operator L and any pair of real numbers a, b, we define the linear operator $F_{ab}(L)$ by

$$F_{ab}(L)f = \frac{1}{2\pi}\int_a^b e^{iH_0 s} L e^{-iH_0 s} f. \tag{2}$$

The integrand being strongly continuous in s, the integral may be defined as a strong limit of approximating Riemann sums.

$F_{a\infty}(L)$ is defined by

$$F_{a\infty}(L)f = \text{s-}\lim_{b\to\infty} F_{ab}(L)f,$$

for any f such that the limit exists. $F_{-\infty,b}(L)$ is defined in a similar way, and we also define

$$F_{-\infty\,\infty}(L) = F_{-\infty\,0}(L) + F_{0\,\infty}(L).$$

It is not difficult to show that

$$e^{iH_0 t} F_{ab}(L) e^{-iH_0 t} = F_{a+t,b+t}(L), \tag{3}$$

so that, for example,

$$e^{iH_0 t}F_{a\infty}(L)e^{-iH_0 t} = F_{a+t,\infty}(L). \tag{3'}$$

In the following sections, we shall have to deal with the F operators where L is of trace class. Since any trace class operator may uniformly be approximated by sums of operators of rank 1, a first step is to study the case where L is of rank 1.

If $L = |\phi\rangle\langle\psi|$, we have

$$F_{ot}(|\phi\rangle\langle\psi|)f = \frac{1}{2\pi}\int_0^t e^{iH_0 s}\,\phi(\psi, e^{-iH_0 s}f)ds. \tag{4}$$

Applying the spectral theorem, we find, for any element y of the space,

$$(y, F_{ot}(|\phi\rangle\langle\psi|)f) = \frac{1}{2\pi}\int_0^t ds(\psi, e^{-iH_0 s}f) \times\left(\int_{-\infty}^{\infty} e^{i\lambda s}d(y, E_\lambda\phi)\right). \tag{5}$$

The interchange of λ and s integrations may readily be justified, giving

$$(y, F_{0t}(|\phi\rangle\langle\psi|)f) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}(\psi, (\lambda - H_0)^{-1} \times [e^{i(\lambda - H_0)t} - 1]f)d_\lambda(y, E_\lambda\phi) \tag{6}$$

Although $(\lambda - H_0)^{-1}$ is not defined whenever λ is an eigenvalue of H_0, we understand the integrand of the rhs in that case to mean

$$\int_{-\infty}^{\infty}\Gamma(\mu)d(\psi, E_\mu f),$$

where

$$\Gamma(\mu) = \frac{e^{i(\lambda-\mu)t} - 1}{(\lambda - \mu)}, \quad \lambda \neq \mu$$
$$= \quad it \quad , \quad \lambda = \mu.$$

Thus Γ is a continuous function.

If f is orthogonal to every eigenvector of H_0, an application of the Lebesgue dominated-covergence theorem shows that

$$(\psi, (\lambda - H_0)^{-1}[e^{i(\lambda-H_0)t} - 1]f) = \lim_{\epsilon\to 0}(\psi, (\lambda - H_0 + i\epsilon)^{-1}[e^{i(\lambda-H_0)t} - 1]f), \tag{7}$$

and, with further application of the theorem, we deduce from Eq. (6) that

$$2\pi i(y, F_{0t}(|\phi\rangle\langle\psi|)f) = \lim_{\epsilon\to 0}\int_{-\infty}^{\infty} (\psi, (\lambda - H_0 + i\epsilon)^{-1}[e^{i(\lambda-H_0)t} - 1)f)d_\lambda(y, E_\lambda\phi). \tag{8}$$

Using this expression, we shall construct a set of elements belonging to the domain of $F_{-\infty\,\infty}(|\phi\rangle\langle\psi|)$ [i.e., these elements will belong to the domain both of $F_{0\infty}(|\phi\rangle\langle\psi|)$ and of $F_{-\infty\,0}(|\phi\rangle\langle\psi|)$].

Denoting by $M_{\text{a.c.}}(H_0)$ the absolutely continuous subspace, let g be any element belonging to $M_{\text{a.c.}}(H_0)$ so that, for some $\gamma \in L_1$,

$$(g, E_\lambda g) = \int_{-\infty}^{\lambda}\gamma(s)ds. \tag{9}$$

We have, apart from a set of values λ having Lebesgue measure zero,

$$\frac{d}{d\lambda}(g, E_\lambda g) = \gamma(\lambda). \tag{9'}$$

For any positive numbers ϵ, M, N we define the following measurable sets of real numbers:

$$S(N) = \{k;\ |1 - I_\alpha^{\pm}(k)| \leq N \text{ for all } \alpha \text{ in the interval } 0 < \alpha \leq 1\}, \tag{10}$$

$$\Sigma(M, N) = \{k; |\gamma(k)| \leq M\} \cap S(N), \tag{10'}$$

where

$$I_\alpha^{\pm}(k) \equiv (\phi, [k - H_0 \pm i\alpha]^{-1}\,\phi). \tag{10''}$$

Although $\gamma(k)$ is not unique, any two representatives of the element γ of $L_1(-\infty, \infty)$ differ only on a set of (Lebesgue) measure zero; we shall say that $\Sigma(M, N)$ is denoted *modulo* a set of measure zero.

Given any measurable set G, there is a corresponding projection operator $E_0(G)$ defined by

$$(y, E_0(G)x) = \int_{-\infty}^{\infty}\chi_G(\lambda)d(y, E_\lambda x), \tag{11}$$

where χ_G is the characteristic function of G. We denote the corresponding projection operator defined in terms of H instead of H_0 by $E_1(G)$.

If G is only given modulo a set of measure zero, the restriction of $E_0(G)$ to $M_{\text{a.c.}}(H_0)$ is nevertheless uniquely defined.

Theorem 1: For any element g belonging to $M_{\text{a.c.}}(H_0)$, the element $g_{M,N} \equiv E_0(\Sigma(M, N))g$ belongs to the domain of $F_{-\infty\,\infty}(|\phi\rangle\langle\psi|)$.

Proof: Define, for $\epsilon \neq 0$, the one-parameter family of operators Ψ_ϵ by

$$\Psi_\epsilon f = \int_{-\infty}^{\infty}(\psi, [\lambda - H_0 + i\epsilon]^{-1}[1 - I_\epsilon^{-}(H_0)]^{-1}f)dE_\lambda\phi. \tag{12}$$

For fixed ϵ, $[1 - I_\epsilon^{-}(\mu)]^{-1}$ is a bounded continuous function of μ, so that $\int_{-\infty}^{\infty}[1 - I_\epsilon^{-}(\mu)]^{-1}dE_\mu f$ exists.

$\|\Psi_\epsilon f\|^2$ is the triple integral

$$\|\Psi_\epsilon f\|^2 = \iiint \frac{d_\lambda(E_\lambda\phi, \phi)d_\mu(f, E_\mu\psi)d_\nu(\psi, E_\nu f)}{(\lambda - \mu - i\epsilon)(\lambda - \nu + i\epsilon)} \times \frac{1}{[1 - I_\epsilon^{+}(\mu)][1 - I_\epsilon^{-}(\nu)]} \tag{13}$$

Substituting

$$(\lambda - \mu - i\epsilon)^{-1}(\lambda - \nu + i\epsilon)^{-1} = (\mu - \nu + 2i\epsilon)^{-1} \times [(\lambda - \mu - i\epsilon)^{-1} - (\lambda - \nu + i\epsilon)^{-1}],$$

we may carry out the λ integration, and, using Eq. (10''), we have

$$\|\Psi_\epsilon f_\epsilon\|^2 = \iint \frac{d_\mu(f, E_\mu\psi)d_\nu(\psi, E_\nu f)}{(\mu - \nu + 2i\epsilon)} \times\left(\frac{1}{1 - I_\epsilon^{-}(\nu)} - \frac{1}{1 - I_\epsilon^{+}(\mu)}\right). \tag{14}$$

Now replace f by the one-parameter family $f_\epsilon \equiv [1 - I_\epsilon^{-}(H_0)]g_{M,N}$.

Then

$$\|\Psi_\epsilon f_\epsilon\|^2 = 2\text{Re}\left(\iint \frac{d_\mu(f_\epsilon, E_\mu\psi)d_\nu(\psi, E_\nu g_{M,N})}{(\mu - \nu + 2i\epsilon)}\right) \tag{15}$$

Now

$$(g_{M,N}, E_\nu g_{M,N}) = \int_{-\infty}^{\nu} \chi_{\Sigma(M,N)}(s)\, \frac{d}{ds}\,(g, E_s g)ds,$$

so that for almost all ν we have

$$\begin{aligned}\frac{d}{d\nu}(g_{M,N}, E_\nu g_{M,N}) &= \chi_{\Sigma(M,N)}(\nu)\, \frac{d}{d\nu}(g, E_\nu g)\\ &= \chi_{\Sigma(M,N)}(\nu)\gamma(\nu) \leq M, \text{ since } |\gamma(\nu)|\\ &\leq M \text{ for } \nu \in \Sigma(M,N).\end{aligned}$$

Denoting by $\psi_{a.c.}$ the orthogonal projection of ψ onto the subspace $M_{a.c.}(H_0)$, we have

$$|\frac{d}{d\nu}(\psi, E_\nu g_{M,N})|^2 \leq \left(\frac{d}{d\nu}(g_{M,N}, E_\nu g_{M,N})\right) \times \left(\frac{d}{d\nu}(\psi_{a.c.}, E_\nu \psi)\right), \quad (16)$$

so that, writing $b(\nu) = (d/d\nu)(\psi, E_\nu g_{M,N})$, we have

$$(\|b\|_2)^2 \equiv \int_{-\infty}^{\infty} |b(\nu)|^2 d\nu \leq M\|\psi\|^2. \quad (17)$$

Similarly, if

$$C(\lambda) = \frac{d}{d\lambda}(g_{M,N}, E_\lambda \phi), \quad (17')$$

we find that

$$(\|C\|_2)^2 \leq M\|\phi\|^2. \quad (17'')$$

Writing $a_\epsilon^*(\mu) = (d/d\mu)(f_\epsilon, E_\mu \psi)$, we have, provided $|\epsilon| \leq 1$,

$$\begin{aligned}|a_\epsilon^*(\mu)|^2 &= \left|[1 - I_\epsilon^+(\mu)]\, \frac{d}{d\mu}(g_{M,N}, E_\mu \psi)\right|^2\\ &\leq |1 - I_\epsilon^+(\mu)|^2 \chi_{\Sigma(M,N)}(\mu)\gamma(\mu)\, \frac{d}{d\mu}(\psi_{a.c.}, E_\mu \psi)\\ &\leq MN^2\, \frac{d}{d\mu}(\psi_{a.c.}, E_\mu \psi) \quad (18)\end{aligned}$$

since $|1 - I_\epsilon^+(\mu)| \leq N$ for $\mu \in \Sigma(M,N)$.

Hence

$$(\|a_\epsilon\|_2)^2 \leq MN^2\|\psi\|^2, \quad (19)$$

and Eq. (15) becomes

$$\|\Psi_\epsilon f_\epsilon\|^2 = 2\mathrm{Re}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{a_\epsilon^*(\mu)b(\nu)d\mu d\nu}{(\mu - \nu + 2i\epsilon)}\right) \quad (20)$$

Similarly we may evaluate $\|\Psi_\epsilon e^{-iH_0 t} f_\epsilon\|$, to obtain

$$\|\Psi_\epsilon e^{-iH_0 t} f_\epsilon\|^2 = 2\mathrm{Re}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{a_\epsilon^*(\mu)b(\nu)e^{i(\mu-\nu)t}}{(\mu - \nu + 2i\epsilon)} d\mu d\nu\right). \quad (21)$$

If we take $\epsilon > 0$ and note that both a_ϵ and b belong to $L_1(-\infty, \infty)$, we have

$$\begin{aligned}&e^{-2\epsilon t}\|\Psi_\epsilon e^{-iH_0 t} f_\epsilon\|^2\\ &\quad = -2\mathrm{Re}\left[i \int_t^{\infty} ds \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_\epsilon^*(\mu)b(\nu)e^{i(\mu-\nu+2i\epsilon)s} d\mu d\nu\right)\right],\end{aligned}$$

i.e.,

$$\begin{aligned}&\|\Psi_\epsilon e^{-iH_0 t} f_\epsilon\|^2\\ &\quad = 4\pi\, \mathrm{Im}\left(\int_t^{\infty} [\hat{a}_\epsilon(s)]^*\, \hat{b}(s)e^{-2\epsilon(s-t)} ds\right), \quad (22)\end{aligned}$$

where $\hat{a}_\epsilon$ and $\hat{b}$ denote the Fourier transforms of a_ϵ and b, respectively, and are elements of $L_2(-\infty, \infty)$. Now, in the limit as $\epsilon \to 0+$, $\lim_{\epsilon \to 0+} I_\epsilon^+(\mu)$ exists for almost all μ, being the boundary value, on the real axis, of a function analytic in the upper half-plane and having negative imaginary part. [See Ref. 11, pp. 69ff, and Appendix 2 of Ref. 10. The mapping $w = (i - z)/(i + z)$ maps the upper half-plane onto the (open) unit disc, and by the sectorial limit theorem (Ref. 11, p. 105) we may prove the existence of a sectorial limit for almost all μ.]

Hence $a_\epsilon(\mu)$ converges *pointwise* a.e. to a limit as $\epsilon \to 0+$. But if follows from Eq. (18) that $|a_\epsilon(\mu)|$ is bounded above by an element of $L_2(-\infty, \infty)$ which is independent of ϵ. Hence, using the Lebesque dominated-convergence theorem, we deduced that, as $\epsilon \to 0+$, $a_\epsilon(\mu)$ converges *in the mean* to a limit in $L_2(-\infty, \infty)$. Moreover, if this limit is denoted by a_0, than $\hat{a}_\epsilon$ converges strongly to $\hat{a}_0$ and we have, from Eq. (22),

$$\lim_{\epsilon \to 0+} \|e^{iH_0 t}\Psi_\epsilon e^{-iH_0 t} f_\epsilon\|^2 = 4\pi\, \mathrm{Im}\left(\int_t^{\infty} [\hat{a}_0(s)]^* \hat{b}(s)ds\right), \quad (23)$$

where

$$\|\hat{a}_0\|_2 \leq M^{1/2}N\|\psi\|, \quad (24)$$

$$\|\hat{b}\|_2 \leq M^{1/2}\|\psi\|. \quad (24')$$

Comparison of Eqs. (12), with $f = [1 - I_\epsilon^-(H_0)]g_{M,N}$, and Eq. (8) shows that

$$\mathrm{w\text{-}lim}_{\epsilon \to 0+}(e^{iH_0 t}\Psi_\epsilon e^{-iH_0 t} f_\epsilon - \Psi_\epsilon f_\epsilon) = 2\pi i\, F_{0t}(|\phi\rangle\langle\psi|)g_{M,N}. \quad (25)$$

Hence

$$\begin{aligned}\mathrm{w\text{-}lim}_{\epsilon \to 0+}(e^{iH_0 t}\Psi_\epsilon e^{-iH_0 t} f_\epsilon - e^{iH_0 s}\Psi_\epsilon e^{-iH_0 s} f_\epsilon)\\ = 2\pi i\, F_{st}(|\phi\rangle\langle\psi|)g_{M,N}. \quad (26)\end{aligned}$$

Therefore

$$\begin{aligned}&2\pi\|F_{st}(|\phi\rangle\langle\psi|)g_{M,N}\|\\ &\quad \leq \lim_{\epsilon > 0} \sup \|e^{iH_0 t}\Psi_\epsilon e^{-iH_0 t} f_\epsilon - e^{iH_0 s}\Psi_\epsilon e^{-iH_0 s} f_\epsilon\|\\ &\quad \leq \left[4\pi\, \mathrm{Im}\left(\int_t^{\infty} [\hat{a}_0(s)]^* \hat{b}(s)\, ds\right)\right]^{1/2}\\ &\qquad + \left[4\pi\, \mathrm{Im}\left(\int_s^{\infty} [\hat{a}_0(t)]^* \hat{b}(t)dt\right)\right]^{1/2}, \quad (27)\end{aligned}$$

from Eq. (23).

Hence

$$\lim_{s,t \to \infty} \|F_{st}(|\phi\rangle\langle\psi|)g_{M,N}\| = 0,$$

so that by the completeness of the space it follows that $F_{0t}(|\phi\rangle\langle\psi|)g_{M,N}$ converges strongly to a limit as $t \to \infty$; i.e., $g_{M,N}$ is in the domain of $F_{0\infty}(|\phi\rangle\langle\psi|)$. We may similarly consider Ψ_ϵ for negative values of ϵ, and by analogous arguments show that $g_{M,N}$ belongs to the domain of $F_{-\infty 0}(|\phi\rangle\langle\psi|)$; this completes the proof of Theorem 1.

We also have, from Eq. (8),

$$\begin{aligned}&(g_{M,N}, F_{st}(|\phi\rangle\langle\psi|)g_{M,N})\\ &\quad = \frac{1}{2\pi i} \lim_{\epsilon \to 0} \iint \frac{e^{i(\lambda-\mu)t} - e^{i(\lambda-\mu)s}}{(\lambda - \mu + i\epsilon)}\, d_\lambda(g_{M,N}, E_\lambda \phi)\\ &\qquad \times d_\mu(\psi, E_\mu g_{M,N})\end{aligned}$$

$$= \frac{1}{2\pi i} \lim_{\epsilon \to 0+} \iint \frac{e^{i(\lambda-\mu)t} - e^{i(\lambda-\mu)s}}{(\lambda - \mu + i\epsilon)} C^*(\lambda) b(\mu) d\lambda d\mu$$

$$= \int_s^t \hat{b}(k)[\hat{C}(k)]^* dk, \tag{28}$$

on using arguments similar to those used to derive Eq. (23).

Using Schwarz's inequality and Eq. (17), we find

$$|(g_{M,N}, F_{st}(|\phi\rangle\langle\psi|) g_{M,N})| \leq M^{1/2}\|\psi\| \left(\int_s^t |\hat{c}(k)|^2 dk\right)^{1/2}. \tag{29}$$

With Eq. (17″), it follows that

$$|(g_{M,N}, F_{st}(|\phi\rangle\langle\psi|) g_{M,N})| \leq M\|\phi\| \, \|\psi\|. \tag{29′}$$

It is important for the subsequent arguments to note that

(i) The bound in Eq. (29′) is independent of N.

(ii) For a given element g belonging to $M_{\text{a.c.}}(H_0)$, and for given values of M, N, the element $g_{M,N}$ and the measurable set $\Sigma(M,N)$ defined in Eq. (10′) both depend only on ϕ, but not on ψ. Hence, for a given $\phi, g_{M,N}$ belongs to the domain of $F_{-\infty\,\infty}(|\phi\rangle\langle\psi|)$ for *all* ψ.

Corollary to Theorem 1: The domain of $F_{-\infty\,\infty}(|\phi\rangle\langle\psi|)$ is dense in the subspace $M_{\text{a.c.}}(H_0)$.

Proof: Almost all real numbers k belong to $\Sigma(M,N)$ for *some* M, N, and hence for *all* M, N sufficiently large. The measure of the complement of $\Sigma(M,N)$ tends to zero as $M, N \to \infty$. Since $g \in M_{\text{a.c.}}(H_0)$, this remains true if we replace Lebesgue measure by the Lebesgue–Stieltjes measure generated by $(E_\lambda g, g)$, since the latter is absolutely continuous with respect to the former.

Hence

$$\underset{M,N\to\infty}{\text{s-lim}}\, g_{M,N} = \underset{M,N\to\infty}{\text{s-lim}}\, E_0(\Sigma(M,N))g = g. \tag{30}$$

Since each element $g_{M,N}$ belongs to the domain of $F_{\infty\,\infty}(|\phi\rangle\langle\psi|)$ and g is a general element of $M_{\text{a.c.}}(H_0)$, the conclusion of the corollary follows.

2. THE EVOLUTION OPERATOR $e^{iHt}e^{-iH_0t}$

Let G_1, G_2 be two bounded, measurable subsets of the real line, with corresponding projection operators $E_k(G_1), E_k(G_2)$, $k = 0, 1$. $E_0(G)$ is defined by Eq. (11), and $E_1(G)$ is defined similarly in terms of the resolution of the identity for H. Throughout this paper we shall use the convention that, e.g., $E_0(G_1)VE_1(G_2)$ denotes an operator defined on the *whole space* by

$$E_0(G_1)VE_1(G_2) \equiv E_0(G_1)HE_1(G_2) - H_0E_0(G_1)E_1(G_2). \tag{31}$$

For any complex number λ, with $\text{Im}\lambda \neq 0$, we define an operator $g(\lambda)$ by

$$g(\lambda) = E_0(G_1)VE_1(G_2)(\lambda - H)^{-1}E_1(G_2)VE_0(G_1). \tag{32}$$

If Σ is some closed interval of the real line containing $G_1 \cup G_2$, we define in the obvious way $g(\lambda)$ for values of λ in the entire complex plane cut along Σ, so that $g(\lambda)$ is strongly continuous, and indeed operator analytic, in λ in the cut plane. If C is a closed contour containing Σ in its interior, we define $\Delta(l)$, for real values of l, by

$$\Delta(t) = \int_C e^{i\lambda t} g(\lambda) d\lambda. \tag{33}$$

$\Delta(l)$ is a one-parameter family of bounded operators, and is independent of the precise shape of the contour C (which however, may conveniently be taken to be rectangular and to intersect the real axis in two points ξ, μ).

We define the bounded linear operator T by

$$Tf = \iint_{0 \leq t_1 \leq t_2 \leq t} dt_1 dt_2 \, e^{iH_0t_1}\Delta(t_2 - t_1)e^{-iH_0t_2}f, \tag{34}$$

and, as in Ref. 10, obtain an identity for the evolution operator by evaluating T in two ways.

We have, first of all, on using the definition of Δ and carrying out the t_1 integration,

$$Tf = \int_C d\lambda \int_0^t dt_2 \, i(\lambda - H_0)^{-1}(e^{-i(\lambda - H_0)t_2} - 1) \times g(\lambda)e^{i(\lambda - H_0)t_2}f,$$

with the convention that the integrand is taken to be strongly continuous in λ where C intersects the real axis.

We deduce that

$$Tf = \int_0^t dt_2 \, ie^{iH_0t_2}\left(\int_C (\lambda - H_0)^{-1} g(\lambda) d\lambda\right) e^{-iH_0t_2}f - \int_C d\lambda (\lambda - H_0)^{-1} g(\lambda - H_0)^{-1}(e^{i(\lambda - H_0)t} - 1)f. \tag{35}$$

Now $(\lambda - H_0)^{-1}g(\lambda)$ is strongly continuous in λ (in the cut plane), and, writing

$$E_0(G_1)VE_1(G_2) = (\lambda - H_0)E_0(G_1)E_1(G_2) - E_0(G_1)(\lambda - H)E_1(G_2),$$

we find, from Eq. (32),

$$(\lambda - H_0)^{-1}g(\lambda) = [E_0(G_1)(\lambda - H)^{-1}E_1(G_2) - (\lambda - H_0)^{-1}E_0(G_1)E_1(G_2)]E_1(G_2)VE_0(G_1) \tag{36}$$

Using the techniques of Ref. 10, to evaluate contour integrals of operator valued functions, we have

$$\int_C (\lambda - H_0)^{-1} g(\lambda) d\lambda \, g = 2\pi i[E_0(G_1)E_1(G_2) - E_0(G_1)E_1(G_2)]E_1(G_2)VE_0(G_1)g = 0 \quad \text{for any element } g$$

so that the first term on the rhs of Eq. (35) vanishes identically. To determine the second term, we have, from Eq. (36),

$$\begin{aligned}(\lambda - H_0)^{-1} g(\lambda)(\lambda - H_0)^{-1} &= E_0(G_1)(\lambda - H)^{-1}[E_1(G_2)(\lambda - H_0)E_0(G_1) \\ &\quad - (\lambda - H)E_1(G_2)E_0(G_1)](\lambda - H_0)^{-1} \\ &\quad - (\lambda - H_0)^{-1}E_0(G_1)E_1(G_2)VE_0(G_1)(\lambda - H_0)^{-1} \\ &= E_0(G_1)(\lambda - H)^{-1}E_1(G_2)E_0(G_1) \\ &\quad - E_0(G_1)E_1(G_2)(\lambda - H_0)^{-1}E_0(G_1) \\ &\quad - (\lambda - H_0)^{-1}E_0(G_1)E_1(G_2)VE_0(G_1)(\lambda - H_0)^{-1}.\end{aligned} \tag{37}$$

By operating on the right by $e^{i(\lambda-H_0)t}-1$ and carrying out the λ integration, Eq. (35) becomes

$$Tf = -2\pi i E_0(G_1)E_1(G_2)[e^{iHt}e^{-iH_0t}-1]E_0(G_1)f + \int_C d\lambda(\lambda-H_0)^{-1}E_0(G_1)E_1(G_2)VE_0(G_1)(\lambda-H_0)^{-1} \times [e^{i(\lambda-H_0)t}-1]f. \quad (38)$$

We can write

$$(\lambda-H_0)^{-1}[e^{i(\lambda-H_0)t}-1]f = i\int_0^t ds e^{i(\lambda-H_0)s}f, \quad (39)$$

so that by integrating with respect to λ the contour integral becomes

$$-2\pi\int_0^t ds e^{iH_0s}E_0(G_1)E_1(G_2)VE_0(G_1)e^{-iH_0s}f,$$

and with the notation of Eq. (2) we have

$$Tf = -2\pi i E_0(G_1)E_1(G_2)[e^{iHt}e^{-iH_0t}-1]E_0(G_1)f - (2\pi)^2F_{0t}(E_0(G_1)E_1(G_2)VE_0(G_1)). \quad (40)$$

Using Eqs. (32) and (33), we find

$$\Delta(t) = 2\pi i E_0(G_1)VE_1(G_2)e^{iHt}E_1(G_2)VE_0(G_1). \quad (41)$$

Thus $\Delta(t)$ is continuous in t, with the operator norm. [It may readily be deduced from the spectral theorem that $e^{iHt}E_1(G_2)$ is continuous in t.]

We may therefore define the one-parameter families of operators A_t, B_t, and L_t by

$$A_t = \int_0^t \Delta(u)e^{-iH_0u}du, \quad (42)$$

$$B_t = \int_0^t e^{-iH_0u}\Delta(u)\,du, \quad (42')$$

and

$$L_t = \int_0^t dv\, F_{0,t-v}(\Delta(v)e^{-iH_0v}) - F_{0t}(B_t). \quad (42'')$$

Now, from Eq. (3) we have

$$L_t = \int_0^t dv e^{-iH_0v}F_{v,t}(\Delta(v)) - F_{0t}(B_t) = \int_0^t dv F_{0t}(e^{-iH_0v}\Delta(v)) - F_{0t}(B_t) - \int_0^t dv F_{0v}(e^{-iH_0v}\Delta(v)). \quad (43)$$

Writing the matrix element $([\int_0^t dv\, F_{0t}(e^{-iH_0v}\Delta(v))]f, g)$ as a double integral, it is easily seen that we may interchange orders of integration and deduce that

$$\int_0^t dv\, F_{0t}(e^{-iH_0v}\Delta(v)) = F_{0t}\left(\int_0^t dv\, e^{-iH_0v}\Delta(v)\right) = F_{0t}(B_t),$$

so that we have the alternative formula for L_t,

$$L_t = -\int_0^t dv\, F_{0v}(e^{-iH_0v}\Delta(v)). \quad (44)$$

Theorem 2: Suppose L_t is compact, and let f be any element of the space belonging to $M_{\text{a.c.}}(H_0)$. Then

$$E_0(G_1)E_1(G_2)[e^{iHt}e^{-iH_0t}-1]E_0(G_1)f = 2\pi i F_{0t}(E_0(G_1)E_1(G_2)VE_0(G_1))f + iF_{0\infty}(A_t-B_t)f + iF_{0t}(B_t)f. \quad (45)$$

Proof: From Eq. (44), we have

$$L_t - e^{iH_0s}L_te^{-iH_0s} = \int_0^t dv[-F_{0v}(e^{-iH_0v}\Delta(v)) + e^{iH_0s}F_{0v}(e^{-iH_0v}\Delta(v))e^{-iH_0s}].$$

The integrand may be written

$$\begin{aligned} &-F_{0v}(e^{-iH_0v}\Delta(v)) + e^{iH_0s}F_{0,v-s}(e^{-iH_0v}\Delta(v))e^{-iH_0s} + e^{iH_0s}F_{v-s,v}(e^{-iH_0v}\Delta(v))e^{-iH_0s} \\ &= -F_{0v}(e^{-iH_0v}\Delta(v)) + e^{iH_0s}F_{0,v-s}(e^{-iH_0v}\Delta(v))e^{-iH_0s} + e^{iH_0(s-v)}F_{v-s,v}(\Delta(v)e^{-iH_0v})e^{-iH_0(s-v)} \\ &= -F_{0v}(e^{-iH_0v}\Delta(v)) + F_{s,v}(e^{-iH_0v}\Delta(v)) + F_{0s}(\Delta(v)e^{-iH_0v}) = F_{0s}(\Delta(v)e^{-iH_0v} - e^{-iH_0v}\Delta(v)). \end{aligned}$$

Hence

$$L_t - e^{iH_0s}L_te^{-iH_0s} = \int_0^t dv\, F_{0s}(\Delta(v)e^{-iH_0v} - e^{-iH_0v}\Delta(v)) = F_{0s}\left(\int_0^t dv\,\Delta(v)e^{-iH_0v} - \int_0^t dv\, e^{-iH_0v}\Delta(v)\right), \quad (46)$$

by the same argument as we used above.

Therefore $L_tf - e^{iH_0s}L_te^{-iH_0s}f = F_{0s}(A_t-B_t)f. \quad (47)$

Since L_t is compact and f belongs to $M_{\text{a.c.}}(H_0)$, it follows that $\lim_{s\to\pm\infty}\|L_te^{-iH_0s}f\| = 0$. An immediate consequence is that

$$L_tf = F_{0\infty}(A_t-B_t)f = F_{0,-\infty}(A_t-B_t)f. \quad (48)$$

Now, from Eq. (34), we have, by the change of variables

$$t_1 = u, \quad t_2 - t_1 = v,$$

$$Tf = \int_{\substack{u+v\le t \\ u,v\ge 0}} \int dudv\, e^{iH_0u}\Delta(v)e^{-iH_0u}e^{-iH_0v}f,$$

which on carrying out the u integration gives

$$Tf = 2\pi\int_0^t dv\, F_{0,t-v}(\Delta(v)e^{-iH_0v})f = 2\pi[L_t + F_{0t}(B_t)]f = 2\pi F_{0\infty}(A_t-B_t)f + 2\pi F_{0t}(B_t)f, \quad (49)$$

on using Eqs. (43) and (48) and applying Eq. (3). Comparing this with Eq. (40), we may deduce Eq. (45), and this completes the proof of the theorem.

3. PROOF OF THE EXISTENCE OF THE WAVE OPERATORS

We now restrict the discussion to scattering systems for which there exist (nontrivial) bounded measurable subsets G_1, G_2 of the real line such that $E_1(G_2)VE_0(G_1)$ is of trace class.

Suppose then that $E_1(G_2)VE_0(G_1)$ is of trace class. Then $E_1(G_2)VE_0(G_1)$ has a (nonunique) representation of the form

$$E_1(G_2)VE_0(G_1) = \sum_{i=1}^{\infty}\alpha_i|\phi_i\rangle\langle\phi_i|, \quad (50)$$

where $\|\phi_i\| = 1$ and the α's are complex numbers such that $\sum_{i=1}^{\infty} |\alpha_i| < \infty$. The rhs of Eq. (50) converges in the Banach space of bounded linear operators on the Hilbert space.

Equation (41) shows that $\Delta(t)$ is also of trace class. Hence $e^{-iH_0 v}\Delta(v)$ is compact. But $F_{0v}(e^{-iH_0 v}\Delta(v))$ may be expressed as a limit (in the Banach space of bounded linear operators) of approximating Riemann sums, and since any limit of compact operators is itself compact, it follows that $F_{0v}(e^{-iH_0 v}\Delta(v))$ is compact. By the same argument, it follows from Eq. (44) that L_t is compact. Hence the conditions of Theorem 2 are satisfied, so that Eq. (45) holds for any f belonging to $M_{\text{a.c.}}(H_0)$.

From Eq. (50) it follows that

$$E_0(G_1)E_1(G_2)VE_0(G_1) = \sum_{i=1}^{\infty} \alpha_i |\tilde{\phi}_i\rangle\langle\tilde{\phi}_i|, \tag{50'}$$

where $\tilde{\phi}_i = E_0(G_1)\phi_i$.

Similar representations may be derived for A_t and B_t. Thus we have, for any element x,

$$\int_0^t du\, e^{iHu}E_1(G_2)VE_0(G_1)e^{-iH_0 u}x$$

$$= -i\int_0^t du\, \frac{d}{du}[e^{iHu}E_1(G_2)E_0(G_1)e^{-iH_0 u}x]$$

$$= -iE_1(G_2)(e^{iHt}e^{-iH_0 t} - 1)E_0(G_1)x,$$

where "d/du" denotes a strong derivative.

Hence from Eqs. (41) and (42) we find

$$A_t = 2\pi E_0(G_1)VE_1(G_2)(e^{iHt}e^{-iH_0 t} - 1)E_0(G_1), \tag{51}$$

so that

$$A_t = \sum_{i=1}^{\infty} \alpha_i^* |\phi_i\rangle\langle\psi_i(t)|, \tag{52}$$

where

$$\psi_i(t) = 2\pi E_0(G_1)[e^{iH_0 t}e^{-iHt} - 1]\phi_i, \tag{53}$$

and

$$\|\psi_i(t)\| \leq 4\pi. \tag{54}$$

It follows similarly that

$$B_t = A_{-t}^* = \sum_{i=1}^{\infty} \alpha_i |\psi_i(-t)\rangle\langle\phi_i|. \tag{55}$$

Now let M be any positive number and let $\{N_i\}$, $i = 1, 2, 3, \cdots$, be any sequence of positive numbers. Given any element g belonging to $M_{\text{a.c.}}(H_0)$, we define $\gamma(\lambda)$ by Eq. (9'). For each value of i, we take $\phi = \phi_i$, $N = N_i$ in Eqs. (10), (10'), and (10") and make the following definitions:

$$S_i(N_i) = \{k; |1 - I^{\pm}_{\alpha,i}(k)| \leq N_i \text{ for all } \alpha \text{ in the interval } 0 < \alpha \leq 1\} \tag{56}$$

$$\Sigma_i(M, N_i) = \{k; |\gamma(k)| \leq M\} \cap S_i(N_i), \tag{56'}$$

where $I^{\pm}_{\alpha,i}(k) \equiv (\phi_i, [k - H_0 \pm i\alpha]^{-1}\phi_i)$. (56")

Each $\Sigma_i(M, N_i)$ is defined modulo a set of measure zero. We denote the corresponding sets, obtained by replacing ϕ_i by $\tilde{\phi}_i$ in the definition of $I^{\pm}_{\alpha,i}(k)$, by $\tilde{S}_i$ and $\tilde{\Sigma}_i(M, N_i)$, respectively.

The following lemma enables us to deduce the existence of the wave operator:

Lemma 1: Suppose that $E_1(G_2)VE_0(G_1)$ is of trace class, and let g belong to $M_{\text{a.c.}}(H_0)$.

Define, for any $M, \{N_i\}$,

$$g_M = E_0\left(\bigcap_{i=1}^{\infty} [\Sigma_i(M, N_i) \cap \tilde{\Sigma}_i(M, N_i)]\right)g. \tag{57}$$

Then

$$\lim_{s,t\to\pm\infty} (E_0(G_1)e^{-iH_0 t}g_M, E_1(G_2) \times [e^{iH(s-t)}e^{-iH_0(s-t)} - 1]E_0(G_1)e^{-iH_0 t}g_M) = 0. \tag{58}$$

Proof: We consider first the limit as $s, t \to +\infty$. The proof will depend on the following result:

Let $\{C_i(s,t)\}$, $i = 1, 2, 3, \cdots$, be a sequence of complex-valued functions of two real variables s, t. Suppose that there exists a sequence $\{d_i\}$ of real positive constants such that

$$|C_i(s,t)| < d_i \text{ for all } s, t$$

and

$$\sum_1^{\infty} d_i < \infty.$$

Then if $\lim_{s,t\to\infty} C_i(s,t) = C_i$, it follows that $\lim_{s,t\to\infty} \sum_1^{\infty} C_i(s,t) = \sum_1^{\infty} C_i$.

We shall also use the corresponding result for functions of a single real variable. Both results are special cases of the Lebesgue dominated-convergence theorem.

By making the substitutions $t \to (s - t)$, $f \to e^{-iH_0 t}g_M$ in Eq. (45) and taking the inner product with $e^{-iH_0 t}g_M$, the lhs of Eq. (58) becomes, on using Eqs. (3) and (3'),

$$\lim_{s,t\to+\infty} [-2\pi i(g_M, F_{s,t}(E_0(G_1)E_1(G_2)VE_0(G_1))g_M) + i(g_M, F_{t\infty}(A_{s-t} - B_{s-t})g_M) - i(g_M, F_{s,t}(B_{s-t}))g_M]. \tag{58'}$$

Now

$$(g_M, F_{t\infty}(A_{s-t} - B_{s-t})g_M) = \lim_{u\to\infty} (g_M, F_{t,u}(A_{s-t} - B_{s-t})g_M)$$

$$= \lim_{u\to\infty} [(g_M, F_{tu}(A_{s-t})g_M) - (g_M, F_{tu}(B^*_{s-t})g_M)^*]$$

$$= \lim_{u\to\infty}\left(\sum_{i=1}^{\infty} \alpha_i^*(g_M, F_{tu}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M) - \sum_{i=1}^{\infty} \alpha_i(g_M, F_{tu}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)^*\right). \tag{59}$$

Now $g_M = E_0(\Sigma_i(M, N_i))g_M$, so that from Theorem 1 we know that g_M belongs to the respective domains of $F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)$ and $F_{t\infty}(|\phi_i\rangle\langle\psi_i(t-s)|)$.

We also have, from Eqs. (29') and (54),

$$|(g_M, F_{tu}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M)| \leq M\|\phi_i\|\cdot\|\psi_i(s-t)\| \leq 4\pi M \tag{60}$$

and

$$|(g_M, F_{tu}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)| \leq 4\pi M. \tag{60'}$$

Moreover, $\sum_{i=1}^{\infty} |\alpha_i| < \infty$, so that we may take the limit under the summation sign, to obtain

$$\lim_{u\to\infty}\Big(\sum_{i=1}^{\infty}\alpha_i^*(g_M,F_{tu}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M) - \sum_{i=1}^{\infty}\alpha_i(g_M,F_{tu}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)^*\Big)$$
$$= \sum_{i=1}^{\infty}[\alpha_i^*(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M) - \alpha_i(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)^*], \quad (61)$$

where again we have the inequalities

$$|(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M)| \le 4\pi M,$$
$$|(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)| \le 4\pi M.$$

From Eq. (29), we have

$$|(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M)| \le 4\pi M^{1/2}\times\Big(\int_t^{\infty}|\hat{C}(k)|^2dk\Big)^{1/2},$$

where from Eq. (17′) it may be verified that $\hat{C}(k)$ depends on ϕ_i, but not on s or t.

Hence

$$\lim_{s,t\to\infty}(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M) = 0.$$

Similarly

$$\lim_{s,t\to\infty}(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M) = 0.$$

We may again take the limit under the integral sign, so that

$$\lim_{s,t\to\infty}(g_M,F_{t\infty}(A_{s-t}-B_{s-t})g_M) = \lim_{s,t\to\infty}\Big(\sum_{i=1}^{\infty}\alpha_i(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(s-t)|)g_M) - \sum_{i=1}^{\infty}\alpha_i(g_M,F_{t\infty}(|\phi_i\rangle\langle\psi_i(t-s)|)g_M)^*\Big) = 0. \quad (62)$$

Similarly,

$$\lim_{s,t\to\infty}(g_M,F_{st}(B_{s-t})g_M) = 0; \quad (62')$$

and, by noting that $g_M = E_0(\tilde{\Sigma}_i(M,N_i))g_M$ and using the representation (50′), we also have

$$\lim_{s,t\to\infty}(g_M,F_{st}(E_0(G_1)E_1(G_2)VE_0(G_1))g_M) = 0 \quad (62'')$$

so that the limit in (58′) vanishes, and we have shown that Eq. (58) holds in the limit as $s,t\to\infty$. The proof that Eq. (58) holds in the limit as $s,t\to-\infty$ is very similar, except that we use Eq. (48) to substitute $F_{t,-\infty}(A_{s-t}-B_{s-t})$ for $F_{t\infty}(A_{s-t}-B_{s-t})$ in (58′).

This completes the proof of Lemma 1.

Corollary to Lemma 1: Suppose that $E_1(G_2)VE_0(G_1)$ is of trace class, and let g belong to $M_{\text{a.c.}}(H_0)$. Then

$$\lim_{s,t\to\pm\infty}(E_0(G_1)e^{-iH_0t}g, E_1(G_2)\times[e^{iH(s-t)}e^{-iH_0(s-t)}-1]E_0(G_1)e^{-iH_0t}g) = 0. \quad (63)$$

Proof: The arguments used to derive the Corollary to Theorem 1 show also that if $h\in M_{\text{a.c.}}(H_0)$, then

$$\text{s-}\lim_{M\to\infty}E_0(\Gamma(M))h = h,$$

where $\Gamma(M) = \{k;\ |\gamma(k)|\le M\}$,

and that, for each value of i,

$$\text{s-}\lim_{N_i\to\infty}E_0(S_i(N_i)\cap\tilde{S}_i(N_i))h = h.$$

Now let ϵ be any positive number.

Define inductively $g^{(i)}$, $i=1,2,3,\cdots$, by

$$g^{(1)} = E_0(S_1(N_1)\cap\tilde{S}_1(N_1))g,$$
$$g^{(i+1)} = E_0(S_{i+1}(N_{i+1})\cap\tilde{S}_{i+1}(N_{i+1}))g^{(i)},$$

and choose successively the positive numbers N_1, $N_2,\cdots$ such that

$$\|g^{(1)}-g\| < \epsilon/2,$$
$$\|g^{(i+1)}-g^{(i)}\| < \epsilon/2^{i+1}, \quad i=1,2,3,\cdots.$$

Then

$$\|E_0\Big(\bigcap_1^{\infty}[S_i(N_i)\cap\tilde{S}_i(N_i)]\Big)g-g\| = \|\text{s-}\lim_{i\to\infty}g^{(i)}-g\| < \epsilon\sum_0^{\infty}\frac{1}{2^{i+1}} = \epsilon.$$

If we now choose M such that

$$\|E_0(\Gamma(M))E_0\Big(\bigcap_1^{\infty}[S_i(N_i)\cap\tilde{S}_i(N_i)]\Big)\times g - E_0\Big(\bigcap_1^{\infty}[S_i(N_i)\cap\tilde{S}_i(N_i)]\Big)g\| < \epsilon,$$

we have, from Eqs. (56′) and (57),

$$\|g_M-g\| < 2\epsilon;$$

i.e., given any $\epsilon>0$, we can choose M and $\{N_i\}$ such that $\|g_M-g\|<2\epsilon$.

We then choose $N_0>0$ such that

$$|(E_0(G_1)e^{-iH_0t}g_M, E_1(G_2)[e^{iH(s-t)}e^{-iH_0(s-t)}-1]E_0(G_1)e^{-iH_0t}g_M)| < \epsilon,$$

provided that $s,t>N_0$ (or alternatively $s,t<-N_0$).

Noting that $\|e^{iH(s-t)}e^{-iH_0(s-t)}-1\|\le 2$ it is a simple exercise, using the Schwarz inequality, to show that if $s,t>N_0$, then

$$|(E_0(G_1)e^{-iH_0t}g, E_1(G_2)[e^{iH(s-t)}e^{-iH_0(s-t)}-1]\times E_0(G_1)e^{-iH_0t}g)| < \epsilon(1+8\|g\|). \quad (64)$$

Since ϵ is arbitrary, we have verified Eq. (63), and this completes the proof of the corollary.

We are now ready to state the main theorem of the paper. We adopt the notation that if G is a subset of the real line, then $\bar{G}$ denotes the complement of G, so that, for example, $E_0(\bar{G}) = 1-E_0(G)$.

Theorem 3: Let g be any element belonging to $M_{\text{a.c.}}(H_0)$. Suppose that, for any $\epsilon>0$ there exists a positive number n and bounded measurable subsets G_1, G_2 of the real line (n, G_1, G_2 depending on ϵ) such

that

(i) $E_1(G_2)VE_0(G_1)$ is of trace class,

(ii) $\|E_0(\bar{G}_1)g\| < \epsilon$,

(iii) $\|E_1(\bar{G}_2)e^{-iH_0t}g\| < \epsilon$ for all $t > n$.

Then g belongs to the domain of the wave operator Ω_-. [If (iii) holds for all $t < -n$, then g belongs to the domain of Ω_+.]

Proof: Given any $\epsilon > 0$, choose n, G_1, G_2 such that (i), (ii), (iii) are satisfied. Then

$$\|E_0(G_1)e^{-iH_0t}g - e^{-iH_0t}g\| = \|e^{-iH_0t}E_0(\bar{G}_1)g\| < \epsilon \quad (65)$$

and

$$\begin{aligned}\|E_1(G_2)E_0(G_1)e^{-iH_0t}g - e^{-iH_0t}g\| &= \|(1 - E_1(\bar{G}_2))(1 - E_0(\bar{G}_1))e^{-iH_0t}g - e^{-iH_0t}g\| \\ &\leq 2\|E_0(\bar{G}_1)e^{-iH_0t}g\| + \|E_1(\bar{G}_2)e^{-iH_0t}g\| < 3\epsilon. \quad (66)\end{aligned}$$

But, from Eq. (63), we may choose s, t, sufficiently large that

$$|(E_0(G_1)e^{-iH_0t}g, E_1(G_2)[e^{iH(s-t)}e^{-iH_0(s-t)} - 1] \times E_0(G_1)e^{-iH_0t}g)| < \epsilon,$$

and we may deduce that

$$|(e^{-iH_0t}g, [e^{iH(s-t)}e^{-iH_0(s-t)} - 1]e^{-iH_0t}g)| < \epsilon(1 + 8\|g\|). \quad (67)$$

Since ϵ is arbitrary, it follows that

$$\lim_{s,t\to\infty} (e^{-iH_0t}g, [e^{iH(s-t)}e^{-iH_0(s-t)} - 1]e^{-iH_0t}g) = 0. \quad (68)$$

Now in Ref. 9 we proved the elementary identity

$$\begin{aligned}\|(e^{iHs}e^{-iH_0s} - e^{iHt}e^{-iH_0t})g\|^2 &= -2\mathrm{Re}\{(e^{-iH_0t}g, [e^{iH(s-t)}e^{-iH_0(s-t)} - 1]e^{-iH_0t}g)\}, \quad (69)\end{aligned}$$

so that we now have

$$\lim_{s,t\to\infty} \|e^{iHs}e^{-iH_0s}g - e^{iHt}e^{-iH_0t}g\| = 0. \quad (70)$$

By the completeness of the Hilbert space, it follows that $\text{s-lim}_{t\to\infty}\, e^{iHt}e^{-iH_0t}g$ exists, so that g belongs to the domain of Ω_-.

This completes the proof of Theorem 3.

Corollary 1: In this statement of Theorem 3, condition (iii) may be replaced by

(iii'): $\|E_1(\bar{G}_2)E_0(G_1)e^{-iH_0t}g\| < \epsilon$ for all $t > n$.

Proof: If

$$\|E_0(\bar{G}_1)g\| < \epsilon'/2 \text{ and } \|E_1(\bar{G}_2)E_0(G_1)e^{-iH_0t}g\| < \epsilon'/2,$$

then

$$\begin{aligned}\|E_1(\bar{G}_2)e^{-iH_0t}g\| &\leq \|E_1(\bar{G}_2)E_0(G_1)e^{-iH_0t}g\| \\ &+ \|E_1(\bar{G}_2)E_0(\bar{G}_1)e^{-iH_0t}g\| < \epsilon'/2 + \epsilon'/2 = \epsilon'. \quad (71)\end{aligned}$$

Hence if n, G_1, G_2 may be chosen, for each ϵ, so that (i), (ii), (iii') are satisfied, then we can also satisfy (i), (ii), (iii) for each ϵ and the conclusion of Theorem 3 remains valid.

Corollary 2: Suppose that, for *every* bounded measurable set G, $E_1(G)VE_0(G)$ is of trace class. Then a necessary and sufficient condition for an element g belonging to $M_{\text{a.c.}}(H_0)$ to belong to the domain of Ω_- is that: Given any $\epsilon > 0$ there exists a positive number n and a bounded measurable set G such that

$$\|E_1(\bar{G})e^{-iH_0t}g\| < \epsilon \quad \text{for all } t > n.$$

Proof: That this condition is *sufficient* follows from Theorem 3. Conversely, suppose g belongs to the domain of Ω_-. Then

$$\begin{aligned}\lim_{t\to\infty} \|E_1(\bar{G})e^{-iH_0t}g\| &= \lim_{t\to\infty} \|E_1(\bar{G})e^{iHt}e^{-iH_0t}g\| \\ &= \|E_1(\bar{G})\Omega_-g\|.\end{aligned}$$

Take G to be the interval $(-a, a)$, and take a sufficiently large so that $\|E_1(\bar{G})\Omega_-g\| < \epsilon/2$.

Then n may be found such that $\|E_1(\bar{G})e^{-iH_0t}g\| < \epsilon$ for all $t > n$, and this completes the proof of Corollary 2.

CONCLUSION

A number of special cases of Theorem 3 are important in practice. Three examples are as follows:

(i) Suppose $E_1(G)VE_0(G)$ is of trace class. Then the element $g = E_0(G)f$, where $f \in M_{\text{a.c.}}(H_0)$, belongs to the domain of Ω_- if and only if

$$\text{s-}\lim_{t\to\infty} E_1(\bar{G})e^{-iH_0t}g = 0.$$

(ii) Suppose $E_1(G)VE_0(G)$ is of trace class and that $E_1(\bar{G})E_0(G)$ is compact. Then every element g of the form $g = E_0(G)f$, where $f \in M_{\text{a.c.}}(H_0)$, belongs to the domain of $\Omega_\pm$.

(iii) Suppose $E_1(G)VE_0(G)$ is of trace class for *every* bounded measurable subset G of the real line and that, for every element f in the domain of H_0,

$$\|Hf\| < a\|H_0f\| + b\|f\|, \quad \text{for some constants } a, b.$$

In that case, $(H + i)(H_0 + i)^{-1}$ is bounded, and, setting $g = (H_0 + i)f$, we have

$$\begin{aligned}\|E_1(\bar{G}_2)e^{-iH_0t}f\| &= \|E_1(\bar{G}_2)(H + i)^{-1}[(H + i)(H_0 + i)^{-1}]e^{-iH_0t}g\| \\ &\leq \text{const}\,\|E_1(\bar{G}_2)(H + i)^{-1}\| \cdot \|g\|.\end{aligned}$$

Taking G_2 to be the interval $(-\alpha, +\alpha)$, we have

$$\|E_1(\bar{G}_2)(H + i)^{-1}\| \leq (1 + \alpha^2)^{-1/2}.$$

Hence

$$\lim_{\alpha\to\infty} \|E_1(\bar{G}_2)e^{-iH_0t}f\| = 0,$$

and we may verify that the conditions of Theorem 3 are satisfied for f to belong to the domain of $\Omega_\pm$.

Since the domain of H_0 is dense, it follows that in this case $\Omega_\pm$ is defined on the entire Hilbert space.

The applications of the results of this paper to particular scattering systems will be considered in a subsequent paper.

[1] T. Kato, Proc. Japan. Acad. **33**, 260 (1957).
[2] S. T. Kuroda, J. Math. Soc. Japan **11**, 247 (1959).
[3] C. R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer-Verlag, Berlin, 1967), p. 96 *et seq.*
[4] M. Rosenblum, Pacific J. Math. **7**, 997 (1957).
[5] T. Kato, Pacific J. Math. **15**, 171 (1965).
[6] S. T. Kuroda, Nuovo Cimento **12**, 431 (1969).
[7] M. S. Birman, Sov. Math. Dokl. **5**, 1505 (1964).
[8] D. B. Pearson, Nuovo Cimento **2A**, 853 (1971).
[9] D. B. Pearson, J. Math. Phys. **11**, 2425 (1970).
[10] M. A. Grubb and D. B. Pearson, J. Math. Phys. **11**, 2415 (1970).
[11] M. Heins, *Selected Topics in the Classical Theory of Functions of a Complex Variable* (Holt, Rinehart, and Winston, New York, 1962).

On the Connection between Levinson's Theorem and Singular Integral Equations

C. M. Becchi and R. Collina*
Istituto di Scienze Fisiche dell'Università, Genova, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy
(Received 22 February 1972)

Levinson's theorem is deduced from a general property of singular integral equations in the case in which both the unperturbed and the total Hamiltonian have a finite number of discrete eigenvalues. We also discuss the conditions of validity of the theorem.

1. INTRODUCTION

Dyson, Fairlie and Polkinghorne, and Lee and Klein[1] have studied the scattering problem for a system whose unperturbed Hamiltonian H_0 has discrete eigenvalues.

The purpose of these authors was to clairfy the physical meaning of the redundant solutions of Low's scattering equation.[2] They analyzed a variety of models in which the scattering amplitudes satisfy the same Low equation and can be calculated directly by solving the Schrödinger equation. They could show that every CDD solution[3] of the Low equation is the physically correct scattering amplitude for a particular model and that there is a one-to-one correspondence between the discrete eigenvalues of H_0 and the CDD poles.

Discussing the solutions of the Low equation, Haag[4] showed that the Levinson theorem should be written, in the nonrelativistics case, in the form

$$[\delta(0) - \delta(\infty)]/\pi = n - N, \tag{1}$$

where n is the number of bound states of the system and N the number of CDD poles.

In the particular case of the models discussed by Dyson et al, Haag's result can be rewritten in the following way:

$$[\delta(0) - \delta(\infty)]/\pi = n - n_0, \tag{2}$$

where n and n_0 are the numbers of the discrete eigenvalues of the total Hamiltonian and of the unperturbed Hamiltonian, respectively.

More recently, by using the channel inelastic N/D equation, a generalized form of the Levinson theorem has been proved for analytic scattering amplitudes.[5] This generalized form can be written in the same way as formula (1) where n is the total number of bound states and ghosts.

Our purpose is to discuss the validity of the extension of the Levinson theorem given in formula (2) for the case in which both the unperturbed Hamiltonian H_0 and the total Hamiltonian $H = H_0 + V$ have a finite number of discrete eigenvalues.

For the sake of simplicity we shall start treating the elastic scattering of two spinless particles in a single partial wave. However the extension to the scattering in any finite number of two-body channels can be given in a completely analogous way.

We shall assume the spectra of both H and H_0 to be orthonormal and complete. Then we shall deduce the Levinson theorem directly from the orthogonality condition of the continuous and discrete eigenvectors of H.[6] This condition, in the representation in which H_0 is diagonal, is a singular integral equation (s.i.e.), which has as many independent solutions as the discrete eigenvectors of H. The Levinson theorem follows directly from classical results of the theory of s.i.e.

Our approach to the problem will be completely formal at the beginning. But before using the theory of s.i.e. we shall assume some appropriate mathematical conditions on the reaction matrix. Those conditions will also guarantee the validity of the previous deductions.[7,8]

In Sec. 2 we shall give a brief description of the formalism, and then we shall deduce the singular integral equation. Its independent solutions will be counted in Sec. 3.

In Sec. 4 we shall prove the Levinson theorem, and we shall state the assumed mathematical properties of the reaction matrix. The corresponding conditions for the potential are given in Sec. 5 together with the extension of our treatment to multichannel scattering.

2. DEDUCTION OF THE SINGULAR INTEGRAL EQUATION

We denote the discrete eigenvectors and the corresponding eigenvalues of H_0 by $|\phi_i\rangle$ and $E_i^{(0)}$, respectively $(i = 1, \ldots, n_0)$, and the continuous eigenvectors of H_0 by $|\phi(E)\rangle$ for $0 \leq E < \infty$. In an analogous way we denote the discrete eigenvectors and eigenvalues of H by $|\psi_i\rangle$ and $E_i (i = 1, \ldots, n)$. The kets $|\psi^+(E)\rangle$ for $0 \leq E < \infty$ are the continuous eigenvectors of H and are normalized as incoming waves.

Expanding $|\psi^+(E)\rangle$ in terms of the complete set of the eigenstates of H_0, we obtain

$$|\psi^+(E)\rangle = \int_0^\infty dE' \psi(E', E) |\phi(E')\rangle + \sum_{i=1}^{n_0} \psi_i(E) |\phi_i\rangle, \tag{3}$$

where $\psi_i(E)$ are square integrable functions, and $\psi(E', E) = \langle \phi(E') | \psi^+(E) \rangle$ is connected to the matrix element $R(E'; E)$ of the reaction operator by the relation[9]

[1] T. Kato, Proc. Japan. Acad. **33**, 260 (1957).
[2] S. T. Kuroda, J. Math. Soc. Japan **11**, 247 (1959).
[3] C. R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer-Verlag, Berlin, 1967), p. 96 *et seq.*
[4] M. Rosenblum, Pacific J. Math. **7**, 997 (1957).
[5] T. Kato, Pacific J. Math. **15**, 171 (1965).
[6] S. T. Kuroda, Nuovo Cimento **12**, 431 (1969).
[7] M. S. Birman, Sov. Math. Dokl. **5**, 1505 (1964).
[8] D. B. Pearson, Nuovo Cimento **2A**, 853 (1971).
[9] D. B. Pearson, J. Math. Phys. **11**, 2425 (1970).
[10] M. A. Grubb and D. B. Pearson, J. Math. Phys. **11**, 2415 (1970).
[11] M. Heins, *Selected Topics in the Classical Theory of Functions of a Complex Variable* (Holt, Rinehart, and Winston, New York, 1962).

On the Connection between Levinson's Theorem and Singular Integral Equations

C. M. Becchi and R. Collina*
Istituto di Scienze Fisiche dell'Università, Genova, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy
(Received 22 February 1972)

Levinson's theorem is deduced from a general property of singular integral equations in the case in which both the unperturbed and the total Hamiltonian have a finite number of discrete eigenvalues. We also discuss the conditions of validity of the theorem.

1. INTRODUCTION

Dyson, Fairlie and Polkinghorne, and Lee and Klein[1] have studied the scattering problem for a system whose unperturbed Hamiltonian H_0 has discrete eigenvalues.

The purpose of these authors was to clairfy the physical meaning of the redundant solutions of Low's scattering equation.[2] They analyzed a variety of models in which the scattering amplitudes satisfy the same Low equation and can be calculated directly by solving the Schrödinger equation. They could show that every CDD solution[3] of the Low equation is the physically correct scattering amplitude for a particular model and that there is a one-to-one correspondence between the discrete eigenvalues of H_0 and the CDD poles.

Discussing the solutions of the Low equation, Haag[4] showed that the Levinson theorem should be written, in the nonrelativistics case, in the form

$$[\delta(0) - \delta(\infty)]/\pi = n - N, \tag{1}$$

where n is the number of bound states of the system and N the number of CDD poles.

In the particular case of the models discussed by Dyson et al, Haag's result can be rewritten in the following way:

$$[\delta(0) - \delta(\infty)]/\pi = n - n_0, \tag{2}$$

where n and n_0 are the numbers of the discrete eigenvalues of the total Hamiltonian and of the unperturbed Hamiltonian, respectively.

More recently, by using the channel inelastic N/D equation, a generalized form of the Levinson theorem has been proved for analytic scattering amplitudes.[5] This generalized form can be written in the same way as formula (1) where n is the total number of bound states and ghosts.

Our purpose is to discuss the validity of the extension of the Levinson theorem given in formula (2) for the case in which both the unperturbed Hamiltonian H_0 and the total Hamiltonian $H = H_0 + V$ have a finite number of discrete eigenvalues.

For the sake of simplicity we shall start treating the elastic scattering of two spinless particles in a single partial wave. However the extension to the scattering in any finite number of two-body channels can be given in a completely analogous way.

We shall assume the spectra of both H and H_0 to be orthonormal and complete. Then we shall deduce the Levinson theorem directly from the orthogonality condition of the continuous and discrete eigenvectors of H.[6] This condition, in the representation in which H_0 is diagonal, is a singular integral equation (s.i.e.), which has as many independent solutions as the discrete eigenvectors of H. The Levinson theorem follows directly from classical results of the theory of s.i.e.

Our approach to the problem will be completely formal at the beginning. But before using the theory of s.i.e. we shall assume some appropriate mathematical conditions on the reaction matrix. Those conditions will also guarantee the validity of the previous deductions.[7,8]

In Sec. 2 we shall give a brief description of the formalism, and then we shall deduce the singular integral equation. Its independent solutions will be counted in Sec. 3.

In Sec. 4 we shall prove the Levinson theorem, and we shall state the assumed mathematical properties of the reaction matrix. The corresponding conditions for the potential are given in Sec. 5 together with the extension of our treatment to multichannel scattering.

2. DEDUCTION OF THE SINGULAR INTEGRAL EQUATION

We denote the discrete eigenvectors and the corresponding eigenvalues of H_0 by $|\phi_i\rangle$ and $E_i^{(0)}$, respectively $(i = 1, \ldots, n_0)$, and the continuous eigenvectors of H_0 by $|\phi(E)\rangle$ for $0 \leq E < \infty$. In an analogous way we denote the discrete eigenvectors and eigenvalues of H by $|\psi_i\rangle$ and $E_i (i = 1, \ldots, n)$. The kets $|\psi^+(E)\rangle$ for $0 \leq E < \infty$ are the continuous eigenvectors of H and are normalized as incoming waves.

Expanding $|\psi^+(E)\rangle$ in terms of the complete set of the eigenstates of H_0, we obtain

$$|\psi^+(E)\rangle = \int_0^\infty dE' \psi(E', E) |\phi(E')\rangle + \sum_{i=1}^{n_0} \psi_i(E) |\phi_i\rangle, \tag{3}$$

where $\psi_i(E)$ are square integrable functions, and $\psi(E', E) = \langle \phi(E') | \psi^+(E) \rangle$ is connected to the matrix element $R(E'; E)$ of the reaction operator by the relation[9]

$$\langle\phi(E)\,|\,\psi^{\dagger}(E')\rangle = \delta(E-E') + [R(E,E')/(E'-E+i\epsilon)]. \tag{4}$$

In the case where H_0 has only a continuous spectrum, the diagonal matrix elements $R(E;E)$ are connected with the phase shifts by the relation

$$R(E,E) = -[e^{i\delta(E)}\sin\delta(E)]/\pi \tag{5}$$

following from unitarity of the scattering matrix. Our case is more general since H_0 has some discrete eigenvalues and the discrete eigenvalues, both of H and H_0, can be embedded in the continuum, as a result the scattering matrix can no more be defined in the usual way, namely,

$$S = \lim_{t\to+\infty}\lim_{t'\to-\infty} e^{iH_0t}e^{-iH(t-t')}e^{-iH_0t'}. \tag{6}$$

However, if the matrix elements $R(E,E')$ are bounded and Hölder-continuous functions in the two variables E and E', we may define a unitary scattering operator on thecontinuous subspace of H_0 in a way analogous to (6). The matrix elements of this scattering operator are given by

$$\delta(E-E')e^{2i\delta(E)} = \langle\phi(E)|S|\phi(E')\rangle = \delta(E-E')[1-2\pi iR(E,E)]. \tag{7}$$

From the last relation, formula (5) follows again.

We see now how to count the discrete independent eigenvectors of H. Let us call discrete subspace of H the subspace spanned by its discrete eigenvectors.

Using the orthogonality and the completeness of the spectrum of H we have that the necessary and sufficient condition for a vector $|\psi\rangle$ to belong to the discrete subspace of H is

$$\langle\psi^{\dagger}(E)\,|\,\psi\rangle = 0, \tag{8}$$

for any $0 \le E < \infty$.

Now we expand the vector $|\psi\rangle$ in the complete set of the eigenvectors of H_0:

$$|\psi\rangle = \int_0^{\infty} dE'\rho(E')|\phi(E')\rangle + \sum_{i=1}^{n_0} x_i|\phi_i\rangle, \tag{9}$$

where $\rho(E)$ is a square integrable function. From (3) and (9), Eq. (8) can be rewritten in the form

$$\int_0^{\infty}\psi^*(E',E)\rho(E')dE' = -\sum_{i=1}^{n_0} x_i\psi_i^*(E). \tag{10}$$

Then, from (4), we get

$$\rho(E) - \int_0^{\infty}\frac{R^*(E',E)}{E-E'+i\eta}\rho(E')dE' = \sum_{i=1}^{n_0} x_i\psi_i^*(E). \tag{11}$$

Equations (11) and (10) are inhomogeneous s.i.e. The term at the right-hand side depends on n_0 independent parameters [namely $x_j (j = 1,\ldots,n_0)$], and the number of nontrivial independent solutions is n. Here we call trivial solutions those for which the functions $\rho(E)$ and the parameters x_i vanish simultaneously.

3. DISCUSSION OF THE EQUATION

The nontrivial solutions of Eq. (10) can be classified into three classes:

(I) Solutions for which $\rho(E) \equiv 0$ and some x_i do not vanish.

(II) Solutions for which $\rho(E) \neq 0$ and all the x_i vanish.

(III) Solutions for which $\rho(E)$ is not identically zero and some $x_i \neq 0$.

We consider separately the number of the linearly independent solutions belonging to the three above mentioned classes. The sum of these three numbers is equal to the number "n" of the independent discrete eigenstates of H.

Let us consider the solutions of Eq. (10) belonging to the class I; such solutions exist if some nontrivial linear combination of the quantities $\psi_i^*(E)$ vanish identically [this happens in particular if some one of the $\psi_i^*(E)$ is identically zero].

If the largest number of the $\psi_i^*(E)$ which are linearly independent is $h \le n_0$; then there are $n_0 - h$ nontrivial independent linear relations among the $\psi_i^*(E)$:

$$\sum_{i=1}^{n_0} f_i^{(l)}\psi_i^*(E) \equiv 0, \qquad l = 1,\ldots,n_0 - h; \tag{12}$$

the coefficients of each linear combination (12) provide a solution of the Eq. (10) belonging to class I.

The number of the independent solutions of class II is the same as that of the independent solutions of the homogeneous equation

$$\int_0^{\infty}\psi^*(E',E)\rho(E')dE' = 0. \tag{13}$$

Let us indicate it by a

We must now find the number of the independent solutions belonging to class III. In this case the right-hand side of Eq. (10) does not vanish. Then Eq. (10) is inhomogeneous and has solutions if and only if the term at the right-hand side is orthogonal to all the solutions of the corresponding adjoint homogeneous equation. This is a well-known property of the integral equations.

The homogeneous equation adjoint to (10) is

$$\int_0^{\infty}\psi(E,E')\eta(E')dE' = 0. \tag{14}$$

Let us suppose that Eq. (14) has "b" linearly independent solutions; we indicate them by $\eta i(E)(i = 1,\ldots,b)$. Then Eq. (10) has solutions if and only if all the following orthogonality conditions are satisfied:

$$\sum_{i=1}^{n_0} x_i\int_0^{\infty}\psi_i^*(E)\eta_j^*(E)dE = 0, \qquad j = 1,\ldots,b. \tag{15}$$

The number of the linearly independent solutions of Eq. (10) belonging to class III coincides with that of the nonvanishing linearly independent combinations of the $\psi_i^*(E)$ satisfying condition (15).

Let $\eta(E)$ be a linear combination of the function $\eta_i(E)$; we define the vector

$$\int_0^{\infty}\eta(E)|\psi^{\dagger}(E)\rangle\,dE = |\phi\rangle, \tag{16}$$

and we have

$$\langle\phi\,|\,\phi\rangle = \int_0^{\infty}|\eta(E)|^2dE. \tag{17}$$

Using the completeness of spectrum of H_0, we obtain

$$\langle\phi|\phi\rangle = \int_0^\infty dE\langle\phi|\phi(E)\rangle\langle\phi(E)|\phi\rangle + \sum_{i=1}^{n_0}\langle\phi|\phi_i\rangle\langle\phi_i|\phi\rangle. \quad (18)$$

From (16), (14), and (3) we have

$$\langle\phi(E)|\phi\rangle = \int_0^\infty dE'\langle\phi(E)|\psi^\dagger(E')\rangle\eta(E') = \int_0^\infty \psi(E,E')\eta(E')dE' = 0. \quad (19)$$

Then, by taking into account (18) and (19), Eq. (17) becomes

$$\int_0^\infty |\eta(E)|^2 dE = \sum_{i=1}^{n_0} |\langle\phi|\phi_i\rangle|^2. \quad (20)$$

Now, by (3) and (16), we have

$$\langle\phi_i|\phi\rangle = \int_0^\infty \eta(E)\psi_i(E)dE \quad (21)$$

and, by (20) and (21),

$$\int_0^\infty |\eta(E)|^2 dE = \sum_{i=1}^{n_0} \left|\int_0^\infty \eta(E)\psi_i(E)dE\right|^2. \quad (22)$$

This shows that a linear combination of the $\eta_i(E)$ is orthogonal to all $\psi_i^*(E)$, if and only if it vanishes identically.

If we now denote by $\mathfrak{M}$ the linear subspace spanned by the functions ψ_i^* and by $\mathfrak{M}_T$, it orthogonal complement, then any η_i can be written as

$$\eta_i = \xi_i + \zeta_i \quad \text{where} \quad \xi_i \in \mathfrak{M} \text{ and } \zeta_i \in \mathfrak{M}_T.$$

The ξ_i are linearly independent. Indeed, if a nontrivial linear combination of the ξ_i is orthogonal to all the ψ_i^* (and therefore equal to zero), the corresponding linear combination of the η_i is orthogonal to $\mathfrak{M}$, thus by formula (22), it vanishes too. Hence the linear manifold $\mathfrak{M}$ contains "b" linear independent vectors and its dimension "h" cannot be smaller than "b." In the case where $h > b$ there are $h - b$ vectors belonging to $\mathfrak{M}$ which are orthogonal to the ξ_i and to the η_i. Consequently there are $h - b$ linear combinations of the ψ_i^* which are orthogonal to all the η_i. Thus we have $h - b$ particular solutions of Eq. (10) belonging to class III. These solutions are independent of each other and of those belonging to classes I and II.

Now we can conclude that the vectors belonging to the discrete subspace of H are $n_0 - h$ belonging to class I; a belonging to class II and $h - b$ belonging to class III. Then we have

$$n - n_0 = -h + a + h - b = a - b. \quad (23)$$

Thus the difference between the number of the discrete eigenvectors of H and that of the discrete eigenvectors of H_0 is equal to the difference $\chi = a - b$ between the number of the independent solutions of the homogeneous equations (13) and (14). These equations, can be put in the form

$$\rho(E) - \int_0^\infty \frac{R^*(E',E)}{E - E' + i\epsilon}\rho(E')dE' = 0, \quad (13')$$

$$\eta(E) + \int_0^\infty \frac{R(E,E')}{E - E' + i\epsilon}\eta(E')dE' = 0. \quad (14')$$

The number χ is given by the theory of the singular integral equations as we shall see in Sec. 4.

4. DEDUCTION OF THE THEOREM

In this section we prove that the relation (23) is equivalent to the Levinson theorem [formula (2)]. This follows directly by applying the theory of the s.i.e. to Eqs. (13') and (14'). However this needs some slight extension of the known theory, in which the integration path consists of a finite number of smooth contours and the considered solutions are only those Hölder-continuous.[10] In our case the integration path is the real axis and the solutions of interest are square integrable functions. However, the theory can be extended following step by step the treatment of Ref. 10 and changing only some details.[11]

So as not to break the continuity of our treatment, we shall give only the general results of the extended theory and apply them to our case.

Let us indicate by $\mathfrak{R}(\alpha,\beta)$ the class of functions of real variable $f(t)$ which satisfy the Hölder condition[10] with index β and vanish for large $|t|$ faster than $|t|^{-\alpha}$; that is

$$|f(t) - f(t')| \leq C\ |t - t'|^\beta, \quad (24)$$

$$|f(t)| \leq C/(1 + |t|^\alpha), \quad (25)$$

where $0 \leq \beta,\ \alpha < 1$. We say that a two-variable function $k(t,t')$ belongs to the class $\mathfrak{R}^2(\alpha,\beta)[k(t,t') \in \mathfrak{R}^2(\alpha,\beta)]$, if it satisfies conditions (24) and (25) simultaneously in the two variables t and t'.

The equation

$$(K\varphi)(t) = \varphi(t) + \frac{1}{\pi i}\int_{-\infty}^{+\infty} dt' \frac{k(t,t')}{t - t' + i\epsilon}\varphi(t') = 0, \quad (26)$$

if $k(t,t') \in \mathfrak{R}^2(\alpha,\beta)$ and $1 - 2k(t,t) \neq 0$ is an s.i.e. The adjoint equation is

$$(K^\dagger\lambda)(t) = \lambda(t) + \frac{1}{\pi i}\int_{-\infty}^{+\infty} dt' \frac{k^*(t',t)}{t - t' + i\epsilon}\lambda(t') = 0. \quad (27)$$

Let us also consider the s.i.e. (of the dominant type)[10]

$$(K_0^+\mu)(t) = \mu(t) + \frac{k^*(t,t)}{\pi i}\int_{-\infty}^{+\infty} dt' \frac{\mu(t')}{t - t' + i\epsilon} = 0 \quad (28)$$

and the corresponding adjoint

$$(K_0\gamma)(t) = \gamma(t) + \frac{1}{\pi i}\int_{-\infty}^{+\infty} dt' \frac{k(t',t')}{t - t' + i\epsilon}\gamma(t') = 0. \quad (29)$$

The main results of the theory of the s.i.e. are

(I) The difference between the number of linearly independent solutions of Eqs. (26) and (27) is equal to that between the number of solutions of (28) and (29).

(II) This difference is equal to the index χ of the operator K:

$$\chi = \frac{1}{2\pi i}\{\log[1 + 2k(\infty,\infty)]^{-1} - \log[1 + 2k(-\infty,-\infty]^{-1}\}. \quad (30)$$

The only difference among (13'), (14') and (26), (27) lies in the integration path. However, if $k(t,t') = 0$ for t or $t' < 0$, it is evident that Eqs. (26) and (27) have

nonzero solutions only on the positive real axis. In this case the index of Eq. (26) can be written in the form

$$\chi = \frac{1}{2\pi i}\{\log[1 + 2k(\infty,\infty)]^{-1} - \log[1 + 2k(0,0)]^{-1}\}. \quad (31)$$

On the other hand, $R(E,E')$ can be considered as a function defined on the whole real axis, vanishing if E or E' are negative. Thus Eqs. (13′) and (14′) can be extended to the entire axis. From these remarks it follows that the above theorems can be directly applied to Eqs. (13′) and (14′) if $R(E,E')$ belongs to the class $\mathcal{R}^2(\alpha,\beta)$ and $1 - 2\pi i\, R(E,E) \neq 0$. Now the last condition when $R(E,E') \in \mathcal{R}^2(\alpha,\beta)$ is implied by (5) [indeed, $1 - 2\pi i\, R(E,E) = e^{2i\delta(E)} \neq 0$].

Therefore, we conclude that if $R(E,E') \in \mathcal{R}^2(\alpha,\beta)$, the difference between the number of the discrete independent eigenstates of H and H_0 is

$$\begin{aligned} n - n_0 = \chi &= (1/2\pi i)\{\log[1 + 2\pi i R^*(\infty,\infty)] \\ &\quad - \log[1 + 2\pi i R^*(0,0)]\} \\ &= [\delta(0) - \delta(\infty)]/\pi. \end{aligned} \quad (32)$$

which proves Eq. (2)

We remark that the required conditions of continuity imply that $R(E,E')$ must be Hölder-continuous at the threshold. For this reason our treatment does not include the case in which the matrix R is a threshold singularity. In the last case the usual form of Levinson's theorem is notoriously not true.[12]

5. FINAL REMARKS

We have proved Levinson's theorem by assuming appropriate conditions of continuity on the matrix elements of the reaction operator $R(E,E')$. It may be interesting to connect these conditions to some properties of the potential. Namely, we may investigate under which conditions on the potential there exists a reaction matrix with the desired properties. This problem is widely discussed in a book by Friedrichs[7] and in other papers therein quoted.[8,13] In particular, Rejto[8] shows that if the potential belongs to the class $\mathcal{R}^2(\alpha,\beta)$, then the reaction matrix exists and usually belongs to the same class. For the deduction of this result and for an analysis of the possible exceptions we refer to the literature.

Finally we briefly consider how our treatment could be extended to the case of the scattering in any finite number N_c of two body channels. In this case the eigenvectors of the unperturbed Hamiltonian H_0 are labelled by a discrete channel index ($i = 1, \ldots, N_C$) besides the energy. So the matrix elements of the reaction operator $R(E,E')$ [formulas (4) and (5)] are themselves matrices in a N_C-dimensional space, and the coefficients $\rho(E)$ and x_i of Eq. (10) are vectors in the same space. Equations (11), (13′) and (14′) become systems of singular integral equations. Taking this into account, the proof of Levinson's theorem can be given in the same way as before. By applying the result of the theory of systems of s.i.e., the theorem becomes

$$(1/2\pi i)\{\operatorname{logdet}|S_{ij}(0)| - \operatorname{logdet}|S_{ij}(\infty)|\} = n - n_0, \quad (33)$$

where $S_{ij}(E) = \delta_{ij} - 2\pi i R_{ij}(E,E)$ are the scattering matrix elements; i and j are channel labels.

However, in this case of Ref. 11, the continuity conditions on the matrix elements R are slightly more restrictive than those considered above. Indeed (24) and (25) are replaced by the Hölder-continuity condition at infinity,

$$|f(t) - f(t')| \leq C\, |t - t'|^{\alpha} / |1 - it|^{\alpha} |1 - it'|^{\alpha}, \quad \text{where} \quad 0 \leq \alpha < 1, \quad (34)$$

and we consider functions $R(E,E')$ for which (34) is simultaneously valid in the two variables.

---

* Supported by a fellowship of the Consiglio Nazionale delle Ricerche.

1 F. J. Dyson, Phys. Rev. **106**, 157 (1957); D. B. Fairlie and J. C. Polkinghorne, Nuovo Cimento **8**, 346, 555 (1958); B. W. Lee and A. Klein, *ibid.* **13**, 891 (1959).
2 F. E. Low, Phys. Rev. **97**, 1392 (1955).
3 L. Castillejo, R. Dalitz, and F. Dyson, Phys. Rev. **101**, 453 (1956).
4 R. Haag, Nuovo Cimento **5**, 203 (1957).
5 R. L. Warnok, in *Lectures in Theoretical High Energy Physics*, edited by H. H. Aly (Interscience, New York, 1968), p. 315.
6 This starting point is the one used by M. J. Jauch to prove the primitive form of the Levinson theorem: Helv. Phys. Acta **30**, Nos. 2–3, 143 (1957).
7 K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space*, Lectures in Applied Mathematics, Proceedings of the Summer Seminar, Boulder, Colorado (1960) (Amer. Math. Soc., Providence, R.I., 1965).
8 P. A. Rejto, Commun. Pure Appl. Math. **16**, 279 (1963); **17**, 257 (1964) (Papers I and II).
9 S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961).
10 N. I. Mushkelishvili, *Singular Integral Equations*, translated from Russian by J. R. Rodok (Noordhoff, Groningen, 1953).
11 In the case of systems of s.i.e. more restrictive conditions on the reaction matrix are needed.
12 R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
13 J. Schwartz, Commun. Pure Appl. Math. **13**, 609 (1960).

On the Generalized Exchange Operators for $SU(n)$

A. Partensky
Institut de Physique Nucléaire, Université Claude Bernard de Lyon (Institut National de Physique Nucléaire et de Physique des Particules), 69-Villeurbanne, France
(Received 20 March 1972)

By following the work of Biedenharn we have redefined the k-particles generalized exchange operators (g.e.o.'s) and studied their properties. By a straightforward but cumbersome calculation we have derived the expression, in terms of the $SU(n)$ Casimir operators, of the 2-, 3-, or 4-particle g.e.o. acting on the A-particle states which span an irreducible representation of the grown $SU(n)$. A striking and interesting result is that the eigenvalues of each of these g.e.o.'s do not depend on the n in $SU(n)$ but only on the Young pattern associated with the irreducible representation considered. For a given g.e.o., the eigenvalues corresponding to two conjugate Young patterns are the same except for the sign which depends on the parity of the g.e.o. considered. Three appendices deal with some related problems and, more specifically, Appendix C contains a method of obtaining the eigenvalues of Gel'fand invariants in a new and simple way.

I. INTRODUCTION

The interest attributed by physicists to the Lie groups has been continuously growing since the pioneer works of Weyl,[1,2] Wigner,[3] and Racah.[4] In particular, unitary groups have been used successfully in various domains of physics, although the physical reasons for their introduction have not always been completely explained. The use of unitary groups in physical problems has in turn given rise to a lot of works devoted to the study of these groups considered as entities.[5] In this paper, we shall use the group $U(n)$ to define, to investigate the properties of, and to calculate the eigenvalues of the so-called generalized exchange operators (g.e.o.'s). This research opening was suggested by Biedenharn,[6] who defined a g.e.o. $P_k^{a_1 a_2 \cdots a_k}$ for k particles and relative to $U(n)$ as an extension of the two-particle exchange operator. [*Note added in proof:* Dr. Klaus Lezuo (University of Mainz) had independently carried out research rather similar to that in this paper. I should like to thank Professor L. C. Biedenharn for his calling the Lezuo paper to my attention.] The two-particle exchange operator has proved to be very useful in nuclear physics (cf. the spin exchange operator or Bartlett forces,[7] the space exchange operator or Majorana forces,[8] and the charge exchange operator or Heisenberg potential[9]). These two-particle exchange operators are a simple transcription of two-body forces. In a similar way, the k-particle g.e.o.'s ($k > 2$) derive from k-body forces. For a long time, the existence of many-body forces among nucleons has been postulated.[10] Recently, some papers[11] have been devoted to three-body forces in an attempt to define their importance relative to two-body forces and Harter,[12] using a Young diagram, derived sum rules for k-body operator spectra.

The material will be organized as follows. In Sec. 2, we study the properties of the g.e.o. $P_k^{a_1 a_2 \cdots a_k}$ defined in a somewhat different manner to the one of Biedenharn.[6] In Sec. 3, we extend this definition to produce an operator P_k^A for A particles ($A \geq k$). Section 4 is devoted to the three particular cases P_2^A, P_3^A and P_4^A. By a direct derivation, we get P_2^A, P_3^A, and P_4^A as functions of the invariant operators for $SU(n)$. For each of the operators P_2^A, P_3^A, and P_4^A, we then obtain the eigenvalues relative to an A-particle state which carries an irreducible representation of $SU(n)$. Making use of the Young diagram corresponding to the representation, we show that the eigenvalues of P_2^A, P_3^A, and P_4^A do not depend on n in $SU(n)$. In Sec. 5, we write the g.e.o.'s in Weyl basis. As is well-known, the Weyl basis makes the study of $U(n)$ easier. The study of our g.e.o.'s is simplified with this basis. The Weyl basis, is, however, less convenient for physical problems such as the one suggested by Biedenharn[13,14] and used in the first three parts of this paper. In particular, Biedenharn's basis enables us to render apparent, through the weight operators H_i, quantities of physical interest, while Weyl's basis does not.

Finally, three appendices deal with problems connected with the main body of the paper, although they are more or less self-contained. Appendix A discusses the algebra whose elements are the structure constants and the coupling coefficients for $U(n)$. Appendix B is devoted with sums of type

$$\sum_{a_1 \neq a_2 \neq \cdots \neq a_n}^{A} X_1^{(a_1)} X_2^{(a_2)} \cdots X_n^{(a_n)}.$$

Last of all, Appendix C gives a general expression for the eigenvalues of Gel'fand's invariants.[15]

For the purpose of unifying the notation, we follow the same definitions and notations as those used in a previous paper,[16] with one exception: Summation of the generators X_A (or x_A) always includes the number operator H_0 (or h_0). As a consequence, the coupling coefficient $[AB^C]$ is now defined by

$$[x_A, x_B]_+ = [AB^C] x_C. \tag{I.1}$$

In other words, we extend the metric for $SU(n)$, to the one of $U(n)$, by letting $g_{AB} = \delta_A^{-B}$ with $g_{A0} = \delta_A^0$.

II. DEFINITION AND SIGNIFICANCE OF THE $SU(n)$ k-PARTICLE G.E.O. ACTING ON THE k-PARTICLE STATES

We shall define such an operator by the following formula:

$$P_k^{a_1, a_2, \cdots, a_k} = 2n^{k-1} \sum_{ABC \cdots E} [ABC \cdots E] \times x_A^{(a_1)} x_B^{(a_2)} x_C^{(a_3)} \cdots x_E^{(a_k)}, \tag{II.1}$$

which differs from that of Biedenharn[6] by the norm and the coefficient $[ABC \cdots E]$. The interests of our definition will appear below.

In Eq. (II.1) $a_1, a_2, \ldots, a_k$ stand for the k-particles, whereas $x_A, x_B, \ldots, x_E$ are the generators of the fundamental representation $[1, 0]$ of $SU(n)$ and, finally, $[ABC \cdots E]$ is the completely symmetric coefficient in all indices $A, B, \ldots, E$ obtained from the coefficient $[AB - M][M \cdots] \cdots [\cdots E]$. More precisely, $[AB]$ (or g_{AB}) and $[ABC]$, which correspond to $k = 2$ and $k = 3$, respectively, are completely symmetric by definition.[17] For $k \geq 4$, the coefficient $[ABC \cdots E]$ can be defined via the following recurrency formula:

$$[A_1A_2\cdots A_k] = \{2/k(k-1)\}\sum_{i>j}^{k}[A_iA_j - M] \times [MA_1\cdots A_{j-1}A_{j+1}\cdots A_{i-1}A_{i+1}\cdots A_k]. \quad \text{(II. 2)}$$

It should be emphasized that the summation in Eq. (II. 1) has to be extended over all the values of the indices $AB\cdots E$ including the value 0 which corresponds to the unit generator.

Let us denote, for the particle a_i, by $\phi_n^m(a_i)$ the mth state ($1 \leq m \leq n$) of the fundamental representation of $SU(n)$. The generator x_A ($x_A = h_i$ or e_α) acting on ϕ_n^m gives

$$h_i\phi_n^m = (2n)^{-1/2}\lambda_m^{(i)}\phi_n^m, \quad e_{\alpha(ij)}\phi_n^m = (2n)^{-1/2}\delta_{jm}\phi_n^i. \quad \text{(II. 3)}$$

The index n is superfluous and we shall omit it in what follows.

Let us now examine the significance of $P_k^{a_1,a_2,\cdots,a_k}$ operating on a tensorial product of such one-particle functions. For $k = 2$, Eq. (II. 1) gives

$$P_2^{a_1a_2} = 2n\sum_A x_A^{(a_1)}x_{-A}^{(a_2)}, \quad \text{(II. 4)}$$

and by making use of Eq. (II. 3) we get

$$P_2^{a_1a_2}\phi^{m_1}(a_1)\phi^{m_2}(a_2) = \phi^{m_2}(a_1)\phi^{m_1}(a_2), \quad \text{(II. 5)}$$

as is well-known; i.e., $P_2^{a_1a_2}$ simply exchanges the particles a_1 and a_2.

For $k = 3$, first using the definition relation of the coupling coefficient $[ABC]$, written here in the form

$$[x_{-A}^{(a_3)}, x_{-B}^{(a_3)}]_+ = [ABC]x_C^{(a_3)}, \quad \text{(II. 6)}$$

and, secondly, by taking advantage of the useful relation

$$\sum_A x_{-A}^{(a_3)}x_{+A}^{(a_1)}\phi^{m_1}(a_1) = \frac{1}{2n}\sum_{i=1}^{n} e_{m_1i}^{(a_3)}\phi^i(a_1), \quad \text{(II. 7)}$$

where $e_{m,i}$ is the $n \times n$ matrix consisting of unity in the (m, i) position and zeros elsewhere, we should obtain

$$P_3^{a_1a_2a_3}\phi^{m_1}(a_1)\phi^{m_2}(a_2)\phi^{m_3}(a_3) = \tfrac{1}{2}\{\phi^{m_2}(a_1)\phi^{m_3}(a_2)\phi^{m_1}(a_3) + \phi^{m_3}(a_1)\phi^{m_1}(a_2)\phi^{m_2}(a_3)\}. \quad \text{(II. 8)}$$

As a matter of fact, this operator induces on the three particles all the permutations which do not leave any of them unchanged. To be more precise, those permutations are, in cyclic notation, (123) and (132), i.e., form the 3-cycle class of the symmetric group S_3.

For $k = 4$, we shall use the following relation definition for the coupling coefficient:

$$[[x_{-A}^{(a_4)}, x_{-B}^{(a_4)}]_{+1}x_{-C}^{(a_4)}]_+ = [AB - M][MCD]x_D^{(a_4)}; \quad \text{(II. 9)}$$

and with Eq. (II. 7) we thus get

$$\sum_{ABMCD}[AB - M][MCD] \times x_A^{(a_1)}x_B^{(a_2)}x_C^{(a_3)}x_D^{(a_4)}\phi^{m_1}(a_1)\phi^{m_2}(a_2)\phi^{m_3}(a_3)\phi^{m_4}(a_4) = (2n)^{-3}\{\phi^{m_4}(a_1)\phi^{m_1}(a_2)\phi^{m_2}(a_3)\phi^{m_3}(a_4) + \phi^{m_2}(a_1)\phi^{m_4}(a_2)\phi^{m_1}(a_3)\phi^{m_3}(a_4) + \phi^{m_3}(a_1)\phi^{m_1}(a_2)\phi^{m_4}(a_3)\phi^{m_2}(a_4) + \phi^{m_2}(a_1)\phi^{m_3}(a_2)\phi^{m_4}(a_3)\phi^{m_1}(a_4)\}. \quad \text{(II. 10)}$$

Here we do not get all the permutations leading none of the particle unchanged, because the coefficient $[AB - M][MCD]$ is not completely symmetric in all indices. This point leads us to define our g.e.o. by introducing a coefficient $[ABCD]$ completely symmetric in all indices. Thus, for $P_4^{a_1a_2a_3a_4}$ we obtain

$$P_4^{a_1a_2a_3a_4}\phi^{m_1}(a_1)\phi^{m_2}(a_2)\phi^{m_3}(a_3)\phi^{m_4}(a_4) = \tfrac{1}{6}\{\phi^{m_4}(a_1)\phi^{m_1}(a_2)\phi^{m_2}(a_3)\phi^{m_3}(a_4) + \phi^{m_2}(a_1)\phi^{m_4}(a_2)\phi^{m_1}(a_3)\phi^{m_3}(a_4) + \phi^{m_3}(a_1)\phi^{m_1}(a_2)\phi^{m_4}(a_3)\phi^{m_2}(a_4) + \phi^{m_2}(a_1)\phi^{m_3}(a_2)\phi^{m_4}(a_3)\phi^{m_1}(a_4) + \phi^{m_4}(a_1)\phi^{m_3}(a_2)\phi^{m_1}(a_3)\phi^{m_2}(a_4) + \phi^{m_3}(a_1)\phi^{m_4}(a_2)\phi^{m_2}(a_3)\phi^{m_1}(a_4)\}. \quad \text{(II. 11)}$$

The significance of the $P_4^{a_1a_2a_3a_4}$ operator is now clear. It induces on the 4-particles the six permutations (1234), (1243), (1324), (1342), (1423), (1432) which belong to the 4-cycle class of S_4. By an inductive process, we might generalize the results obtained for $k = 2, 3$, and 4. The significance of a k-particle g.e.o. $P_k^{a_1a_2\cdots a_k}$ for $SU(n)$ ($n \geq k$) acting on a k-particle state function corresponds to the permutations on the k-objects belonging to the class defined by the cycles of length k.

III. g.e.o. AND A PARTICLES

A. Definition of the $SU(n)$ k-Particle g.e.o. Acting on the A-Particle States ($A \geq k$)

We shall define it through

$$P_k^A = \frac{k!}{A(A-1)\cdots(A-k+1)}\sum_{1\ a_1>a_2>\cdots>a_k}^{A} P_k^{a_1a_2\cdots a_k}, \quad \text{(III. 1)}$$

which is a generalization of Eq. (II. 1). (One must verify, of course, that, for the particular case $A = k$, we have $P_k^k = P_k^{a_1a_2\cdots a_k}$.) By expanding Eq. (III. 1), we get

$$P_k^A = \frac{2n^{k-1}k!}{A(A-1)\cdots(A-k+1)}\sum_{1\ a_1>a_2>\cdots>a_k}^{A}\ \sum_{ABC\cdots E}[ABC\cdots E]x_A^{(a_1)}x_B^{(a_2)}\cdots x_E^{(a_k)}. \quad \text{(III. 2)}$$

It is clear that the quantity $\sum_{AB\cdots E}[AB\cdots E]x_A^{(a_1)}x_B^{(a_2)}\cdots x_E^{(a_k)}$ is completely symmetric in indices $a_1, a_2, \ldots, a_k$. This provides interesting consequences as it will be seen in Appendix A.

B. Eigenvalues of g.e.o. in the Particular Case Relative to a Maximum Weight State

Let us suppose that the a_ith particle is in the state of higher weight denoted by ϕ_n^1. We thus have

$$h_i^{(a_l)}\phi_n^1(a_l)=(1/\sqrt{2n})\lambda_1^{(i)}\phi_n^1(a_l), \quad e_\alpha^{(a_l)}\phi_n^1(a_l)=0. \tag{III.3}$$

By coupling such one-particle functions, we would get the A-particle functions ψ^A. The transcription of Eq. (III. 3) for ψ^A yields

$$\psi^A=\prod_{a_l=1}^{A}\phi_n^1(a_l). \tag{III.4}$$

By remembering that all the x's of Eq. (III. 2) commute, we only have to consider that part of P_k^A which contains products of type $h_i^{(a_1)}h_j^{(a_2)}\dots h_p^{(a_k)}$ so that

$$P_k^A\psi^A=\frac{2n^{k-1}k!}{A(A-1)\dots(A-k+1)}\sum_{a_1>\dots>a_k}^{A}\sum_{i,j,k\dots l}[ijk\dots l]h_i^{(a_1)}h_j^{(a_2)}\dots h_l^{(a_k)}\prod_{a_l=1}^{A}\phi_n^1(a_l).$$

We easily get

$$\sum_{ijk\dots l}^{n-1}[ijk\dots l]\lambda_1^{(i)}\lambda_1^{(j)}\dots\lambda_1^{(l)}=\left(\frac{2}{n}\right)^{(k/2)-1} \tag{III.5}$$

and finally

$$P_k^A\psi^A=\psi^A. \tag{III.6}$$

In other words, the eigenvalue of P_k^A relative to a symmetric tensorial product of maximum weight is unity. This is an *a posteriori* justification for the normalization coefficient introduced in Eq. (III. 1).

IV. EIGENVALUES OF P_2^A, P_3^A, AND P_4^A

A. Expression of the g.e.o. P_2^A, P_3^A, and P_4^A in Terms of Invariants

Before we undertake the explicit calculation of the eigenvalues for P_2^A, P_3^A, and P_4^A, we mention two preliminary remarks:

(1) If in the definition of the invariant operators (cf. Refs. 13 and 16) we allow the summation to include the index zero, we thus define operators, denoted as $\mathring{I}_k$, for the group $U(n)$, which are expressible in terms of the invariant operators I_k of $SU(n)$, and of the operators H_0 for $U(1)$. In what follows, we use the three first operators:

$$\begin{aligned}\mathring{I}_2&=I_2+H_0^2,\\ \mathring{I}_3&=I_3+(3\sqrt{2}/n)H_0I_2+(\sqrt{2}/n)H_0^3,\\ \mathring{I}_4&=I_4+(2/n^2)I_2^2+(4\sqrt{2}/n)H_0I_3+(12/n^2)H_0^2I_2+(2/n^2)H_0^4.\end{aligned} \tag{IV.1}$$

(2) From the fundamental representation of algebra spanned by one-particle generators, we may span a more general representation by defining polyparticle generators:

$$X_B=\sum_{a_1=1}^{A}x_B(a_1). \tag{IV.2}$$

1. Operator P_2^A

The general definition of P_k^A gives in that particular case

$$P_2^A=\frac{4n}{A(A-1)}\sum_{a_1>a_2}^{A}\sum_B x_B^{(a_1)}x_{-B}^{(a_2)}.$$

After the utilization of Eq. (A9) from Appendix A, this formula reads

$$P_2^A=\frac{2n}{A(A-1)}\sum_B\left(\sum_{a_1=1}^{A}\sum_{a_2=1}^{A}x_B^{(a_1)}x_{-B}^{(a_2)}-\sum_{a_1=1}^{A}x_B^{(a_1)}x_{-B}^{(a_1)}\right).$$

Introduction of the relation (III. 2) in P_2^A leads to

$$P_2^A=\{2n/A(A-1)\}\{\mathring{I}_2-(n/\sqrt{2})H_0\}$$

or

$$P_2^A=\{2n/A(A-1)\}\{I_2+H_0^2-(n/\sqrt{2})H_0\}. \tag{IV.3}$$

The eigenvalue of P_2^A then follows from that of I_2,[16] and that of H_0 (i.e., $A/n\sqrt{2}$). In abbreviated form, we have

$$P_2^A=\{2n/A(A-1)\}\,(I_2+A^2/2n^2-A/2), \tag{IV.4}$$

a well-known result.

2. Operator P_3^A

By following the same procedure as P_2^A [and using Eq. (A9)] we easily get

$$P_3^A=\frac{2n^2}{A(A-1)(A-2)}\times\left(\mathring{I}_3-3\sum_{ABC}[ABC]\sum_{a_1=1}^{A}\sum_{a_2=1}^{A}\mathcal{S}(x_A^{(a_1)}x_B^{(a_1)})x_C^{(a_2)}+\frac{n^2+1}{n\sqrt{2}}H_0\right),$$

where $\mathcal{S}$ renders in all the indices A, B, and C symmetric. Owing to the symmetry of $[ABC]$, we can write

$$\sum_{ABC}[ABC]\sum_{a_1=1}^{A}\sum_{a_2=1}^{A}\mathcal{S}(x_A^{(a_1)}x_B^{(a_1)})x_C^{(a_2)}=\sum_{ABC}[ABC]\sum_{a_1=1}^{A}\sum_{a_2=1}^{A}x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_2)}.$$

Or by using Eq. (B1) of Appendix B, we have

$$\sum_{ABC}[ABC]\sum_{a_1=1}^{A}\sum_{a_2=1}^{A}x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_2)}=\frac{1}{2}(\mathring{I}_2+H_0^2) \tag{IV.5}$$

so that P_3^A is expressible as a function of $\mathring{I}_3, \mathring{I}_2$, and H_0. The transcription of this result in terms of I_3, I_2, and H_0, via (IV. 1), gives (in abbreviated form)

$$P_3^A=\frac{2n^2}{A(A-1)(A-2)}\left\{I_3+3\left(\frac{A}{n^2}-\frac{1}{2}\right)I_2+\frac{A^3}{2n^4}-\frac{3A^2}{2n^2}+\frac{n^2+1}{2n^2}A\right\}. \tag{IV.6}$$

3. Operator P_4^A

By following the same procedure as for P_3^A, we have to calculate here quantities of type $[ABCD]x_Ax_Bx_Cx_D$. The calculation of such quantitites is achieved by using the general formula

$$\sum_{AB}[ABCD][AB^M]=\tfrac{1}{3}(2+\delta_C^0+\delta_D^0+2\delta_M^0)[MCD], \tag{IV.7}$$

which is derived from Eq. (B11) of Appendix B. In the detail, we thus get

$$[ABCD]x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_2)}x_D^{(a_3)} = \tfrac{1}{3}[ABC]x_A^{(a_1)}x_B^{(a_2)}x_C^{(a_3)} + (\sqrt{2}/6n)(2h_0^{(a_1)}x_A^{(a_2)}x_{-A}^{(a_3)} + h_0^{(a_2)}x_A^{(a_1)}x_{-A}^{(a_3)} + h_0^{(a_3)}x_A^{(a_1)}x_{-A}^{(a_2)}),$$

$$[ABCD]x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_2)}x_D^{(a_2)} = \{(n^2+2)/6n^2\}x_A^{(a_1)}x_{-A}^{(a_2)} + \tfrac{1}{2}h_0^{(a_1)}h_0^{(a_2)},$$

$$[ABCD]x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_1)}x_D^{(a_2)} = \{(n^2+3)/6n^2\}x_A^{(a_1)}x_{-A}^{(a_2)} + \tfrac{1}{3}h_0^{(a_1)}h_0^{(a_2)}. \quad \text{(IV.8)}$$

We further have

$$\sum_{ABCD}[ABCD]X_AX_BX_CX_D = \mathring{I}_4 - \tfrac{1}{12}\mathring{I}_2 + \tfrac{1}{12}H_0^2,$$

$$\sum_{ABCD}[ABCD]x_A^{(a_1)}x_B^{(a_1)}x_C^{(a_1)}x_D^{(a_1)} = \{(n^2+5)/12n^2\}I^{(a_1)}. \quad \text{(IV.9)}$$

Introducing Eqs. (IV.8) and (IV.9) in the expression for P_4^A, we get

$$P_4^A = \frac{2n^3}{A(A-1)(A-2)(A-3)} \times \left(\mathring{I}_4 - 2\mathring{I}_3 + \frac{7n^2+20-16A}{4n^2}\mathring{I}_2 + \frac{17A^2}{8n^2} - \frac{(n^2+5)}{2n^2}A\right) \quad \text{(IV.10)}$$

or finally

$$P_4^A = \frac{2n^3}{A(A-1)(A-2)(A-3)}\left(I_4 + \frac{2}{n^2}I_2^2 - \frac{2(n^2-2A)}{n^2}I_3 + \frac{7n^4+20n^2-40n^2A+24A^2}{4n^4}I_2 + \frac{A^4-6n^2A^3+(6n^4+5n^2)A^2-(n^6+5n^4)A}{2n^6}\right). \quad \text{(IV.11)}$$

The same procedure might be extended to P_k^A with $k \geq 5$, leading then to cumbersome and fastidious calculations.

B. Expression of the Eigenvalue of Each g.e.o. as a Function of the Young Pattern Characterizing the IR $[\mu_1\mu_2\cdots\mu_p]$

In Eqs. (IV.4), (IV.6), and (IV.11) giving each g.e.o. as a function of the invariant operators for $U(n)$, (n arbitrary, but fixed) the figure n appears explicitly. We shall see in what follows that the eigenvalue for each of our g.e.o.'s does not depend on the index n, but only on the higher weight of the considered irreducible representation.

To make clear this point, it is sufficient to express the eigenvalues of P_2^A, P_3^A, and P_4^A in terms of the μ's relative to $[\mu_1,\mu_2,\ldots,\mu_p] = [\mu]$. This is achieved by the following transformation:

$$\langle[\mu]|H_i|[\mu]\rangle = \frac{1}{\sqrt{2n}}\sum_{l=1}^{n}\lambda_l^{(i)}\mu_l, \quad \text{(IV.12)}$$

where, of course,

$$A = \sum_{l=1}^{n}\mu_l. \quad \text{(IV.13)}$$

Thus, we get

$$P_2^A|[\mu]\rangle = \{1/A(A-1)\}\{\mu_l^2 - 2l\mu_l + A\}|[\mu]\rangle$$

$$P_3^A|[\mu]\rangle = \{1/A(A-1)(A-2)\}\{\mu_l^3 - 3l\mu_l^2 + \tfrac{3}{2}\mu_l^2 + 3l^2\mu_l - 3l\mu_l + 2A - \tfrac{3}{2}A^2\}|[\mu]\rangle$$

$$P_4^A|[\mu]\rangle = \{1/A(A-1)(A-2)(A-3)\}\{\mu_l^4 - 2(2l-1)\mu_l^3 + [6l(l-1) - 4A + 7]\mu_l^2 - 2(2l^2 - 3l - 4A + 7)l\mu_l - 4A^2 + 6A\}|[\mu]\rangle. \quad \text{(IV.14)}$$

In those equations the summations over l have to be assumed; from a certain rank $\mu_l = 0$, so that the sum over l has to be extended on all the l's for which $\mu_l \neq 0$.

C. Relation between the Eigenvalues Corresponding to the Associated IR's Pattern $[\mu]$ and $[\tilde{\mu}]$

As is well known, the Young pattern associated with a given pattern is obtained from the first one by exchanging its rows and columns, so that, given the Young pattern characterized by the partition

$$[\mu] = [\mu_1,\mu_2,\ldots,\mu_p].$$

The associated pattern corresponds to

$$[\tilde{\mu}] = [\tilde{\mu}_1,\tilde{\mu}_2,\ldots,\tilde{\mu}_k],$$

where

$$[\tilde{\mu}] = [\underbrace{l\ldots l}_{\mu_l}, \underbrace{l-1\ldots l-1}_{\mu_{l-1}-\mu_l}, \ldots, \underbrace{k\ldots k}_{\mu_k-\mu_{k+1}}, \ldots, \underbrace{1\ldots 1}_{\mu_1-\mu_2}]. \quad \text{(IV.15)}$$

Owing to the identity

$$\sum_k k^\alpha\tilde{\mu}_k^\beta = \sum_{k=1}^{l}\sum_{q=1}^{\mu_k} q^\alpha[k^\alpha - (k-1)^\beta], \quad \text{(IV.16)}$$

it is a simple matter of calculation to get

$$P_2^A|[\tilde{\mu}]\rangle = -P_2^A|[\mu]\rangle,$$
$$P_3^A|[\tilde{\mu}]\rangle = P_3^A|[\mu]\rangle, \quad \text{(IV.17)}$$
$$P_4^A|[\tilde{\mu}]\rangle = -P_4^A|[\mu]\rangle.$$

Consequently, the eigenvalues of P_2^A, P_3^A, and P_4^A relative to two associated Young patterns are in the ratios $-1, +1$, and -1, respectively, so that the eigenvalues for P_2^A and P_4^A relative to a self-associated Young pattern are zero. As a check of Eq. (III.6) we easily obtained 1 for the eigenvalues of P_2^A, P_3^A, and P_4^A relative to the eigenfunction characterized by $[A,\mathring{0}]$. By applying Eq. (IV.17) the eigenvalues of P_2^A, P_3^A, and P_4^A relative to $[1^A]$, the associated partition of the preceding one, are $-1, +1$ and -1, respectively.

V. GENERALIZED EXCHANGE OPERATORS EXPRESSED IN WEYL'S BASIS

The above definition of the g.e.o.'s involves the group $U(n)$ rather than the group $SU(n)$. Further, the eigenvalue problem of those operators is more easily tractable when resolved via the use of Young diagram $[\mu_1,\mu_2,\ldots,\mu_l]$. Consequently, we shall rewrite the g.e.o.'s by using the n^2 Weyl generators E_{ij}, which obey the commutation relation

$$[E_{ij}, E_{kl}] = \delta_{jk}E_{il} - \delta_{li}E_{kj}. \quad \text{(V.1)}$$

Let us denote by e_{ij} those operators in the fundamental representation. (The so-called $n \times n$ matrix units, in which 1 appears in row i and column j, and 0 appears elsewhere). The basis vectors of this representation are the function ϕ^m, and we have

$$l_{ij}\phi^m = \delta_{jm}\phi^i. \tag{V.2}$$

In this basis, formula (II.1) reads

$$P_k^{a_1 a_2 \cdots a_k} = \frac{1}{(k-1)!} \sum_{i_1, i_2, \ldots, i_k}^{n} e_{i_1 i_2}^{(a_1)} \left(e_{i_2 i_3}^{(a_2)} \cdots e_{i_k i_1}^{(a_k)}\right)_S. \tag{V.3}$$

The parenthesis lower-indexed by S means that the inside quantity is fully symmetrical on all the permutations of the symbols $a_2, a_3, \ldots, a_k$. The development of $(e_{i_2 i_3}^{(a_2)} e_{i_3 i_4}^{(a_3)} \cdots e_{i_k i_1}^{(a_k)})_S$ then contains $(k-1)!$ terms.

The significance of the $P_k^{a_1 a_2 \cdots a_k}$ operating on a tensorial product of one-particle function ϕ^m is quite clear. Indeed, owing to Eq. (V.2), we derive

$$\sum_{i_1 i_2 \cdots i_k} e_{i_1 i_2}^{(a_1)} e_{i_2 i_3}^{(a_2)} \cdots e_{i_{k-1} i_k}^{(a_{k-1})} e_{i_k i_1}^{(a_k)} \phi^{m_1}(a_1)\phi^{m_2}(a_2) \cdots \phi^{m_k}(a_k) = \phi^{m_k}(a_1)\phi^{m_1}(a_2)\phi^{m_2}(a_3) \cdots \phi^{m_{k-1}}(a_k). \tag{V.4}$$

By using the last equation and the symmetry on the indices $a_1, a_2, \ldots, a_k$ occurring in Eq. (V.3) we deduce that the effect of $P_k^{a_1 a_2 \cdots a_k}$ operating on a k-particle state function corresponds to the permutations on the k-objects belonging to the class defined by the cycle of length k.

In a similar way, it is possible to rewrite P_k^A as function of the e_{ij}. Formula (III.2) gives

$$P_k^A = \frac{k}{A(A-1)\ldots(A-k+1)} \times \sum_{i_1 i_2 \cdots i_k}^{n} \sum_{a_1 > a_2 > \cdots > a_k}^{A} e_{i_1 i_2}^{(a_1)} (e_{i_2 i_3}^{(a_2)} \cdots e_{i_k i_1}^{(a_k)})_S \tag{V.5}$$

or

$$P_k^A = \frac{1}{A(A-1)\ldots(A-k+1)} \times \sum_{i_1 i_2 \cdots i_k}^{n} \sum_{a_1 > \cdots > a_k}^{A} (e_{i_1 i_2}^{(a_1)} e_{i_2 i_3}^{(a_2)} \cdots e_{i_k i_1}^{(a_k)})_S. \tag{V.6}$$

Let us write

$$E_{ij} = \sum_{a_1=1}^{A} e_{ij}^{(a_1)}. \tag{V.7}$$

Equation (A8) of Appendix A, the relation commutation (V.1) and the following relation,

$$e_{ij}e_{kl} = e_{il}\delta_{jk}, \tag{V.8}$$

specific to the fundamental representation, enables us to express P_k^A from the Gel'fand invariants:

$$I_k^{(n)} = \sum_{i_1 i_2 \cdots i_k}^{n} E_{i_1 i_2} E_{i_2 i_3} \cdots E_{i_k i_1}. \tag{V.9}$$

The eigenvalue of P_k^A is then easily obtained from those of Gel'fand invariants (cf. Appendix C).

ACKNOWLEDGMENT

I wish to thank Dr. M. Kibler for detailed discussion of the results of this paper.

APPENDIX A

Let us consider the quantity $x_1^{(a_1)} x_2^{(a_2)} \cdots x_n^{(a_n)}$, where each variable x_i depends on the upper indices a_i. We shall write $(x_1^{(a_1)} x_2^{(a_2)} \ldots x_n^{(a_n)})_S$ the polynomial obtained by adding the $n!$ terms resulting from all the permutations on the indices $1, 2, \ldots n$ of $x_1^{(a_1)} x_2^{(a_2)} \ldots x_n^{(a_n)}$. Let us suppose now that each a_i can take A $(A \geq n)$ distinct values. We then have the following identities:

$$\sum_{a_1 \neq a_2 \neq \cdots \neq a_n}^{A} x_1^{(a_1)} x_2^{(a_2)} \cdots x_n^{(a_n)} = \sum_{a_1 > a_2 > \cdots > a_n}^{A} (x_1^{(a_1)} x_2^{(a_2)} \cdots x_n^{(a_n)})_S = \sum_{a_1=n}^{A} \sum_{a_2=n-1}^{a_1-1} \cdots \sum_{a_{n-1}=2}^{a_{n-2}-1} \sum_{a_n=1}^{a_{n-1}-1} (x_1^{(a_1)} x_2^{(a_2)} \cdots x_n^{(a_n)})_S = \sum_{a_1, a_2, \ldots, a_n}^{A} x_1^{(a_1)} x_2^{(a_2)} \ldots x_n^{(a_n)} \prod_{i<j}^{n} F_{a_i a_j}, \tag{A1}$$

where $F_{a_i a_j}$ stands for $1 - \delta_{a_i a_j}$ ($\delta_{a_i a_j}$ being the Kronecker symbol). $F_{a_i a_j}$ has two important properties, viz.,

$$(F_{a_i a_j})^2 = F_{a_i a_j}, \qquad \delta_{a_i a_j} F_{a_i a_j} = 0. \tag{A2}$$

Let us define:

$$F^{(n)} = \prod_{i<j}^{n} (1 - \delta_{a_i a_j}). \tag{A3}$$

Clearly $F^{(n)}$ can be written as a determinant of order n. It is merely sufficient to multiply the jth column of

$$\begin{vmatrix} \delta_{a_1 a_1} & \cdots & \delta_{a_1 a_n} \\ \cdots & \cdots & \cdots \\ \delta_{a_i a_j} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \delta_{a_n a_1} & \cdots & \delta_{a_n a_n} \end{vmatrix}$$

by

$$\prod_{k=j+1}^{n} (1 - \delta_{a_k a_j})$$

and to use the properties (A2) to get

$$F^{(n)} = \det \| \delta_{a_i a_j} \|. \tag{A4}$$

The development of the obtained determinant as

$$F^{(n)} = \sum_{P \in S_n} \chi(P) \delta_{a_1 a_{P_1}} \delta_{a_2 a_{P_2}} \cdots \delta_{a_n a_{P_n}} \tag{A5}$$

[where $P = \binom{1,2 \ldots n}{P_1, P_2 \ldots P_n}$ and $\chi(P)$ denotes the parity of p] can be written in a somewhat more "condensed" form.

For that purpose let us define

$$\delta^{(0)}_{a_i a_i} = \delta_{a_i=1},$$
$$\delta^{(1)}_{a_i a_j} = \begin{cases} 1 & \text{if } a_i = a_j \\ 0 & \text{if } a_i \neq a_j \end{cases}, \tag{A6}$$

$$\delta^{(k)}_{a_1 a_2 \dots a_k a_{k+1}} = \begin{cases} 1 & \text{if } a_1 = a_2 = \dots = a_k = a_{k+1} \\ 0 & \text{otherwise} \end{cases}.$$

For each partition $[\lambda] = [\lambda_1, \lambda_2, \dots, \lambda_n]$ of n,

$$(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0; \lambda_1 + \lambda_2 + \dots + \lambda_n = n),$$

let us define the elements

$$\delta^{[\lambda]}_i = \delta^{(\lambda_1 - 1)}_{a_{i_1} a_{i_2} \dots a_{i_{\lambda_1}}} \delta^{(\lambda_2 - 1)}_{a_{i_{\lambda_1+1}} \dots a_{i_{\lambda_1+\lambda_2}}} \delta \dots \delta^{(\lambda_k - 1)}_{\dots a_{i_n}} \tag{A7}$$

(with $\lambda_k \neq 0$ and $\lambda_{k+1} = 0$).

Due to the symmetry of the indices each element $\delta^{[\lambda]}_i$ has the multiplicity

$$(\lambda_1 - 1)!(\lambda_2 - 1)! \dots (\lambda_k - 1)!.$$

The parity $\chi(P)$ being a class function we can write (A5) as:

$$F^{(n)} = \sum_{[\lambda], i} \chi(\lambda)(\lambda_1 - 1)!(\lambda_2 - 1)! \dots (\lambda_k - 1)! \delta^{[\lambda]}_i, \tag{A8}$$

where the sum over $[\lambda]$ runs on all partitions of n and, for a given partition $[\lambda]$ of n, the summation has to be extended to all i which give different $\delta^{[\lambda]}_i$.

For the symmetric groups S_2, S_3, and S_4, (A8) reads

$$F^{(2)} = 1 - \delta_{a_1 a_2},$$

$$F^{(3)} = 1 - (\delta_{a_1 a_2} + \delta_{a_1 a_3} + \delta_{a_2 a_3}) + 2\delta_{a_1 a_2 a_3},$$

$$\begin{aligned} F^{(4)} = 1 &- (\delta_{a_1 a_2} + \delta_{a_1 a_3} + \delta_{a_1 a_4} \\ &+ \delta_{a_2 a_3} + \delta_{a_2 a_4} + \delta_{a_3 a_4}) \\ &+ 2(\delta_{a_1 a_2 a_3} + \delta_{a_1 a_2 a_4} + \delta_{a_2 a_3 a_4} + \delta_{a_1 a_3 a_4}) \\ &+ (\delta_{a_1 a_2}\delta_{a_3 a_4} + \delta_{a_1 a_3}\delta_{a_2 a_4} + \delta_{a_1 a_4}\delta_{a_2 a_3}) - 6\delta_{a_1 a_2 a_3 a_4}. \end{aligned} \tag{A9}$$

The relation (A8) enables us to express (A1) with relations where all the summation indices run from 1 to A.

To close this appendix, we offer two expressions which are straightforward applications of the formulas (A1) and (A8)

(1) In Eq. (A1), let us put all the $x_i^{(a_i)}$ equal to 1. We then easily get

$$\sum_{a_1, a_2, \dots, a_n}^{A} 1^{(a_1)} 1^{(a_2)} \dots 1^{(a_n)} F^{(n)} = A(A-1) \dots (A-n+1) \tag{A10}$$

or

$$\sum_{a_1=n}^{A} \sum_{a_2=n-1}^{a_1 - 1} \dots \sum_{a_{n-1}=2}^{a_{n-2}-1} \sum_{a_n=1}^{a_{n-1}-1} 1^{(a_n)} = \binom{A}{n}, \tag{A11}$$

where $\binom{A}{n}$ is the binomial coefficient used in the normalization of our g.e.o.'s. This decomposition of $\binom{A}{n}$ is interesting of itself: By an iterative process it enables us to compute any sum of the type $\sum_{x=1}^{q} x^{\alpha}$, where α and q are positive integers.

(2) The introduction of Eq. (A8) in Eq. (A10) yields the striking formula:

$$\sum_{(\nu)} (-1)^{\sum_k \nu_{2k}} \frac{n!}{\nu_1! \nu_2! \dots \nu_n! 1^{\nu_1} 2^{\nu_2} \dots n^{\nu_n}} A^{\sum_{k=1}^{n} \nu_k} = A(A-1) \dots (A-n+1), \tag{A12}$$

where $\nu_1, \nu_2, \dots, \nu_n$ denote a cycle structure of S_n (i.e., $\sum_i i\nu_i = n$) and the summation on (ν) has to be extended on all the cycle structure of S_n. In the machinery of the symmetric groups S_n, the terms

$$(-1)^{\sum_k \nu_{2k}}, \quad \frac{n!}{\nu_1! \nu_2! \dots \nu_n! 1^{\nu_1} 2^{\nu_2} \dots n^{\nu_n}} \quad \text{and} \quad \sum_k \nu_k$$

are, respectively, the parity, the number of elements, and the number of disjoint cycles for the class $(1^{\nu_1}, 2^{\nu_2}, \dots, n^{\nu_n})$.

APPENDIX B

As is well known, the structure constants allow us to set up the adjoint representation. The purpose of this appendix is to construct an algebra from the structure constants and the coupling coefficients for $U(n)$. This will provide us with useful relations between the elements of this algebra.

It is a simple calculation to prove that

$$\begin{aligned} \sum_{AB} (\bar{A}A^B)(\bar{B}B^A) &= (1 - \delta^{\bar{A}}_0 \delta^{\bar{B}}_0)\delta^{\bar{B}}_{\bar{A}}, \\ \sum_{AB} [\bar{A}A^B][\bar{B}B^A] &= (1 + \delta^{\bar{A}}_0 \delta^{\bar{B}}_0)\delta^{\bar{B}}_{\bar{A}}. \end{aligned} \tag{B1}$$

Let us now define the matrices $\mathfrak{A}_A$ and $\mathfrak{B}_A$ by the elements:

$$\begin{aligned} |\mathfrak{A}_A|_{BC} &= -(AB^C), \\ |\mathfrak{B}_A|_{BC} &= [AB^C]. \end{aligned} \tag{B2}$$

We then have the symmetries

$$\begin{aligned} |\mathfrak{A}_A|_{BC} &= -|\mathfrak{A}_A|_{-C,-B}, \\ |\mathfrak{B}_A|_{BC} &= |\mathfrak{B}_A|_{-C,-B}, \end{aligned} \tag{B3}$$

from which it follows that $\mathfrak{A}_A$ and $\mathfrak{B}_A$ are linearly independent. Further, the symmetry of the matrices $\mathfrak{A}_A$ and $\mathfrak{B}_A$ induces the useful following relation:

$$\begin{aligned} \operatorname{Tr} M_1^{\alpha_1} M_2^{\alpha_2} \dots M_{j-1}^{\alpha_{j-1}} M_j^{\alpha_j} \\ = (-1)^{\sum_{i=1}^{j} \alpha_i} \operatorname{Tr} M_j^{\alpha_j} M_{j-1}^{\alpha_{j-1}} \dots M_2^{\alpha_2} M_1^{\alpha_1}, \end{aligned} \tag{B4}$$

where $M_i^{\alpha_i}$ stands for $\mathfrak{A}_A$ or $\mathfrak{B}_A$, according to whether α is 1 or 0, respectively.

The introduction of these matrices in Eqs. (B1) yields

$$\begin{aligned} |\mathfrak{A}_M \mathfrak{A}_{-M}|_{A,-B} &= \operatorname{Tr} \mathfrak{A}_A \mathfrak{A}_B = (1 - \delta^A_0 \delta^B_0)\delta^{-B}_A, \\ |\mathfrak{B}_M \mathfrak{B}_{-M}|_{A,-B} &= \operatorname{Tr} \mathfrak{B}_A \mathfrak{B}_B = (1 + \delta^A_0 \delta^B_0)\delta^{-B}_A. \end{aligned} \tag{B5}$$

Let us go now to the algebra spanned by our matrices. By rewriting the Jacobi identity verified by the structure constants in terms of the $\mathfrak{A}$'s, we get

$$[\mathfrak{A}_A, \mathfrak{A}_B]_- = (AB^C)\,\mathfrak{A}_C. \tag{B6}$$

From the remarkable identity

$$[X_A,[X_B,X_C]_+]_+ - [X_B,[X_C,X_A]_+]_+ = [[X_A,X_B]_-X_C]_-$$

we easily obtain

$$[BC^M][AM^E] - [CA^M][BM^E] = (AB^M)(MC^E)$$

Or, in terms of the $\mathfrak{B}$'s,

$$[\mathfrak{B}_A, \mathfrak{B}_B]_- = (AB^C)\mathfrak{A}_C. \tag{B7}$$

Finally, explicating the remarkable identity

$$[X_A,[X_B,X_C]_+]_- + [X_B,[X_C,X_A]_+]_- + [X_C,[X_A,X_B]_+]_- \equiv 0,$$

we find, by rearranging the letters and using the symmetry property of coefficients,

$$-[CB^D](ED^A) + (EB^D)[CD^A] = -[DB^A](EC^D),$$
$$[BC^D](AD^E) + [AC^D](BD^E) = [AB^D](DC^E).$$

In matrix notation, these equations read

$$[\mathfrak{A}_A, \mathfrak{B}_B]_- = (AB^C)\mathfrak{B}_C, \tag{B8}$$

$$\mathfrak{B}_A\mathfrak{A}_B + \mathfrak{B}_B\mathfrak{A}_A = [AB^C]\mathfrak{A}_C. \tag{B9}$$

From these basic commutation relations, we are now in a position to derive some useful relations involving products of a higher number of matrices $\mathfrak{A}_A$ and/or $\mathfrak{B}_A$. As a first example, by manipulating Eq. (B5) we get

$$\begin{aligned}
&|\mathfrak{B}_M\mathfrak{B}_{-M}\mathfrak{B}_A|_{CD} + |\mathfrak{B}_A\mathfrak{B}_M\mathfrak{B}_{-M}|_{CD} \\
&\quad = |\mathfrak{B}_M\mathfrak{B}_{-M}|_{CE}|\mathfrak{B}_A|_{ED} + |\mathfrak{B}_A|_{CE}|\mathfrak{B}_M\mathfrak{B}_{-M}|_{ED} \\
&\quad = (1+\delta_0^C\delta_0^E)\delta_C^E|\mathfrak{B}_A|_{ED} + (1+\delta_0^E\delta_0^D)\delta_E^D|\mathfrak{B}_A|_{CE} \\
&\quad = (2+\delta_0^C+\delta_0^D)|\mathfrak{B}_A|_{CD},
\end{aligned}$$

which after introduction of Eq. (B7) reads

$$\begin{aligned}
&2|\mathfrak{B}_M\mathfrak{B}_A\mathfrak{B}_{-M}|_{CD} + (ME^A)|\mathfrak{B}_M\mathfrak{A}_E + \mathfrak{A}_M\mathfrak{B}_E|_{CD} \\
&\quad = (2+\delta_0^C+\delta_0^D)|\mathfrak{B}_A|_{CD}.
\end{aligned}$$

The second term of the lhs of this Eq. is easily shown to be

$$(AB^C)(\mathfrak{A}_A\mathfrak{B}_B + \mathfrak{B}_A\mathfrak{A}_B) = (1-\delta_0^C)\mathfrak{B}_C \tag{B10}$$

[multiply (B8) by (AB^D) and sum over A and B], so that

$$2|\mathfrak{B}_M\mathfrak{B}_A\mathfrak{B}_{-M}|_{BC} = (1+\delta_0^A+\delta_0^B+\delta_0^C)|\mathfrak{B}_A|_{BC} \tag{B11}$$

can be rewritten as

$$2\mathrm{Tr}\mathfrak{B}_A\mathfrak{B}_B\mathfrak{B}_C = (1+\delta_0^A+\delta_0^B+\delta_0^C)[ABC]. \tag{B12}$$

In a similar way, we would obtain

$$2\mathrm{Tr}\mathfrak{A}_A\mathfrak{A}_B\mathfrak{A}_C = (ABC), \tag{B13}$$

$$2\mathrm{Tr}\mathfrak{B}_A\mathfrak{B}_B\mathfrak{A}_C = (ABC), \tag{B14}$$

$$2\mathrm{Tr}\mathfrak{B}_A\mathfrak{A}_B\mathfrak{A}_C = (1+\delta_0^A-\delta_0^B-\delta_0^C)[ABC]. \tag{B15}$$

As a second example, by multiplying both sides of (B7) by $\mathfrak{B}_C$ and $\mathfrak{A}_D$ and taking the trace, we get

$$2\mathrm{Tr}[\mathfrak{B}_A, \mathfrak{B}_B]_-\mathfrak{B}_C\mathfrak{A}_D = (AB-M)[MCD](1+\delta_0^C-\delta_0^D).$$

The symmetry property (B4) yields

$$\begin{aligned}
4\mathrm{Tr}\mathfrak{B}_A\mathfrak{B}_B\mathfrak{B}_C\mathfrak{A}_D &= (AB-M)[MCD] + [AB-M](MCD) \\
&\quad + (\sqrt{2}/n)\{(BCD)\delta_0^A + (ACD)\delta_0^B + (ABD)\delta_0^C \\
&\quad - (ABC)\delta_0^D\}.
\end{aligned} \tag{B16}$$

The same type of calculation would lead to

$$\begin{aligned}
4\mathrm{Tr}\mathfrak{B}_A\mathfrak{B}_B\mathfrak{B}_C\mathfrak{B}_D &= [AB-M][MCD] + [AD-M][MCB] \\
&\quad - [AC-M][MBD] + (2/n^2)\{\delta_A^B\delta_C^D + \delta_A^D\delta_C^B + \delta_A^C\delta_B^D\} \\
&\quad (\sqrt{2}/n)\{\delta_0^A[BCD] + \delta_0^B[ACD] + \delta_0^C[ADB] + \delta_0^D[ACB]\}.
\end{aligned}$$

APPENDIX C: EIGENVALUES OF THE GEL'FAND INVARIANTS FOR $U(n)$

The determination of the eigenvalues of the Gel'fand invariants[2] has given rise to several papers. We shall restrict ourselves to the paper by Popov and Perelomov[18] and to the one by Louck and Biedenharn.[19] (For more details see the references quoted in those two articles.) Those authors give a complete answer to this problem. The aim of this appendix is to give a simpler derivation of these eigenvalues.

Following Louck and Biedenharn, let us define from the generators of $U(n)$, the vector operator $V_{ij}(q)$ thus:

$$\begin{aligned}
&V_{ij}(1) = E_{ij}, \\
&V_{ij}(2) = \sum_{i_1=1}^{n} E_{ii_1}E_{i_1j}, \\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots \\
&V_{ij}(q) = \sum_{i_1,i_2,\ldots,i_{q-1}}^{n} E_{ii_1}E_{i_1i_2}\cdots E_{i_{q-1}j},
\end{aligned} \tag{C1}$$

which verify the commutation relations

$$[E_{ij}, V_{kl}(q)] = \delta_{jk}V_{il}(q) - \delta_{il}V_{kj}(q). \tag{C2}$$

For a given representation, let $|\overline{m}\rangle$ be the vector of highest weight. We then have

$$E_{ii}|\overline{m}\rangle = m_{in}|\overline{m}\rangle \tag{C3}$$

whereas

$$E_{ij}|\overline{m}\rangle = 0 \quad \text{with } i<j. \tag{C4}$$

Let us now investigate the quantity

$$P = \sum_{i=1}^{n} f_\alpha^{(n)}(i)V_{ii}(q)|\overline{m}\rangle, \tag{C5}$$

where $f_\alpha^{(n)}(i)$ is a scalar function depending on i. By using Eqs. (C3) and (C4) and

$$V_{ii}(q) = \sum_{j=1}^{n} V_{ij}(q-1)E_{ji},$$

Eq. (C5) reads

$$P = \sum_{i=1}^{n} f_\alpha^{(n)}(i) \left\{ m_{in} V_{ii}(q-1) + \sum_{j>i}^{n}{}_1 [V_{ij}(q-1), E_{ji}] \right\} |\overline{m}\rangle$$
$$= \sum_{i=1}^{n} f_\alpha^{(n)}(i) \left\{ m_{in} V_{ii}(q-1) + \sum_{j>i}^{n}{}_1 (\delta_{jj} V_{ii}(q-1) - \delta_{ii} V_{jj}(q-1)) \right\} |\overline{m}\rangle. \quad \text{(C6)}$$

We note that

$$\sum_{j>i}^{n}{}_1 f_\alpha^{(n)}(i)\delta_{ii} V_{jj}(q-1)|\overline{m}\rangle = \sum_{i=1}^{n} \sum_{j=1}^{i-1} f_\alpha^{(n)}(j) V_{ii}(q-1)|\overline{m}\rangle$$

so that

$$\sum_{i=1}^{n} f_\alpha^{(n)}(i) V_{ii}(q)|\overline{m}\rangle = \sum_{i=1}^{n} \left\{ (m_{in} - i + n + 1) f_\alpha^{(n)}(i) - \sum_{j=1}^{i} f_\alpha^{(n)}(j) \right\} V_{ii}(q-1)|\overline{m}\rangle, \quad \text{(C7)}$$

from which we get the recurrent relation

$$f_{q+1}^{(n)}(i) = (m_{in} - i + n + 1) f_q^{(n)}(i) - \sum_{j=1}^{i} f_q^{(n)}(j) \quad \text{(C8)}$$

ending by

$$f_1^{(n)}(i) = 1. \quad \text{(C9)}$$

Let us now go to the Gel'fand invariant

$$I_k^{(n)} = \sum_{i=1}^{n} V_{ii}(k).$$

It is then straightforward to see that its eigenvalue is

$$I_k^{(n)} = \sum_{i=1}^{n} m_{in} f_k^{(n)}(i), \quad \text{(C10)}$$

where $f_k^{(n)}(i)$ is completely determined by Eqs. (C8) and (C9). By introducing the hook length

$$P_{in} = m_{in} - i + n, \quad \text{(C11)}$$

Eq. (C8) reads

$$f_{q+1}^{(n)}(i) = P_{in} f_q^{(n)}(i) - \sum_{j=1}^{i-1} f_q^{(n)}(j). \quad \text{(C12)}$$

[1] H. Weyl, *The Theory of Groups and Quantum Mechanics*, translated by H. P. Robertson (Methuen, London, 1931).
[2] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946).
[3] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra*, translated by J. J. Griffin (Academic, New York, 1959).
[4] G. Racah, Lectures notes, Institute for Advanced Study, Princeton, New Jersey, 1951 (unpublished).
[5] J. D. Louck, Amer. J. Phys. **38**, 3 (1970), and references quoted here.
[6] L. C. Biedenharn, in *Lectures in Theoretical Physics, Boulder, 1962* (Interscience, New York, 1963), Vol. 5.
[7] J. H. Bartlett, Jr., Phys. Rev. **49**, 102 (1936).
[8] E. Majorana, Z. Phys. **82**, 137 (1933).
[9] W. Heisenberg, Z. Phys. **77**, 1 (1932).
[10] H. Primakoff and T. Holstein, Phys. Rev. **55**, 218 (1938); L. Janossy, Proc. Cambridge Phil. Soc. **35**, 616 (1939).
[11] R. K. Bhaduri, Y. Nogami, and C. K. Ross, Phys. Rev. C **2**, 2082 (1970); B. A. Loiseau, Y. Nogami, and C. K. Ross, Nucl. Phys. **A165**, 601 (1971); B. H. J. McKellar and R. Rajaraman, Phys. Rev. C **3**, 1877 (1971); P. C. Y. Yip, Y. Nogami, and C. K. Ross, Nucl. Phys. **A176**, 505 (1971).
[12] W. G. Harter, Phys. Rev. A **3**, 1891 (1971).
[13] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[14] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[15] I. M. Gel'fand, Mat. Sb. **26**, 103 (1950).
[16] A. Partensky, J. Math. Phys. **13**, 621 (1972).
[17] Cf. Ref. 16 and Appendix B.
[18] V. S. Popov and A. M. Perelomov, Yad. Fiz. **7**, 460 (1968) [Sov. J. Nucl. Phys. **7**, 290 (1968)].
[19] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970), see Appendix B.

Steady-State Sound Propagation in Continuous, Statistically Isotropic Media

Jerome A. Neubert

Naval Undersea Research and Development Center, Pasadena Laboratory, Pasadena, California 91107

(Received 6 April 1972)

A stochastic Eulerian–Lagrangian procedure is applied to steady-state sound propagation from a small, collimated acoustic source to an omnidirectional point receiver imbedded in an infinite, continuous, statistically isotropic medium. An analytic procedure is developed for obtaining the Lagrangian measure function $B(\mathbf{x}, \xi \mid s)$ from its characteristic function $\phi(\mathbf{k}, \xi \mid s)$ for stochastic-Fermat media. The results include a coefficient of intensity variation V that evinces a frequency-dependent, phase-dominance region and a frequency-independent, amplitude-dominance region. The methods employed in this study are new to the problem of sound propagation through continuous stochastic media and avoid three common difficulties: (1) range limitations due to cumulative phase effects, (2) discrete scattering assumptions, and (3) restriction to an Eulerian path.

INTRODUCTION

In a previous paper,[1] it was shown that, for very low turbulent Mach numbers u/c and monochromatic transmissions, steady-state sound propagation in a turbulent fluid can be adequately represented by the stochastic Helmholtz equation

$$\nabla^2 p + k_0^2 \mu^2 p = 0, \quad (1)$$

where p is the pressure wave, k_0 is the free-space wavenumber, and μ is the refractive index, if the acoustic wavelength λ statisfies the condition

$$\lambda \lesssim \lambda_g / [10(u/c)(\sigma_\kappa)^{1/2}]^{1/2}, \quad (2)$$

where λ_g is the Taylor microscale for velocity fluctuations, u is the rms turbulent velocity, c is the sound propagation velocity, and σ_κ is the Prandtl number. For the purposes of this study, it will be assumed that Eq. (1) is valid for steady-state sound propagation through continuous, statistically isotropic media and that μ describes the stochastic nature of the media.

Assume that μ is statistically homogeneous and has the mean value

$$P = \sum_{i=1}^{n} f_\alpha^{(n)}(i) \left\{ m_{in} V_{ii}(q-1) + \sum_{j>i}^{n}{}_1 [V_{ij}(q-1), E_{ji}] \right\} |\overline{m}\rangle$$
$$= \sum_{i=1}^{n} f_\alpha^{(n)}(i) \left\{ m_{in} V_{ii}(q-1) + \sum_{j>i}^{n}{}_1 (\delta_{jj} V_{ii}(q-1) - \delta_{ii} V_{jj}(q-1)) \right\} |\overline{m}\rangle. \quad \text{(C6)}$$

We note that

$$\sum_{j>i}^{n}{}_1 f_\alpha^{(n)}(i)\delta_{ii} V_{jj}(q-1)|\overline{m}\rangle = \sum_{i=1}^{n} \sum_{j=1}^{i-1} f_\alpha^{(n)}(j) V_{ii}(q-1)|\overline{m}\rangle$$

so that

$$\sum_{i=1}^{n} f_\alpha^{(n)}(i) V_{ii}(q)|\overline{m}\rangle = \sum_{i=1}^{n} \left\{ (m_{in} - i + n + 1) f_\alpha^{(n)}(i) - \sum_{j=1}^{i} f_\alpha^{(n)}(j) \right\} V_{ii}(q-1)|\overline{m}\rangle, \quad \text{(C7)}$$

from which we get the recurrent relation

$$f_{q+1}^{(n)}(i) = (m_{in} - i + n + 1) f_q^{(n)}(i) - \sum_{j=1}^{i} f_q^{(n)}(j) \quad \text{(C8)}$$

ending by

$$f_1^{(n)}(i) = 1. \quad \text{(C9)}$$

Let us now go to the Gel'fand invariant

$$I_k^{(n)} = \sum_{i=1}^{n} V_{ii}(k).$$

It is then straightforward to see that its eigenvalue is

$$I_k^{(n)} = \sum_{i=1}^{n} m_{in} f_k^{(n)}(i), \quad \text{(C10)}$$

where $f_k^{(n)}(i)$ is completely determined by Eqs. (C8) and (C9). By introducing the hook length

$$P_{in} = m_{in} - i + n, \quad \text{(C11)}$$

Eq. (C8) reads

$$f_{q+1}^{(n)}(i) = P_{in} f_q^{(n)}(i) - \sum_{j=1}^{i-1} f_q^{(n)}(j). \quad \text{(C12)}$$

[1] H. Weyl, *The Theory of Groups and Quantum Mechanics*, translated by H. P. Robertson (Methuen, London, 1931).
[2] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946).
[3] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra*, translated by J. J. Griffin (Academic, New York, 1959).
[4] G. Racah, Lectures notes, Institute for Advanced Study, Princeton, New Jersey, 1951 (unpublished).
[5] J. D. Louck, Amer. J. Phys. **38**, 3 (1970), and references quoted here.
[6] L. C. Biedenharn, in *Lectures in Theoretical Physics, Boulder, 1962* (Interscience, New York, 1963), Vol. 5.
[7] J. H. Bartlett, Jr., Phys. Rev. **49**, 102 (1936).
[8] E. Majorana, Z. Phys. **82**, 137 (1933).
[9] W. Heisenberg, Z. Phys. **77**, 1 (1932).
[10] H. Primakoff and T. Holstein, Phys. Rev. **55**, 218 (1938); L. Janossy, Proc. Cambridge Phil. Soc. **35**, 616 (1939).
[11] R. K. Bhaduri, Y. Nogami, and C. K. Ross, Phys. Rev. C **2**, 2082 (1970); B. A. Loiseau, Y. Nogami, and C. K. Ross, Nucl. Phys. **A165**, 601 (1971); B. H. J. McKellar and R. Rajaraman, Phys. Rev. C **3**, 1877 (1971); P. C. Y. Yip, Y. Nogami, and C. K. Ross, Nucl. Phys. **A176**, 505 (1971).
[12] W. G. Harter, Phys. Rev. A **3**, 1891 (1971).
[13] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[14] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[15] I. M. Gel'fand, Mat. Sb. **26**, 103 (1950).
[16] A. Partensky, J. Math. Phys. **13**, 621 (1972).
[17] Cf. Ref. 16 and Appendix B.
[18] V. S. Popov and A. M. Perelomov, Yad. Fiz. **7**, 460 (1968) [Sov. J. Nucl. Phys. **7**, 290 (1968)].
[19] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970), see Appendix B.

Steady-State Sound Propagation in Continuous, Statistically Isotropic Media

Jerome A. Neubert

Naval Undersea Research and Development Center, Pasadena Laboratory, Pasadena, California 91107

(Received 6 April 1972)

A stochastic Eulerian–Lagrangian procedure is applied to steady-state sound propagation from a small, collimated acoustic source to an omnidirectional point receiver imbedded in an infinite, continuous, statistically isotropic medium. An analytic procedure is developed for obtaining the Lagrangian measure function $B(\mathbf{x}, \xi \mid s)$ from its characteristic function $\phi(\mathbf{k}, \xi \mid s)$ for stochastic-Fermat media. The results include a coefficient of intensity variation V that evinces a frequency-dependent, phase-dominance region and a frequency-independent, amplitude-dominance region. The methods employed in this study are new to the problem of sound propagation through continuous stochastic media and avoid three common difficulties: (1) range limitations due to cumulative phase effects, (2) discrete scattering assumptions, and (3) restriction to an Eulerian path.

INTRODUCTION

In a previous paper,[1] it was shown that, for very low turbulent Mach numbers u/c and monochromatic transmissions, steady-state sound propagation in a turbulent fluid can be adequately represented by the stochastic Helmholtz equation

$$\nabla^2 p + k_0^2 \mu^2 p = 0, \quad (1)$$

where p is the pressure wave, k_0 is the free-space wavenumber, and μ is the refractive index, if the acoustic wavelength λ statisfies the condition

$$\lambda \lesssim \lambda_g / [10(u/c)(\sigma_\kappa)^{1/2}]^{1/2}, \quad (2)$$

where λ_g is the Taylor microscale for velocity fluctuations, u is the rms turbulent velocity, c is the sound propagation velocity, and σ_κ is the Prandtl number. For the purposes of this study, it will be assumed that Eq. (1) is valid for steady-state sound propagation through continuous, statistically isotropic media and that μ describes the stochastic nature of the media.

Assume that μ is statistically homogeneous and has the mean value

$$\langle\mu\rangle = \langle\mu(\mathbf{x})\rangle = 1. \tag{3}$$

It is convenient to represent μ by

$$\mu = \mu(\mathbf{x}) = 1 + \alpha n(\mathbf{x}) = 1 + \alpha n, \tag{4}$$

where αn is the fluctuation of μ about its mean value and α is the rms fluctuation of μ,

$$\alpha \equiv [\langle(\mu - 1)^2\rangle]^{1/2} = [\langle\mu^2\rangle - 1]^{1/2} \geq 0, \tag{5}$$

so that

$$\langle n\rangle = \langle n(\mathbf{x})\rangle = 0 \tag{6}$$

and

$$\langle n^2\rangle = \langle n^2(\mathbf{x})\rangle = 1. \tag{7}$$

It is mathematically convenient and physically accurate[2-8] to assume that

$$\alpha \ll 1; \tag{8}$$

$\mu(\mathbf{x})$ is then only *weakly inhomogeneous*.

In Ref. 2, it was concluded that a two-variable expansion is sufficient for representing sound propagation through a continuous, weakly inhomogeneous medium and that the Debye asymptotic expansion[9,2,5]

$$p(\mathbf{x}) = [A(\mathbf{x}) + (1/ik_0)A^{(1)}(\mathbf{x}) + \dots]e^{ik_0\psi(\mathbf{x})}, \tag{9}$$

where $\psi, A, A^{(1)}, \cdots$ are real-valued functions of $\mathbf{x}$ and time t, is a proper two-variable expansion. Applying Eq. (9) to Eq. (1) and separating into real and imaginary parts yields, to order[2] $\alpha\sigma_\kappa\sqrt{R_t}/\lambda_g^2 k_0^2$ for a turbulent fluid, the eikonal equation,

$$\nabla\psi\cdot\nabla\psi - \mu^2 = 0, \tag{10}$$

and the transport equation,

$$2\nabla A\cdot\nabla\psi + A\nabla^2\psi = 0, \tag{11}$$

where R_t is the turbulent Reynolds number

$$R_t \equiv u\lambda_g/\nu, \tag{12}$$

and ν is the kinematic viscosity. Equation (10) will be assumed throughout this paper, although other extremum relations can also be treated in an exactly analogous fashion.

By standard procedures,[5,9-11] Eq. (10) produces the Fermat relation

$$\mu_{,i} = \frac{d}{ds}\left(\mu\frac{dX_i}{ds}\right) \tag{13}$$

and

$$\psi[\mathbf{X}(s,\boldsymbol{\xi})] = \psi(s) = \psi(\boldsymbol{\xi}) + \int_0^s ds'\mu(s'), \tag{14}$$

where s is the curvilinear path distance from the initial point $\boldsymbol{\xi} = \mathbf{X}(0,\boldsymbol{\xi})$ on the initial surface S_0 to the terminal point $\mathbf{X} = \mathbf{X}(s,\boldsymbol{\xi})$. Note that $\mu(s)$ and $\psi(s)$ are Lagrangian functionals since they must be integrated along the particular path $\mathbf{X}(s,\boldsymbol{\xi})$, dictated by the particular Eulerian field $\mu(\mathbf{x})$ and the pertinent initial conditions, from $\boldsymbol{\xi}$ to $\mathbf{X}$ for each given s.[12] Equations (10) and (11) render[2,3,5,13,14] the Lagrangian relation for the sound-pressure wave after having traversed a curvilinear distance s from $\boldsymbol{\xi}$ to $\mathbf{X}(s,\boldsymbol{\xi})$ over the continuous field $\mu(\mathbf{x})$:

$$p[\mathbf{X}(s,\boldsymbol{\xi})] = A(s)e^{ik_0\psi(s)} \tag{9'}$$

$$= p_0(\boldsymbol{\xi})\exp\left(-\frac{1}{2}\int_0^s \frac{ds'}{\mu(s')}\nabla^2\psi + ik_0\int_0^s ds'\mu(s')\right) \tag{15}$$

$$= p_0(\boldsymbol{\xi})\exp\left[-\frac{1}{2}\int_0^s \frac{ds'}{\mu(s')}\left(\mu\frac{d\mathbf{X}_i}{ds}\right)_{,i} + ik_0\int_0^s ds'\mu(s')\right] \tag{16}$$

$$= p_0(\boldsymbol{\xi})\exp\left(-\frac{1}{2}\int_0^s \frac{ds'}{\mu(s')}\int_0^{s'} ds''\mu_{,ii}(s'') + ik_0\int_0^s ds'\mu(s')\right), \tag{17}$$

where

$$p_0(\boldsymbol{\xi}) = A(\boldsymbol{\xi})e^{ik_0\psi(\boldsymbol{\xi})}. \tag{18}$$

The occurrence of Eq. (13) implies that *stochastic-Fermat* media are under consideration.

In one dimension, Eq. (16) produces[2,5] the coefficient of intensity fluctuation [see Eq. (84)]

$$V^2 \sim e^{2\alpha^2 k_0^2 L_e x} - 1, \tag{19}$$

where L_e is the Eulerian integral scale

$$L_e \equiv \lim_{x\to\infty}\int_0^x d\rho_1(1 - \rho_1/x)P_e(\rho_1) \tag{20}$$

and

$$P_e(\rho_1) \equiv \langle n(x')n(x'')\rangle, \tag{21a}$$

$$\rho_1 \equiv |x' - x''|. \tag{21b}$$

Equation (19) reduces to

$$V^2 \sim 2\alpha^2 k_0^2 L_e x \tag{22}$$

when the condition

$$1 \gg 2\alpha^2 k_0^2 L_e x \tag{23}$$

is imposed.

In Ref. 2, it was implied that Mintzer[6] had reduced his stochastic Eulerian–Lagrangian sound propagation problem[15] to its Eulerian equivalent by considering the refractive index fluctuations along the source-to-receiver line since L_e equals $\int_0^\infty d\rho_1 N(\rho_1)$ of Sec. IV when the latter exists. In addition, it was demonstrated that there exists no k_0-independent behavior for large k_0 in one dimension, although Bergmann,[16] Mintzer,[6] and Potter and Murphy[17] all indicate such behavior occurs in three dimensions. Therefore, this k_0-independent behavior (if it does occur in an infinite, continuous, statistically isotropic medium) is either a three-dimensional (i.e., geometric) effect or is due to some mechanism other than turbulence (e.g., discrete impurities). This paper addresses itself to the former effect.

In this paper, the stochastic Eulerian–Lagrangian approach of Ref. 15 is applied to the Lagrangian pressure wave relation of Eq. (17) for steady-state sound propagation from a small, collimated acoustic source to an omnidirectional point receiver imbedded in an infinite, continuous, statistically isotropic, stochastic-Fermat medium. An analytic procedure for obtaining the Lagrangian measure function[15] $B(\mathbf{x},\boldsymbol{\xi}\,|\,s)$ from its

characteristic function $\phi(\mathbf{k}, \boldsymbol{\xi} \mid s)$ is illustrated by considering a plane transducer face perpendicular to the x axis and located at the origin for a Cartesian coordinate system. The initial path angles at every point $\boldsymbol{\xi}$ on this initial surface S_0 are represented by $\theta_0(\boldsymbol{\xi})$, $\phi_0(\boldsymbol{\xi})$. Specifying the initial pressure amplitude $p_0(\boldsymbol{\xi})$ and path angles $\theta_0(\boldsymbol{\xi})$, $\phi_0(\boldsymbol{\xi})$ for Eq. (17) permits the inclusion of beam patterns for the transducer; however, this study shall be restricted to the case of a uniform, collimated beam by choosing

$$p(\boldsymbol{\xi}) = p_0, \quad \text{constant and real,} \tag{24}$$

$$\theta_0(\boldsymbol{\xi}) = \theta_0 = \pi/2, \tag{25a}$$

$$\phi_0(\boldsymbol{\xi}) = \phi_0 = 0, \tag{25b}$$

so that the vector expressing the initial path orientation is given by

$$\left[\frac{dX_j}{ds}(0)\right] \equiv \left[\frac{dX}{ds}(0), \frac{dY}{ds}(0), \frac{dZ}{ds}(0)\right] \tag{26}$$

$$= [\sin\theta_0 \cos\phi_0, \sin\theta_0 \sin\phi_0, \cos\theta_0] \tag{26'a}$$

$$= [1, 0, 0]. \tag{26'b}$$

It will become apparent in Sec. I that this choice of geometric orientation will greatly simplify the form and facilitate the intepretation of $B(x_i - \xi_i \mid s)$; this does not alter its physical content. The complex, interwoven development of Sec. I is typical of the involved asymptotic Lagrangian evaluations[3,5,18] that often result when the theory of Refs. 15, 18, and 5 is applied to real world sound propagation problems. The main preoccupation of Sec. II is the evaluation of $\langle p(\mathbf{x}) \rangle$ from $p[\mathbf{X}(s, \boldsymbol{\xi})]$ via Eq. (14) of Ref. 15. In the course of this evaluation, it is shown that a saddle-point treatment (cf. Appendix D) of the resulting integral provides an adequate approximation of $\langle p(\mathbf{x}) \rangle$ for this type of problem. In Sec. III, $\langle |p(\mathbf{x})|^2 \rangle$ is determined by a similar, double saddle-point technique and Sec. IV develops a useful relation for the coefficient of intensity fluctuation V for steady-state sound propagation through an infinite, continuous stochastic medium. A brief comparison of this result with the classical treatments of Bergmann[16] and Mintzer[6,7] is also presented.

I. DETERMINATION OF THE LAGRANGIAN MEASURE FUNCTION

Once the physical problem has been clearly stated, the Lagrangian measure function $B(x_i - \xi_i \mid s)$ can be determined[15,5] from its Lagrangian characteristic function $\phi(\mathbf{k}, \boldsymbol{\xi} \mid s)$ which, in Cartesian coordinates, is defined by

$$\phi(\mathbf{k}, \boldsymbol{\xi} \mid s) \equiv E\{e^{ik_j(X_j - \xi_j)}\} \tag{27}$$

$$= E\left\{ \exp\left(i k_j \int_0^s ds' \frac{dX_j}{ds}(s')\right)\right\}, \tag{27'}$$

where

$$\frac{dX_j}{ds}(s) \equiv \frac{dX_j}{ds}[\mathbf{X}(s, \boldsymbol{\xi})], \tag{28}$$

$$X_j - \xi_j \equiv X_j(s, \boldsymbol{\xi}) - \xi_j = \int_0^s ds' \frac{dX_j}{ds}(s'), \tag{29}$$

$$X_1 - \xi_1 = X - \xi \equiv X(s, \boldsymbol{\xi}) - \xi = \int_0^s ds' \frac{dX}{ds}(s'), \tag{29a}$$

$$X_2 - \xi_2 = Y - \eta \equiv Y(s, \boldsymbol{\xi}) - \eta = \int_0^s ds' \frac{dY}{ds}(s'), \tag{29b}$$

$$X_3 - \xi_3 = Z - \zeta \equiv Z(s, \boldsymbol{\xi}) - \zeta = \int_0^s ds' \frac{dZ}{ds}(s'). \tag{29c}$$

Here, $k_j = k_1, k_2, k_3$ are dummy variables, and only stochastic quantities, such as $\int_0^s ds' (dX_j/ds)(s')$, have pertinence in $\phi(\mathbf{k}, \boldsymbol{\xi} \mid s)$ since deterministic factors are canceled out in Eq. (40). Note that the $X_j(s, \boldsymbol{\xi})$ take on all their possible values in Eq. (27) and later, in Eq. (40), $B(x_i - \xi_i \mid s)$ picks the desired value for each. Therefore, the Lagrangian ensemble, rather than subensemble, expectation[18,5] must be applied in Eq. (27) since $B(x_i - \xi_i \mid s)$ must be permitted to treat all possible values of $X_j(s, \boldsymbol{\xi})$ in order to be capable of determining the measure per unit s per unit $\boldsymbol{\xi}$ of the Lagrangian subensemble under consideration in Eq. (40) for each $\boldsymbol{\xi}$, s, $\mathbf{x}$. For analytical convenience, a plane transducer S_0 at $\xi = 0$ and the initial orientation of Eq. (26b) will be chosen.

Define the following quantities using Eq. (26b) and the stochastic-Fermat relation of Eq. (13):

$$u_1 = u_1[\mathbf{X}(s, \boldsymbol{\xi})] \equiv \frac{dX}{ds}(s) - m_1 \tag{30a}$$

$$= \left[\frac{\mu(0)}{\mu(s)} - E\left\{\frac{\mu(0)}{\mu(s)}\right\}\right] \frac{dX}{ds}(0) + \int_0^s ds' \frac{\mu_{,1}(s')}{\mu(s)} \tag{31a}$$

$$= \left[\frac{\mu(0)}{\mu(s)} - E\left\{\frac{\mu(0)}{\mu(s)}\right\}\right] + \int_0^s ds' \frac{\mu_{,1}(s')}{\mu(s)}, \tag{32a}$$

$$u_2 = u_2[\mathbf{X}(s, \boldsymbol{\xi})] \equiv \frac{dY}{ds}(s) - m_2 \tag{30b}$$

$$= \left[\frac{\mu(0)}{\mu(s)} - E\left\{\frac{\mu(0)}{\mu(s)}\right\}\right] \frac{dY}{ds}(0) + \int_0^s ds' \frac{\mu_{,2}(s')}{\mu(s)} \tag{31b}$$

$$= \int_0^s ds' \frac{\mu_{,2}(s')}{\mu(s)}, \tag{32b}$$

$$u_3 = u_3[\mathbf{X}(s, \boldsymbol{\xi})] \equiv \frac{dZ}{ds}(s) - m_3 \tag{30c}$$

$$= \left[\frac{\mu(0)}{\mu(s)} - E\left\{\frac{\mu(0)}{\mu(s)}\right\}\right] \frac{dZ}{ds}(0) + \int_0^s ds' \frac{\mu_{,3}(s')}{\mu(s)} \tag{31c}$$

$$= \int_0^s ds' \frac{\mu_{,3}(s')}{\mu(s)}, \tag{32c}$$

where[18,5]

$$\mu(s) \equiv \mu[\mathbf{X}(s, \boldsymbol{\xi})] \tag{33}$$

$$\mu(0) \equiv \mu[\mathbf{X}(s = 0, \boldsymbol{\xi})] = \mu(\boldsymbol{\xi}), \tag{33'}$$

$$m_1 \equiv E\left\{\frac{dX}{ds}(s)\right\} \tag{34a}$$

$$= E\left\{\frac{\mu(0)}{\mu(s)}\right\} \frac{dX}{ds}(0) \tag{35a}$$

$$= 1 - \alpha^2 \int_0^s d\sigma (s - \sigma)[R(\sigma) - 0(\alpha)] \tag{36a}$$

$$R(\sigma) = R(s', s'') \equiv E\{n_{,i}(s') n_{,i}(s'')\}, \quad \sigma \equiv |s' - s''|, \tag{37}$$

$$m_2 \equiv E\left\{\frac{dY}{ds}(s)\right\} \tag{34b}$$

$$= 0, \tag{36b}$$

and

$$m_3 \equiv E\left\{\frac{dZ}{ds}(s)\right\} \tag{34c}$$

$$= 0. \tag{36c}$$

Note that the geometric condition of Eq. (26b) gives $m_2 \equiv m_3$ as expected for a randomly isotropic medimedium. However, the Lagrangian ensemble expectation $E\{(dX/ds)(s)\}$ deviates from $E\{(dX/ds)(0)\}$ as explained in Refs. 18 and 5. In the nonstochastic limit, i.e., $\alpha \to 0$,

$$x = E\{X(s,\xi)\} = \int_0^s ds' E\left\{\frac{dX}{ds}(s')\right\} = s, \tag{38a}$$

$$y - \eta = E\{Y(s,\xi) - \eta\} = \int_0^s ds' E\left\{\frac{dY}{ds}(s')\right\} = 0, \tag{38b}$$

$$z - \zeta = E\{Z(s,\xi) - \zeta\} = \int_0^s ds' E\left\{\frac{dZ}{ds}(s')\right\} = 0, \tag{38c}$$

as required in a uniform medium where $\mu(\mathbf{x}) = 1$.

The characteristic function can now be expressed, via Eq. (15) of Ref. 15 and Eqs. (27)–(36c), in the form form[15,5]

$$\phi(\mathbf{k},\xi\,|\,s) = E\left\{\exp\left(ik_j \int_0^s ds' \frac{dX_j}{ds}(s')\right)\right\} \tag{27'a}$$

$$\sim \exp\left(-\tfrac{1}{2}k_i k_j \int_0^s ds' \int_0^s ds'' E\{u_i(s')u_j(s'')\} + ik_j \int_0^s ds' m_j(s')\right). \tag{39}$$

Note that $u_i(s')$ and $u_j(s'')$ are along the same $\mathbf{X}(s,\xi)$ path for each field realization $\mu_\beta(\mathbf{x})$; they are assumed to satisfy some suitable set of sufficiency conditions[19,5,15] for the validity of Eqs. (39) and (44). The $\mathbf{x}$ dependence is implicit in $\phi(\mathbf{k},\xi\,|\,s)$ but becomes explicit in $B(x_i - \xi_i\,|\,s)$; see Eq. (40). Since s is fixed and, therefore, nonstochastic, the integrals in Eq. (39) are Lagrangian stochastic integrals.[18,5] Thus, the Lagrangian measure function, which is actually a joint probability density for this statistically isotropic case, is given by

$$B(x_i - \xi_i\,|\,s) = B(x, y-\eta, z-\zeta\,|\,s) = (2\pi)^{-3}\int_{-\infty}^{\infty} d\mathbf{k}\, \phi(\mathbf{k},\xi\,|\,s)e^{-ik_j(x_j-\xi_j)}. \tag{40}$$

Equation (14) of Ref. 15 can now be expressed by

$$\langle F(\mathbf{x})\rangle = \int_0^\infty ds \int_{S_0} d\eta\, d\zeta\, \tilde{E}\{F[\mathbf{X}(s,\xi)]\} B(x, y-\eta, z-\zeta\,|\,s). \tag{41}$$

This relation denotes a weighted synthesis of all the paths from S_0 which terminate at $\mathbf{x}$.

This analysis will be restricted to stochastic-Fermat paths (the theory of Lagrangian stochastic analysis for Fermat paths was developed in Refs. 3, 5, 15 and 18). Applying Eq. (39) to Eq. (40) yields

$$B(x, y-\eta, z-\zeta\,|\,s) = \frac{1}{8\pi^3}\iiint_{-\infty}^{\infty} dk_1\, dk_2\, dk_3\, \phi(\mathbf{k},\xi\,|\,s)e^{-ik_1 x - ik_2(y-\eta) - ik_3(z-\zeta)} \tag{40'}$$

$$\sim \frac{1}{8\pi^3}\iiint_{-\infty}^{\infty} dk_1\, dk_2\, dk_3 \exp\left[-\tfrac{1}{2}k_1 k_j U_{ij} - ik_1\left(x - \int_0^s m_1 ds'\right)\right] e^{-ik_2(y-\eta) - ik_3(z-\zeta)} \tag{42}$$

$$= \frac{\exp[-\frac{1}{2}U_{ij}^{-1}(x_i - \xi_i - m_{0i})(x_j - \xi_j - m_{0j})]}{2\pi\sqrt{2\pi\|U_{ij}\|}} \tag{43}$$

$$\sim \gamma(x - s + \tfrac{1}{2}\alpha^2 s^2 R_0 A\,|\,s)\gamma(y-\eta\,|\,s)\gamma(z-\zeta\,|\,s) \tag{44}$$

$$\equiv \gamma(\mathbf{x},\xi\,|\,s), \tag{45}$$

where

$$U_{ij} \equiv \int_0^s ds' \int_0^s ds'' E\{u_i(s')u_j(s'')\}. \tag{46}$$

The Lagrangian spreading matrix[15] $[U_{ij}]$, whose components are given by Eq. (46), is now expressed in terms of its principal axes due to the convenient choice of orientation expressed by Eq. (26b) and the statistical isotropy of $\mu(\mathbf{x})$; see Eqs. (A23) through (A25). The determinant of $[U_{ij}]$ is represented by $\|U_{ij}\|$ [see Eqs. (A14)–(A15)] and the components of the inverse matrix of $[U_{ij}]$ are denoted by U_{ij}^{-1} [see Eqs. (A17a)–(A19)]. The transformation to Eq. (43) follows when $[U_{ij}]$ is positive-definite and the components m_{0i} are given in Eqs. (A13a)–(A13c). Equation (44) follows asymptotically after considerable detailed calculation; see Appendix A, especially Eqs. (A39)–(A41).

Equation (45) then permits Eq. (41) to be expressed asymptotically as

$$\langle F(x)\rangle \sim \int_0^\infty ds \int_{S_0} d\eta d\xi\, \tilde{E}\{F[\mathbf{X}(s,\xi)]\}\gamma(\mathbf{x},\xi\,|\,s). \tag{47}$$

It can be shown[20,21] that, for a positive parameter $\mathbf{b}$, inside an integral including $x = 0$

$$\lim_{\mathbf{b}\to 0} \frac{1}{\sqrt{2\pi\mathbf{b}}} \exp\left(-\frac{x^2}{2\mathbf{b}^2}\right) = \delta(x), \tag{48}$$

i.e., this limit behaves like a delta function. Thence,[15]

$$\lim_{\alpha\to 0} \gamma(x - s + \tfrac{1}{2}\alpha^2 s^2 R_0 A\,|\,s) = \delta(x-s), \tag{49a}$$

$$\lim_{\alpha\to 0} \gamma(y-\eta\,|\,s) = \delta(y-\eta), \tag{49b}$$

$$\lim_{\alpha\to 0} \gamma(z-\zeta\,|\,s) = \delta(z-\zeta). \tag{49c}$$

Thus,

$$\int_{-\eta_0}^{\eta_0} d\eta\, \gamma(y-\eta\,|\,s) \xrightarrow[\alpha\to 0]{} \int_{-\eta_0}^{\eta_0} d\eta\, \delta(y-\eta) = \begin{cases} 1, & |y| \le \eta_0 \\ 0, & |y| > \eta_0, \end{cases} \tag{50a}$$

$$\int_{-\zeta_0}^{\zeta_0} d\zeta\, \gamma(z-\zeta\,|\,s) \xrightarrow[\alpha\to 0]{} \int_{-\zeta_0}^{\zeta_0} d\zeta\, \delta(z-\zeta) = \begin{cases} 1, & |z| \le \zeta_0 \\ 0, & |z| > \zeta_0, \end{cases} \tag{50b}$$

and, from Eqs. (49a)–(49c) and (45), Eq. (47) yields

$$\lim_{\alpha\to 0}\langle F(\mathbf{x})\rangle \sim \int_0^\infty ds \int_{-\eta_0}^{\eta_0} d\eta \int_{-\zeta_0}^{\zeta_0} d\zeta\, \delta(x-s)\delta(y-\eta) \times \delta(z-\zeta)\tilde{F}[(\mathbf{X}(s,\zeta)] \tag{51'}$$

$$= \begin{cases} F(x,\eta,\zeta), & -\eta_0 \le \eta \le \eta_0,\ -\zeta_0 \le \zeta \le \zeta_0 \\ 0, & \text{otherwise,} \end{cases} \tag{51}$$

when the initial surface S_0 is a rectangle $2\eta_0$ by $2\zeta_0$. Equation (51) means that in the nonstochastic limit, i.e., $\alpha \to 0, \gamma(\mathbf{x}, \xi \mid s)$ behaves like a delta function, i.e.,

$$\lim_{\alpha \to 0} \gamma(\mathbf{x}, \xi \mid s) = \delta(x - s)\,\delta(y - \eta)\,\delta(z - \zeta), \tag{52}$$

so that Eq. (47) yields the necessary nonstochastic result which would occur in a uniform, nondissipative medium.[15]

II. DETERMINATION OF $\langle p(\mathbf{x}) \rangle$

In this section, the sound pressure wave expectation $\langle p(\mathbf{x}) \rangle$ is calculated by applying the saddle-point method (see Appendix D) which seems natural for treating sound propagation problems via Eq. (47). To obtain $\langle p(\mathbf{x}) \rangle$ [see Eq. (57)], $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ is first found by the analytical procedures of Refs. 18 and 5. Applying the Lagrangian subensemble expectation of Sec. V of Ref. 18 to Eq. (17) for μ statistically isotropic and employing Eq. (15) of Ref. 15 gives

$$\tilde{E}\{p[\mathbf{X}(s, \xi)]\} = p_0(\xi)\,\tilde{E}\left\{\exp\left(ik_0 \int_0^s ds'\mu(s') - \tfrac{1}{2}\int_0^s ds' \int_0^{s'} ds'' \frac{\mu_{,ii}(s'')}{\mu(s')}\right)\right\} \tag{53}$$

$$\sim p_0(\xi)\,\exp(-\tfrac{1}{2}k_i k_j \tilde{V}_{ij} + ik_j \int_0^s ds' \tilde{v}_j(s')) \tag{54}$$

$$\sim p_0(\xi) e^{-\alpha^2(k_0^2 \tilde{L} s - s^3 \tilde{Q}_0 \tilde{E}/12) + ik_0(1+\alpha^2 s\tilde{D})s} \tag{55}$$

$$\xrightarrow[\alpha \to 0]{} p_0(\xi) e^{ik_0 s}, \tag{56}$$

remembering Eqs. (8) and (26b), where $i, j = 1, 2$, with $k_1 = k_0, k_2 = i/2$. Equation (15) of Ref. 15 and Eqs. (B1)–(B6) produce Eq. (54) which goes asymptotically to Eq. (55) via Eq. (B7), (B9), (B14), (B20), (B23), and (B26). The development of Eq. (55) is quite involved and is given in detail in Appendix B. Equation (56) represents the proper nonstochastic limit [see Eq. (51)] and can be obtained from both Eqs. (53) and (55). It is apparent from Eq. (55) that the phase factor,[18] represented by $\tilde{L}$, dominates $|\tilde{E}\{p[\mathbf{X}(s, \xi)]\}|$ when $s^2\tilde{Q}_0\tilde{E}/12k_0^2\tilde{L} \ll 1$ and that the amplitude factor,[18] represented by $\tilde{E}$, dominates when $s^2\tilde{Q}_0\tilde{E}/12k_0^2\tilde{L} \gg 1$; this behavior is a direct consequence of the choice of a two-variable expansion[2] in Eq. (9) and will be discussed further in Sec. IV. Note that, due to the statistical isotropy of $\mu(\mathbf{x})$, Eq. (55) does not depend on ξ except through $p_0(\xi)$.

Although $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ of Eq. (55) may seem to be an unusual sound pressure relation, the traditional, i.e., nonstochastic Eulerian,[2,3] behavior should not be expected for this (asymptotic) Lagrangian (subensemble) expectation since Eq. (47) has not yet been applied. Note, for example, that due to the varying s power dependence of the exponential, the diffusion-like character of $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ changes with s. Initially, $\tilde{E}\{p(\mathbf{X}(s, \xi)]\}$ decreases in magnitude from its initial value $p_0(\xi)$ with s until $k_0^2\tilde{L} = \tilde{Q}\tilde{E}s^2/12$; in this region, the phase factor is dominant. Thereafter, $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ increases with s since the amplitude fluctuations dominate. This behavior is directly attributable to the application of a two-variable expansion[2] for $p(\mathbf{x})$. Note also that $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ decreases with k_0 [but see Eq. (86)]. None of this "unorthodox" behavior represents a physical contradiction since $\gamma(\mathbf{x}, \xi \mid s)$, defined in Eq. (45), controls the ultimate value of $\langle p(\mathbf{x}) \rangle$; see, for example, Eq. (57). It will be observed later that these apparent anomalies resolve themselves when $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ (and $\tilde{E}\{p[\mathbf{X}(s_1, \xi_1)]p^*[\mathbf{X}(s_2, \xi_2)]\}$ in Sec. III is related to the Eulerian expectation $\langle p(\mathbf{X}) \rangle$ ($\langle |p|^2 \rangle$ in Sec. III) via Eq. (47).

Note that in order to obtain the Lagrangian subensemble expectation $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ via Eqs. (53)–(55), traditional knowledge of the details of the geometric spreading factor and the divergence from this spreading due to the inhomogeneities of the medium, which are manifested by $\{-\frac{1}{2}\int_0^s [ds'/\mu(s')]\,(\mu\, dX_i/ds)_{,i}(s')\}$ of Eq. (16) for each realization μ_β,[3,5] had to be sacrificed. In place of these more physically intuitive (Eulerian) concepts, $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ has been expressed (asymptotically) in terms of the Lagrangian phase and amplitude integral scales and, therefore, the relative importance of the amplitude (transport) and phase (eikonal) factors can still be discerned. This loss of traditional (Eulerian) physical detail is unavoidable when treating an intrinsically Lagrangian stochastic phenomenon in its natural Lagrangian form. To compensate for this, care has been taken to express $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ explicitly in terms of k_0 ("large"), α ("small"), and s (variable but "large") as well as the Lagrangian integral scales. In general, this permits consideration of these terms in order to simplify the resulting mathematical relations and, thereby, facilitate the analytical treatment of complex physical problems.

When Eq. (24), which represents a uniform source, is applied to Eq. (55), $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$ has no explicit dependence, in a statistically isotropic medium, on its initial ξ position and can be represented by $\tilde{E}\{p\}(s)$. Therefore, the Eulerian ensemble expectation of the sound pressure wave, when both the uniform, collimated source and point receiver are imbedded in an infinite, continuous statistically isotropic medium, is given asymptotically by

$$\langle p(x) \rangle \sim \int_0^\infty ds \int_{S_0} d\eta d\zeta\; \tilde{E}\{p\}(s)\gamma(\mathbf{x}, \xi \mid s) \tag{57}$$

$$= \int_0^\infty ds\, \tilde{E}\{p\}(s)\gamma(x - s + \tfrac{1}{2}\alpha^2 s^2 R_0 A \mid s) \times \int_{-\eta_0}^{\eta_0} d\eta\, \gamma(y - \eta \mid s) \int_{-\zeta_0}^{\zeta_0} d\zeta\, \gamma(z - \zeta \mid s) \tag{58}$$

$$\xrightarrow[\alpha \to 0]{} p_0 e^{ik_0 x}. \tag{59}$$

Equation (47) gives Eq. (57), where $\gamma(\mathbf{x}, \xi \mid s)$ alone determines the ξ dependence of the integrand, and Eq. (44) produces Eq. (58), where, for analytical simplicity, the acoustic source S_0 is assumed to be a rectangular transducer of dimensions $2\eta_0$ by $2\zeta_0$. Note that the required null-scattering result occurs naturally in the nonstochastic limit of Eq. (59) via Eqs. (49a)–(49c). The integrals $\int_{-\eta_0}^{\eta_0} d\eta\, \gamma(y - \eta \mid s)$ and $\int_{-\zeta_0}^{\zeta_0} d\zeta\, \gamma(z - \zeta \mid s)$ in Eq. (58) are tabulated functions *(error functions*[22]) so, for analytic convenience (otherwise, numerical integration techniques are necessary to perform the s interval integration), this treatment will be restricted to terminal locations such that, for a given S_0,

$$|y| \gg \eta_0, \qquad |z| \gg \zeta_0. \tag{60}$$

This means that only energy diverted away from the direct source-to-receiver path will be considered. [In practice, Eqs. (60) and (81) imply a very small surface S_0 with $\mathbf{x}$ located off-axis at large x.] Equation (60) permits

$$\int_{-\eta_0}^{\eta_0} d\eta\, \gamma(y-\eta\,|\,s) \int_{-\zeta_0}^{\zeta_0} d\zeta\, \gamma(z-\zeta\,|\,s) \simeq S_0 \gamma(y\,|\,s)\,\gamma(z\,|\,s), \quad (61)$$

where, from Eqs. (A40), (A41), and (60),

$$\gamma(y\,|\,s)\gamma(z\,|\,s) = e^{-(y^2+z^2)/2\alpha^2 R_0 A_3 s^3}/2\pi\alpha^2 R_0 A_3 s^3 = e^{-\rho^2/2\alpha^2 R_0 A_3 s^3}/2\pi\alpha^2 R_0 A_3 s^3 \equiv \gamma(\rho\,|\,s), \quad (62)$$

where

$$\rho^2 \equiv y^2 + z^2. \quad (63)$$

Therefore,

$$\langle p(\mathbf{x})\rangle \sim p_0 S_0 \int_0^\infty ds\, e^{-\alpha^2(k_0^2\tilde{L}s - s^3\tilde{Q}_0\tilde{E}/12) + ik_0(1+\alpha^2 s\tilde{D})s} \times \frac{e^{-(x-s+\frac{1}{2}\alpha^2 s^2 R_0 A)^2/2\alpha^2 R_0(2A_3+\alpha^2 s R_0 B_4)s^3}}{[2\pi\alpha^2 R_0(2A_3+\alpha^2 s R_0 B_4)s^3]^{1/2}} \times \frac{e^{-\rho^2/2\alpha^2 R_0 A_3 s^3}}{2\pi\alpha^2 R_0 A_3 s^3}, \quad (64)$$

via Eqs. (58), (61), (62), (55), (24), and (A39). Since $0<\alpha\ll 1$, the factor $\gamma(x-s+\frac{1}{2}\alpha^2 s^2 R_0 A\,|\,s)\gamma(\rho\,|\,s)$ dominates the integral; in fact, it can be viewed as a sort of *path guiding* or *subensemble selecting* function. Analytically, this suggests a saddle-point evaluation (see Appendix D) and physically it has the following interpretation: the initial $\boldsymbol{\xi}$ and terminal $\mathbf{x}$ locations determine the distribution of s values and then the factor $\gamma(x-s+\frac{1}{2}\alpha^2 s^2 R_0 A\,|\,s)\gamma(\rho\,|\,s)$ weighs $\tilde{E}\{p\}(s)$ to give $\langle p(\mathbf{x})\rangle$ via Eq. (64).

III. DETERMINATION OF $\langle|p|^2\rangle$

Paralleling the analysis of Sec. II, $\langle|p|^2\rangle$ can be calculated for the same physical situation by incorporating the concept of the Lagrangian subensemble crosspath integral scale which was developed in Sec. V of Ref. 18. Thus,

$$\tilde{E}\{p[\mathbf{X}(s_1,\boldsymbol{\xi}_1)]p^*[\mathbf{X}(s_2,\boldsymbol{\xi}_2)]\} \equiv \tilde{E}\{p(s_1)p^*(s_2)\} = p_0^2\tilde{E}\left\{\exp\left(ik_0\int_0^{s_1} ds_1'\,\mu(s_1') - ik_0\int_0^{s_2} ds_2'\,\mu(s_2')\right) \times \exp\left(-\frac{1}{2}\int_0^{s_1} ds_1' \int_0^{s_1'} ds_1'' \frac{\mu_{,ii}(s_1'')}{\mu(s_1')} - \frac{1}{2}\int_0^{s_2} ds_2' \int_0^{s_2'} ds_2'' \frac{\mu_{,ii}(s_2'')}{\mu(s_2')}\right)\right\} \quad (65)$$

$$\sim p_0^2 \exp\left(-\tfrac{1}{2}k_i k_j \tilde{W}_{ij} + ik_j \int_0^{s_{1,2}} ds_{1,2}'\,\tilde{\nu}_j\right) \quad (66)$$

$$\sim p_0^2\, e^{2\alpha^2[k_0^2 s_0 \tilde{P}_a(\delta) + s_0^3 \tilde{Q}_a(\delta)\tilde{J}(\delta)/12]} \times e^{-2\alpha^2(k_0^2\tilde{L}s_0 - s_0^3\tilde{Q}_0\tilde{E}/12)} \quad (\text{when } s_1 = s_0 = s_2) \quad (67)$$

$$\xrightarrow[\alpha\to 0]{} p_0^2 \quad (\text{when } s_1 = s_2), \quad (68)$$

remembering Eqs. (8) and (26b), where $i,j = 1,2,3,4$ with $k_1 = k_0, k_3 = -k_0$, $k_2 = k_4 = i/2$. (The mathematical interpretation of the operator $\int_0^{s_{1,2}} ds_{1,2}$ will soon be clear.) Equations (65) and (C1)–(C11) produce Eq. (66), which goes asymptotically to Eq. (67) via Eqs. (C12)–(C21), (C26), and (C43). The development of Eq. (67) is quite involved and is given in detail in Appendix C. Equation (68) represents the proper deterministic limit [see Eq. (51)], and can be obtained from both Eqs. (65) and (67). Note that, due to Eq. (24) and the statistical isotropy of $\mu(\mathbf{x})$, Eq. (67) does not depend explicitly on the initial point $\boldsymbol{\xi}$.

The Eulerian sound intensity ensemble expectation $\langle|p(\mathbf{x})|^2\rangle$, when both the uniform, collimated, rectangular transducer and omnidirectional point receiver are imbedded in an infinite, continuous, statistically isotropic medium, can now be determined. Since the Lagrangian subensemble crosspath expectation $\tilde{E}\{p(s_1)p^*(s_2)\}$ involves two paths in almost all realizations, Eqs. (41) and (47), respectively, produce

$$\langle|p(\mathbf{x})|^2\rangle = \int_0^\infty ds_1 \int_0^\infty ds_2 \int_{S_0} d\boldsymbol{\xi}_1 \int_{S_0} d\boldsymbol{\xi}_2\, \tilde{E}\{p(s_1)p^*(s_2)\} \times B(x,y-\eta_1,z-\zeta_1\,|\,s_1) \times B(x,y-\eta_2,z-\zeta_2\,|\,s_2) \quad (69)$$

$$\sim \int_0^\infty ds_1 \int_0^\infty ds_2 \int_{S_0} d\boldsymbol{\xi}_1 \int_{S_0} d\boldsymbol{\xi}_2\, \tilde{E}\{p(s_1)p^*(s_2)\} \times \gamma(\mathbf{x},\boldsymbol{\xi}_1\,|\,s_1)\gamma(\mathbf{x},\boldsymbol{\xi}_2\,|\,s_2) \quad (70)$$

$$\xrightarrow[\alpha\to 0]{} p_0^2 \quad (71)$$

$$\simeq \int_0^\infty ds_1\, \gamma(\rho\,|\,s_1)\gamma(x-s_1\,|\,s_1) \int_0^\infty ds_2\, \gamma(\rho\,|\,s_2) \times \gamma(x-s_2\,|\,s_2) \int_{S_0} d\boldsymbol{\xi}_1 \int_{S_0} d\boldsymbol{\xi}_2\, \tilde{E}\{p(s_1)p^*(s_2)\} \quad (72)$$

$$\sim |\langle p(x,\rho)\rangle_e|^2 \frac{1}{S_0^2} \int_{S_0} d\boldsymbol{\xi}_1 \int_{S_0} d\boldsymbol{\xi}_2 \times e^{2\alpha^2[k_0^2 s_0 \tilde{P}_a(\delta)\tilde{H}(\delta) + s_0^3\tilde{Q}_a(\delta)\tilde{J}(\delta)/12]} \quad (73)$$

$$= |\langle p(s,\rho)\rangle_e|^2 \frac{1}{S_0} \int_{S_0} d\boldsymbol{\delta} \times e^{2\alpha^2[k_0^2 s_0 \tilde{P}_a(\delta)\tilde{H}(\delta) + s_0^3\tilde{Q}_a(\delta)\tilde{J}(\delta)/12]} \quad (74)$$

$$\simeq |\langle p(x,\rho)\rangle_e|^2 \left(1 + 2\alpha^2 \frac{1}{S_0} \times \int_{S_0} d\boldsymbol{\delta}[k_0^2 s_0 \tilde{H}\tilde{P}_a + s_0^3 \tilde{Q}_a \tilde{J}/12]\right) \quad (75)$$

$$\equiv |\langle p(x,\rho)\rangle_e|^2 \left(1 + 2\alpha^2[k_0^2 s_0 \tilde{H}\overline{P_a} + s_0^3 \overline{\tilde{Q}_a \tilde{J}}/12]\right), \quad (76)$$

where

$$\boldsymbol{\delta} = \boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, \quad (77)$$

$$\overline{P_a} \equiv \frac{1}{S_0}\int_{S_0} d\boldsymbol{\delta}\, \tilde{P}_a(\delta), \quad (78a)$$

and

$$\overline{\tilde{Q}_a \tilde{J}} \equiv \frac{1}{S_0}\int_{S_0} d\boldsymbol{\delta}\, \tilde{Q}_a(\delta)\tilde{J}(\delta). \quad (78b)$$

Equation (45) renders Eq. (70) from Eq. (69) and then Eqs. (49a)–(52) yield Eq. (71). Equations (D2a) and (D2c) reduce Eq. (A39) to

$$\gamma(x-s\,|\,s) = e^{-(x-s)^2/4\alpha^2 R_0 A_3 s^3}/(4\pi\alpha^2 R_0 A_3 s^3)^{1/2} \quad (79)$$

so that Eqs. (44), (60), (62), and (79) produce Eq. (72). Equations (D1)–(D4), (67), and (D9′) then yield Eq. (73), via a saddle-point evaluation of each integral in Eq. (72) which is exactly similar to that which produced Eq. (D9); note that the two saddle points are identical and given by Eq. (D4). Introducing the change of variables $\xi_1 - \xi_2 = \delta$, $\xi_1 + \xi_2 = 2\Delta$ to Eq. (73) and carrying out the integration with respect to Δ gives Eq. (74). The condition

$$2\alpha^2[k_0^2 s_0 \tilde{L} + s_0^3 \tilde{Q}_0 \tilde{E}/12] \ll 1, \tag{80}$$

reduces Eq. (74) to Eq. (75) via Eqs. (C37) and (C54). This condition serves to restrict α, k_0, and r; it will be discussed further in Sec. IV. Assuming that the maximum lateral dimension of S_0, say ρ_0, obeys

$$\rho_0 < 3\tilde{L}, \tag{81}$$

$\tilde{H}(\delta) \simeq \tilde{L}$, via Eq. (C36); although Eq. (81) will be assumed throughout the remainder of this analysis, the symbol $\tilde{H}$ $(=\tilde{L})$ will be employed for purposes of identification. Equations (78a) and (78b) represent convenient averages over the total transducer surface. Because of Eqs. (C37′) and (C54), respectively,

$$1 > \overline{P_a} > 0 \tag{82}$$

and

$$\tilde{Q}_0 \tilde{E} > \overline{Q_a J} \geq 0. \tag{83}$$

Note that if $\tilde{L} = \tilde{P}_a(\delta)\tilde{H}(\delta)$, k_0 would completely disappear from Eq. (73) in contrast with experimental evidence [see Sec. IV and Baerg and Schwarz[23]] so that Eqs. (C37) and (82) are well justified. The relative importance of k_0 and s_0 is investigated further in Refs. 5 and 24.

The fact that functionals on two different paths in each realization must be correlated via $\tilde{E}\{p(s_1)p^*(s_2)\}$ in Eq. (69), in order to obtain an intensity type relation like $\langle |p|^2 \rangle$, should be no surprise since, in the usual "ray-bundle" approach[3,10,11] for estimating acoustic intensity, the divergence between two different rays as a function of their curvilinear distance in space is considered. In the present analysis, each pair of paths that leaves the initial surface S_0, at $\theta_0(\xi) = \pi/2$, is correlated while in the other approach, a ray bundle is created by allowing a slight span of initial angles $\Delta\theta_0$ from a single point ξ. It makes no *physical* sense to speak of intensity along a single path; in fact, it is often convenient to remove the dependence on the mean value $|\langle p \rangle_e|^2$ from stochastic intensity relations as is done in Sec. IV.

IV. THE COEFFICIENT OF INTENSITY FLUCTUATION V

When studying sound propagation through extended stochastic regions, the coefficient of intensity fluctuation V, where

$$V^2 \equiv (\langle |p|^2 \rangle - |\langle p \rangle|^2)/|\langle p \rangle|^2, \tag{84}$$

has often proved to be a useful measure of the random fluctuation of the signal strength.[2,3,5,6,16–18,24] Substituting Eq. (76) into Eq. (84) yields

$$V^2 \sim 2\alpha^2[k_0^2 \tilde{H}\overline{P_a} s_0 + s_0^3 \overline{Q_a J}/12] \tag{85}$$

$$\simeq 2\alpha^2[k_0^2 \tilde{H}\overline{P_a} x[2 - (1 - 6\tan^2\Theta)^{1/2}) + \overline{Q_a J} x^3[2 - (1 - 6\tan^2\Theta)^{1/2}]^{3/12}] \tag{86}$$

$$\simeq \begin{cases} 2\alpha^2 k_0^2 \tilde{H}\overline{P_a} x[2 - (1 - 6\tan^2\Theta)^{1/2}], & \sigma^2 \ll 1, \quad (86a) \\ 2\alpha^2(\overline{Q_a J}/12)x^3[2 - (1 - 6\tan^2\Theta)^{1/2}]^3, & \sigma^2 \gg 1, \quad (86b) \end{cases}$$

where

$$\sigma \equiv x[2 - (1 - 6\tan^2\Theta)^{1/2}]/k_0[\tilde{H}\overline{P_a}/(\overline{Q_a J}/12)]^{1/2}. \tag{87}$$

Equation (D5) produces Eq. (86) which has a phase-dominance region, Eq. (86a), and an amplitude-dominance region, Eq. (86b), as determined by the parameter σ of Eq. (87); remember the Θ restriction of Eq. (D8).

Via Eq. (D5d), Eqs. (86)–(87) reduce, respectively, to

$$V^2 \sim 2\alpha^2[k_0^2 \overline{P_a}\tilde{H} r + (\overline{Q_a J}/12)r^3] \tag{88}$$

$$\simeq \begin{cases} 2\alpha^2 k_0^2 \overline{P_a}\tilde{H} r, & \sigma_0^2 \ll 1, \quad (88a) \\ 2\alpha^2(\overline{Q_a J}/12)r^3, & \sigma_0^2 \gg 1, \quad (88b) \end{cases}$$

where

$$\sigma_0 \equiv r/k_0[\overline{P_a}\tilde{H}/(\overline{Q_a J}/12)]^{1/2} \tag{87′}$$

and

$$0 < \Theta < 4°. \tag{89}$$

Also, Eq. (80) reduces to

$$2\alpha^2 k_0^2 \tilde{L} r \ll 1, \quad \sigma_0^2 \ll 1, \tag{90}$$

via Eq. (D5d); Eq. (23) is the one-dimensional analog of Eq. (90).

Equations (88a) and (88b) are very similar to, respectively, the results

$$V^2 \simeq 2\alpha^2 k_0^2 r \int_0^\infty d\rho_1 \, N(\rho_1) \tag{91}$$

and

$$V^2 \simeq 2\alpha^2 r^3 \tfrac{1}{120} \int_0^\infty d\rho_1 [\nabla^2\nabla^2 N(\rho)]_{\rho=\rho_1}, \tag{92}$$

where

$$N(\rho) = \langle n(\mathbf{x}) n(\mathbf{x} + \boldsymbol{\rho}) \rangle; \tag{93}$$

$$\rho = (\rho_1^2 + \rho_2^2 + \rho_3^2)^{1/2} \tag{94}$$

is the separation of the points, $\mathbf{x}$, $\mathbf{x} + \boldsymbol{\rho}$, and ρ_1 is taken along the source-to-receiver line. Equation (91) was found by Mintzer,[6,7] and Eq. (92) was derived by Bergmann[16] from Eqs. (10) and (11) via a variational approach (see also Sec. VI of Ref. 18).

V. CONCLUSION

The stochastic Eulerian-Lagrangian methods of Refs. 15 and 18 applied to steady-state sound propagation from a small, collimated acoustic source to an omnidirectional point receiver imbedded in an infinite, continuous, statistically isotropic, stochastic-Fermat medium have predicted that V becomes frequency independent for the *low*-frequency, amplitude-dominance region, Eq. (88b), in contrast to the frequency independence at *high*-frequency predicted by Potter

and Murphy[17] (however, consider Ref. 24). Finally, it should be emphasized that the methods presented in this study are new to the problem of sound propagation through continuous stochastic media (however, see Lumley[19]) and avoid three principal difficulties that have plagued earlier[4,6-8,16,17] studies in this area, namely,

(1) range limitations due to cumulative phase effects,
(2) discrete scattering assumptions, and
(3) restriction to an Eulerian path.

APPENDIX A: THE CALCULATION OF $\gamma(\mathbf{x}, \xi \mid s)$

The detailed calculation of the asymptotic Lagrangian measure function $\gamma(\mathbf{x}, \xi \mid s)$, Eq. (45), for stochastic-Fermat media follows. Consider

$$\frac{dX_i}{ds}(s, \xi) = \frac{\mu(0)}{\mu(s)} \frac{dX_i}{ds}(0) + \int_0^s ds' \frac{\mu_{,i}(s')}{\mu(s)} \tag{A1}$$

from Eq. (13),

$$\begin{aligned}\mu(s) - \mu(0) &= \int_0^s ds' \, \mu_{,i}(s') \frac{dX_i}{ds'}(s, \xi) \\ &= \mu(0) \frac{dX_i}{ds}(0) \int_0^s ds' \frac{\mu_{,i}(s')}{\mu(s')} \\ &\quad + \int_0^s ds' \frac{\mu_{,i}(s')}{\mu(s')} \int_0^{s'} ds'' \, \mu_{,i}(s''), \end{aligned} \tag{A2}$$

and thence[18,5]

$$E\left\{\frac{\mu(0)}{\mu(s)}\right\} = 1 - \alpha^2 \int_0^s ds' \int_0^{s'} ds'' \times E\{n_{,i}(s')n_{,i}(s'') - O(\alpha)\rangle \tag{A3}$$

$$= 1 - \alpha^2 s R_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right)\frac{R(\sigma)}{R_0} + O(\alpha^3 s) \tag{A4}$$

$$\sim 1 - \alpha^2 s R_0 A, \tag{A5}$$

where

$$\sigma = s' - s''. \tag{A6}$$

Defining

$$R(\sigma) = R(s', s'') \equiv E\{n_{,i}(s')n_{,i}(s'')\}, \tag{A7}$$

via Eq. (14) of Ref. 18, gives Eq. (A4) which goes asymptotically to Eq. (A5), where the Lagrangian integral scale corresponding to the Lagrangian autopath correlation $R(\sigma)$ is defined by

$$A \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right)\frac{R(\sigma)}{R_0}, \tag{A8}$$

with the intensity factor

$$R_0 \equiv E\{n_{,i}(\xi)n_{,i}(\xi)\} \tag{A9'}$$

$$= \langle n_{,i}(\xi)n_{,i}(\xi)\rangle; \tag{A9}$$

A is assumed to exist and cannot be a function of s. In Eq. (42),

$$\int_0^s m_1 ds' = \frac{dX}{ds}(0) \int_0^s ds' \, E\left\{\frac{\mu(0)}{\mu(s')}\right\} \tag{A10}$$

$$= s - \tfrac{1}{2}\alpha^2 s^2 R_0 \int_0^s d\sigma \left(1 - \frac{2\sigma}{s} + \frac{\sigma^2}{s^2}\right)\frac{R(\sigma)}{R_0} + O(\alpha^3 s^2) \tag{A11}$$

$$\sim s - \tfrac{1}{2}\alpha^2 s^2 R_0 A. \tag{A12}$$

In Eq. (43),

$$m_{01} s \equiv \int_0^s m_1 ds' \tag{A13a}$$

$$\sim s - \alpha^2 s^2 R_0 A/2, \tag{A12}$$

$$m_{02} \equiv 0, \tag{A13b}$$

$$m_{03} \equiv 0; \tag{A13c}$$

see Eqs. (36b) and (36c).

Since $[U_{ij}]$ is in terms of its principle axes, the determinant $\|U_{ij}\|$ can be reduced as follows:

$$\|U_{ij}\| \equiv \begin{vmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{vmatrix} \tag{A14}$$

$$= U_{11} U_{22} U_{33} \tag{14'}$$

$$\sim \alpha^6 s^9 R_0^3 (2A_3 + \alpha^2 s R_0 B_4) A_3^2, \tag{A15}$$

where Eq. (A15) follows asymptotically via Eqs. (A23) and (A24); A, R_0 and B_4 are given by Eqs. (A8), (A9), and (A37), respectively. For notational simplicity, let

$$A_3 = \tfrac{2}{9} A. \tag{A16}$$

Equations (A23)-(A25) and (A14') give the components of U_{ij}^{-1} to be

$$U_{11}^{-1} = \frac{U_{22}U_{33}}{\|U_{ij}\|} = \frac{1}{U_{11}} \tag{A17a}$$

$$\sim \frac{1}{\alpha^2 s^3 R_0 (2A_3 + \alpha^2 s R_0 B_4)}, \tag{A18a}$$

$$U_{22}^{-1} = \frac{U_{11}U_{33}}{\|U_{ij}\|} = \frac{1}{U_{22}} \tag{A17b}$$

$$\sim \frac{1}{\alpha^2 s^3 R_0 A_3}, \tag{A18b}$$

$$U_{33}^{-1} = \frac{U_{11}U_{22}}{\|U_{ij}\|} = \frac{1}{U_{33}} \tag{A17c}$$

$$\sim \frac{1}{\alpha^2 s^3 R_0 A_3}, \tag{A18c}$$

$$U_{ij}^{-1} = 0 \quad \text{for } i \neq j. \tag{A19}$$

The Lagrangian stochastic integrals U_{ij} must be evaluated with considerable care[18,5]:

$$U_{ij} \equiv \int_0^s ds_1 \int_0^s ds_2 \, E\{u_i(s_1)u_j(s_2)\} \tag{46}$$

$$= \int_0^s ds_1 \int_0^s ds_2 \, E\left\{\left[\int_0^{s_1} ds_1' \frac{\mu_{,i}(s_1')}{\mu(s_1)}\right.\right.$$

$$+\left(\frac{\mu(0)}{\mu(s_1)}-E\left\{\frac{\mu(0)}{\mu(s_1)}\right\}\right)\frac{dX_i}{ds}(0)\Bigg]$$
$$\times\left[\int_0^{s_2}ds_2'\,\frac{\mu_{,j}(s_2')}{\mu(s_2)}+\left(\frac{\mu(0)}{\mu(s_2)}\right.\right.$$
$$\left.\left.-E\left\{\frac{\mu(0)}{\mu(s_2)}\right\}\right)\frac{dX_j}{ds}(0)\right]\Bigg\} \tag{A20}$$

$$=\int_0^s ds_1\int_0^s ds_2\left[E\left\{\int_0^{s_1}ds_1'\,\frac{\mu_{,i}(s_1')}{\mu(s_1)}\int_0^{s_2}ds_2'\right.\right.$$
$$\times\left.\frac{\mu_{,j}(s_2')}{\mu(s_2)}\right\}+E\left\{\left(\frac{\mu(0)}{\mu(s_1)}-E\left\{\frac{\mu(0)}{\mu(s_1)}\right\}\right)\left(\frac{\mu(0)}{\mu(s_2)}\right.\right.$$
$$\left.\left.-E\left\{\frac{\mu(0)}{\mu(s_2)}\right\}\right)\right\}\frac{dX_i}{ds}(0)\,\frac{dX_j}{ds}(0)\Bigg] \tag{A21}$$

$$\sim\alpha^2 s^3R_0\left[\delta_{ij}A_3+(A_3+\alpha^2sR_0B_4)\frac{dX_i}{ds}(0)\right.$$
$$\left.\times\frac{dX_j}{ds}(0)\right], \tag{A22}$$

so that

$$U_{11}\sim\alpha^2s^3R_0(2A_3+\alpha^2sR_0B_4), \tag{A23}$$

$$U_{22}=U_{33}\sim\alpha^2s^3R_0A_3, \tag{A24}$$

$$U_{ij}=0\quad\text{for } i\neq j. \tag{A25}$$

Equations (31a)–(31c) and Eq. (51) of Ref. 18 yield Eqs. (A20) and (A21), respectively. The asymptotic result of Eq. (A22) follows, after some complicated analysis of the type illustrated in Sec. IV of Ref. 18, from Eqs. (A27) and (A32). Equations (A22) and (26b) produce Eqs. (A23)–(A25). Equation (A26) in Eq. (A21) gives $U_{22}=U_{33}$ and Eq. (A25); this shows that the matrix $[U_{ij}]$ is expressed in terms of its principal axes, as mentioned earlier, and is a direct consequence of statistical isotropy, via Eqs. (51) and (55) of Ref. 18, and the convenient choice of the initial path orientation, represented by Eq. (26b). The presence of the null cross-correlations, i.e., $U_{ij}=0$ for $i\neq j$, permits the expeditious reduction of Eq. (43) to Eq. (44). This decoupling of the joint spacial probability density greatly facilitates the analysis of Sections II and III.

The steps from Eq. (A21) to Eq. (A22) are now given in detail employing the procedures developed in Refs. 18 and 5. The first term in Eq. (A21) is treated as follows:

$$\int_0^s ds_1\int_0^s ds_2\,E\left\{\int_0^{s_1}ds_1'\,\frac{\mu_{,i}(s_1')}{\mu(s_1)}\int_0^{s_2}ds_2'\,\frac{\mu_{,1}(s_2')}{\mu(s_2)}\right\}$$
$$=\alpha^2\delta_{ij}s^3R_0$$
$$\times\left[\int_0^s d\sigma\left(\frac{2}{3}-\frac{\sigma}{s}+\frac{\sigma^3}{2s^3}\right)\frac{1}{3}\frac{R(\sigma)}{R_0}-O(\alpha s^3)\right] \tag{A26}$$

$$\sim\alpha^2\delta_{ij}s^3R_0A_3. \tag{A27}$$

Equation (55) of Ref. 18 produces Eq. (A26), where

$$\sigma=s_1'-s_2',\qquad S=s_1'+s_2', \tag{A28}$$

$$s_1'=\tfrac{1}{2}(S+\sigma),\qquad s_2'=\tfrac{1}{2}(S-\sigma), \tag{A29}$$

which goes asymptotically to Eq. (A27) with

$$A_3\equiv\lim_{s\to\infty}\tfrac{2}{9}\int_0^s d\sigma\left(1-\frac{3\sigma}{2s}+\frac{3\sigma^3}{4s^3}\right)\frac{R(\sigma)}{R_0} \tag{A30}$$
$$=\tfrac{2}{9}A. \tag{A16}$$

The second term in Eqs. (A21) is developed as follows:

$$\int_0^s ds_1\int_0^s ds_2\,E\left\{\left(\frac{\mu(0)}{\mu(s_1)}-E\left\{\frac{\mu(0)}{\mu(s_1)}\right\}\right)\right.$$
$$\left.\times\left(\frac{\mu(0)}{\mu(s_2)}-E\left\{\frac{\mu(0)}{\mu(s_2)}\right\}\right)\right\}$$
$$=\alpha^2\int_0^s ds_1\int_0^s ds_2\left(\int_0^{s_1}ds_1'\int_0^{s_2}ds_2'\right.$$
$$\times\left[\tfrac{1}{3}R(s_1'-s_2')+O(\alpha)\right]\delta_{ij}\frac{dX_i}{ds}(0)\,\frac{dX_j}{ds}(0)$$
$$+\alpha^2\int_0^{s_1}ds_1'\int_0^{s_1'}ds_1''\int_0^{s_2}ds_2'\int_0^{s_2'}ds_2''$$
$$\times\left[R_2(s_1',s_1'',s_2',s_2'')-R(s_1'-s_1'')R(s_2'-s_2'')\right.$$
$$\left.\left.+O(\alpha)\right]\right) \tag{A31}$$

$$\sim\alpha^2s^3R_0(A_3+\alpha^2sR_0B_4). \tag{A32}$$

Invoking Eqs. (A2), (8), (A7), and (A33), plus Eqs. (51) and (55) of Ref. 18, yields Eq. (A31). Applying Eq. (A27) to the first term in Eq. (A31) yields, asymptotically, the first term in Eq. (A32). In Section 3.15 of Lumley,[19] it is shown that if $R(\sigma)$ behaves suitably as $\sigma\to\infty$ [e.g., an ergodic process provides a suitable $R(\sigma)$],

$$R_2(s_1',s_1'',s_2',s_2'')\equiv E\{n_{,i}(s_1')n_{,i}(s_1'')n_{,j}(s_2')n_{,j}(s_2'')\} \tag{A33}$$

$$\sim R(s_1'-s_1'')R(s_2'-s_2'')+\tfrac{1}{3}R(s_1'-s_2')R(s_1''-s_2'')$$
$$+\tfrac{1}{3}R(s_1'-s_2'')R(s_2'-s_1''), \tag{A34}$$

via Eq. (55) of Ref. 18. Thence,

$$\int_0^s ds_1\int_0^s ds_2\int_0^{s_1}ds_1'\int_0^{s_1'}ds_1''\int_0^{s_2}ds_2'\int_0^{s_2'}ds_2''$$
$$\times[R_2(s_1',s_1'',s_2',s_2'')-R(s_1'-s_1'')R(s_2'-s_2'')]$$
$$\sim\tfrac{1}{3}\int_0^s ds_1\int_0^s ds_2\int_0^{s_1}ds_1'\int_0^{s_2}ds_2'R(s_1'-s_2')$$
$$\times\int_0^{s_1'}ds_1''\int_0^{s_2'}ds_2''R(s_1''-s_2'')$$
$$+\tfrac{1}{3}\int_0^s ds_1\int_0^s ds_2\int_0^{s_2}ds_2'\int_0^{s_1'}ds_1''R(s_2'-s_1'')$$
$$\times\int_0^{s_1}ds_1'\int_0^{s_2'}ds_2''R(s_1'-s_2'') \tag{A35}$$

$$\sim s^4R_0^2B_4, \tag{A36}$$

which gives the second term in Eq. (A32), where the appropriate integral scale is given by

$$B_4\equiv\lim_{s\to\infty}\frac{1}{3}\left[\int_0^s\frac{ds_1}{s}\int_0^s\frac{ds_2}{s}\frac{1}{s}\int_0^{s_1}ds_1'\int_0^{s_2}ds_2'\right.$$
$$\times\frac{R(s_1'-s_2')}{R_0}\frac{1}{s}\int_0^{s_1'}ds_1''\int_0^{s_2'}ds_2''\frac{R(s_1''-s_2'')}{R_0}$$
$$+\int_0^s\frac{ds_1}{s}\int_0^s\frac{ds_2}{s}\int_0^{s_1}\frac{ds_1'}{s}\int_0^{s_2}\frac{ds_2'}{s}\int_0^{s_2'}ds_1''$$

$$\times \frac{R(s_2' - s_1'')}{R_0} \int_0^{s_1'} ds_2'' \frac{R(s_1' - s_2'')}{R_0}\Bigg] \tag{A37}$$

$$= \tfrac{1}{6}\left(\int_0^\infty d\sigma \frac{R(\sigma)}{R_0}\right)^2 \equiv \tfrac{1}{6}A^2, \tag{A38}$$

via Eq. (A8); B_4 is assumed to exist and cannot be a function of s. Note that the second term in Eq. (A31) must reach its asymptotic form much more slowly than the first term because of the additional $\alpha^2 s$ factor.

Equations (A27) and (A32) give Eq. (A22) which renders Eqs. (A14′)–(A19). Therefore, Eq. (43) goes asymptotically to Eq. (44), where

$$\gamma(x - s + \tfrac{1}{2}\alpha^2 s^2 R_0 A \,|\, s) \equiv e^{-(x-s+\frac{1}{2}\alpha^2 s^2 R_0 A)^2/2\alpha^2 R_0 (2A_3 + \alpha^2 s R_0 B_4) s^3} / [2\pi\alpha^2 R_0 (2A_3 + \alpha^2 s R_0 B_4) s^3]^{1/2}, \tag{A39}$$

$$\gamma(y - \eta \,|\, s) \equiv e^{-(y-\eta)^2/2\alpha^2 R_0 A_3 s^3}/(2\pi\alpha^2 R_0 A_3 s^3)^{1/2}, \tag{A40}$$

and

$$\gamma(z - \zeta \,|\, s) \equiv e^{-(z-\zeta)^2/2\alpha^2 R_0 A_3 s^3}/(2\pi\alpha^2 R_0 A_3 s^3)^{1/2}, \tag{A41}$$

via Eqs. (A13a)–(A19).

Equation (45) displays several properties that are characteristic of diffusionlike phenomena. Because the equation which governs the paths generated in each realization μ_β, i.e., Eq. (13), is known in principle, this study can proceed further than many other diffusionlike analyses. Equations (A39)–(A41) and (49a)–(49c), respectively, indicate that when $\alpha \neq 0$, there exists (asymptotically) a Lagrangian spread of possible X, Y, Z values about their respective null-scattering (or $\alpha = 0$) values of s, η, ζ [see Eq. (51)] and that these spreads reduce to zero as $\alpha \to 0$. This is in agreement with the discussion in Sec. IV of Ref. 18; see especially Eq. (69) of Ref. 18 which should be compared with Eqs. (A39)–(A41). By specifying the terminal location (observation point) $\mathbf{x}$ for $\alpha \neq 0$ and integrating over all possible values of s, η, ζ, the (asymptotic) Lagrangian measure of all those $\mathbf{X}(s, \xi)$ paths which reach $\mathbf{x}$ from the initial surface S_0 can be determined for steady-state, collimated sound transmission, i.e.,

$$M\{\mu_\beta \,|\, \mathbf{X}(s, \xi) \approx \mathbf{x}, \eta \in [-\eta_0, \eta_0], \zeta \in [-\zeta_0, \zeta_0], s \in [0, \infty)\} = \int_0^\infty ds \int_{-\eta_0}^{\eta_0} d\eta \int_{-\zeta_0}^{\zeta_0} d\zeta \, B(x_i - \zeta_i \,|\, s) \tag{A42}$$

$$\sim \int_0^\infty ds \int_{-\eta_0}^{\eta_0} d\eta \int_{-\zeta_0}^{\zeta_0} d\zeta \, \gamma(\mathbf{x}, \xi \,|\, s), \tag{A43}$$

via Eq. (45) and Eq. (6) of Ref. 15.

For a given initial point ξ and terminal point $\mathbf{x}$, one would expect an optimum value of s, say s_m, to exist in the sense that a maximum number of realizations from $\{\mu_\beta\}$ give $\mathbf{X}(s, \xi) \approx \mathbf{x}$. This property is characterized by a maximum in the (asymptotic) Lagrangian measure function $\gamma(\mathbf{x}, \xi \,|\, s)$. Therefore, the behavior of $\gamma(\mathbf{x}, \xi \,|\, s)$ with respect to s is now examined for $\mathbf{x}, \xi$ fixed. First, note that

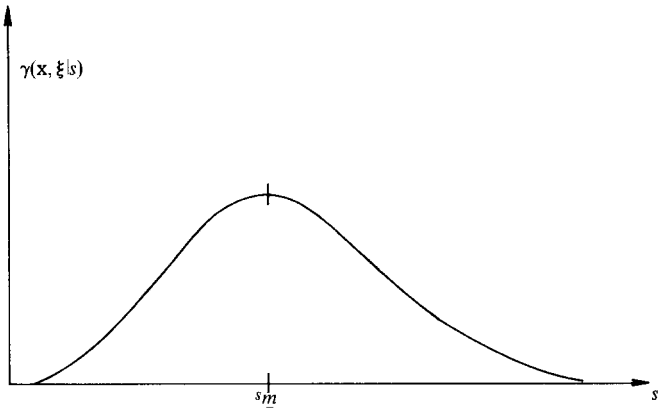


FIG. 1. Lagrangian measure function $\gamma(\mathbf{x}, \xi \,|\, s)$ vs s.

$$\gamma(\mathbf{x}, \xi \,|\, s) \xrightarrow[s \to 0]{} 0, \tag{A44a}$$

as it must since no realization μ_β can produce $\mathbf{X}(s, \xi) \approx \mathbf{x}$ paths that reach $[x^2 + (y - \eta)^2 + (z - \zeta)^2]^{1/2} > 0$ for $s = 0$. The value of $\gamma(\mathbf{x}, \xi \,|\, s)$ then increases with s which means that more and more realization μ_β can produce $\mathbf{X}(s, \xi) \approx \mathbf{x}$ paths. Eventually, an optimum value of $s = s_m$ should be obtained where

$$\left[\frac{\partial}{\partial s}\gamma(\mathbf{x}, \xi \,|\, s)\right]_{s_m} = 0. \tag{A44b}$$

Thereafter, $\gamma(\mathbf{x}, \xi \,|\, s)$ should decrease with s until

$$\gamma(\mathbf{x}, \xi \,|\, s) \xrightarrow[s \to \infty]{} 0 \tag{A44c}$$

while the spread of possible terminal locations increases without limit; see Eq. (69) of Ref. 18. This behavior is illustrated in Fig. 1.

APPENDIX B: THE CALCULATION OF $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$

The detailed asymptotic evaluation of the Lagrangian subensemble autopath relation for $\tilde{E}\{p[\mathbf{X}(s, \xi)]\}$, Eq. (55), follows. In Eq. (54), the Lagrangian subensemble autopath stochastic integrals[18,5] are given by

$$\tilde{V}_{ij} \equiv \int_0^s ds' \int_0^s ds'' \, \tilde{E}\{v_i(s') v_j(s'')\}, \quad i, j = 1, 2, \tag{B1}$$

where

$$v_1 = v_1[\mathbf{X}(s', \xi)] \equiv \mu(s') - \tilde{\nu}_1 \tag{B2′}$$

$$= \alpha[n(s') - \tilde{E}\{n(s')\}], \tag{B2}$$

$$v_2 = v_2[\mathbf{X}(s', \xi)] \equiv \int_0^{s'} ds'' \frac{\mu_{,ii}(s'')}{\mu(s')} - \tilde{\nu}_2 \tag{B3′}$$

$$= \alpha \int_0^{s'} ds'' \left[\frac{n_{,ii}(s'')}{\mu(s')} - \tilde{E}\left\{\frac{n_{,ii}(s'')}{\mu(s')}\right\}\right], \tag{B3}$$

$$\mu(s') = \mu[\mathbf{X}(s', \xi)] = 1 + \alpha n(s'), \quad 0 \leq \alpha \ll 1, \tag{B4}$$

$$\tilde{\nu}_1 \equiv \tilde{E}\{\mu[\mathbf{X}(s', \xi)]\} = \tilde{E}\{\mu(s')\} \tag{B5}$$

$$= 1 + \alpha^2 \int_0^{s'} d\sigma (s' - \sigma)[\tilde{R}(\sigma) - O(\alpha)] \tag{B5′}$$

[cf. Eq. (36a) and Eq. (93) of Ref. 18], and

$$\tilde{\nu}_2 \equiv \tilde{E}\left\{\int_0^{s'} ds'' \frac{\mu_{,ii}(s'')}{\mu(s')}\right\}. \tag{B6}$$

Note that v_1, v_2 are stochastic variables measured relative to their expectations so that the $\tilde{E}\{v_i(s')v_j(s'')\}$ are correlations of centered random variables; $\tilde{R}(\sigma)$ is the subensemble equivalent of $R(\sigma)$ of Eq. (37).

In Eq. (54), via Eq. (B5'),

$$\int_0^s ds'\, \tilde{\nu}_1(s') = s + \alpha^2 s^2 \tilde{R}_0 \tfrac{1}{2} \int_0^s d\sigma \left(1 - \frac{2\sigma}{s} + \frac{\sigma^2}{s^2}\right) \times \frac{\tilde{R}(\sigma)}{\tilde{R}_0} + O(\alpha^3 s^2) \quad \text{(B7')}$$

$$\sim s + \tfrac{1}{2}\alpha^2 s^2 \tilde{R}_0 \tilde{A}, \quad \text{(B7)}$$

where

$$\tilde{A} \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{2\sigma}{s} + \frac{\sigma^2}{s^2}\right) \frac{\tilde{R}(\sigma)}{\tilde{R}_0} \quad \text{(B8)}$$

$$= \lim_{s\to\infty} \int_0^s d\sigma\, \frac{\tilde{R}(\sigma)}{\tilde{R}_0}, \quad \text{(B8')}$$

if the latter exists; $\tilde{A}$ is the subensemble equivalent of Eq. (A8). Furthermore,

$$\int_0^s ds'\, \tilde{\nu}_2(s') = \alpha \int_0^s ds' \int_0^{s'} ds''\, \tilde{E}\left\{\frac{n_{,ii}(s'')}{\mu(s')}\right\} \sim \alpha^2 \tilde{B}_1 \tilde{b}_1 s \quad \text{(B9)}$$

[cf. Eq. (60) of Ref. 18], where

$$\alpha^2 \tilde{b}_1 \equiv \lim_{s\to\infty} \int_0^s d(s'-s'') \left(1 - \frac{(s'-s'')}{s}\right) \times \tilde{E}\left\{\frac{\mu_{,ii}(s'')}{\mu(s')}\right\} / \tilde{B}_1 \quad \text{(B10)}$$

is assumed to exist and must be independent of s, and

$$\tilde{B}_1 \equiv \tilde{E}\{\mu_{,ii}(\xi)/\mu(\xi)\} \quad \text{(B11')}$$

$$= \langle \mu_{,ii}(\xi)/\mu(\xi)\rangle. \quad \text{(B11)}$$

Since $s\tilde{B}_1\tilde{b}_1 \ll s^3\tilde{Q}_0\tilde{E}$, in the sense that $\tilde{B}_1\tilde{b}_1$ and $\tilde{Q}_0\tilde{E}$ are bounded while s is not, $\exp(-\alpha^2\tilde{B}_1\tilde{b}_1 s)$ makes a negligible contribution to Eq. (55).

The Lagrangian subensemble autopath stochastic integrals $\tilde{V}_{ij}$ are evaluated asymptotically as follows by the methods of Refs. 18 and 5. Define the *phase correlation*

$$\tilde{P}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) \equiv \tilde{E}\{[n(s_1') - \tilde{E}\{n(s_1')\}] \times [n(s_2') - \tilde{E}\{n(s_2')\}]\} \quad \text{(B12)}$$

$$= \tilde{P}(\sigma, \boldsymbol{\xi}, \mathbf{x}), \quad \text{(B12')}$$

where Eq. (90) of Ref. 18 gives Eq. (B12'). Therefore,

$$\tilde{V}_{11} \equiv \alpha^2 \int_0^s ds_1' \int_0^s ds_2'\, \tilde{P}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) \quad \text{(B13)}$$

$$= 2\alpha^2 s \tilde{P}_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{P}(\sigma, \boldsymbol{\xi}, \mathbf{x})}{\tilde{P}_0} \quad \text{(B13')}$$

$$\sim 2\alpha^2 s \tilde{L}, \quad \text{(B14)}$$

where the Lagrangian subensemble *phase integral scale* $\tilde{L}$ is defined by

$$\tilde{L} \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{P}(\sigma)}{\tilde{P}_0} \quad \text{(B15)}$$

$$\geq L_e \quad \text{(B15')}$$

and where the scaling factor is

$$\tilde{P}_0 \equiv \tilde{E}\{n^2(\xi)\} \quad \text{(B16')}$$

$$= \langle n^2(\xi)\rangle = 1, \quad \text{(B16)}$$

cf. Eqs. (A8) and (A9). The phase integral scale $\tilde{L}$, which is assumed to exist and must be independent of s, is a Lagrangian subensemble measure of the curvilinear range of strong phase correlation of the pressure wavefronts. Because of the path curvature that occurs for almost every μ, the Lagrangian subensemble integral scale $\tilde{L}$ and its corresponding full ensemble integral scale L are both longer than the corresponding Eulerian integral scale L_e of Eq. (20) and $L = L_e$ can occur only when $\alpha = 0$; cf. Eq. (21) of Ref. 18. Likewise, since the path curvature for $\tilde{L}$ is less than that for L (see Figs. 2 and 3 of Ref. 18)

$$L > \tilde{L}. \quad \text{(B17)}$$

Define the *phase-amplitude interaction correlation* by

$$\tilde{P}_{12}(s_1', s_2'', \boldsymbol{\xi}, \mathbf{x}) \equiv \tilde{E}\{[n(s_1') - \tilde{E}\{n(s_1')\}] \times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}]\} \quad \text{(B18)}$$

$$= \tilde{P}_{12}(\sigma, \boldsymbol{\xi}, \mathbf{x}), \qquad \sigma = s_1' - s_2'', \quad \text{(B18')}$$

where Eq. (90) of Ref. 18 gives Eq. (B18'). Therefore,[18,5]

$$\tilde{V}_{12} = \tilde{V}_{21} = \alpha^2 \int_0^s ds_1' \int_0^s ds_2' \int_0^{s_2'} ds_2'' \times \tilde{E}\{[n(s_1') - \tilde{E}\{n(s_1')\}] \times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}] - O(\alpha)\} \quad \text{(B19)}$$

$$= \alpha^2 s^2 \tilde{C}_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{P}_{12}(\sigma, \boldsymbol{\xi}, \mathbf{x})}{\tilde{C}_0} - O(\alpha^3 s^2) \quad \text{(B20')}$$

$$\sim \alpha^2 s^2 \tilde{C}_0 \tilde{C}_2, \quad \text{(B20)}$$

where the *phase-amplitude interaction integral scale* $\tilde{C}_2$ is defined by

$$\tilde{C}_2 \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{P}_{12}(\sigma)}{\tilde{C}_0} \quad \text{(B21)}$$

and where the scaling factor is

$$\tilde{C}_0 \equiv \tilde{E}\{n(\xi) n_{,ii}(\xi)\} \quad \text{(B22')}$$

$$= \langle n(\xi) n_{,ii}(\xi)\rangle. \quad \text{(B22)}$$

The integral scale $\tilde{C}_2$ is assumed to exist and cannot be dependent on s. For convenience, define

$$\tilde{D} \equiv \tfrac{1}{2}(\tilde{R}_0\tilde{A} - \tilde{C}_0\tilde{C}_2), \quad \text{(B23)}$$

which, in Eq. (55), represents a weak frequency shift factor.

Define the *amplitude correlation* by

$$\tilde{Q}(s_1'', s_2'', \boldsymbol{\xi}, \mathbf{x}) \equiv \tilde{E}\{[n_{,ii}(s_1'') - \tilde{E}\{n_{,ii}(s_1'')\}]$$

$$\times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}]\} \quad \text{(B24)}$$

$$= \tilde{Q}(\sigma, \xi, \mathbf{x}), \qquad \sigma = s_1'' - s_2'', \quad \text{(B24')}$$

where Eq. (90) of Ref. 18 gives Eq. (B24'). Therefore,[18,5]

$$\tilde{V}_{22} = \alpha^2 \int_0^s ds_1' \int_0^{s_1'} ds_1'' \int_0^s ds_2' \int_0^{s_2'} ds_2''$$
$$\times \tilde{E}\{[n_{,ii}(s_1'') - \tilde{E}\{n_{,ii}(s_1'')\}]$$
$$\times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}] - O(\alpha)\} \quad \text{(B25)}$$

$$= \alpha^2 s^3 \tilde{Q}_0 \tfrac{2}{3} \int_0^s d\sigma \left(1 - \frac{\sigma}{2s} - \frac{\sigma^2}{2s^2} + \frac{3\sigma^3}{4s^3}\right)$$
$$\times \frac{\tilde{Q}(\sigma, \xi, \mathbf{x})}{\tilde{Q}_0} - O(\alpha^3 s^3) \quad \text{(B26')}$$

$$\sim \tfrac{2}{3} \alpha^2 s^3 \tilde{Q}_0 \tilde{E}. \quad \text{(B26)}$$

Equation (B26') goes asymptotically to Eq. (B26), where the Lagrangian subensemble *amplitude integral scale* $\tilde{E}$ is defined by

$$\tilde{E} \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{2s} - \frac{\sigma^2}{2s^2} + \frac{3\sigma^3}{4s^3}\right) \frac{\tilde{Q}(\sigma, \xi, \mathbf{x})}{\tilde{Q}_0} \quad \text{(B27)}$$

and where the scaling factor is

$$\tilde{Q}_0 \equiv \tilde{E}\{n_{,ii}(\xi) n_{,ii}(\xi)\} \quad \text{(B28')}$$

$$= \langle n_{,ii}(\xi) n_{,ii}(\xi) \rangle. \quad \text{(B28)}$$

The amplitude integral scale $\tilde{E}$, which is assumed to exist and must be independent of s, is a Lagrangian subensemble measure of the curvilinear range of strong amplitude correlation of the pressure wavefronts.

APPENDIX C: THE CALCULATION OF $\tilde{E}\{p(s_1)p^*(s_2)\}$

The detailed asymptotic evaluation of the Lagrangian subensemble relation for $\tilde{E}\{p(s_1)p^*(s_2)\}$ [Eq. (67)] follows. In Eq. (66), the Lagrangian subsemble crosspath stochastic integrals[18,5] are given by

$$\tilde{W}_{ij} \equiv \int_0^{s_{1,2}} ds_{1,2}' \int_0^{s_{1,2}} ds_{1,2}'' \, \tilde{E}\{w_i(s_{1,2}') w_j(s_{1,2}'')\},$$
$$i,j = 1, 2, 3, 4, \quad \text{(C1)}$$

where [compare with Eqs. (B1)–(B6)]

$$w_1 = w_1[\mathbf{X}(s_1', \xi_1)] \equiv \mu(s_1') - \tilde{\nu}_1 \quad \text{(C2a')}$$
$$= \alpha[n(s_1') - \tilde{E}\{n(s_1')\}], \quad \text{(C2)}$$

$$w_2 = w_2[\mathbf{X}(s_1', \xi_1)] \equiv \int_0^{s_1'} ds_1'' \frac{\mu_{,ii}(s_1'')}{\mu(s_1')} - \tilde{\nu}_2 \quad \text{(C3')}$$

$$= \alpha \int_0^{s_1'} ds_1'' \left[\frac{n_{,ii}(s_1'')}{\mu(s_1')} - \tilde{E}\left\{\frac{n_{,ii}(s_1'')}{\mu(s_1')}\right\}\right] \quad \text{(C3)}$$

$$\mu(s_1') = \mu(\mathbf{X}(s_1', \xi_1)] = 1 + \alpha n(s_1'), \qquad 0 \le \alpha \ll 1, \quad \text{(C4)}$$

$$w_3 = w_3[\mathbf{X}(s_2', \xi_2)] \equiv \mu(s_2') - \tilde{\nu}_3 \quad \text{(C5')}$$

$$= \alpha[n(s_2') - \tilde{E}\{n(s_2')\}], \quad \text{(C5)}$$

$$w_4 = w_4[\mathbf{X}(s_2', \xi_2)] \equiv \int_0^{s_2'} ds_2'' \frac{\mu_{,ii}(s_2'')}{\mu(s_2')} - \tilde{\nu}_4 \quad \text{(C6')}$$

$$= \alpha \int_0^{s_2'} ds_2'' \left[\frac{n_{,ii}(s_2'')}{\mu(s_2')} - \tilde{E}\left\{\frac{n_{,ii}(s_2'')}{\mu(s_2')}\right\}\right], \quad \text{(C6)}$$

$$\mu(s_2') = \mu[\mathbf{X}(s_2', \xi_2)] \equiv 1 + \alpha n(s_2'), \qquad 0 \le \alpha \ll 1, \quad \text{(C7)}$$

$$\tilde{\nu}_1 \equiv \tilde{E}\{\mu[\mathbf{X}(s_1', \xi_1)]\} = \tilde{E}\{\mu(s_1')\} \quad \text{(C8')}$$

$$= 1 + \alpha^2 \int_0^{s_1'} d\sigma (s_1' - \sigma)\,[\tilde{R}(\sigma) - O(\alpha)], \quad \text{(C8)}$$

$$\tilde{\nu}_2 \equiv \tilde{E}\left\{\int_0^{s_1'} ds_1'' \frac{\mu_{,ii}(s_1'')}{\mu(s_1')}\right\}, \quad \text{(C9)}$$

$$\tilde{\nu}_3 \equiv \tilde{E}\{\mu[\mathbf{X}(s_2', \xi_2)]\} = \tilde{E}\{\mu(s_2')\} \quad \text{(C10')}$$

$$= 1 + \alpha^2 \int_0^{s_2'} d\sigma (s_2' - \sigma)[\tilde{R}(\sigma) - O(\sigma)], \quad \text{(C10)}$$

and

$$\tilde{\nu}_4 \equiv \tilde{E}\left\{\int_0^{s_2'} ds_2'' \frac{\mu_{,ii}(s_2'')}{\mu(s_2')}\right\}. \quad \text{(C11)}$$

Note that w_1, w_2, w_3, w_4 are stochastic variables measured relative to their expectations so that the $\tilde{E}\{w_i(s_{1,2}')w_j(s_{1,2}'')\}$ are correlations of centered random variables; $\tilde{R}(\sigma)$ is the subensemble equivalent of $R(\sigma)$ of Eq. (37).

In Eqs. (66) and (67),

$$\int_0^{s_1} ds_1' \, \tilde{\nu}_1(s_1') \sim s_1 + \tfrac{1}{2}\alpha^2 s_1^2 \tilde{R}_0 \tilde{A}, \quad \text{(C12)}$$

$$\int_0^{s_1} ds_1' \, \tilde{\nu}_2(s_1') \sim \alpha^2 \tilde{B}_1 \tilde{b}_1 s_1, \quad \text{(C13)}$$

$$\int_0^{s_2} ds_2' \, \tilde{\nu}_3(s_2') \sim s_2 + \tfrac{1}{2}\alpha^2 s_2^2 \tilde{R}_0 \tilde{A}, \quad \text{(C14)}$$

and

$$\int_0^{s_2} ds_2' \, \tilde{\nu}_4(s_2') \sim \alpha^2 \tilde{B}_1 \tilde{b}_1 s_2, \quad \text{(C15)}$$

from Eqs. (C8)–(C11) and (B7)–(B11). The Lagrangian subensemble, autopath, stochastic integrals in Eqs. (66) and (67) are

$$\tilde{W}_{11} \sim 2\alpha^2 s_1 \tilde{L} \quad \text{(C16)}$$

and

$$\tilde{W}_{33} \sim 2\alpha^2 s_2 \tilde{L}, \quad \text{(C17)}$$

via Eqs. (C1), (C2), (C5), and (B14),

$$\tilde{W}_{12} = \tilde{W}_{21} \sim \alpha^2 s_1^2 \tilde{C}_0 \tilde{C}_2 \quad \text{(C18)}$$

and

$$\tilde{W}_{34} = \tilde{W}_{43} \sim \alpha^2 s_2^2 \tilde{C}_0 \tilde{C}_2, \quad \text{(C19)}$$

via Eqs. (C1)–(C7) and (B20), and

$$\tilde{W}_{22} \sim \tfrac{2}{3}\alpha^2 s_1^2 \tilde{Q}_0 \tilde{E} \quad \text{(C20)}$$

and

$$\tilde{W}_{44} \sim \tfrac{2}{3}\alpha^2 s_2^2 \tilde{Q}_0 \tilde{E}, \quad \text{(C21)}$$

via Eqs. (C1), (C4), (C7), and (B26).

The Lagrangian subensemble crosspath stochastic integrals $\tilde{W}_{13} = \tilde{W}_{31}$, $\tilde{W}_{14} = \tilde{W}_{41}$, $\tilde{W}_{23} = \tilde{W}_{32}$, and $\tilde{W}_{24} = \tilde{W}_{42}$ are evaluated asymptotically as follows by the methods of Refs. 18 and 5. Define the *crosspath phase correlation*, via Eqs. (C2) and (C5), by

$$\tilde{P}(s_1', s_2', \xi_1, \xi_2, \mathbf{x}) \equiv \tilde{E}\{[n(s_1') - \tilde{E}\{n(s_1')\}] \times [n(s_2') - \tilde{E}\{n(s_2')\}]\} \quad (C22)$$

$$= \tilde{P}(\sigma, S, \xi_1, \xi_2, \mathbf{x}), \quad (C22')$$

where Eq. (103) of Ref. 18 gives Eq. (C22'). Therefore, Eqs. (C1), (C2), (C5), and (C22) yield

$$W_{13} = W_{31} = \alpha^2 \int_0^{s_1} ds_1' \int_0^{s_2} ds_2' \, \tilde{P}(s_1', s_2', \xi_1, \xi_2, \mathbf{x}) \quad (C23)$$

$$\xrightarrow[s_1 \to s_2 = s_0]{} 2\alpha^2 s_0 \tilde{P}_0(0, s_0, \xi_1, \xi_2, \mathbf{x}) \int_0^{s_0} d\sigma \, \frac{1}{2s} \int_\sigma^{2s_0 - \sigma} dS \times \frac{\tilde{P}(\sigma, S, \xi_1, \xi_2, \mathbf{x})}{\tilde{P}_0(0, s_0, \xi_1, \xi_2, \mathbf{x})} \quad (C24)$$

$$\xrightarrow[\delta \to 0]{} 2\alpha^2 s_0 \tilde{P}_0 \int_0^{s_0} d\sigma \; 1\left(-\frac{\sigma}{s_0}\right) \frac{\tilde{P}(\sigma, \xi, \mathbf{x})}{\tilde{P}_0} \quad (C25)$$

$$\sim 2\alpha^2 s_0 \tilde{P}_a(\delta) \tilde{H}(\delta) \quad (C26)$$

$$\xrightarrow[\delta \to 0]{} 2\alpha^2 s_0 \tilde{L}, \quad (C27)$$

where

$$\tilde{P}_0(0, s_0, \xi_1, \xi_2, \mathbf{x}) \equiv \frac{1}{2s_0} \int_0^{2s_0} dS \tilde{P}(0, S, \xi_1, \xi_2, \mathbf{x}) \quad (C28)$$

$$\sim \tilde{P}_0(0, s_0, \delta) \simeq \tilde{P}_a(\delta). \quad (C29)$$

Since the saddle-point evaluation of $\langle |p(\mathbf{x})|^2 \rangle$, given in Appendix D, results in $s_1 = s_2 = s_0$ (the saddle-point), Eq. (C24) and, consequently, Eq. (C26), suffices for this study; the more general situation is given by Eqs. (107) and (111) of Ref. 18. When $\delta \to 0$, Eq. (C24) reduces to Eq. (C25), i.e., Eq. (B13'), and Eq. (C26) reduces to Eq. (C27), i.e., Eq. (B14). Equation (C24) goes asymptotically to Eq. (C26), where the Lagrangian subensemble crosspath phase integral scale $\tilde{H}(\delta)$,

$$\tilde{H}(\delta) \equiv \lim_{s_0 \to \infty} \int_0^{s_0} d\sigma \, \frac{1}{2s_0} \int_\sigma^{2s_0 - \sigma} dS \, \frac{\tilde{P}(\sigma, S, \delta)}{\tilde{P}_0(0, s_0, \delta)} \quad (C30)$$

$$\xrightarrow[\delta \to 0]{} \tilde{L}, \quad (C31)$$

[cf. Eqs. (102) and (116) of Ref. 18] is assumed to exist and must be independent of s_0 and where $\tilde{P}_a(\delta)$ represents the curvilinear average of $\tilde{P}_0(0, s_0, \delta)$ over the s_0 *range of experimental interest* (see Appendix B of Ref. 18). Note that the intensity factor at $s_0 = 0$ is

$$\tilde{P}_0(\delta) \equiv \tilde{E}\{n[\mathbf{X}(0, \xi_1)] n[\mathbf{X}(0, \xi_2)]\} \quad (C32)$$

$$= \langle n(\xi_1) n(\xi_2) \rangle \gtrsim \tilde{P}_a(\delta) \quad (C33)$$

$$\xrightarrow[\delta \to 0]{} \tilde{P}_0 = 1 \quad (C34)$$

$$\xrightarrow[\delta \to \infty]{} 0. \quad (C35)$$

Since, via Eqs. (113a) and (96a) of Ref. 18,

$$\frac{1}{2(s_0 - \sigma)} \int_\sigma^{2s_0 - \sigma} dS \, \frac{\tilde{P}(\sigma, S, \delta)}{\tilde{P}_0(0, s_0, \delta)} \xrightarrow[\delta \to 0]{} \frac{\tilde{P}(\sigma)}{\tilde{P}_0}, \quad (C36')$$

$\tilde{H}(\delta)$ is probably nearly equal to $\tilde{L}$ for all s_0 of interest and δ less than, say $3\tilde{L}$, i.e.,

$$\tilde{H}(\delta) \simeq \tilde{L}, \quad \delta < 3\tilde{L}, \quad (C36)$$

[cf. Eqs. (121a) and (121) of Ref. 18]. However, via Eqs. (C32)–(C35),

$$1 > \tilde{P}_0(\delta) \gtrsim \tilde{P}_a(\delta) > 0, \quad (\delta > 0, s_0 > 0) \quad (C37')$$

so that

$$\tilde{L} > \tilde{P}_a(\delta) \tilde{H}(\delta) \geq 0 \quad (\delta > 0, s_0 > 0) \quad (C37)$$

in Eq. (67). In Sec. IV, $\tilde{P}_a(\delta)\tilde{H}(\delta)$ proves to be a dominant factor in the coefficient of intensity fluctuation V in the phase-dominance region via Eq. (75).

Since $k_1 = k_0 = -k_3$ while $k_2 = i/2 = k_4$, the factors involving $\tilde{W}_{14}$ and $\tilde{W}_{41}$ cancel the factors involving $\tilde{W}_{23}$ and $\tilde{W}_{32}$ from Eq. (67) when $s_1 = s_2 = s_0$. Since this is the only case considered in this study, the Lagrangian subensemble crosspath phase-amplitude integral scale contribution will not be considered.

Define the *crosspath amplitude correlation*, via Eqs. (C3) and (C6), by

$$\tilde{Q}(s_1'', s_2'', \xi_1, \xi_2, \mathbf{x}) \equiv \tilde{E}\{[n_{,ii}(s_1'') - \tilde{E}\{n_{,ii}(s_1'')\}] \times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}]\} \quad (C38)$$

$$= \tilde{Q}(\sigma, S, \xi_1, \xi_2, \mathbf{x}), \quad (C38')$$

where Eq. (103) of Ref. 18 gives Eq. (C38'). Therefore, applying Eqs. (C3) and (C6) to Eq. (C1) produces

$$\tilde{W}_{24} = \tilde{W}_{42} = \alpha^2 \int_0^{s_1} ds_1' \int_0^{s_1'} ds_1'' \int_0^{s_2} ds_2' \int_0^{s_2'} ds_2'' \times [\tilde{Q}(s_1'', s_2'', \xi_1, \xi_2, \mathbf{x}) - O(\alpha)] \quad (C39)$$

$$\xrightarrow[s_1 \to s_2 = s_0]{} \alpha^2 \int_0^{s_0} ds_1''(s_0 - s_1'') \int_0^{s_0} ds_2''(s_0 - s_2'') \times \tilde{Q}(s_1'', s_2'', \xi_1, \xi_2, \mathbf{x}) - O(\alpha^3 s^3) \quad (C40)$$

$$\xrightarrow[\delta \to 0]{} \alpha^2 \int_0^{s_0} ds_1''(s_0 - s_1'') \int_0^{s_0} ds_2''(s_0 - s_2'') \times \tilde{Q}(s_1'', s_2'', \xi, \mathbf{x}) - O(\alpha^2 s^3) \quad (C41)$$

$$= \alpha^2 s_0^3 \tilde{Q}_3(0, s_0, \xi_1, \xi_2, \mathbf{x}) \tfrac{2}{3} \int_0^{s_0} d\sigma \times \left(1 - \frac{\sigma}{2s_0} - \frac{\sigma^2}{2s_0^2} + \frac{3\sigma^3}{4s_0^3}\right) \times \tilde{Q}_3(\sigma, s_0, \xi_1, \xi_2, \mathbf{x}) / \tilde{Q}_3(0, s_0, \xi_1, \xi_2, \mathbf{x}) - O(\alpha^2 s^3) \quad (C42)$$

$$\sim \tfrac{2}{3} \alpha^2 s_0^3 \tilde{Q}_a(\delta) \tilde{J}(\delta) \quad (C43)$$

$$\xrightarrow[\delta \to 0]{} \tfrac{2}{3} \alpha^2 s_0^3 \tilde{Q}_0 \tilde{E}. \quad (C44)$$

Defining

$$\left(\frac{2s_0^3}{3} - \frac{\sigma s_0^2}{3} - \frac{\sigma^2 s_0}{3} + \frac{\sigma^3}{2}\right) \tilde{Q}_3(\sigma, s_0, \xi_1, \xi_2, \mathbf{x}) \equiv \int_\sigma^{2s_0 - \sigma} dS[s - \tfrac{1}{2}(S + \sigma)][s - \tfrac{1}{2}(S - \sigma)] \times \tilde{Q}(\sigma, S, \xi_1, \xi_2, \mathbf{x}) \quad (C45)$$

$$\sim s_0^3 \tfrac{2}{3} \tilde{Q}_3(\sigma, s_0, \delta) \quad (C46)$$

[cf. Eq. (B26') and Eqs. (102) through (104) of Ref. 18]

yields Eq. (C42) by standard procedures. Equation (C42) goes asymptotically to Eq. (C43), where the Lagrangian subensemble crosspath amplitude integral scale $\tilde{J}(\delta)$,

$$\tilde{J}(\delta) \equiv \lim_{s_0\to\infty} \int_0^{s_0} d\sigma \left(1 - \frac{\sigma}{2s_0} - \frac{\sigma^2}{2s_0^2} + \frac{3\sigma^3}{4s_0^3}\right) \frac{\tilde{Q}_3(\sigma, s_0, \delta)}{\tilde{Q}_3(0, s_0, \delta)} \tag{C47}$$

$$\xrightarrow[\delta\to 0]{} \tilde{E}, \tag{C48}$$

[cf. Eq. (116) of Ref. 18] is assumed to exist and cannot be dependent on s_0 and where $\tilde{Q}_a(\delta)$ represents the curvilinear average of $\tilde{Q}_3(0, s_0, \delta)$ over the s_0 range of experimental interest. Note that the intensity factor at $s_0 = 0$ is

$$\tilde{Q}_0(\delta) \equiv \tilde{E}\{n_{,ii}[\mathbf{X}(0, \xi_1)] n_{,ii}[\mathbf{X}(0, \xi_2)]\} \tag{C49}$$

$$= \langle n_{,ii}(\xi_1) n_{,ii}(\xi_2)\rangle \gtrsim \tilde{Q}_a(\delta) \tag{C50}$$

$$\xrightarrow[\delta\to 0]{} \tilde{Q}_0 \tag{C51}$$

$$\xrightarrow[\delta\to\infty]{} 0. \tag{C52}$$

Since, via Eqs. (113a) and (96a) of Ref. 18,

$$\tilde{Q}_3(\sigma, s_0, \delta)/\tilde{Q}_3(0, s_0, \delta) \xrightarrow[\delta\to 0]{} \tilde{Q}(\sigma)/\tilde{Q}_0, \tag{C53'}$$

$\tilde{J}(\delta)$ is probably nearly equal to $\tilde{E}$ for all s_0 of interest and δ less than, say, $3\tilde{E}$, i.e.,

$$\tilde{J}(\delta) \simeq \tilde{E}, \quad \delta < 3\tilde{E}, \tag{C53}$$

[cf. Eqs. (121a) and (121) of Ref. 18]. However, via Eqs. (C49) through (C52),

$$\tilde{Q}_0 > \tilde{Q}_0(\delta) \gtrsim \tilde{Q}_a(\delta) \geq 0 \quad (\delta > 0, s_0 > 0) \tag{C54'}$$

so that

$$\tilde{Q}_0\tilde{E} > \tilde{Q}_a(\delta)\tilde{J}(\delta) \geq 0 \quad (\delta > 0, s_0 > 0) \tag{C54}$$

in Eq. (67). In Sec. IV, $\tilde{Q}_a(\delta)\tilde{J}(\delta)$ proves to be a dominant factor in the coefficient of intensity fluctuation V in the amplitude-dominance region.

APPENDIX D: SADDLE-POINT APPROXIMATIONS

The saddle-point approximation[10,22] gives

$$\int_0^\infty ds g(s) e^{f(s)/\alpha^2} \simeq [-2\pi\alpha^2/f''(s_0)]^{1/2} g(s_0) e^{f(s_0)/\alpha^2}, \tag{D1}$$

where $g(s)$ is a relatively slowly changing function compared to $\exp[f(s)/\alpha^2]$ near the saddle point s_0. The saddle-point evaluation of Eq. (64) proceeds as follows. Assume

$$\tfrac{1}{2}\alpha^2 s R_0 A \ll 1, \tag{D2a}$$

$$\alpha^2 s \tilde{D} \ll 1, \tag{D2b}$$

$$\tfrac{1}{2}\alpha^2 s R_0 B_4/A_3 = \tfrac{3}{8}\alpha^2 s R_0 A \ll 1, \tag{D2c}$$

via Eqs. (A16) and (A38), and define

$$f(s) \equiv -[(s-x)^2 + 2\rho^2]/4R_0 A_3 s^3 \tag{D3}$$

so that

$$0 = \frac{\partial f}{\partial s}\Big]_{s_0} = [-2(s_0 - x)s_0 + 3(s_0 - x)^2 + 6\rho^2]/ \times 4R_0 A_3 s_0^4 \tag{D3'}$$

gives the saddle-point value of s:

$$s_0 = 2x - (x^2 - 6\rho)^{1/2} \tag{D4}$$

$$\xrightarrow[\rho\to 0]{} x \tag{D4a}$$

$$\simeq x + 3\rho^2/x, \quad \rho \ll x, \tag{D4b}$$

$$= x[2 - (1 - 6\tan^2\Theta)^{1/2}] \tag{D5}$$

$$\xrightarrow[\Theta\to 0]{} x \tag{D5a}$$

$$\xrightarrow[\Theta\to\pi/2]{} \infty \tag{D5b}$$

$$\simeq x + 3x\tan^2\Theta, \quad \Theta < 5°, \tag{D5c}$$

$$\simeq x \simeq r, \quad \Theta < 4°. \tag{D5d}$$

The assumption of Eqs. (D2a)–(D2c) simplify the form of Eq. (64) so that Eq. (D3') is no worse to solve than a quadratic. They can be justified as follows for a turbulent fluid [see, also, the discussion of (4.6-16) of Ref. 5]. This is demonstrated by the order of magnitude techniques employed in Ref. 1. Since the turbulent microscale λ_θ is related to λ_g by[1,2,5,24,25]

$$\lambda_\theta^2 = \lambda_g^2/\sigma_\kappa, \tag{D6a}$$

where $\sigma_\kappa \simeq 7$ in turbulent water,[18,5]

$$\tilde{R}_0 = R_0 = \langle n_{,i}(\xi) n_{,i}(\xi)\rangle \approx 3\langle n^2\rangle/\lambda_\theta^2 \simeq 21/\lambda_g^2. \tag{D6b}$$

Equation (56) of Ref. 1 provides

$$L_u/\lambda_g \approx R_t/10 \tag{D6c}$$

so that

$$\tfrac{1}{2}\alpha^2 s\tilde{R}_0\tilde{A} < 10\alpha^2(L_u/\lambda_g^2)s \tag{D6d}$$

$$\approx \tfrac{1}{10}\alpha^2 R_t^2(s/L_u) \tag{D6e}$$

with $\tilde{A} < L_u$, the turbulent macroscale defined in Eq. (44) of Ref. 1 (probably $\tilde{A} \ll L_u$). Since

$$\tfrac{1}{2}\alpha^2 s\tilde{C}_0\tilde{C}_2 < \tfrac{1}{2}\alpha^2 s\tilde{R}_0\tilde{A} \tag{D6f}$$

[via Eqs. (B21) and (B8)] and

$$\tfrac{1}{2}\alpha^2 s R_0 A < \tfrac{1}{2}\alpha^2 s\tilde{R}_0\tilde{A} \tag{D6g}$$

(via Sec. V of Ref. 18), Eqs. (D2a)–(D2c) follow for

$$s \lesssim L_u/(\alpha R_t)^2 \tag{D6h}$$

or

$$s \lesssim \lambda_g^2/(100\alpha^2 L_u), \tag{D6i}$$

remembering Eq. (B23). For the Stone and Mintzer experiments,[26,27] where[1,5] $\alpha \simeq 1.6 \times 10^{-4}$ and $R_t \approx 58$, Eq. (D6h) permits s to be as large as $10^4 L_u$. In the case of the turbulent upper ocean discussed in Ref. 1, where $\alpha \approx 10^{-4}$ and $L_u/\lambda_g^2 \approx 1\text{m}^{-1}$, Eq. (D6i) provides for $s \lesssim 10^6$m. In Eq. (D5), the scattering angle Θ has been defined by

$$\tan\Theta = \rho/x \tag{D7}$$

so that

$$x = r\cos\Theta, \tag{D7a}$$

where

$$r = (x^2 + \rho^2)^{1/2} \tag{D7b}$$

is the source-to-receiver range. Since $\rho \to 0$ and $\Theta \to 0$ when $\alpha \to 0$, Eqs. (D4a) and (D5a) give the required nonstochastic limit and Eqs. (D4) and (D5) show how the saddle-point curvilinear distance s_0 varies from its value x in a nonstochastic, uniform medium by a factor that depends on the scattering geometry via $\tan\Theta$. When Eqs. (D2a)–(D2c) are not imposed, s_0 depends on the statistical nature of the medium, via α and R_0A, as well as the scattering geometry. Equation (D5b) shows that this saddle-point approximation is limited to $\Theta < \pi/2$, while Eq. (60) imposes the limit $\Theta > 0$, since Eq. (55) increases without limit as $s_0 \to \infty$ due to the dominance of $\exp(\alpha^2\tilde{Q}_0\tilde{E}s_0^3/12)$ for large s_0, i.e.,

$$0 < \Theta < \pi/2. \tag{D8}$$

The saddle-point estimate of Eq. (64), under the assumption of Eqs. (D2a)–(D2c), is, therefore,

$$\langle p(\mathbf{x})\rangle \sim p_0 S_0 \left(\frac{2\pi\alpha^2}{-f''(s_0)}\right)^{1/2} \times \frac{e^{-\alpha^2(k_0^2\tilde{L}s_0 - s_0^3\tilde{Q}_0\tilde{E}/12) + ik_0s_0}}{2\pi\alpha^2R_0A_3s_0^3(4\pi\alpha^2R_0A_3s_0^3)^{1/2}}\, e^{f(s_0)/\alpha^2} \tag{D9}$$

$$\equiv \langle p(x,\rho)\rangle_e \tag{D9'}$$

$$\xrightarrow[\alpha\to 0]{} 0. \tag{D10}$$

Equation (D1) produces Eq. (D9), where, despite the considerable analytical simplification that has preceded in this analysis, $f(s_0)$ of Eq. (D3) and

$$f''(s_0) \equiv \left[\frac{\partial^2 f}{\partial s^2}\right]_{s_0} = \frac{s_0^2 - 6s_0x + 6(x^2 + 2\rho^2)}{-2R_0A_3s_0^5} \tag{D11}$$

are complicated functions of x and ρ when Eq. (D4) is applied. However, these factors will cancel out later in Sec. IV. Equation (D10) follows from Eq. (D9) because of the simplifying conditions behind Eq. (D1) disallow consideration of the null-deviation path result in the nonstochastic limit; the correct nonstochastic limit of Eq. (59) must follow from Eq. (58) since the saddle-point attack requires $\alpha > 0$ as well as $\alpha \ll 1$. It is noted that the peak of $\exp[f(s_0)/\alpha^2]$ broadens as s_0 increases. However, it can be shown that the saddle-point approximation is quite good by demonstrating that

$$\{[h''(s)/h(s)](2\alpha^2R_0A_3s^3)\}^{1/2} \ll 1,$$

where

$$h(s) = e^{-\alpha^2k_0^2\tilde{L}s}/2\pi\alpha^2R_0A_3s^3;$$

this proves $h(s)$ slowly changing over the spread, $2\alpha^2R_0A_3s^3$, of $\exp[f(s)/\alpha^2]$. For $\alpha^2k_0^2\tilde{L}s \ll 1$,

$$[h''(s)/h(s)](2\alpha^2R_0A_3s^3) \approx 4\alpha^2R_0A_3s \lesssim 8\times 10^{-5}(s/L_u) \ll 1$$

when $s \lesssim 10^4L_u$ for the Stone and Mintzer[26,27] experiment.[2] Likewise, for the turbulent upper ocean case discussed in Ref. 2,

$$[h''(s)/h(s)](2\alpha^2R_0A_3s^3) \approx 8\times 10^{-7}(s/1\text{ meter}) \ll 1$$

for $s \ll 10^5$m.

1 J. A. Neubert and J. L. Lumley, J. Acoust. Soc. Amer. **48**, 1212 (1970).
2 J. A. Neubert, J. Acoust. Soc. Amer. **48**, 1203 (1970).
3 J. A. Neubert, J. Acoust. Soc. Amer. **51**, 310 (1972).
4 V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).
5 J. A. Neubert, Ph.D. thesis (Pennsylvania State University, 1970).
6 D. Mintzer, J. Acoust. Soc. Amer. **25**, 922 (1953).
7 D. Mintzer, J. Acoust. Soc. Amer. **25**, 1107 (1953).
8 D. Mintzer, J. Acoust. Soc. Amer. **26**, 186 (1954).
9 Y. A. Kravtsov, Akust. Zh. **14**, 1 (1968) [Sov. Phys. Acoust. **14**, 1 (1968)].
10 C. B. Officer, *Introduction to the Theory of Sound Transmissions* (McGraw-Hill, New York, 1958).
11 I. Tolstoy and C. S. Clay, *Ocean Acoustics* (McGraw-Hill, New York, 1966).
12 J. L. Lumley, J. Math. Phys. **3**, 309 (1962).
13 U. Frisch, *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bhrucha-Reid (Academic, New York, 1968).
14 P. Uginčvius, J. Acoust. Soc. Amer. **45**, 206 (1969).
15 J. A. Neubert, J. Math. Phys. **12**, 1030 (1971).
16 R. G. Bergmann, Phys. Rev. **70**, 486 (1946).
17 D. Potter and S. Murphy, J. Acoust. Soc. Amer. **29**, 197 (1957).
18 J. A. Neubert and J. L. Lumley, J. Math. Phys. **13**, 169 (1972).
19 J. L. Lumley, *Stochastic Tools in Turbulence* (Academic, New York, 1971).
20 F. Reif, *Fundamentals of Statistical and Thermal Physis* (McGraw-Hill, New York, 1965).
21 J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
22 J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, New York, 1964).
23 W. Baerg and W. H. Schwarz, J. Acoust. Soc. Amer. **39**, 1125 (1966).
24 J. A. Neubert and J. L. Lumley, "The Nature of Sound Fluctuation in the Upper Ocean" (in preparation).
25 H. Tennekes and J. L. Lumley, *Turbulent Flow* (Prentice-Hall, New York, 1971).
26 R. G. Stone and D. Mintzer, J. Acoust. Soc. Amer. **34**, 647 (1962).
27 R. G. Stone and D. Mintzer, J. Acoust. Soc. Amer. **38**, 843 (1965).

Recurrence Formula for the Veneziano Model N-Point Functions

Koichi Mano
Air Force Cambridge Research Laboratories, Bedford, Massachusetts 01730
(Received 1 May 1972)

A recurrence formula is derived for a function which reduces to the Veneziano model $(n+3)$-point function. It is shown that the formula is equivalent to, but is more self-contained than, the Hopkinson and Plahte formula in that it does not require the prescription for the parameters involved.

The extension to the n-point functions of the Veneziano's four-point function was accomplished either by generalizing the integral representation for the beta function which comprises the essential ingredient of the four-point function[1,2] or by generating a recurrence formula for the n-point function.[3] In the latter approach, the authors attempted to justify the formula for arbitrary value of n after showing, through introduction of the integral representation for the beta function, that the formula produces the already known integral expressions for the cases of $n = 5, 6,$ and 7. The recurrence formula which has apparently been discovered on a heuristic basis is not necessarily very transparent, as the authors themselves admit it, especially in connection with the definition of their variables x'_{ij}.

Recently it was pointed out that the generalized Veneziano amplitudes may be regarded as the boundary values of a class of generalized hypergeometric functions that are Radon transforms of products of linear forms.[4] In a work by the present author which shows that the amplitudes possess a structure similar to that of the Lauricella's hypergeometric functions,[5] he has made an iterative use of a recurrence formula for the amplitude.[6]

The purpose of this note is to point out that the author's recurrence formula obtains itself in a quite natural manner such that for a special choice of the variables, it reproduces the formula proposed in Ref. 3 without requiring any prescription for the parameters involved therein.

To begin our discussion, let us consider a function V_n of variables w_{ij} defined as below:

$$\begin{aligned} &V_n(\alpha_{01},\alpha_{02},\ldots,\alpha_{0n};\alpha_{11},\alpha_{12},\ldots,\alpha_{1n};\\ &\quad \alpha_{21},\alpha_{22},\ldots,\alpha_{2,n-1};\alpha_{31},\ldots,\alpha_{3,n-2};\ldots;\\ &\quad \alpha_{n-1,1},\alpha_{n-1,2};\alpha_{n1}\,|w_{02},w_{03},\ldots,w_{0n};\\ &\quad w_{13},\ldots,w_{1n};\ldots;w_{n-2,n})\\ &= \int_0^1\cdots\int_0^1 \prod_{i=1}^{n}\Big\{du_i u_i^{\alpha_{0i-1}}(1-u_i)^{\alpha_{1i-1}}\\ &\quad\times\prod_{k=2}^{i}1-u_i\Big(w_{i-k,i}\prod_{j=0}^{k-2}u_{i-j-1}\Big)\Big]^{\alpha_{k,i-k+1}}\Big\}. \end{aligned} \tag{1}$$

When the parameters α_{ij} are regarded as functions of the momenta p_i, $i = 1,2,\ldots,n+3$, of the external particles, it will be seen that V_n for $w_{ij} = 1$ can readily be related to the known integral representation for the $(n+3)$-point function.[7]

In carrying out the multiple integrations in Eq. (1), use has been made in Ref. 6 of the following recurrence formula:

$$\begin{aligned} &V_n(\alpha_{01},\ldots,\alpha_{0n};\alpha_{11},\ldots,\alpha_{1n};\ldots;\alpha_{n1}\\ &\quad|w_{02},\ldots,w_{0n};w_{13},\ldots,w_{1n};\ldots;w_{n-2,n})\\ &= B(\alpha_{0n},\alpha_{1n})\sum\frac{(\alpha_{0n},\beta_{n,n-1})}{(\alpha_{0n}+\alpha_{1n},\beta_{n,n-1})}\\ &\quad\times\frac{(-\alpha_{2,n-1},r_{2,n-1})\cdots(-\alpha_{n1},r_{n1})}{(1,r_{2,n-1})\cdots(1,r_{n1})}\\ &\quad\times(w_{n-2,n})^{r_{2,n-1}}(w_{n-3,n})^{r_{3,n-2}}\ldots(w_{0n})^{r_{n1}}\\ &\quad\times V_{n-1}(\alpha_{01}+\beta_{n1},\ldots,\alpha_{0,n-1}+\beta_{n,n-1};\alpha_{11},\ldots,\\ &\quad\alpha_{1,n-1};\ldots;\alpha_{n-1,1}\,|w_{02},\ldots,w_{0,n-1};w_{13},\cdots,w_{1,n-1};\\ &\quad\cdots;w_{n-3,n-1}). \end{aligned} \tag{2}$$

In Eq. (2) the summation is over the integers between 0 and ∞ of the indexes r_{ij}, B stands for the beta function, (α,r) under the summation symbol is written for $\Gamma(\alpha+r)/\Gamma(\alpha)$, and β_{pg} are given by

$$\beta_{pq} = \sum_{k=p-q+1}^{p} r_{k,p-k+1} \quad \text{for} \quad \begin{cases} p = 2,3,\ldots,n \\ q = 1,2,\ldots,p-1. \end{cases} \tag{3}$$

Our task in what follows is to show that when $w_{ij} = 1$, Eq. (2) reduces to the recurrence formula for the N-point function $B_N(x)$ in Ref. 3 as given below:

$$\begin{aligned} B_N(x) = \sum_{k_{i,N-1}=0}^{\infty}\Big[\prod_{i=2}^{N-3}(-1)^{k_{i,N-1}}\binom{z_{i,N-1}}{k_{i,N-1}}\Big]\\ \times B_4\Big(x_{N-2,N-1},x_{N-1,N}+\sum_{i=2}^{N-3}k_{i,N-1}\Big)\,B_{N-1}(x'), \end{aligned} \tag{4}$$

where

$$x_{ij} = -\alpha(s_{ij}) \quad \text{with} \quad s_{ij} = (p_i + p_{i+1} + \cdots + p_j)^2,$$
$$z_{ij} = x_{ij} - x_{i+1,j} - x_{i,j-1} + x_{i+1,j-1},$$

and x'_{ij} are defined according to certain rules (given in a tabular form) which will not be reproduced here. More noteworthy of the present formula is the fact that Eq. (2) is self-contained such that in contrast to Ref. 3, there is required no prescription for defining the parameters of the function V_{n-1} which corresponds to B_{N-1} of Eq. (4).

In order to achieve the above we have to establish the relation between our parameters α_{ij} and those of Ref. 3. For this purpose let us note first that the external lines are labeled $1,2,\ldots,n+3$ both for the $(n+3)$-point function in Ref. 1 and for V_n in the present paper, while they are labeled $0,1,\ldots,n+2$ for the $(n+3)$-point function B_{n+1} of Ref. 7. Further, we note that the integration variables $u_1,u_2,\ldots,u_n$ in Ref. 7 and the present paper may be made to correspond to $u_{12},u_{13},\ldots,u_{1,n+1}$ of Ref. 1. With this in mind one can compare Eq. (1) with the corresponding expression that follows from the representation for $B_{n+3}(x)$ of Ref. 1 through rearrangement of the integrand. Namely, by introducing $x_{ij} = -\bar{\alpha}_{ij} - 1$ from Ref. 1, where we write $\bar{\alpha}_{ij}$ for $\alpha_{ij} = \alpha(s_{ij})$ of Ref. 1, it becomes possible to express our α_{ij} in terms of $\bar{\alpha}_{km}$. If we further write $\xi_{ij} = -\bar{\alpha}_{ij}$ with $\xi_{ii} \equiv 0$ and

$$\zeta_{ij} = \xi_{ij} - \xi_{i+1,j} - \xi_{i,j-1} + \xi_{i+1,j-1}, \tag{5}$$

where ξ_{ij} and ζ_{ij} stand for x_{ij} and z_{ij}, respectively, of Ref. 3, it follows that

$$\left.\begin{aligned}\alpha_{0i} &= \xi_{i,i+1}\\ \alpha_{1i} &= \xi_{i+1,i+2}\end{aligned}\right\} \quad \text{for} \quad i = 1,2,\ldots,n \tag{6}$$

and

$$\alpha_{ki} = \zeta_{i+1,i+k+1} \quad \text{for} \quad k = 2,3,\ldots,n$$
$$\text{and } i = 1,2,\ldots,n-k+1. \tag{7}$$

That the integral in Eq. (1) reduces for $w_{ij} = 1$ to B_{n+1} of Ref. 7 can be seen from the observation that our $p_i, i = 1,2,\ldots,n+3$ correspond to $p_i, i = 0,1,\ldots, n+2$ of Ref. 7 and through specialization of the relation $\bar{\alpha}(s_{ij}) = \alpha' s_{ij} + \alpha_{ij}^0$ of Ref. 1 to the form $bs_{ij} + a$ as is done in Ref. 7.

Although we have connected our parameters to those of Ref. 3 in Eqs. (6) and (7), the precise correspondence between Eqs. (2) and (4) cannot be considered complete until the arrangement of ξ_{ij} in $B_N(\xi)$ of Eq. (4) is unambiguously established. [In Ref. 3 this arrangement has been left out unstated, which fact is responsible in part for requiring the somewhat troublesome rules for determining x' in $B_{N-1}(x')$ which should have really been unnecessary, as will be shown below.]

Let us suppose that the correspondence between V_n in this paper and B_N, for $N = n+3$, of Ref. 3 is given by the following:

$$\begin{aligned}&V_n(\alpha_{01},\ldots,\alpha_{0n};\alpha_{11},\ldots,\alpha_{1n};\alpha_{21},\ldots,\alpha_{2,n-1};\ldots;\\ &\quad \alpha_{n1}\,|w_{02},\ldots,w_{0n};\ldots;w_{n-2,n})\\ &\quad \longleftrightarrow B_{n+3}(\xi_{12},\xi_{23},\ldots,\xi_{n+1,n+2};\xi_{13},\xi_{24},\ldots,\\ &\quad \xi_{n,n+2};\xi_{14},\xi_{25},\ldots,\xi_{n-1,n+2};\ldots;\xi_{1,n+1},\xi_{2,n+2}).\end{aligned} \tag{8}$$

Then the transition from V_n to B_{n+3} and vice versa can be effected on a firm basis by referring to Eqs. (6) and (7).

With the help of Eq. (8) we now can translate Eq. (2) into a formula which is given in terms of the function B_N:

$$\begin{aligned}&B_{n+3}(\xi_{12},\xi_{23},\ldots,\xi_{n+1,n+2};\xi_{13},\xi_{24},\ldots,\\ &\quad \xi_{n,n+2};\ldots;\xi_{1,n+1},\xi_{2,n+2})\\ &= \sum \frac{(-\zeta_{n,n+2},r_{2,n-1})\cdots(-\zeta_{2,n+2},r_{n1})}{(1,r_{2,n-1})\cdots(1,r_{n1})}\\ &\quad \times B_4(\xi_{n+1,n+2},\xi_{1,n+1}+\beta_{n,n-1})\\ &\quad \times B_{n+2}(\xi_{12}+\beta_{n1},\xi_{23},\ldots,\xi_{n,n+1};\xi_{13}+\beta_{n2},\\ &\quad \xi_{24},\ldots,\xi_{n-1,n+1};\ldots;\xi_{1n}+\beta_{n,n-1},\xi_{2,n+1}).\end{aligned} \tag{9}$$

Note that we wrote B_4 for B, and use was made of the following relation in obtaining Eq. (9):

$$\begin{aligned}&B(\alpha_{0n},\alpha_{1n})\,(\alpha_{0n},\beta_{n,n-1})/(\alpha_{0n}+\alpha_{1n},\beta_{n,n-1})\\ &\qquad = B(\alpha_{1n},\alpha_{0n}+\beta_{n,n-1}).\end{aligned}$$

That Eq. (9) is identical to Eq. (4) with $N = n+3$, $x \to \xi$, and $z \to \zeta$, can be checked easily. This establishes, therefore, that the order in which x_{ij} appears in $B_N(x)$ of Eq. (4), which was not stated explicitly in Ref. 3, should be exactly as is displayed in B_{n+3} of Eq. (8)

We emphasize that the recurrence formula for the $(n+3)$-point function, Eq. (9), as derived from Eq. (2) is complete as it stands and requires no rules for defining the parameters of the function B_{n+2}. In connection with the table for x'_{ij} of $B_{N-1}(x')$ in Ref. 3, we note that not all the entries in the table are actually needed for the recurrence formula. In fact, what is needed is only that portion of the table for $i = 1$, $j < N-2$ and $i > 1$, $j \leq N-2$ because, as may be seen from the arguments of B_{n+2} in Eq. (9), we require only ξ_{1j} for $j < N-2 = n+1$ and ξ_{ij} with $i > 1$ for $j \leq N-2 = n+1$. Moreover, there arises no need for including in the table the relation $x'_{1j} = x_{1j} + \sum_2^{j-1} k_{\alpha,N-1}$, for $j = N-2$, unless we unnecessarily rewrite the argument $x_{N-1,N} + \sum_2^{N-3} k_{i,N-1}$ in B_4 of Eq. (4) as $x_{1,N-2} + \sum_2^{N-3} k_{i,N-1} = x'_{1,N-2}$. Finally, it is noted further that x'_{ij} for $i > 1$ and $j = N-1$ should have not been included in the table since no such variables are actually involved in the recurrence formula for $B_N(x)$.

1 Chan Hong-Mo and Tsou Sheung Tsun, Phys. Letters **28B**, 485 (1969).
2 C. J. Goebel and B. Sakita, Phys. Rev. Letters **22**, 257 (1969).
3 J. F. L. Hopkinson and E. Plahte, Phys. Letters **28B**, 489 (1969).
4 A. C. T. Wu, J. Math. Phys. **12**, 2035 (1971).
5 P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques—polynomes d'Hermite* (Gauthier-Villars, Paris, 1926), p. 114.
6 K. Mano, J. Math. Phys. **13**, 1136 (1972).
7 K. Bardakci and H. Ruegg, Phys. Rev. **181**, 1884 (1969).

Linear Inequalities for Density Matrices. II

Walter B. McRae
Department of Chemistry, University of Georgia, Athens, Georgia 30601

and

Ernest R. Davidson
Department of Chemistry, University of Washington, Seattle, Washington 98195
(Received 11 February 1972; Revised Manuscript Received 5 June 1972)

A general method is described for finding linear inequalities relating physical properties of an ensemble which can exist only in a finite number of states. This algorithm is used to solve a few examples of the fixed-N and variable-N Slater hull problem and the fixed-N Boson hull problem. Interpretations of the results are made in terms of boundary conditions for fermion and boson reduced density matrices, pair distributions of lattice vacancies, and pair distributions of spins in an Ising model.

INTRODUCTION

The distribution of particles among orbitals in a wavefunction, the distribution of pairs of particles among pairs of orbitals, the Ising model of magnetism, the reaction rate of adsorbed atoms on a surface, and the distribution of vacancies in a lattice are quite dissimilar problems. In each of these cases, however, if one asks what could conceivably happen rather than what does happen, a similar mathematical problem arises. Solution of this problem gives certain bounds on the results which may be of interest in cases where exact solutions are not available.

In order to formulate the mathematical problem, let us envisage a system which can exist in only a finite number of states labeled by $K = 1, 2, \ldots, \mathcal{K}$. An example of such a system would be a set of sites labeled by $\mu = 1, 2, \ldots, r$ where each site may exist in a finite number of discrete states $k_\mu = 1, 2, \ldots, k_\mu$. The permitted configurations for this example can be labeled by $K = \{k_1, k_2, \ldots, k_r\}$ where restrictions in the physics of the system may prevent certain combinations of the k_μ from occurring.

If there exists a set of properties (operators in quantum mechanics) $\{\lambda^*_\alpha, \alpha = 1, 2, \ldots, d\}$ of interest, one can arrange these into a column vector $\boldsymbol{\lambda}^*$. Further, if $\boldsymbol{\lambda}_K$ denotes the value of $\boldsymbol{\lambda}^*$ in configuration K, a matrix $\boldsymbol{\Lambda}$ may be formed with the $\boldsymbol{\lambda}_K$ as columns.

In an ensemble each configuration will occur with probability w_K determined by the physical laws governing the system. These probabilities determine the ensemble average $\boldsymbol{\lambda}$ of $\boldsymbol{\lambda}^*$ according to the rule

$$\lambda_\alpha = \sum_K w_K \lambda_{\alpha,K} \tag{1}$$

or

$$\boldsymbol{\lambda} = \boldsymbol{\Lambda}\mathbf{w}, \tag{2}$$

where $\mathbf{w}$ is a column vector of the w_K.

The problem to which this paper is addressed may now be stated as follows. What, if anything, can be said about $\boldsymbol{\lambda}$ independent of any knowledge of the w_K? Or more precisely, what are the limitations on $\boldsymbol{\lambda}$ so that it is compatible with Eq. (2) subject only to the conditions

$$w_K \geq 0 \tag{3}$$

and

$$\sum w_K = 1\,? \tag{4}$$

One restriction is immediately apparent. Since λ_α is an average of the $\lambda_{\alpha,K}$,

$$\min_K\{\lambda_{\alpha,K}\} \leq \lambda_\alpha \leq \max_K\{\lambda_{\alpha,K}\}. \tag{5}$$

That there are usually other restrictions is also apparent from the consideration that, if λ_1 is at its maximum value, then λ_2 is restricted to lie in the range spanned by the $\lambda_{2,K}$ for which $\lambda_{1,K}$ equals λ_1. Two properties would be independent only if every value of one property occurred for each value of the other.

REDUCTION TO STANDARD FORM

The problem, as stated, is well known in the theory of convex sets.[1] No general solution is possible; but there are several algorithms available which will solve any specific case of interest. In order to use these algorithms, however, a slight modification of $\boldsymbol{\Lambda}$ is usually necessary.

The set of values of $\boldsymbol{\lambda}$ defined by (2) lie in a d-dimensional convex polytope. That is, this set is convex so that if $\boldsymbol{\lambda}_a$ and $\boldsymbol{\lambda}_b$ are two possible values of $\boldsymbol{\lambda}$, then every $\boldsymbol{\lambda}$ of the form

$$\boldsymbol{\lambda} = a\boldsymbol{\lambda}_a + b\boldsymbol{\lambda}_b$$

with $a, b \geq 0$ and $a + b = 1$ is also a possible value of $\boldsymbol{\lambda}$. Moreover, because this convex set is the closed convex hull of the $\boldsymbol{\lambda}_K$ which are finite in number, it is a polytope whose vertices are a subset of the $\boldsymbol{\lambda}_K$.

Now suppose an additional variable λ_{d+1} is introduced in the $\boldsymbol{\lambda}$ vector with $\lambda_{d+1,K} = 1$ for all K. Then the condition $\sum w_K = 1$ is equivalent to

$$\lambda_{d+1} = 1. \tag{6}$$

Hence the convex set of interest may be regarded as the intersection of the convex polyhedral cone defined by

$$\boldsymbol{\Lambda}\mathbf{w} = \boldsymbol{\lambda}, \tag{7}$$

$$\mathbf{w} \geq 0 \tag{8}$$

(where $\mathbf{w} \geq 0$ means every $w_K \geq 0$) in a $(d + 1)$-dimensional space with the hyperplane $\lambda_{d+1} = 1$. Since $\boldsymbol{\lambda} = 0$ does not lie in the hyperplane $\lambda_{d+1} = 1$, this convex cone is pointed with apex at the origin.[2]

The condition that $\boldsymbol{\lambda}$ lies in the desired convex cone may now be expressed as a set of linear inequalities by a well-known theorem from linear programming[1] which states that there exists a nonnegative solution $\mathbf{w}$ satisfying

$$\boldsymbol{\Lambda}\mathbf{w} = \boldsymbol{\lambda} \tag{9}$$

if and only if for every $\mathbf{y}$ satisfying

$$\mathbf{y}^T\boldsymbol{\Lambda} \geq 0, \tag{10}$$

necessarily

$$\mathbf{y}^T\boldsymbol{\lambda} \geq 0. \tag{11}$$

The necessity of this condition is obvious since multiplication of (7) by $\mathbf{y}^T$ gives

$$\mathbf{y}^T\boldsymbol{\Lambda}\mathbf{w} = \mathbf{y}^T\boldsymbol{\lambda},$$

and $\mathbf{y}^T\boldsymbol{\Lambda}$ and $\mathbf{w}$ are both nonnegative. Sufficiency can be established by noticing that $\mathbf{y}^T\boldsymbol{\lambda} = 0$ is the equation of a hyperplane which divides the $(d+1)$-dimensional space into two half spaces. Equations (10) and (11) then state that each $\boldsymbol{\lambda}_K$ is in the same half space with $\boldsymbol{\lambda}$. If $\boldsymbol{\lambda}$ lies outside the cone defined by the $\boldsymbol{\lambda}_K$, then it is always possible to construct a hyperplane with all of the $\boldsymbol{\lambda}_K$ on one side and $\boldsymbol{\lambda}$ on the other. The $\mathbf{y}$ defining this plane would then satisfy $\mathbf{y}^T\boldsymbol{\Lambda} \geq 0$ and $\mathbf{y}^T\boldsymbol{\lambda} < 0$ contrary to the hypothesis.

If a $\mathbf{y}$ exists such that $\mathbf{y}^T\boldsymbol{\Lambda} = 0$, then both $\mathbf{y}$ and $-\mathbf{y}$ satisfy (10). Equation (11) then becomes

$$\mathbf{y}^T\boldsymbol{\lambda} = 0. \tag{12}$$

This is equivalent to saying that if

$$\sum y_\alpha \lambda_{\alpha,K} = 0$$

for all K, then the ensemble average must also satisfy this equality. Thus, if some of the λ^*_α are linearly dependent, they may be eliminated since the corresponding λ_α must be given by a known linear combination of the independent λ's. Let us assume, henceforth that this reduction has been made. If Λ was originally of rank s, then $d+1-s$ of the λ_α can be eliminated. After the reduction the dimension of λ is s, and the number $\mathcal{K}$ of configurations K is not less than s. Further, at least one subset consisting of s of the $\mathcal{K}$ columns of $\boldsymbol{\Lambda}$ are linearly independent, and one such subset may be assumed to be the first s columns of Λ. The problem is thus reduced to the form

$$\boldsymbol{\Lambda}\mathbf{w} = \boldsymbol{\lambda} \tag{13}$$

where $\boldsymbol{\Lambda}$ is an $s \times \mathcal{K}$ matrix ($\mathcal{K} \geq s$) and the first s columns of $\boldsymbol{\Lambda}$ are linearly independent. The convex cone of these $\boldsymbol{\lambda}$ with apex at the origin is pointed and of full dimension (i.e., it occupies a nonzero volume in this s-dimensional space). Even though many of the $\boldsymbol{\lambda}_K$ may be interior to this cone and, hence, not essential to its definition, no method exists for elimination of these $\boldsymbol{\lambda}_K$ in advance of solving the problem.

ALGORITHMS

The actual algorithm used for the calculations presented here will be published separately. The general considerations involved in constructing an algorithm are of some interest, however.

If two vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ satisfy (10), then so does $a_1\mathbf{y}_1 + a_2\mathbf{y}_2$ with $a_1, a_2 \geq 0$. Hence the vectors y also form a convex polyhedral cone (called the polar to the λ-cone). Since $\mathbf{y}_1^T\boldsymbol{\lambda} \geq 0$ and $\mathbf{y}_2^T\boldsymbol{\lambda} \geq 0$ imply $(a_1\mathbf{y}_1 + a_2\mathbf{y}_2)^T\boldsymbol{\lambda} \geq 0$, only the extreme rays of the y cone generate nonredundant inequalities (extreme rays may be defined by the condition that they are not expressible as nonnegative linear combinations of other rays of the y cone). Hence a complete set of nonredundant conditions on $\boldsymbol{\lambda}$ can be found by computing the extreme rays of the y cone.

It is easily seen that the extreme rays of the y cone are normal to the $(s-1)$-dimensional faces, i.e., facets, of the λ cone. In other words, there is one nonredundant inequality corresponding to each facet of the λ cone. The inequality associated with a particular facet simply says that $\boldsymbol{\lambda}$ lies on the same side of that facet as do all of the $\boldsymbol{\lambda}_K$. Thus this procedure restricts $\boldsymbol{\lambda}$ by requiring that it be interior to all of the facets of the λ cone.

A heuristic algorithm for finding all of the facets of the λ cone would be to construct the hyperplanes defined by each $s-1$ independent columns of $\boldsymbol{\Lambda}$ together with the origin. Even though the intersection of most such hyperplanes with the λ cone would be interior to this convex cone, a search through all $\binom{\mathcal{K}}{s-1}$ such hyperplanes would produce all the facets. Polytopes, and thus convex polyhedral cones, are known with the number of facets as low as s [$(s-1)$ − simplices] to as large as $\sim 2\binom{\mathcal{K}-s/2-1}{s/2}$ for cyclic polytopes.[2] If the $\boldsymbol{\lambda}_K$ are chosen as random numbers, then almost all cones have $2 + (s-2)(\mathcal{K}-s+1)$ facets. Since the number of facets is clearly a negligible fraction of the number of $s-1$ subsets of the columns of $\boldsymbol{\Lambda}$, this algorithm is not feasible.

Viewing the extreme rays of the y cone as normals to the facets of the λ cone suggests that the column vector $\boldsymbol{\eta} = \boldsymbol{\Lambda}^T\mathbf{y}$ should be given further consideration since $\boldsymbol{\eta} \geq 0$ and $\eta_K = 0$ if the facet contains $\boldsymbol{\lambda}_K$. It is easily established that $\mathbf{y}$ is extreme if and only if there exists no other $\mathbf{y}'$ such that $\{i \mid \eta_i = 0\} \subset \{i \mid \eta'_i = 0\}$. That is, $\mathbf{y}$ is extreme if the set $Z = \{i \mid \eta_i = 0\}$ is maximal.[1,3-5]

A variant of the double-description algorithm based on the maximal property of Z is the most efficient known algorithm for constructing the facets.[4,5] This method proceeds by calculating the η corresponding to extreme $\mathbf{y}$ rays. The main defect of this algorithm is that it generates all of the η vectors simultaneously so that the memory capacity of computers becomes a severe limitation when the number of extreme rays exceeds 10^5. Also, any symmetry present in $\boldsymbol{\Lambda}$ cannot be used to simplify the problem.

A third algorithm which allows sequential calculation of the facets of the λ cone has been devised.[6,7] This is based on the fact that each $(s-2)$-dimensional "edge" of a facet is common to exactly one adjacent facet. Rotation around the edges of a facet will produce a list of adjacent facets. Rotation around the edges of these will produce still more facets. Continuation of this process eventually yields all facets. While this process is redundant since each facet will be generated as many times as it has adjacent facets, the algorithm has the advantage that the facets are found sequentially so that some can be found even if all cannot. Also, if symmetry is present so that facets fall into equivalence classes, only adjacent facets to one facet from each equivalence class need be found in order to guarantee finding representative facets from all equivalence classes.

APPLICATIONS

A. The N-Representability Problem for Fermions

A wavefunction for a system of N indistinguishable fermions may be approximated by a finite linear combination of Slater determinants. These determinants

may be represented in terms of a set of r orthonormal spin-orbitals as

$$\phi_K = (N!)^{-1/2}\det\{\phi_{k_1}\ \phi_{k_2}\ \cdots\ \phi_{k_N}\}, \tag{14}$$

where

$$K = \{1 \le k_1 < k_2 \cdots < k_N \le r\}$$

or as

$$\phi_K = |n_1 n_2 \cdots n_r\rangle, \tag{15}$$

where $n_i = 1$ if $i \in K$ and 0 otherwise. The reduced second-order density matrix Γ_ψ and first-order density ρ_ψ for a wavefunction ψ are defined by

$$\Gamma_\psi(1,2;1',2') = \binom{N}{2}\int \psi(1,2,3,\cdots,N) \times \psi(1',2',3,\ldots,N)dX_3\cdots dX_N, \tag{16}$$

$$\rho_\psi = 2(N-1)^{-1}\int \Gamma_\psi(1,2;1',2)dX_2, \tag{17}$$

where

$$\psi = \sum C_K\phi_K \tag{18}$$

and

$$\sum |C_K|^2 = 1. \tag{19}$$

If the Hamiltonian for the system is a linear combination of only one-body and two-body operators, the energy is easily expressed in terms of Γ_ψ. If the boundary conditions for Γ_ψ were known, an arbitrary function Γ obeying these boundary conditions could be substituted into the energy expression and varied to obtain an upper bound on the ground state energy. The problem of determining these boundary conditions is known as the pure-state N-representability problem.

Less stringent boundary conditions result from considering an ensemble rather than pure states. For

$$H = \sum_i h(i) + \sum_{i<j} g(i,j)$$

and

$$G(i,j) = g(i,j) + (N-1)^{-1}[h(i) + h(j)],$$

it is well known that

$$\overline{E}_\psi = \langle\psi|H|\psi\rangle$$

becomes

$$\overline{E}_\psi = \mathrm{Tr}G\Gamma_\psi = \int\int (G\Gamma_\psi)_{\substack{1=1'\\2=2'}} dX_1dX_2.$$

The ensemble average of E over states of the same N gives

$$\overline{E}_N = \sum_\psi w_\psi\langle\psi|H|\psi\rangle,$$

where the probability of finding the system in state ψ is w_ψ ($0 \le w_\psi \le 1; \sum_\psi w_\psi = 1$). In terms of the ensemble density,

$$\Gamma_N = \sum_\psi w_\psi\Gamma_\psi; \tag{20}$$

this energy may be expressed as

$$\overline{E}_N = \mathrm{Tr}G\Gamma_N.$$

Since the minimum of $\overline{E}_N$ is the same as the minimum of $\overline{E}_\psi$ (both equal the lowest eigenvalue of H), it suffices to restrict a trial Γ to obey ensemble conditions in calculating the ground state energy. These boundary conditions are known as the N-representability conditions.[8,9]

A further generalization is possible if $\overline{E}_\psi$ is expressed as

$$\overline{E}_\psi = \mathrm{Tr}g\Gamma_\psi + \mathrm{Tr}h\rho_\psi.$$

Since this expression has no explicit N dependence, the ensemble average can be carried out over wavefunctions of variable N. This gives

$$\overline{E} = \mathrm{Tr}g\Gamma + \mathrm{Tr}h\rho,$$

where the sums

$$\Gamma = \sum_\psi w_\psi\Gamma_\psi, \tag{21}$$

$$\rho = \sum_\psi w_\psi\rho_\psi, \tag{22}$$

extend over wavefunctions of differing values of N. For some systems the minimum of $\overline{E}$ will coincide with the desired minimum of $\overline{E}_\psi$. In every case the family of functions satsifying the representability conditions for variable-N ensembles contains the fixed-N ensemble and pure-state densities.

The density matrices for the form (18) may all be expressed in the form

$$\Gamma = \sum_{i<j}\sum_{k<l}\Gamma_{ijkl}\phi_{ij}(1,2)\phi^*_{kl}(1',2'), \tag{23}$$

where

$$\phi_{ij}(1,2) = 2^{-1/2}[\phi_i(1)\phi_j(2) - \phi_j(1)\phi_i(2)]. \tag{24}$$

For Γ_ψ the Γ_{ijkl} can be evaluated as

$$\Gamma_{ijkl} = \sum_K\sum_J C_KC^*_J\epsilon^{(K)}_{i,j,(K-i-j)}\epsilon^{k,l(J-k-l)}_{(J)}\epsilon^{(K-i-j)}_{(J-k-l)}, \tag{25}$$

where $\epsilon^{k_1\cdots k_N}_{i_1\cdots i_N}$ is the generalized Kronecker delta defined as 0 if the set of subscripts differ from the set of superscripts and as $(-1)^P$ if P transpositions are required to arrange the two sets in identical order.

Alternatively, for Γ_ψ,

$$\Gamma_{ijkl} = \langle\psi|a_k^\dagger a_l^\dagger a_j a_i|\psi\rangle,$$

where a_i is the usual annihilation operator associated with orbital i. The representability of a density matrix clearly depends only on the Γ_{ijkl} rather than on any particular properties of ϕ_{ij}. Since the Γ_{ijkl} are expressible as average values of the operators $a_k^\dagger a_l^\dagger a_j a_i$, the formalism introduced previously applies to these quantities. Hence, the Γ_{ijkl} form a convex set whose interior may be described by a set of inequalities.

The difficulty with finding the boundary conditions on the Γ_{ijkl} in this formulation arise because the "con-

figurations" in this ensemble are the wavefunctions ψ and, hence, are nondenumerable. Actually, from the introductory discussion it is clear that only the extreme rays λ_K of the λ cone need be denumerable (provided some way were available of enumerating just these and omitting all λ_K lying in the interior of the cone). Enough progress has been made in enumerating these extreme rays to show that they are nondenumerable.[10]

Nevertheless some progress can be made. For every Hermitian operator M,

$$M = \sum_{ijkl} m_{ijkl} a_k^\dagger a_l^\dagger a_j a_i, \tag{26}$$

with maximum eigenvalue $m_>$ and minimum $m_<$,

$$m_< \leq \mathrm{Tr} M\Gamma \leq m_>$$

gives two linear constraints on the Γ_{ijkl}. For example, for $ij \neq kl$,

$$M = \tfrac{1}{2}(a_k^\dagger a_l^\dagger a_j a_i + a_i^\dagger a_j^\dagger a_l a_k)$$

has eigenfunctions

(i) $\phi_K \pm \phi_J$, $\quad ij \in K, kl \notin K, J = \{K - i - j + k + l\}$

and

(ii) Φ_L, $\quad L$ not appearing as J or K in (i).

Some eigenfunctions of type (i) have eigenvalues $\pm \frac{1}{2}$ while the rest of these and all of type (ii) have eigenvalue 0. Hence, the real part of Γ_{ijkl}, $\mathrm{Re}(\Gamma_{ijkl})$, is bounded by

$$-\tfrac{1}{2} \leq \mathrm{Re}(\Gamma_{ijkl}) \leq \tfrac{1}{2}. \tag{27}$$

Similarly

$$M = \tfrac{1}{2}i(a_k^\dagger a_l^\dagger a_j a_i - a_i^\dagger a_j^\dagger a_l a_k)$$

gives a bound on the imaginary part of Γ_{ijkl},

$$-\tfrac{1}{2} \leq \mathrm{Im}(\Gamma_{ijkl}) \leq \tfrac{1}{2}. \tag{28}$$

For $ij = kl$, $a_k^\dagger a_l^\dagger a_j a_i$ becomes a product of number operators $\eta_i = a_i^\dagger a_i$. Clearly $M = \eta_j \eta_i$ has eigenvalues 1 and 0, so that

$$0 \leq \Gamma_{ijij} \leq 1. \tag{29}$$

The matrix elements Γ_{ijkl} of Γ_N may be shown to be bounded collectively as well as individually. If one defines

$$\Gamma_N^\dagger \cdot \Gamma_N = \int \Gamma_N^*(1,2;1''2'')\Gamma_N(1'',2'';1'2')dX_1'' dX_2'',$$

then

$$\Gamma_N^\dagger \cdot \Gamma_N = \sum_{ijkl} Q_{ijkl}\, \phi_{ij}(1,2)\phi_{k,l}^*(1',2'),$$

where

$$Q_{ijkl} = \sum_{pq} \Gamma_{ijpq}^* \Gamma_{pqkl}.$$

Hence

$$\mathrm{Tr}(\Gamma_N^\dagger \cdot \Gamma_N) = \sum_{ijkl} |\Gamma_{ijkl}|^2.$$

But Γ_N may be regarded as the kernel of an Hermitian integral operator.[11] This allows it to be written in diagonal form as

$$\Gamma_N = \sum \gamma_i f_i(1,2) f_i^*(1',2'), \tag{30}$$

where the γ_i are its eigenvalues and the f_i are its eigenfunctions. Since Γ_N is positive definite by definition, $\gamma_i \geq 0$.

Also

$$\mathrm{Tr}\Gamma_N = \sum \gamma_i = \binom{N}{2}. \tag{31}$$

Then

$$\begin{aligned}\mathrm{Tr}\Gamma_N^\dagger \cdot \Gamma_N &= \sum \gamma_i^2 \\ &= \Big(\sum \gamma_i\Big)^2 - \sum_{i \neq j} \gamma_i \gamma_j \\ &\leq \Big(\sum \gamma_i\Big)^2.\end{aligned}$$

Hence

$$\sum_{ijkl} |\Gamma_{ijkl}|^2 \leq \binom{N}{2}^2. \tag{32}$$

If the fact that $\gamma \leq N$ is taken into account,[12] a stronger result

$$\sum_{ijkl} |\Gamma_{ijkl}|^2 \leq N\binom{N}{2}. \tag{33}$$

is easily derived. Thus if the Γ_{ijkl} are regarded as coordinates in a finite-dimensional unitary space, Γ_N is a bounded convex set which may be taken as closed. Moreover, this bound is independent of the dimension $(\frac{r}{2})^2$ of the space. The variable-N ensemble Γ is unbounded, however, in the limit of large r. Since for fixed finite r, $N \leq r$ for all states in the ensemble, the varaible-N ensemble is bounded by

$$\sum_{ijkl} |\Gamma_{ijkl}|^2 \leq r\binom{r}{2}, \tag{34}$$

There is a close connection between the λ cone and the operator

$$M = \sum m_\alpha \lambda_\alpha^*. \tag{35}$$

In fact the conditions

$$\sum m_\alpha \lambda_\alpha \geq m_<$$

for all m_α generate the complete set of inequalities defining the cone.[3] Further, the extreme rays of the λ cone are contained in the family of eigenfunctions of M corresponding to nondegenerate minimum eigenvalues for all possible choices of m_α.

For the M of Eq. (26) this family of eigenfunctions is nondenumerable. Some progress in finding necessary conditions in the Γ_{ijkl} can be made, however, if a restricted form of M can be found with the property that the family of its nondegenerate ground state eigenfunctions is denumerable. Trivial examples of this were used previously to find bounds on Γ_{ijkl}.

The most interesting restricted form of M found to date is the so-called Slater hull.[8] If M is restricted to the form (where $\eta_i = a_i^\dagger a_i$)

$$M = \sum m_{ij}\eta_i\eta_j, \tag{36}$$

the ϕ_K themselves form the whole set of nondegenerate ground state eigenfunctions. Thus the λ cone with $\lambda_{ij} = \Gamma_{ijij} = \langle \eta_i\eta_j \rangle$ has as its extreme rays

$$\lambda_{ij,K} = \langle \phi_K | \eta_i\eta_j | \phi_K \rangle. \tag{37}$$

The algorithm described previously can be used to generate explicit conditions of the form

$$\sum_{i<j} y_{ij}\lambda_{ij} \geq 0, \tag{38}$$

$$\sum \lambda_{ij} = \binom{N}{2} \tag{39}$$

from this $\mathbf{\Lambda}$ matrix.

The relation between the Slater hull $\{\Gamma_S\}$ and $\{\Gamma_N\}$ and $\{\Gamma_\psi\}$ may be restated as follows. For any set of $\lambda_{ij} = \Gamma_{ijij}$ which obey the Slate hull boundary conditions (38), it is possible to construct an element of $\{\Gamma_N\}$ and $\{\Gamma_\psi\}$ with these values of λ_{ij}. Conversely, all elements of $\{\Gamma_N\}$ and $\{\Gamma_\psi\}$ have diagonal elements which obey the Slater hull conditions. Thus the inequalities (38) contain all of the restrictions on the λ_{ij} which can be stated without reference to the off-diagonal elements Γ_{ijkl}.

These assertions are easily demonstrated.[13] By definition

$$\Gamma_S = \sum w_K \Gamma_K, \tag{40}$$

where Γ_K is the density matrix of ϕ_K. Since the inequalities (37) guarantee that for the given λ_{ij} a set of $w_K \geq 0$ do exist and (39) guarantees $\sum w_K = 1$, any solution to (38) and (39) may be written in the form (40). But a Γ_S of this form is by definition an element of $\{\Gamma_N\}$. Further, all wavefunctions of the form

$$\psi = \sum \sqrt{w_K}\, e^{i\alpha K}\phi_K,$$

where the α_K are real arbitrary numbers give

$$\Gamma_{ijij} = \sum w_K \lambda_{ij,K} = \lambda_{ij}.$$

Conversely, if $\psi = \sum C_K \phi_K$, then

$$\Gamma_{ijij} = \sum |C_K|^2 \lambda_{ij,K}.$$

If $w_K = |C_K|^2$, then Γ_S will have the same diagonal elements as Γ_ψ. Similarly, if

$$\Gamma_N = \sum_\psi w_\psi \Gamma_\psi,$$

then

$$\Gamma_{ijij} = \sum_\psi w_\psi \sum_K |C_{K,\psi}|^2 \lambda_{ij,K} = \sum_K w_K \lambda_{ij,K},$$

where $w_K = \sum_\psi w_\psi |C_{K,\psi}|^2$. Since the w_K obey the definitions of probabilities, $\Gamma_S = \sum w_K \Gamma_K$ will have the same diagonal elements as Γ_N.

In other words, while $\{\Gamma_N\}$, $\{\Gamma_S\}$, and $\{\Gamma_\psi\}$ differ in the full $\binom{r}{2}^2$-dimensional space of the Γ_{ijkl}, the projection of each of these onto the $\binom{r}{2}$-dimensional space of the Γ_{ijij} is identical. Further, a unitary transformation among the basis spin-orbitals ϕ_i generates a rotation of axes in the Γ_{ijkl} space.[14] Such a transformation defines a new $\{\Gamma'_S\}$ whose projection onto the Γ'_{ijij} space coincides with the projections of $\{\Gamma_N\}$ and $\{\Gamma_\psi\}$. Such transformations could be used to generate additional inequalities for the elements of Γ_N and Γ_ψ which would involve their off-diagonal elements.

A special case of such transformations is permutation of the orbitals. Since a permutation merely relabels the orbitals, the inequalities (38) must transform into themselves under permutations of the orbital labels. This may be stated more mathematically by noting that a permutation $\mathcal{P}$ on the orbitals ϕ_i generates a permutation P on the ϕ_{ij} and a permutation Q on the ϕ_K. Hence, $\mathbf{P\Lambda} = \mathbf{\Lambda Q}$. If

$$\mathbf{\Lambda}^T\mathbf{y} = \boldsymbol{\eta}$$

then

$$\mathbf{Q}^T\mathbf{\Lambda}^T\mathbf{y} = \mathbf{Q}^T\boldsymbol{\eta} \quad \text{or} \quad \mathbf{\Lambda}^T(\mathbf{P}^T\mathbf{y}) = (\mathbf{Q}^T\boldsymbol{\eta}).$$

$$\text{Hence} \quad \mathbf{y}' = \mathbf{P}^T\mathbf{y} \quad \text{satisfies} \quad \mathbf{\Lambda}^T\mathbf{y}' = \boldsymbol{\eta}', \tag{41}$$

$$\text{where} \quad \boldsymbol{\eta}' = \mathbf{Q}^T\boldsymbol{\eta}.$$

Clearly if $\mathbf{y}$ is normal to a facet of the λ cone, then so is $\mathbf{y}'$ since $\boldsymbol{\eta}'$ will also have an extremal set of zero elements.

Thus for each of the $r!$ permutations of the permutation group of r objects, the inequality generated from $\mathbf{P}^T\mathbf{y}$ will be included in (38) if $\mathbf{y}$ is included. The vectors $\mathbf{y}$ may, therefore, be grouped into equivalence classes[15] $\{\mathbf{y}_i | \mathbf{P}_i{}^T\mathbf{y} = \mathbf{y}_i\}$ under $\mathcal{P}_r$. The dimension of each class will be no larger than $r!$. Geometrically this corresponds to the fact that in the Slater hull each direction in space is equivalent as are all extreme rays of the cone. The facets of the cone are not all alike, however, and the number of different kinds of facets is the number of equivalence classes.

A further simplification can be made by noting that the configurations for N particles in r orbitals are in one-to-one correspondence with the configurations for $r-N$ particles in r orbitals. This correspondence is most naturally established by the particle-hole transformation which associates

$$\phi_{K,N} = |n_1 n_2 \cdots n_r\rangle$$

with

$$\phi_{K,r-N} = |1-n_1, 1-n_2, \cdots, 1-n_r\rangle.$$

If for the $r-N$ case one examines the properties

$$\lambda'_{ij} = \langle (1-\eta_i)(1-\eta_j) \rangle \tag{42}$$

(the pair occupation number for holes), then

$$\begin{aligned}\lambda'_{ij,K,r-N} &= \langle K, r-N | (1-\eta_i)(1-\eta_j) | K, r-N \rangle \\ &= \langle K, r | \eta_i\eta_j | K, r \rangle \\ &= \lambda_{ij,K,N}\end{aligned} \tag{43}$$

so that Λ'_{r-N} is identical to Λ_N. Thus the inequalities

$$\sum_{i<j} y_{ij,N}\lambda_{ij,N} \geq 0$$

must become

$$\sum_{i<j} y_{ij,N}\lambda'_{ij,r-N} \geq 0 \tag{44}$$

under this transformation. Equation (44) may be rewritten in the standard form

$$\sum_{i<j} y_{ij,r-N}\lambda_{ij,r-N} \geq 0 \tag{45}$$

by the substitutions

$$\begin{aligned} M(M-1)\lambda'_{ij,M} &= \langle M(M-1)\rangle - M\langle (M-1)(\eta_i+\eta_j)\rangle \\ &\quad + M(M-1)\langle \eta_i\eta_j\rangle, \\ \eta &= \sum_i \eta_i, \end{aligned} \tag{46}$$

$$\begin{aligned} M(M-1) &= \langle \eta(\eta-1)\rangle \\ &= 2\sum_{k<l}\langle \eta_k\eta_l\rangle, \end{aligned}$$

or

$$\langle M(M-1)\rangle = 2\sum_{k<l}\lambda_{kl,M}, \tag{47}$$

$$\begin{aligned} \langle (M-1)\eta_i\rangle &= \langle (\eta-1)\eta_i\rangle, \\ &= \sum_{k\neq i}\langle \eta_k\eta_i\rangle, \end{aligned}$$

or

$$\langle (M-1)\eta_i\rangle = \sum_{k<l}\lambda_{ki,M} + \sum_{k>i}\lambda_{ik,M}. \tag{48}$$

Thus the inequalities for $r-N$ particles may be easily derived from those for N particles.

B. Results for the Slater Hull Problem

The matrix elements of Λ for the Slater hull as given by Eq. (37) are easily evaluated by inspection. The $\binom{r}{N}$ configurations may be ordered by defining $K \leq L$ provided that the first element of $k_1 < k_2 \cdots < k_N$ which differs from $l_1 < l_2 \cdots < l_N$ satisfies $k_p < l_p$.

Thus for three particles in four orbitals, the configurations can be ordered as $|1110\rangle$, $|1101\rangle$, $|1001\rangle$, $|0111\rangle$. Similarly the λ_{ij} can be ordered listing λ_{ij} before λ_{kl} if $i < k$ or if $i = k$ and $j < l$. Thus for $r = 4$, the order would be

$$(i,j) = (1,2)\ (1,3)\ (1,4)\ (2,3)\ (2,4)\ (3,4).$$

Then for $r = 4$ and $N = 3$,

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

All other values of r and N may be handled just as simply.

For $N = 0$ or 1, $\Lambda \equiv 0$ and reduction to standard form gives one independent variable (from λ_{d+1}) with the results

$$\lambda_{ij} = 0. \tag{49}$$

For these two cases the variables

$$\lambda_i = \langle \eta_i\rangle = \rho_{ii}$$

cannot be derived from the λ_{ij} by Eq. (48). For larger values of N, Eq. (48) shows that λ_i are linearly dependent on the λ_{ij}. Hence they need not be included in forming Λ since they would be eliminated in the reduction to standard form by the equality

$$(N-1)\lambda_i = \sum_{k<i}\lambda_{ki} + \sum_{k>i}\lambda_{ik}.$$

Also for $N \geq 2$,

$$\lambda_{d+1} = 1 = \binom{N}{2}^{-1}\sum_{k<l}\lambda_{kl}$$

so the supplementary variable λ_{d+1} is not needed.

For $N = 0$, inclusion of the λ_i in Λ still gives $\Lambda = 0$. Reduction to standard form gives

$$\lambda_i = 0 \tag{50}$$

as well as Eq. (49). Mapping of the $N = 0$ case into the $N' = r - N = r$ case by the particle–hole transformation gives

$$\lambda'_{ij,r} = 1 - \lambda_{i,r} - \lambda_{j,r} + \lambda_{ij,r} = 0, \tag{51}$$

$$\lambda'_{i,r} = 1 - \lambda_{i,r} = 0. \tag{52}$$

Solving these equations simultaneously gives the expected result

$$\lambda_{i,r} = 1, \tag{53}$$

$$\lambda_{ij,r} = 1. \tag{54}$$

For $N = 1$, inclusion of the λ_i in Λ gives $\Lambda = (\frac{1}{0})$. Since $\lambda_{d+1} = 1 = \sum\lambda_i$ no supplementary variable need be included in reduction to standard form. Elimination of the linearly dependent variables λ_{ij} gives Eq. (49). The reduced Λ matrix $\Lambda = 1$ then gives the inequalities

$$\lambda_{i,1} \geq 0 \tag{55}$$

with trace condition

$$\sum\lambda_{i,1} = 1.$$

Mapping these into $N' = r - 1$ gives

$$1 - \lambda_{i,r-1} - \lambda_{j,r-1} + \lambda_{ij,r-1} = 0 \tag{56}$$

and

$$1 - \lambda_{i,r-1} \geq 0 \tag{57}$$

which could be rewritten in standard form using (47) and (48).

For $N = 2$, $\Lambda = 1$, and the inequalities become simply

$$\lambda_{ij,2} \geq 0. \tag{58}$$

Mapping Eq. (58) into $N' = r - 2$ gives

$$1 - \lambda_{i,r-2} - \lambda_{j,r-2} + \lambda_{ij,r-2} \geq 0. \tag{59}$$

This result (59) was previously derived by Yoseloff and Kuhn by a much more elaborate procedure.[16]

For $r - 3 \geq N \geq 3$, Λ has more columns than rows. Further, the rank $R(\Lambda)$ of Λ is $\binom{r}{2}$ so no equalities among the λ_{ij} exist.

The fact that $R(\Lambda)$ is maximal in these cases may easily be shown by induction on r. For $r = N + 2$, the induction hypothesis is certainly true because, as has been indicated, $\Lambda_{N+2,N}$ is the particle–hole mapping of the identity matrix of dimension $\binom{N+2}{2}$. It may be shown that (with a slight renumbering of the rows and columns)

$$\Lambda_{r,N} = \begin{pmatrix} \Lambda_{r-1,N} & \Lambda_{r-1,N-1} \\ 0 & \theta_{r-1,N-1} \end{pmatrix}, \tag{60}$$

where θ is the Λ matrix for the variables λ_i. Because $R(\Lambda_{r-1,N})$ is assumed to be maximal, it follows that

$$\begin{aligned} R(\Lambda_{r,N}) &= R(\Lambda_{r-1,N}) + R(\theta_{r-1,N-1}) \\ &= \binom{r-1}{2} + R(\theta_{r-1,N-1}). \end{aligned}$$

But

$$\theta_{S,M} = \begin{pmatrix} \theta_{S-1,M} & \theta_{S-1,M-1} \\ 0 & 1, 1, \ldots, 1 \end{pmatrix}.$$

Observing that $\theta_{M+1,M}$ is the particle–hole transformation of an identity matrix and thus of maximum rank $M + 1$, a secondary inductive argument yields that $R(\theta_{S,M}) = S$. Thus

$$\begin{aligned} R(\Lambda_{r,N}) &= \binom{r-1}{2} + r - 1 \\ &= \binom{r}{2}. \end{aligned}$$

No analytic solution for this general case of $N \geq 3$ is known although the form (60) of Λ suggests that some form of recurrence relation should exist. Some specific examples of Λ have been solved, however. Kuhn has published[17] a solution for $N = 3, r = 6$. More recently Yoseloff[13] has solved the $N = 3, r = 7$ case and has obtained 19 equivalence classes for $N = 3$, $r = 8$.

By application of the algorithm outlined we have verified the previous results for $N = 3, r = 6, 7, 8$ and have obtained partial solutions to the cases $N = 3$, $r = 9$ and $N = 4, r = 8, 9$. The particle–hole transformation was applied to obtain results for $N = 4$, $r = 7$; $N = 5, r = 8, 9$; and $N = 6, r = 9$. The number of equivalence classes and inequalities found for each of these cases is shown in Table I. Question marks in Part B of this table indicate entries believed to be correct even though the calculation could not be completed. For each of these cases the number of equivalence classes was not increased by the last few facets scanned for neighbors. For example, in the $N = 3, r = 9$ case the entire list of 143 classes was found by computing neighbors to the first 58. Scans were completed for 66 additional facets without producing any new ones. For the $N = 4, r = 9$ case no claim is made to completeness as only 195 of the 1 089 classes were scanned.

Part A of Table I is included to show the great reduction which arises from use of permutational symmetry. For the larger values of r and N, not only do the number of equivalence classes increase rapidly; but the fraction of them which have $r!$ distinct inequalities in the class also increases. Part A also emphasizes the essential futility, in hindsight, of this approach. It clearly would be impossible to use the 2×10^8 inequalities from the $r = 9, N = 4$ case for any purpose.

TABLE I. Number of facets for the fermion fixed-N Slater hull.

A. Total number of facets.

r/N	2	3	4	5	6	7
4	6					
5	10	10				
6	15	70	15			
7	21	896	896	21		
8	28	5.2×10^4	5.8×10^5	5.2×10^4	28	
9	36	1.2×10^7	2.3×10^8	2.3×10^8	1.2×10^7	36

B. Number of equivalence classes.

r/N	2	3	4	5	6	7
4	1					
5	1	1				
6	1	4	1			
7	1	7	7	1		
8	1	19?	57?	19?	1	
9	1	143?	> 1089	> 1089	143?	1

TABLE II. Slater hull equivalence classes for $N = 4, r = 7$.

$$\begin{gathered} \lambda_{12} \geq 0 \\ \lambda_1 - \lambda_{12} \geq 0 \\ 1 - \lambda_1 - \lambda_2 + \lambda_{12} \geq 0 \\ 1 - \lambda_1 - \lambda_2 - \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} \geq 0 \\ -1 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \lambda_{12} - \lambda_{13} - \lambda_{24} \geq 0 \\ -3 + 2\lambda_1 + 2\lambda_2 - \lambda_{12} + \lambda_{34} + \lambda_{37} + \lambda_{45} + \lambda_{56} + \lambda_{67} \geq 0 \\ \lambda_1 - \lambda_{12} - \lambda_{13} + \lambda_{23} \geq 0 \end{gathered}$$

The number of inequalities in Table I can be compared with the number expected for various special polyhedral cones. For $N = 4, r = 9$ the Slater hull cone has $\binom{9}{4} = 126$ extreme rays in an $\binom{9}{2} = 36$-dimensional space. For a completely random selection of extreme rays, one would expect ~3 100 facets (all nonequivalent under permutations). The range for special polyhedral cones would be from a low of 36 inequalities (one equivalence class, e.g., the positive orthant) if the λ_{ij} were independent properties to a

TABLE III. Factorable equivalence classes for the variable-N Slater hull.

$$\left\langle \left(\sum_{i=1}^{q} \eta_i - 1\right)\left(\sum_{i=1}^{q} \eta_i - 2\right)\right\rangle \geq 0 \qquad \text{(III. 1)}$$

(each $q = 2, 3, \cdots, r$ occur for a particular $r \geq 2$)

$$\left\langle \left(\sum_{i=1}^{q} \eta_i - 2\right)\left(\sum_{i=1}^{q} \eta_i - 3\right)\right\rangle \geq 0 \qquad \text{(III. 2)}$$

$$\left\langle \left(\sum_{1}^{q} \eta_i + \eta_1 - 2\right)\left(\sum_{1}^{q} \eta_i + \eta_1 - 3\right)\right\rangle \geq 0 \qquad \text{(III. 3)}$$

(III. 2 and III. 3 occur for each $q = 5, 6, \cdots, r$ for a particular $r \geq 5$)

$$\left\langle \left(\sum_{1}^{q} \eta_i + \eta_1 - 3\right)\left(\sum_{1}^{q} \eta_i + \eta_1 - 4\right)\right\rangle \geq 0 \qquad \text{(III. 4)}$$

$$\left\langle \left(\sum_{1}^{q} \eta_i + \eta_1 + \eta_2 - 3\right)\left(\sum_{1}^{q} \eta_i + \eta_1 + \eta_2 - 4\right)\right\rangle \geq 0 \qquad \text{(III. 5)}$$

$$\left\langle \left(\sum_{1}^{q} \eta_i + 2\eta_1 - 3\right)\left(\sum_{1}^{q} \eta_i + 2\eta_1 - 4\right)\right\rangle \geq 0 \qquad \text{(III. 6)}$$

(III. 4, III. 5, III. 6 occur for each $q = 6, 7, \cdots, r$ for a particular $r \geq 6$)

maximum of nearly 10^{20} for the cone obtained from a cyclic polytope.

One serious deficiency, beyond their complexity, with the fixed-N ensemble results is that the conditions for one value of r and N are generally not even necessary for any other values. Since every value of r, N must be treated separately, little is learned from one case which will apply to any other.

C. Variable-N Ensemble and Variable-N Slater Hull

For variable-N ensembles, the $\lambda_{ij} = \langle \eta_i \eta_j \rangle$, $\lambda_i = \langle \eta_i \rangle$, and $\lambda_0 = 1$ are linearly independent variables. Reasoning as for the fixed-N case leads one to consider the convex hull of the 2^r configurations $\phi_K = |n_1 n_2 \cdots n_r\rangle$, $n_i = 0$ or 1, since each of these ϕ_K gives a λ_K on an extreme ray of the λ cone. Clearly all Γ, Γ_N, and Γ_ψ have diagonal elements which obey the variable-N Slater hull conditions. Any set of λ_i and λ_{ij} which obey these Slater hull conditions can be used to construct an element of Γ (but not necessarily of Γ_N or Γ_ψ). Thus these conditions are somewhat weaker than for the fixed-N case.

One relation to the fixed-N case should be noted. If for fixed r, N one considers the properties $\lambda_{ij} = \langle \eta_i \eta_j \rangle$ and $\lambda_i = \langle \eta_i \rangle$ for $i < j < q$, the resulting $\mathbf{\Lambda}$ matrix will be identical to the variable-N Slater hull matrix for q orbitals provided all occupancies of the first q orbitals are possible. That is, for $q \leq N$ and $q \leq r - N$, the variable-N Slater hull conditions give all of the information which can be stated about these λ_{ij} without explicit reference to any λ_{ij} with $i > q$. As a result, it is not surprising that the variable-N inequalities for any q are found as a subset of the fixed-N inequalities for all $N \geq q$ and $r - N \geq q$.

By the reasoning used in the previous paragraph it is easy to establish that the inequalities found for one value of r in the variable-N case are at least necessary for all larger values. Further, the inequalities found so far have the property that each equivalence class of inequalities found for one value of r reappear as equivalence classes for all larger values considered.

Thus the variable-N inequalities are of considerable interest since calculation of these for one r gives a partial solution for all higher r and for some fixed-N, r cases. The solutions obtained also give necessary, even if not extreme, conditions for all density matrices.

There is an additional symmetry feature present in the variable-N case besides the obvious permutational symmetry. The change of variables produced by replacing η_i by $1 - \eta_i$ for $i \in \{i_1 \cdots i_p\}$ everywhere that these η_i appear in the definition of λ_i and λ_{ij} leaves the $\mathbf{\Lambda}$ matrix unchanged except for a permutation Q of the columns. Since these new variables λ' are just linear combinations of the previous λ, one can write $\mathbf{\Lambda}' = \mathbf{R\Lambda} = \mathbf{\Lambda Q}$. Clearly, if γ corresponds to a facet of the λ cone, so does $\mathbf{R}\gamma$ for each of the 2^r possible operators R. Hence an equivalence class may be defined as

$$\{\mathbf{y}' | \mathbf{y}' = \mathbf{RP}\gamma\},$$

where R is generated by a partial particle–hole transformation and P is generated by a permutation $\mathcal{P}$ on orbital labels. A previous paper which reported results for $r = 2, 3, 4$ overlooked this type of symmetry and included more results than necessary.[3]

A further simplification was conjectured in the previous work[3] by noting that the inequalities found for $r \leq 4$ could all be factored into a form permutationally equivalent to

$$\left\langle \left(\sum_{i=1}^{P} \eta_i - \sum_{i=p+1}^{q} \eta_i - t\right)\left(\sum_{i=1}^{p} \eta_i - \sum_{p+1}^{q} \eta_i - t - 1\right)\right\rangle \geq 0. \qquad (61)$$

It was hoped that all inequalities might factor into this form. By judicious use of partial particle–hole transformations, all expressions of this form can be changed to a form chosen as the standard representative of this equivalence class

$$\left\langle \left(\sum_{1}^{q} \eta_i - \beta\right)\left(\sum_{1}^{q} \eta_i - \beta - 1\right)\right\rangle \geq 0, \qquad (62)$$

where

$$0 \leq \beta \leq \tfrac{1}{2}(q - 1).$$

Unfortunately the conjecture of this previous paper is not true. Calculations for $r = 6$ have revealed additional factored forms as well as many inequalities which do not appear to factor. Table III gives representative inequalities from all of the factorable equivalence classes with $q = 6$. Each of these classes was found to occur for all $r \geq q$. Each of these classes, the new ones generated by placing $q = 7$ in the representative forms, and some additional new classes occur for $r = 7$. Table IV gives a list of the nonfactorable classes for $r = q = 6$. No nonfactorable classes occur for $r \leq 5$. The results shown in these tables are certainly complete through $r = 5$ and are probably complete for $r = 6$. A list which is possibly complete for $r = 7$ (100 classes) is available elsewhere.[7]

D. Pair Distributions of Particles in Lattice Sites

If N particles (or vacancies) are distributed in r lattice sites, the distribution of pairs is described by $\lambda_{ij} = \langle \eta_i \eta_j \rangle$ where the brackets imply ordinary en-

TABLE IV. Nonfactorable equivalence classes for the $r = 6$ variable-N Slater hull[a].

y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{23}	y_{24}	y_{25}	y_{26}	y_{34}	y_{35}	y_{36}	y_{45}	y_{46}	y_{56}
6	−3	−5	−4	−3	−3	−3	2	2	1	1	1	3	2	2	2	2	1	2	1	1	1
3	−1	−2	−1	−2	−2	−2	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1
5	−4	−3	−2	−3	−2	−4	2	1	2	2	3	1	1	1	2	1	0	2	1	2	1
3	−3	−2	−1	−2	−2	−2	2	1	2	2	2	1	1	1	1	1	0	1	1	1	1
2	−1	−2	−1	−1	−1	−1	1	0	0	1	1	1	1	1	1	1	0	1	1	0	0
9	−6	−8	−6	−4	−4	−4	5	3	2	2	2	5	3	3	3	2	2	2	1	1	1
6	−5	−6	−3	−3	−3	−3	5	2	2	2	2	3	3	3	3	1	1	1	1	1	1
3	−3	−2	−3	−2	−2	−2	2	3	2	2	2	2	1	1	1	2	2	2	1	1	1
3	0	−3	−2	−2	−2	−1	1	0	1	0	−1	2	2	2	1	1	1	1	1	0	1
5	−2	−4	−3	−3	−2	−2	2	1	1	1	0	2	2	2	1	1	1	1	1	1	0
6	−3	−1	−1	−3	−3	−3	0	1	1	1	1	−1	0	1	1	1	0	0	1	1	1
3	−1	−2	0	−1	−2	−2	1	1	0	1	0	1	0	1	1	−1	0	0	1	1	1

[a] $y_0 + \sum y_i \lambda_i + \sum y_{ij} \lambda_{ij} \geq 0$.

semble averaging over the $\binom{r}{N}$ configurations of the system. Bounds on the λ_{ij} are clearly the same as for the fixed-N Slater hull problem. For large N and r these detailed bounds are of less interest than those which are true for all r and N(N and r are not known exactly anyway if r is of the size of 10^{23}). Hence the variable-N results for low values of q are generally of most concern. These express necessary conditions on the λ_{ij} which must be obeyed by any approximate scheme for constructing the λ_{ij}.

For many ensembles of interest all sites are equivalent and all λ_i are equal. The distances between sites i and j are not all the same, however, so all pairs are not equivalent. In this case the $q = 2$ class of inequalities contains

$$\lambda_{12} \geq 0, \tag{63}$$

$$\lambda_{12} \geq 2\lambda - 1, \tag{64}$$

$$\lambda \geq \lambda_{12}. \tag{65}$$

Combining these gives the weaker result $1 \geq \lambda \geq 0$. The first inequality may be interpreted to mean that the probability of finding a pair of particles in sites 1 and 2 is nonnegative. The second inequality says the probability of a pair of vacancies in these sites is nonnegative. The last inequality says the probability of a particle at one site and a vacancy at the other is nonnegative.

Alternatively, these equations may be interpreted by introducing the deviation from a random distribution

$$\tau_{ij} = \lambda_{ij} - \lambda^2 \tag{66}$$

to give

$$\tau_{ij} \geq -\lambda^2, \tag{67}$$

$$\tau_{ij} \geq -(1-\lambda)^2, \tag{68}$$

$$\lambda(1-\lambda) \geq \tau_{ij}. \tag{69}$$

Clearly for λ small this implies

$$\lambda \geq \tau_{ij} \geq -\lambda^2 \tag{70}$$

and for λ large

$$(1-\lambda) \geq \tau_{ij} \geq -(1-\lambda)^2 \tag{71}$$

so that τ_{ij} approaches 0 for either λ or $1-\lambda$ near zero.

The $q = 3$ class of inequalities is more complicated.

$$\tau_{12} + \tau_{13} + \tau_{23} \geq -3\lambda^2 + 3\lambda - 1, \tag{72}$$

$$\lambda(1-\lambda) + \tau_{12} \geq \tau_{13} + \tau_{23}. \tag{73}$$

For $\frac{2}{3} > \lambda > \frac{1}{3}$, the first of these is stronger than those obtained for $q = 2$. It is related to the fact that, when 3λ requires more than one particle distributed between three sites, all three λ_{ij} cannot be arbitrarily small. Inequality (73) represents a slight strengthening of (69) to account for the fact that λ_{13} and λ_{23} cannot both reach their upper bound of $\lambda(1-\lambda)$ unless λ_{12} is simultaneously large.

Clearly for larger q, the restrictions modify finer details of the distribution while the interpretation gets more difficult. For $q = 4$, for example, the resulting set of restrictions are very difficult to visualize:

$$\tau_{12} + \tau_{13} + \tau_{23} + \tau_{14} + \tau_{24} + \tau_{34} \geq 2(-3\lambda^2 + 3\lambda - 1) \pm (1 - 2\lambda), \tag{74}$$

$$\tau_{12} + \tau_{13} + \tau_{23} - \tau_{14} - \tau_{24} - \tau_{34} \geq -\max(\lambda, 1-\lambda), \tag{75}$$

$$\tau_{12} - \tau_{13} - \tau_{23} - \tau_{14} - \tau_{24} + \tau_{34} \geq -2\lambda(1-\lambda). \tag{76}$$

E. Ising Model

For the simplest Ising model

$$H = 4\sum_{i<j} J_{ij} S_i S_j + 2\sum_i B S_i, \tag{77}$$

where the possible configurations are 2^r states with $S_i = \pm\frac{1}{2}$. The interaction J_{ij} is different for each type of pairs of lattice sites i and j. If the variables $\mu_i = \langle 2S_i \rangle$ and $\mu_{ij} = \langle 4S_i S_j \rangle$ are introduced,

$$E = \sum_{i<j} J_{ij} \mu_{ij} + \sum_i B \mu_i.$$

Now if $\lambda_i = \frac{1}{2}(1 + \mu_i)$ and $\lambda_{ij} = \frac{1}{4}(\mu_{ij} + \mu_i + \mu_j + 1)$, then the configurations generate the variable-N Slater hull Λ matrix. If it is assumed that all μ_i are equal, the inequalities just discussed can be rewritten in terms of μ and the μ_{ij} as

$$1 \geq \mu_{12} \geq -1 + 2|\mu|, \tag{78}$$

$$\mu_{23} \geq -1 + |\mu_{12} + \mu_{13}|, \tag{79}$$

$$\mu_{12} + \mu_{13} + \mu_{23} + \mu_{14} + \mu_{24} + \mu_{34} \geq -2 + 4|\mu|, \tag{80}$$

$$\mu_{12} + \mu_{13} + \mu_{23} - \mu_{14} - \mu_{24} - \mu_{34} \geq -2 + 2|\mu|, \tag{81}$$

$$\mu_{12} - \mu_{13} - \mu_{23} - \mu_{14} - \mu_{24} + \mu_{34} \geq -2. \tag{82}$$

For $B = 0$, μ will usually be zero. In this case perfect alteration of spins on adjacent sites is possible so μ_{ij} can reach -1. As $|\mu|$ increases due to an applied field, the spins align with the field and μ_{ij} must be greater than $2|\mu| - 1$. For any set of three sites, if two pairs have perfect anti-correlation ($\mu_{12} = \mu_{13} = -1$), then necessarily μ_2 and μ_3 are instantaneously parallel so $\mu_{23} = 1$.

F. The N-representability Problem for Bosons

A wavefunction for a system of N indistinguishable bosons may be approximated by a finite linear combination of symmetrized product functions. These products may be represented in terms of a set of r spin-orbitals as

$$\phi_K = (n_1! n_2! \cdots n_r!)^{-1/2} (N!)^{-1/2} \mathcal{S} \phi_{k_1}(1) \cdots \phi_{k_N}(N), \tag{83}$$

where $\mathcal{S}$ is the sum over all permutation operators and the n_i are the occupation numbers of the orbitals. If K is restricted to $K = \{1 \le k_1 \le k_2 \cdots k_N \le r\}$, the ϕ_K form an orthonormal set of size $\binom{r+N-1}{N}$, and ψ may be written as

$$\psi = \sum C_K \phi_K. \tag{84}$$

With the definition of Γ used for fermions,

$$\Gamma = \sum_{i \le j} \sum_{k \le l} \Gamma_{ijkl} \phi_{ij}(1,2) \phi_{kl}(1',2'), \tag{85}$$

where

$$\phi_{ij} = 2^{-1/2}\{\phi_i(1)\phi_j(2) + \phi_j(1)\phi_i(2)\}, \quad i < j, \tag{85'a}$$

$$\phi_{ii} = \phi_i(1)\phi_i(2). \tag{85'b}$$

For Γ_ψ

$$\Gamma_{ijkl} = \sum_K \sum_J C_K C_J^* \epsilon^{KJ}{}_{ijkl}, \tag{86}$$

$$\epsilon^{KJ}{}_{ijkl} = \binom{N}{2} \int \phi^*_{ij}(1,2) \phi_K(1,2,3,\ldots,N) \phi^*_J(1'2'3\cdots N) \times \phi_{kl}(1'2') dX_1 \cdots dX_N dX_1' dX_2',$$

$$\epsilon^{KJ}{}_{ijkl} = \epsilon^K_{ij} \epsilon^J_{kl} \delta_{K-i-j, J-k-l}, \tag{87}$$

$$\epsilon^K_{ij} = (n_i{}^K n_j{}^K)^{1/2}, \quad i < j \tag{88a}$$

$$= [\tfrac{1}{2} n_i{}^K (n_i{}^K - 1)]^{1/2}, \quad i = j. \tag{88b}$$

Most of the methods used for elucidating the structure of fermion density matrices also apply to Bosons. For Γ_N or Γ_ψ, $\Gamma_{ijij} = \lambda_{ij}$ is bounded by

$$\tfrac{1}{2} N(N-1) \ge \lambda_{ii} = \langle \tfrac{1}{2} \eta_i (\eta_i - 1) \rangle \ge 0, \tag{89a}$$

$$\tfrac{1}{4} N^2 \ge \lambda_{ij} = \langle \eta_i \eta_j \rangle \ge 0, \quad i \ne j. \tag{89b}$$

Also

$$-\tfrac{1}{8} N^2 \ge \mathrm{Re}(\Gamma_{ijkl}) \le \tfrac{1}{8} N^2, \quad i < j, k < l, (ij) \ne (kl), \tag{90a}$$

$$-\tfrac{1}{4} N[\tfrac{1}{2} N(N-1)]^{1/2} \le \mathrm{Re}(\Gamma_{iikl}) \le \tfrac{1}{4} N[\tfrac{1}{2} N(N-1)]^{1/2}, \quad k < l, \tag{90b}$$

$$-\tfrac{1}{4} N(N-1) \le \mathrm{Re}(\Gamma_{iijj}) \le \tfrac{1}{4} N(N-1), \quad i \ne j \tag{90c}$$

with similar bounds on $\mathrm{Im}(\Gamma_{ijkl})$. The set of Boson density matrices Γ_N is a bounded convex set with

$$\sum_{i \le j} \sum_{k \le l} |\Gamma_{ijkl}|^2 \le \binom{N}{2}^2. \tag{91}$$

A Boson hull may be defined as the closed convex hull of the $\lambda_{ij,K}$ associated with the ϕ_K for a finite-dimensional spin-orbital basis. It is easily seen that Γ_N and Γ_ψ have diagonal elements which lie in the Boson hull. Conversely, every element of the Boson hull corresponds to some element of $\{\Gamma_N\}$ and $\{\Gamma_\psi\}$.

Clearly the Boson hull has the same permutational properties as the Slater hull. Consequently, its facets may be classified into equivalence classes. The particle–hole transformation, however, does not apply to the Boson case.

G. Results for the Boson Hull Problem

The case of $r = 2$ provides one of the few examples which can be solved analytically. If K ($K = 0, 1, 2, \cdots N$) is the occupation number of ϕ_1 in ϕ_K, the rows $(\lambda_{11}, \lambda_{12}, \lambda_{22})$ of Λ^T are given by

$$\boldsymbol{\lambda}_K{}^T = \left[\binom{K}{2}, K(N-K), \binom{N-K}{2}\right] \tag{92}$$

Since all of the points lie in the plane

$$\lambda_{11} + \lambda_{12} + \lambda_{22} = \binom{N}{2}, \tag{93}$$

it is useful to change variables to the set

$$t = \lambda_{11} + \lambda_{12} + \lambda_{22} - \binom{N}{2}, \tag{94a}$$

$$u = (\lambda_{11} - \lambda_{22})(N-1)^{-1}, \tag{94b}$$

$$v = \tfrac{1}{3}\left[2\lambda_{12} - \lambda_{11} - \lambda_{22} + \binom{N}{2}\right], \tag{94c}$$

which gives

$$\lambda'_K = (t_K, u_K, v_K) = (0, K - \tfrac{1}{2}N, K(N-K)). \tag{95}$$

Hence all of the configurations lie on the parabola

$$v = -u^2 + \tfrac{1}{4} N^2. \tag{96}$$

The points enclosed by this parabolic boundary obey the conditions

$$t = 0, \tag{97}$$

$$v \le \tfrac{1}{4} N^2 - u^2, \tag{98}$$

$$v \ge 0. \tag{99}$$

The polygon enclosed by the convex hull of the λ_K is given by

$$t = 0,$$

$$v \ge 0,$$

$$v \le \tfrac{1}{4} N^2 - 2(u - K + \tfrac{1}{2}N)(K + \tfrac{1}{2}N) - (K - \tfrac{1}{2}N)^2 \quad \text{for } K = 0, 1, 2, \ldots, N-1. \tag{100}$$

Each of these latter inequalities is interior to the parabola only for

$$K - \tfrac{1}{2}N \le u \le K + 1 - \tfrac{1}{2}N, \tag{101}$$

and only in this range of u does the inequality in (100) form a part of the boundary of this Boson hull. The maximum error introduced by replacing the exact boundary by the parabola is $\Delta v = \frac{1}{4}$ which occurs at $u = K + \frac{1}{2} - \frac{1}{2}N$. Since v would be expected to grow like N^2 for most real situations, the relative error $\Delta v/v$ might become acceptable for large N even though Δv is constant.

The case $N = 2$ is also solvable in closed form for any r. This gives simply

$$\lambda_{ij} \geq 0, \quad i \leq j, \tag{102}$$

since Λ is a unit matrix in this case.

The algorithm discussed earlier has been applied to the cases $N = 3, 4, 5$ and $r = 3, 4, 5$. Since no Boson results have previously appeared in the literature, some of these inequalities are given in Table V. As for fermions, these inequalities generally apply only to the r, N case for which they were derived. Table VI gives the number of equivalence classes found for each r, N. Complete tabulations of these results are available elsewhere.[7]

No results for variable N could be obtained by these methods since the Λ matrix has an infinite number of columns if N is unbounded. The variable-N results would be interesting since it would give necessary conditions for all r, N. The conditions $\lambda_{ij} \geq 0$ are probably variable-N inequalities. These cannot be all of the conditions, however, as the obvious result

$$\langle(\eta_1 - \eta_2)^2\rangle = 2\lambda_{11} + \lambda_1 - 2\lambda_{12} + 2\lambda_{22} + \lambda_2 \geq 0$$

cannot be expressed as a positive combination of $\lambda_{ij} \geq 0$.

Fairly strong inequalities of the type found for fermions may be constructed for bosons. If β and α_i are integers, then necessarily

$$\left\langle\left\langle\left(\sum \alpha_i\eta_i - \beta\right)\left(\sum \alpha_i\eta_i - \beta - 1\right)\right\rangle\right\rangle \geq 0 \tag{103}$$

since this expression is nonnegative on every configuration $\phi_K(\sum \alpha_i\lambda_{i,K}$ is an integer so it cannot lie between β and $\beta + 1$). For $\beta = [\langle\alpha_i\eta_i\rangle]$ where $[x]$ stands for the largest integer no larger than x, Eq. (103) is fairly strong as

$$\left\langle\left\langle\sum \alpha_i\eta_i - \beta\right\rangle\right\rangle\left\langle\left\langle\sum \alpha_i\eta_i - \beta - 1\right\rangle\right\rangle \leq 0.$$

Examples of (103) are

$$2\lambda_{11} + \lambda_1 - 2\lambda_{12} + 2\lambda_{22} + \lambda_2 + (2\beta + 1)(\lambda_2 - \lambda_1) + \beta(\beta + 1) \geq 0$$

and

$$2\lambda_{11} + \lambda_1 + 2\lambda_{12} + 2\lambda_{22} + \lambda_2 - (2\beta + 1)(\lambda_2 + \lambda_1) + \beta(\beta + 1) \geq 0$$

for all integer β. For $\beta = [\lambda_1 - \lambda_2]$ and $[\lambda_1 + \lambda_2]$, respectively, these inequalities are rather sharp. It is likely that some of the extreme variable-N Boson inequalities may be of the form (103).

TABLE V. Boson hull equivalence classes.

($N = 2$, any r)

$$\lambda_{11} \geq 0 \tag{V.1}$$

$$\lambda_{12} \geq 0 \tag{V.2}$$

($N = 3, r = 2$) V.1, V.2, and

$$\lambda_{12} \leq 2 \tag{V.3}$$

($N = 4, r = 2$) V.1, V.2, and

$$\lambda_{12} - \lambda_{11} \leq 3 \tag{V.4}$$

($N = 5, r = 2$) V.1, V.2, and

$$\lambda_{12} - 2\lambda_{11} \leq 4 \tag{V.5}$$

$$\lambda_{12} \leq 6 \tag{V.6}$$

($N = 3, r = 3$) V.1, V.2 and

$$\lambda_{12} + 2\lambda_{22} - \lambda_{23} + 2\lambda_{33} \geq 0 \tag{V.7}$$

$$\lambda_{13} + \lambda_{23} \leq 2 \tag{V.8}$$

($N = 3, r = 4$) V.1, V.2, and

$$2\lambda_{11} + 2\lambda_{12} - \lambda_{14} + 2\lambda_{22} - \lambda_{24} + \lambda_{34} + 2\lambda_{44} \geq 0 \tag{V.9}$$

$$2\lambda_{11} + 2\lambda_{12} + \lambda_{13} - \lambda_{14} + 2\lambda_{22} + \lambda_{23} - \lambda_{24} + 2\lambda_{44} \geq 0 \tag{V.10}$$

$$\lambda_{12} + \lambda_{23} + \lambda_{24} \leq 2 \tag{V.11}$$

$$\lambda_{12} + 2\lambda_{22} - \lambda_{24} + \lambda_{34} + 2\lambda_{44} \geq 0 \tag{V.12}$$

$$\lambda_{13} + \lambda_{14} + \lambda_{23} + \lambda_{24} \leq 2 \tag{V.13}$$

$$\lambda_{12} + 2\lambda_{22} + \lambda_{23} - \lambda_{24} + 2\lambda_{44} \geq 0$$

($N = 4, r = 3$) V.1, V.2, and

$$3\lambda_{11} - \lambda_{12} - 3\lambda_{13} + \lambda_{22} + 5\lambda_{23} + 9\lambda_{33} \geq 0$$

$$\lambda_{11} - \lambda_{13} + \lambda_{23} + 3\lambda_{33} \geq 0$$

$$2\lambda_{12} + 2\lambda_{13} + \lambda_{11} \leq 9$$

$$3\lambda_{11} - 2\lambda_{13} + \lambda_{23} + 5\lambda_{33} \geq 0$$

$$6\lambda_{11} - 2\lambda_{12} + 2\lambda_{22} + \lambda_{23} \geq 0$$

$$\lambda_{12} + \lambda_{13} - \lambda_{11} \leq 3$$

TABLE VI. Number of equivalence classes for Boson hull.

r/N	2	3	4	5
2	2	3	3	4
3	2	4	8	15
4	2	8	41	244
5	2	17	589	

ACKNOWLEDGMENT

The authors wish to thank the National Science Foundation for a grant which supported this work.

[1] G. Hadley, *Linear Programming* (Addison-Wesley, Reading, Mass., 1962).
[2] B. Grünbaum, *Convex Polytopes* (Interscience, New York, 1967).
[3] E. R. Davidson, J. Math. Phys. **10**, 725 (1969), Sec. I.
[4] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall, Ann. Math. Studies **28**, 51 (1953).
[5] D. A. Kohler, "Projections of Convex Polyhedral Sets," Ph. D. Thesis, University of California, Berkeley, California (1967).
[6] W. B. McRae and E. R. Davidson (to be published).
[7] W. B. McRae, "A Study of Diagonal N-Representability", Ph. D. Thesis, University of Washington, Seattle, Washington (1971).
[8] H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[9] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 659 (1961).
[10] A. J. Coleman in "Report of the Density Matrix Seminar," edited by A. J. Coleman and R. M. Erdahl, Department of Mathematics, Queen's University, Kingston, Ontario, (1969), p. 6.

[11] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[12] F. Sasaki, Phys. Rev. **138**, 1338 (1965).
[13] M. L. Yoseloff, thesis (Princeton University, 1970).
[14] F. Weinhold and E. B. Wilson, Jr., J. Chem. Phys. **47**, 2298 (1967).
[15] M. Eisen, *Elementary Combinatorial Analysis* (Gordon and Breach, New York (1969), pp. 181-86.
[16] M. L. Yoseloff and H. W. Kuhn, J. Math. Phys. **10**, 703 (1969).
[17] H. W. Kuhn, Proc. Sym. Appl. Math. **10**, 141 (1960).

Stationary Axially Symmetric Fields and the Kerr Metric

R. M. Misra and Deo Bhushan Pandey
Department of Physics, University of Gorakhpur, Gorakhpur, India
(Received 7 April 1971)

In this note the axially symmetric metric for stationary gravitational field, in a slightly general form, is discussed. The vacuum field equations for this metric are given. Specialization of this metric leads to a different form of field equations previously discussed in literature. In particular, the Kerr metric is given in a new form. A justification for interpreting the Kerr metric as an exterior solution corresponding to a spinning rod or a rotating spherical body is given.

1. INTRODUCTION

Axially symmetric gravitational fields within the framework of general relativity have been investigated by various authors from the early days of the theory. Earlier studies of stationary axisymmetric fields were carried out to determine the relativistic effects on the motion of a slowly rotating body. In particular an attempt was made to understand the nature of inertial forces.[1]

Recently, interest in the study of gravitational fields with axial symmetry has been renewed due to discovery of massive astrophysical bodies known as quasars.[2] These bodies release enormous amount of energy, which is believed to be a general relativistic effect of the collapsing star. When one looks for the effect of rotation on the course of collapse and the ultimate fate of the collapsed star, the study of axisymmetric fields becomes important.

There is yet another important reason for the study of this class of fields. Aside from their important role in the study of gravitational radiation,[3] their investigation is necessary for understanding the basic structure of the general relativity itself. Today, it is still a matter of controversy how to give a precise formulation of Mach's principle and whether general relativity includes Mach's principle or needs to be supplemented by boundary conditions or must be modified in order to be consistent with this principle.[4]

In view of the above discussion, we present in this paper some of the properties of the stationary axially symmetric fields. We choose the stationary metric in a slightly general form, which we believe is a new approach to the problem. Special choices of a harmonic function allow the metric with axial symmetry to be expressed in different coordinate systems. In particular, we have shown that the Lewis–Papapetrou canonical metric can be obtained from our general metric.

In Sec. 3 the formalism has been applied to the Kerr solution, which is the only known exact solution of the axially symmetric stationary problem representing the exterior field of a finite body. Discussion of this section throws interesting light on the shape of the material body whose external field is given by the Kerr metric. This has been discussed in the last section.

2. THE METRIC AND THE FIELD EQUATIONS

The metric of a stationary space–time with axial symmetry may be taken as

$$ds^2 = e^{2u}(dt + wd\phi)^2 - e^{2k-2u}[(dx^1)^2 + (dx^2)^2] - h^2e^{-2u}d\phi^2, \quad (1)$$

where u, w, k, and h are functions of x^1 and x^2 only. The justification for choosing the metric in this form is based to some extent on the Newtonian concept of axial symmetry.[5] This line element shows that the metric form is preserved under the simultaneous reflection of ϕ and t coordinates, i.e.,

$$(\phi, t) \to (-\phi, -t). \quad (2)$$

Because of this symmetry the cross terms between x^1x^2 and ϕt part of the metric are eliminated since, for instance, dx^1dt and dx^2dt would change sign under (2). Thus stationary flow is allowed in the ϕ direction only, and flow in x^1 and x^2 directions is excluded. The metric is, therefore, not the most general stationary line element. In addition to this reflection symmetry, the metric form is preserved under the conformal transformation of the x^1x^2 plane. This transformation may be expressed as

$$\bar{z} = x^1 + ix^2 = f(x^1 + ix^2) = f(z). \quad (3)$$

Under this transformation u and h transform like scalars (although their functional form will be changed) and

$$e^{2k'} = \frac{\partial f}{\partial z} \cdot \frac{\partial f}{\partial z} \cdot e^{2k}. \quad (4)$$

In addition, one should also demand that the solutions obtained from (1) be physically meaningful. This requirement imposes two conditions on the metric coefficients, namely asymptotic flatness and elementary flatness.[6]

The vacuum field equations for the metric (1) may be easily set up by calculating the components of the Ricci tensor. By a straightforward but tedious calculation we find that nonvanishing components of the Ricci tensor have the following values:

$$R_{00} = -e^{4u-2k}[u_{,11} + u_{,22} + (u_{,1}h_{,1} + u_{,2}h_{,2}) \times (1/h) + \tfrac{1}{2}e^{4u}(w_{,1}^2 - w_{,2}^2)/h^2], \quad (5)$$

[11] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[12] F. Sasaki, Phys. Rev. **138**, 1338 (1965).
[13] M. L. Yoseloff, thesis (Princeton University, 1970).
[14] F. Weinhold and E. B. Wilson, Jr., J. Chem. Phys. **47**, 2298 (1967).
[15] M. Eisen, *Elementary Combinatorial Analysis* (Gordon and Breach, New York (1969), pp. 181-86.
[16] M. L. Yoseloff and H. W. Kuhn, J. Math. Phys. **10**, 703 (1969).
[17] H. W. Kuhn, Proc. Sym. Appl. Math. **10**, 141 (1960).

Stationary Axially Symmetric Fields and the Kerr Metric

R. M. Misra and Deo Bhushan Pandey
Department of Physics, University of Gorakhpur, Gorakhpur, India
(Received 7 April 1971)

In this note the axially symmetric metric for stationary gravitational field, in a slightly general form, is discussed. The vacuum field equations for this metric are given. Specialization of this metric leads to a different form of field equations previously discussed in literature. In particular, the Kerr metric is given in a new form. A justification for interpreting the Kerr metric as an exterior solution corresponding to a spinning rod or a rotating spherical body is given.

1. INTRODUCTION

Axially symmetric gravitational fields within the framework of general relativity have been investigated by various authors from the early days of the theory. Earlier studies of stationary axisymmetric fields were carried out to determine the relativistic effects on the motion of a slowly rotating body. In particular an attempt was made to understand the nature of inertial forces.[1]

Recently, interest in the study of gravitational fields with axial symmetry has been renewed due to discovery of massive astrophysical bodies known as quasars.[2] These bodies release enormous amount of energy, which is believed to be a general relativistic effect of the collapsing star. When one looks for the effect of rotation on the course of collapse and the ultimate fate of the collapsed star, the study of axisymmetric fields becomes important.

There is yet another important reason for the study of this class of fields. Aside from their important role in the study of gravitational radiation,[3] their investigation is necessary for understanding the basic structure of the general relativity itself. Today, it is still a matter of controversy how to give a precise formulation of Mach's principle and whether general relativity includes Mach's principle or needs to be supplemented by boundary conditions or must be modified in order to be consistent with this principle.[4]

In view of the above discussion, we present in this paper some of the properties of the stationary axially symmetric fields. We choose the stationary metric in a slightly general form, which we believe is a new approach to the problem. Special choices of a harmonic function allow the metric with axial symmetry to be expressed in different coordinate systems. In particular, we have shown that the Lewis–Papapetrou canonical metric can be obtained from our general metric.

In Sec. 3 the formalism has been applied to the Kerr solution, which is the only known exact solution of the axially symmetric stationary problem representing the exterior field of a finite body. Discussion of this section throws interesting light on the shape of the material body whose external field is given by the Kerr metric. This has been discussed in the last section.

2. THE METRIC AND THE FIELD EQUATIONS

The metric of a stationary space–time with axial symmetry may be taken as

$$ds^2 = e^{2u}(dt + wd\phi)^2 - e^{2k-2u}[(dx^1)^2 + (dx^2)^2] - h^2e^{-2u}d\phi^2, \quad (1)$$

where u, w, k, and h are functions of x^1 and x^2 only. The justification for choosing the metric in this form is based to some extent on the Newtonian concept of axial symmetry.[5] This line element shows that the metric form is preserved under the simultaneous reflection of ϕ and t coordinates, i.e.,

$$(\phi, t) \to (-\phi, -t). \quad (2)$$

Because of this symmetry the cross terms between x^1x^2 and ϕt part of the metric are eliminated since, for instance, dx^1dt and dx^2dt would change sign under (2). Thus stationary flow is allowed in the ϕ direction only, and flow in x^1 and x^2 directions is excluded. The metric is, therefore, not the most general stationary line element. In addition to this reflection symmetry, the metric form is preserved under the conformal transformation of the x^1x^2 plane. This transformation may be expressed as

$$\bar{z} = x^1 + ix^2 = f(x^1 + ix^2) = f(z). \quad (3)$$

Under this transformation u and h transform like scalars (although their functional form will be changed) and

$$e^{2k'} = \frac{\partial f}{\partial z} \cdot \frac{\partial f}{\partial z} \cdot e^{2k}. \quad (4)$$

In addition, one should also demand that the solutions obtained from (1) be physically meaningful. This requirement imposes two conditions on the metric coefficients, namely asymptotic flatness and elementary flatness.[6]

The vacuum field equations for the metric (1) may be easily set up by calculating the components of the Ricci tensor. By a straightforward but tedious calculation we find that nonvanishing components of the Ricci tensor have the following values:

$$R_{00} = -e^{4u-2k}[u_{,11} + u_{,22} + (u_{,1}h_{,1} + u_{,2}h_{,2}) \times (1/h) + \tfrac{1}{2}e^{4u}(w_{,1}^2 - w_{,2}^2)/h^2], \quad (5)$$

$$R_{11} = k_{,11} + k_{,22} - u_{,11} - u_{,22} + 2u_{,1}^2 - [h_{,1}(k_{,1} + u_{,1}) + w_{,1}^2(e^{4u}/2h) - h_{,11} - h_{,2}(k_{,2} - u_{,2})]/h, \quad (6)$$

$$R_{22} = k_{,11} + k_{,22} - u_{,11} - u_{,22} + 2u_{,2}^2 + [h_{,1}(k_{,1} - u_{,1}) - w_{,2}^2(e^{4u}/2h) + h_{,22} - h_{,2}(k_{,2} + u_{,2})]/h, \quad (7)$$

$$R_{12} = 2u_{,1}u_{,2} - [k_{,2}h_{,1} + k_{,1}h_{,2} + (w_{,1}w_{,2}e^{4u}/2h) - h_{,12}]/h, \quad (8)$$

$$R_{30} = wR_{00} - [2(u_{,1}w_{,1} + u_{,2}w_{,2}) + \tfrac{1}{2}(w_{,11} + w_{,22}) - (w_{,1}h_{,1} + w_{,2}h_{,2})/2h]e^{4u-2k}, \quad (9)$$

$$R_{33} = 2wR_{30} + R_{00}(h^2e^{-4u} - w^2) + h(h_{,11} + h_{,22})e^{-2k}. \quad (10)$$

We note that

$$R_0^0 + R_3^3 = -(1/h)e^{2u-2k}\nabla^2 h. \quad (11)$$

Now the field equations for empty space can be easily set up. However, not all the components of the Ricci tensor give independent field equations. These are

$$u_{,11} + u_{,22} + [(u_{,1}h_{,1} + u_{,2}h_{,2}) + e^{4u}(w_{,1}^2 + w_{,2}^2)/2h]/h = 0, \quad (12)$$

$$w_{,11} + w_{,22} + 4(w_{,1}u_{,1} + w_{,2}u_{,2}) - (w_{,1}h_{,1} + w_{,2}h_{,2})/h = 0, \quad (13)$$

$$2(u_{,2}^2 - u_{,1}^2) + [2(k_{,1}h_{,1} - k_{,2}h_{,2}) + h_{,22} - h_{,11} + e^{4u}(w_{,1}^2 - w_{,2}^2)/2h]/h = 0, \quad (14)$$

$$2u_{,1}u_{,2} - [k_{,1}h_{,2} + k_{,2}h_{,1} - h_{,12} + w_{,1}w_{,2} \times e^{4u}/2h]/h = 0, \quad (15)$$

$$h_{,11} + h_{,22} = 0. \quad (16)$$

The choice of a solution of (16) is not a restriction on the general solution of the field equations but a coordinate condition. This should be evident from the fact that h is a scalar and it is also a harmonic function. Hence it can be chosen as the imaginary part of the analytic transformation leading to a new coordinate system. The most simple and common choice of h is $h = x^1$, which leads to the canonical coordinate system. If we call $(x^1, x^2) \to (\rho, z)$ (1) is transformed to Lewis–Papapetrou metric[6-8]

$$ds^2 = e^{2u}(dt + wd\phi)^2 - e^{2k-2u}(d\rho^2 + dz^2) - \rho^2 e^{-2u}d\phi^2. \quad (17)$$

Equations (12)–(15) go over to well-known vacuum field equations of stationary fields in canonical coordinates.

3. ON COORDINATE TRANSFORMATION

Equations (12)–(15) in canonical coordinates have been investigated by various authors, but the only known exact solution which is also physically meaningful is that given by Kerr.[9-12] This solution describes the exterior field of some finite rotating body. The solution is algebraically special and contains two parameters, m and "a", which are identified with the mass and angular momentum per unit mass of the source.[13] We now wish to show that the Kerr metric may be transformed to the form (1) by a suitable choice of the harmonic function h.

The Kerr metric is given by[13]

$$ds^2 = -(r^2 + a^2\cos^2\theta)[d\theta^2 + dr^2/(r^2 - 2mr + a^2)] - (r^2 + a^2)\sin^2\theta\, d\phi^2 + dt^2 - [2mr/(r^2 + a^2\cos^2\theta)] \times (dt + a\sin^2\theta\, d\phi)^2; \quad (18)$$

we call r and θ Kerr polar coordinates. Evidently in these coordinates the form (1) is violated. The relation between canonical coordinates ρ, z and polar coordinates r, θ is given by

$$\rho^2 = (r^2 - 2mr + a^2)\sin^2\theta, \quad (19)$$

$$z = (r - m)\cos\theta. \quad (20)$$

If we now relabel the r = const surfaces by the transformation

$$r = R + m + (m^2 - a^2)/4R, \quad (21)$$

we obtain

$$dr^2 = [1 - (m^2 - a^2)/4R^2]^2 dR^2, \qquad r^2 - 2mr + a^2 = [1 - (m^2 - a^2)/4R^2]^2 R^2, \quad (22)$$

and hence

$$\frac{dr^2}{r^2 - 2mr + a^2} = \frac{dR^2}{R^2}. \quad (23)$$

Thus Eq. (18) is transformed to the form

$$ds^2 = -\{[R + m + (m^2 - a^2)/4R]^2 + a^2\cos^2\theta\} \times (d\theta^2 + dR^2/R^2) - \{[R + m + (m^2 - a^2)/4R]^2 + a^2\} \times \sin^2\theta\, d\phi^2 + dt^2 - \frac{2m\{R + m + (m^2 - a^2)/4R\}}{\{R + m + (m^2 - a^2)/4R\}^2 + a^2\cos^2\theta} \times (dt + a\sin^2\theta\, d\phi)^2. \quad (24)$$

It may be noted that, in case of vanishing "a", (24) reduces to Schwarzschild metric in isotropic coordinates. Clearly, now a further transformation $R = \exp(\bar{R})$ will restore (24) to the form (1). For this case,

$$h = [R - (m^2 - a^2)/4R]\sin\theta = [e^{\bar{R}} - (m^2 - a^2)e^{-\bar{R}}/4]\sin\theta. \quad (25)$$

Evidently (25) is a solution of Eq. (16) if $x^1 = \bar{R}$ and $x^2 = \theta$. However, the two choices of h may be considered to be the same, i.e.,

$$h = \rho = [e^{\bar{R}} - (m^2 - a^2)e^{-\bar{R}}/4]\sin\theta. \quad (26)$$

The transformation from Lewis–Papapetrou canonical coordinates to polar coordinates is given by

$$z + i\rho = Re^{i\theta} + (m^2 - a^2)e^{-i\theta}/4R \quad (27)$$

or

$$z = [R + (m^2 - a^2)/4R]\cos\theta,$$
$$\rho = [R - (m^2 - a^2)/4R]\sin\theta. \tag{28}$$

4. ON SOURCES OF THE KERR METRIC

It is well known from the investigations of Erez and Rosen[14] that the Schwarzschild metric is generated by a rod of length $2m$ with mass density $\frac{1}{2}$. Also, the Kerr solution (18) goes over to the Schwarzschild solution in the limit of vanishing angular momentum parameter "a". This consideration leads to the interpretation of the Kerr metric as the exterior gravitational field of a rotating rod. However, it has been shown by Cohen that, to the first order in "a", the source of Kerr metric may be taken a thin spherical mass shell.[15] Further, Boyer and Price pointed out that a slowly rotating solid of perfect fluid can also be the source of Kerr metric to the same approximation in "a".[16]

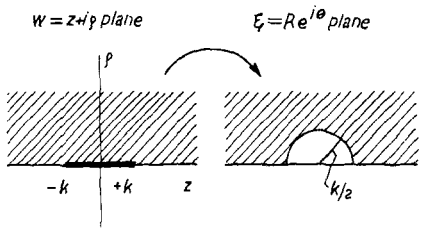


FIG. 1. Mapping from the Lewis-Papapetrou manifold to the Kerr type manifold, $w = \xi + k^2/4\xi$, $k^2 = m^2 - a^2$.

The transformation (27) throws interesting light on the shape of the material body whose exterior field is given by the Kerr metric. If we look upon the transformation (27) as a mapping of the upper half ($\rho \geq 0$) of the complex $z + i\rho$ plane into $R\exp(i\theta)$ complex plane, the line segment $\rho = 0$, $-(m^2 - a^2)^{1/2} \leq z \leq (m^2 - a^2)^{1/2}$ get mapped into the semicircle of radius $1/2(m^2 - a^2)^{1/2}$ (Fig. 1). It is this mapping that explains why the Kerr solution is interpreted to represent the exterior field of a spinning rod or a rotating spherical body.

1 H. Thirring, Z. Physik **19**, 33 (1918).
2 I. Robinson, A. Schild, and E. L. Schucking, *Quasi Stellar Sources and Gravitational Collapse* (U. of Chicago Press, Chicago, 1965).
3 H. Bondi, M. G. J. Vanderberg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
4 R. H. Dicke, *Relativity, Groups and Topology* (Gordon and Breach, New York, 1964).
5 J. L. Synge, *Relativity, The General Theory* (North-Holland, Amsterdam, 1960).
6 T. Lewis, Proc. Roy. Soc. (London) **A136**, 176 (1932).
7 W. J. VanStockum, Proc. Roy. Soc. Edinburgh **57**, 135 (1937).
8 A. Papapetrou, Ann. Physik **12**, 309 (1953).
9 R. P. Kerr, Phys. Rev. Letters **11**, 237 (1963).
10 R. P. Kerr and A. Schild, Proceedings of Galileo Galilei Centenary Meeting on General Relativity (Florence 1965), edited by G. Barbara.
11 F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
12 M. Misra, Proc. Roy. Irish Acad. **A69**, 39 (1970).
13 R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
14 G. Erez and N. Rosen, Bull. Res. Council Israel **8F**, 47 (1959).
15 J. M. Cohen, J. Math. Phys. **8** 1477 (1967).
16 R. H. Boyer and T. G. Price, Proc. Cambridge Phil. Soc. **61**, 531 (1965).

On the Convergence of the Born Series for All Energies

P. J. Bushell
The University of Sussex, Brighton, Sussex, England

(Received 15 May 1972)

Formal solution of the Schrödinger equation for nonrelativistic scattering by a spherically symmetric static potential $-\mu V(r)$ leads to a power series in the real parameter μ for the scattering amplitude (the Born series). It is shown that if $\int_0^\infty r|V(r)|dr < \infty$, $\int_0^\infty r^2|V(r)|dr < \infty$ and if $-\mu|V(r)|$ is too weak to support a bound state, then the Born series converges at all energies. The method gives a lower bound for the radius of convergence of the Born series which is exact if $V \geq 0$.

1. INTRODUCTION

The Schrödinger equation for nonrelativistic scattering by a spherically symmetric static potential $-\mu V(r)$, where μ is a real parameter, may be written as

$$(\nabla^2 + k^2)\psi = -\mu V\psi.$$

The scattering solution is given by

$$\psi(\mathbf{k}_0,\mathbf{r}) = \exp(i\mathbf{k}_0\cdot\mathbf{r}) + \frac{\mu}{4\pi}\int \frac{\exp(ik|\mathbf{r}-\mathbf{s}|)}{|\mathbf{r}-\mathbf{s}|} V(s)\psi(\mathbf{k}_0,\mathbf{s})d\mathbf{s}, \tag{1.1}$$

and the scattering amplitude $F(\mathbf{k},\mathbf{k}_0)$ in a direction $\mathbf{k}$ is given by

$$F(\mathbf{k},\mathbf{k}_0) = \frac{\mu}{4\pi}\int \exp(-i\mathbf{k}\cdot\mathbf{r})V(r)\psi(\mathbf{k}_0,r)dr, \tag{1.2}$$

where $|\mathbf{k}| = |\mathbf{k}_0| = k$.

Formal solution of Eq. (1.1) by iteration yields a power series in μ for $\psi(\mathbf{k}_0,\mathbf{r})$ and if this series is substituted in Eq. (1.2) we obtain the Born series for $F(\mathbf{k},\mathbf{k}_0)$. Several methods have been used to find estimates for the radius of convergence of the Born series at varying energies.[1-5]

In particular Davies[3] and Huby[4] have considered a restricted class of potentials, namely those which are bounded and which have finite range. Davies showed that if the potential $-|\mu V(r)|$ is too weak to support a bound state, then the Born series converges absolutely and uniformly for all $\mathbf{k}$ and $\mathbf{k}_0$ (with $|\mathbf{k}| = |\mathbf{k}_0|$). Huby extended this result to the Born series for the various partial wave scattering amplitudes and phase shifts. The purpose of this note is to extend Davies', and hence Huby's results to a wider class of potentials.

We will prove the following

or

$$z = [R + (m^2 - a^2)/4R]\cos\theta,$$
$$\rho = [R - (m^2 - a^2)/4R]\sin\theta. \tag{28}$$

4. ON SOURCES OF THE KERR METRIC

It is well known from the investigations of Erez and Rosen[14] that the Schwarzschild metric is generated by a rod of length $2m$ with mass density $\frac{1}{2}$. Also, the Kerr solution (18) goes over to the Schwarzschild solution in the limit of vanishing angular momentum parameter "a". This consideration leads to the interpretation of the Kerr metric as the exterior gravitational field of a rotating rod. However, it has been shown by Cohen that, to the first order in "a", the source of Kerr metric may be taken a thin spherical mass shell.[15] Further, Boyer and Price pointed out that a slowly rotating solid of perfect fluid can also be the source of Kerr metric to the same approximation in "a".[16]

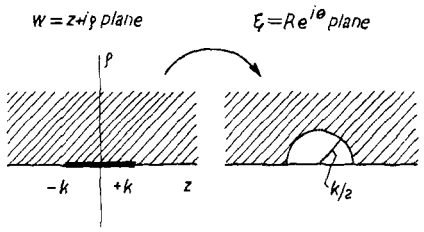


FIG. 1. Mapping from the Lewis-Papapetrou manifold to the Kerr type manifold, $w = \xi + k^2/4\xi$, $k^2 = m^2 - a^2$.

The transformation (27) throws interesting light on the shape of the material body whose exterior field is given by the Kerr metric. If we look upon the transformation (27) as a mapping of the upper half ($\rho \geq 0$) of the complex $z + i\rho$ plane into $R\exp(i\theta)$ complex plane, the line segment $\rho = 0$, $-(m^2 - a^2)^{1/2} \leq z \leq (m^2 - a^2)^{1/2}$ get mapped into the semicircle of radius $1/2(m^2 - a^2)^{1/2}$ (Fig. 1). It is this mapping that explains why the Kerr solution is interpreted to represent the exterior field of a spinning rod or a rotating spherical body.

1 H. Thirring, Z. Physik **19**, 33 (1918).
2 I. Robinson, A. Schild, and E. L. Schucking, *Quasi Stellar Sources and Gravitational Collapse* (U. of Chicago Press, Chicago, 1965).
3 H. Bondi, M. G. J. Vanderberg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
4 R. H. Dicke, *Relativity, Groups and Topology* (Gordon and Breach, New York, 1964).
5 J. L. Synge, *Relativity, The General Theory* (North-Holland, Amsterdam, 1960).
6 T. Lewis, Proc. Roy. Soc. (London) **A136**, 176 (1932).
7 W. J. VanStockum, Proc. Roy. Soc. Edinburgh **57**, 135 (1937).
8 A. Papapetrou, Ann. Physik **12**, 309 (1953).
9 R. P. Kerr, Phys. Rev. Letters **11**, 237 (1963).
10 R. P. Kerr and A. Schild, Proceedings of Galileo Galilei Centenary Meeting on General Relativity (Florence 1965), edited by G. Barbara.
11 F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
12 M. Misra, Proc. Roy. Irish Acad. **A69**, 39 (1970).
13 R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
14 G. Erez and N. Rosen, Bull. Res. Council Israel **8F**, 47 (1959).
15 J. M. Cohen, J. Math. Phys. **8** 1477 (1967).
16 R. H. Boyer and T. G. Price, Proc. Cambridge Phil. Soc. **61**, 531 (1965).

On the Convergence of the Born Series for All Energies

P. J. Bushell
The University of Sussex, Brighton, Sussex, England

(Received 15 May 1972)

Formal solution of the Schrödinger equation for nonrelativistic scattering by a spherically symmetric static potential $-\mu V(r)$ leads to a power series in the real parameter μ for the scattering amplitude (the Born series). It is shown that if $\int_0^\infty r|V(r)|dr < \infty$, $\int_0^\infty r^2|V(r)|dr < \infty$ and if $-\mu|V(r)|$ is too weak to support a bound state, then the Born series converges at all energies. The method gives a lower bound for the radius of convergence of the Born series which is exact if $V \geq 0$.

1. INTRODUCTION

The Schrödinger equation for nonrelativistic scattering by a spherically symmetric static potential $-\mu V(r)$, where μ is a real parameter, may be written as

$$(\nabla^2 + k^2)\psi = -\mu V\psi.$$

The scattering solution is given by

$$\psi(\mathbf{k}_0,\mathbf{r}) = \exp(i\mathbf{k}_0\cdot\mathbf{r}) + \frac{\mu}{4\pi}\int \frac{\exp(ik|\mathbf{r}-\mathbf{s}|)}{|\mathbf{r}-\mathbf{s}|} V(s)\psi(\mathbf{k}_0,\mathbf{s})d\mathbf{s}, \tag{1.1}$$

and the scattering amplitude $F(\mathbf{k},\mathbf{k}_0)$ in a direction $\mathbf{k}$ is given by

$$F(\mathbf{k},\mathbf{k}_0) = \frac{\mu}{4\pi}\int \exp(-i\mathbf{k}\cdot\mathbf{r})V(r)\psi(\mathbf{k}_0,r)dr, \tag{1.2}$$

where $|\mathbf{k}| = |\mathbf{k}_0| = k$.

Formal solution of Eq. (1.1) by iteration yields a power series in μ for $\psi(\mathbf{k}_0,\mathbf{r})$ and if this series is substituted in Eq. (1.2) we obtain the Born series for $F(\mathbf{k},\mathbf{k}_0)$. Several methods have been used to find estimates for the radius of convergence of the Born series at varying energies.[1-5]

In particular Davies[3] and Huby[4] have considered a restricted class of potentials, namely those which are bounded and which have finite range. Davies showed that if the potential $-|\mu V(r)|$ is too weak to support a bound state, then the Born series converges absolutely and uniformly for all $\mathbf{k}$ and $\mathbf{k}_0$ (with $|\mathbf{k}| = |\mathbf{k}_0|$). Huby extended this result to the Born series for the various partial wave scattering amplitudes and phase shifts. The purpose of this note is to extend Davies', and hence Huby's results to a wider class of potentials.

We will prove the following

Sufficient Conditions for Convergence

If $V(r)$ is a piece-wise continuous function of r for $0 < r < \infty$, if

$$\int_0^\infty r|V(r)|dr < \infty \quad \text{and} \quad \int_0^\infty r^2|V(r)|dr < \infty, \qquad \text{(A)}$$

and if $-\mu|V(r)|$ is too weak to support a bound state then the Born series converges for all $\mathbf{k}$ and $\mathbf{k}_0$ (with $|\mathbf{k}| = |\mathbf{k}_0|$).

The basic idea behind our method of proof is due to Davies.[3] We show that the Born series is majorized by a series which can be expressed as the quotient of two analytic functions of μ, and we examine the zeros of the denominator. The details are by no means trivial and they can be found in Ref. 6. In view of the leading term in the series (2.1) below it is clear that condition (A) is the best possible integrability condition for Davies' method.

The method yields a lower bound for the radius of convergence of the Born series for all energies, and in Sec. 3 we give some examples of this bound. We remark that if $V \geq 0$, then the lower bound is precisely the radius of convergence of the Born series at zero energy.

2. DISCUSSION

Following Davies[3] we remark that the Born series for $F(\mathbf{k},\mathbf{k}_0)$ is majorized by the series

$$\frac{\mu}{2\pi}\int q(r)d\mathbf{r} + \left(\frac{\mu}{2\pi}\right)^2 \int q(r)d\mathbf{r}\int \frac{1}{|\mathbf{r}-\mathbf{s}|}q(s)d\mathbf{s} + \cdots,$$

where we have written $q(r) = |V(r)|$. Performing the angular integrations, we obtain the series

$$\begin{aligned} M(\mu) = {} & \mu\int_0^\infty r^2 q(r)dr \\ & + \mu^2\int_0^\infty rq(r)dr\int_0^\infty m(r,s)sq(s)ds \\ & + \mu^3\int_0^\infty rq(r)dr\int_0^\infty m(r,s)sq(s)ds \\ & \times \int_0^\infty m(s,t)tq(t)dt + \cdots, \end{aligned} \qquad (2.1)$$

where $m(r,s)$ denotes the smaller of r and s.

Let $\phi(r,k,\mu)$ denote the reduced wavefunction for S-wave scattering by the potential $-\mu|V(r)|$. Then ϕ is the solution of the differential equation

$$\phi'' + [k^2 + \mu q(r)]\phi = 0, \quad 0 \leq r < \infty,$$

for which

$$\phi(0,k,\mu) = 0 \quad \text{and} \quad \phi'(0,k,\mu) = 1.$$

It is well known that for real nonzero values of k,

$$\phi(r,k,\mu) \sim (1/2ik)[f(k,\mu)e^{ikr} - f(-k,\mu)e^{-ikr}]$$

as $r \to \infty$, where $f(k,\mu)$ is a function analytic in k for $\mathrm{Im}k < 0$. Moreover, the bound states for S-wave scattering correspond exactly to the zeros of $f(k,\mu)$ in the half-plane $\mathrm{Im}k \leq 0$.

It is shown in Ref. 6 that

$$\phi(r,0,\mu) \sim rf(0,\mu) + \frac{\partial f(0,\mu)}{\partial k} \qquad (2.2)$$

as $r \to \infty$.

Now let $\chi(r,\mu)$ denote the solution of

$$\chi'' + \mu q(r)\chi = 0, \quad 0 \leq r < \infty,$$

which is such that

$$\chi(r,\mu) \sim r + C(\mu) \qquad (2.3)$$

as $r \to \infty$. Then χ is the solution of the integral equation

$$\chi(r,\mu) = r + \int_0^\infty m(r,s)q(s)\chi(s,\mu)ds \qquad (2.4)$$

and

$$\chi(r,\mu) \sim r + \mu\int_0^\infty sq(s)\chi(s,\mu)ds \qquad (2.5)$$

as $r \to \infty$.

If the solution of Eq. (2.4) found by formal iteration is substituted in Eq. (2.5) and the resulting series integrated term by term, we find that $C(\mu) = M(\mu)$. By comparing Eqs. (2.2) and (2.3) it follows that

$$M(\mu) = \frac{\partial f(0,\mu)}{\partial k}[f(0,\mu)]^{-1}. \qquad (2.6)$$

It is easy to verify that $f(0,\mu)$ and $\partial f(0,\mu)/\partial k$ are integral functions of a complex variable μ.[7] Moreover, it is shown in Ref. 6 that the zeros of $f(0,\mu)$ are all real and positive, say $0 < \mu_1 < \mu_2 < \cdots$ and that if $N(\mu)$ is the number of bound states for S-wave scattering by the potential $-\mu q(r)$, then

$$n(\mu) = n \quad \text{for } \mu_n < \mu \leq \mu_{n+1}, \quad \text{for } n = 0,1,2,\ldots \quad (\text{where } \mu_0 = 0).$$

Clearly, the radius of convergence of $M(\mu)$ is μ_1. Therefore, if $0 \leq \mu \leq \mu_1$, that is, if the potential $-\mu q(r)$ is too weak to support a bound state, then the series for $M(\mu)$ [and hence the Born series for $F(\mathbf{k},\mathbf{k}_0)$] is absolutely and uniformly convergent with respect to $\mathbf{k}$ and $\mathbf{k}_0$.

3. APPLICATIONS

Let ρ denote the radius of convergence of the Born series $F(\mathbf{k},\mathbf{k}_0)$. In this section we give some examples of the lower bound μ_1 for ρ.

(1) If

$$q(r) = \begin{cases} 1, & 0 \leq r \leq a \\ 0, & a < r, \end{cases}$$

then $\mu_1, \mu_2, \cdots$ are the zeros of

$$J_{-1/2}(a\sqrt{\mu}) = \{2/a\sqrt{\mu}\}^{1/2}\cos a\sqrt{\mu}$$

and hence $\rho \geq \mu_1 = (\pi/2a)^2$.[8]

(2) If $q(r) = e^{-2\alpha r}/(1 + e^{-2\alpha r})^2$, $\alpha > 0$,

then the bound states are

$$E_n = -(\tfrac{1}{2}\sqrt{\mu} - 2\alpha n)^2,$$

and hence $\rho \geq \mu_1 = 2\sqrt{\alpha}$.

(3) If $q(r) = e^{-r}$, it is found that

$$\phi(r,0,\mu) = \alpha J_0(2\sqrt{\mu}\, e^{-r/2}) + \beta Y_0(2\sqrt{\mu}\, e^{-r/2}),$$

where J_0 and Y_0 are Bessel functions of the first and second kind, respectively,

$$\alpha = Y_0(2\sqrt{\mu})/(W\sqrt{\mu}), \qquad \beta = -J_0(2\sqrt{\mu})/(W\sqrt{\mu})$$

and

$$W = J_0(2\sqrt{\mu})Y_0'(2\sqrt{\mu}) - J_0'(2\sqrt{\mu})Y_0(2\sqrt{\mu}).$$

It follows that

$$f(0,\mu) = J_0(2\sqrt{\mu})/2\sqrt{\mu}\,W$$

and hence that $\rho \geq (a/2)^2$, where a is the least positive zero of J_0.

[1] W. Kohn, Rev. Mod. Phys. **26**, 292 (1954).
[2] A. Klein and C. Zemach, Ann. Phys. (N.Y.) **7**, 440 (1959).
[3] H. Davies, Nucl. Phys. **14**, 465 (1959).
[4] R. Huby, Nucl. Phys. **45**, 473 (1963).
[5] I. Manning, Phys. Rev. **139**, B495 (1965).
[6] P. J. Bushell, J. London Math. Soc. (1972) (to be published).
[7] R. D. Kemp and N. Levinson, Proc. Amer. Math. Soc. **10**, 82 (1959).
[8] P. J. Bushell, Quart. J. Math. (Oxford) **14**, 106 (1963), Sec. 5.

Nonmetrical Specification of Space-Time Sources

C. Martin Pereira*
Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 10 November 1971; Revised Manuscript Received 17 May 1972)

The role played by the energy-momentum conservation law in general relativity is examined. It is noted that this law can be interpreted in two ways. It may be thought of as a condition determining the evolution of the energy-momentum tensor density in time, or it may be thought of as a condition determining the metric. In the present paper, the second of these ways of thinking about the energy-momentum conservation law is explored. Einstein's nonvacuum gravitational field equations (which imply the conservation law) are examined. It is shown that given any analytic symmetric contravariant energy-momentum tensor density as a function of the space-time coordinates, a solution to the gravitational field equations always exists. Furthermore, this solution is such that the law of conservation of energy-momentum is satisfied. The proof uses a coordinate transformation method to exploit the covariance of the energy-momentum conservation law. Riquier's existence theorem enters as an important part of the proof, and a general discussion of Riquier's existence theorem from a physical point of view is given. Both the geodesic nature of the trajectories of free particles and the unit magnitude of the velocity 4-vector are discussed. An interpretation of the above-described results is given.

1. INTRODUCTION

Recently, a number of new results have been obtained regarding the existence of solutions to Einstein's nonvacuum field equations[1]

$$(-g)^{1/2} G^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu}. \tag{1.1}$$

The novelty of these results does not lie in the existence per se but rather in the variety of the functions $\tilde{T}^{\mu\nu}(x)$ compatible with this existence.

It will be convenient to consider certain unique features of the general relativistic conservation law

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{1.2}$$

Since the special relativistic form of this law

$$\tilde{T}^{\mu\nu}{}_{,\nu} = 0 \tag{1.3}$$

provides a real restriction on $\tilde{T}^{\mu\nu}$ in the sense that not every tensor density $\tilde{T}^{\mu\nu}$ satisfies (1.3), Eq. (1.2) is usually thought of as restricting $\tilde{T}^{\mu\nu}$. However, since (1.2) involves the metric it can also be thought of as restricting the metric alone while leaving $\tilde{T}^{\mu\nu}$ completely unrestricted.

Einstein's ten nonvacuum gravitational field equations

$$(-g)^{1/2} G^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu} \tag{1.4a}$$

imply the generally covariant conservation law

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{1.4b}$$

It is useful to consider the system of 14 equations (1.4a), (1.4b) which is equivalent to Eq. (1.4a) alone.

The basic result described in the present paper is a proof that the combined system (1.4) admits an unrestricted $\tilde{T}^{\mu\nu}(x)$ just as the generally covariant energy momentum law (1.2) does. The discussion below describes this result in somewhat more detail.

Equation (1.3), the special relativistic form of (1.4b), provides a real restriction on $\tilde{T}^{\mu\nu}$. Because of this, the combination of (1.4a) with (1.4b) is usually thought of as providing a restriction on $\tilde{T}^{\mu\nu}$. The basic mathematical result described in this paper shows that because of the structure of (1.4a) and because (1.4b) involves the metric, the combined gravitational system (1.4) can also be thought of as restricting the metric alone while leaving $\tilde{T}^{\mu\nu}$ completely undetermined.

Thus there are three ways of looking at the system (1.4). First, there is the viewpoint enunciated by Schrödinger[2,3] in which Eq. (1.4) is thought of as defining $\tilde{T}^{\mu\nu}$ in terms of the metric $g_{\mu\nu}$. According to this viewpoint, $\tilde{T}^{\mu\nu}$ is taken to be completely determined while $g_{\mu\nu}$ is completely undetermined. Second, there is a viewpoint most closely associated with the existence proof of Lichnerowicz.[4] This viewpoint takes the g_{ij} and the $\tilde{T}^{4\nu}$ to be determined by (1.4), while the $g_{\nu 4}$ and the $\tilde{T}^{ij}$ remain undetermined. Third, there is the viewpoint presented in the present paper. This viewpoint takes the g_{ij} and the $g_{\nu 4}$ to be determined by (1.4), while the $\tilde{T}^{\nu 4}$ and the $\tilde{T}^{ij}$ remain undetermined.

Riquier's existence theorem, the existence theorem used to prove this new result, is not wellknown. Therefore, a general description of the procedures used in applying the theorem to systems of equations is presented in Sec. 2. In addition, diagrams are pre-

$$\phi(r,0,\mu) = \alpha J_0(2\sqrt{\mu}\, e^{-r/2}) + \beta Y_0(2\sqrt{\mu}\, e^{-r/2}),$$

where J_0 and Y_0 are Bessel functions of the first and second kind, respectively,

$$\alpha = Y_0(2\sqrt{\mu})/(W\sqrt{\mu}), \qquad \beta = -J_0(2\sqrt{\mu})/(W\sqrt{\mu})$$

and

$$W = J_0(2\sqrt{\mu})Y_0'(2\sqrt{\mu}) - J_0'(2\sqrt{\mu})Y_0(2\sqrt{\mu}).$$

It follows that

$$f(0,\mu) = J_0(2\sqrt{\mu})/2\sqrt{\mu}\,W$$

and hence that $\rho \geq (a/2)^2$, where a is the least positive zero of J_0.

[1] W. Kohn, Rev. Mod. Phys. **26**, 292 (1954).
[2] A. Klein and C. Zemach, Ann. Phys. (N.Y.) **7**, 440 (1959).
[3] H. Davies, Nucl. Phys. **14**, 465 (1959).
[4] R. Huby, Nucl. Phys. **45**, 473 (1963).
[5] I. Manning, Phys. Rev. **139**, B495 (1965).
[6] P. J. Bushell, J. London Math. Soc. (1972) (to be published).
[7] R. D. Kemp and N. Levinson, Proc. Amer. Math. Soc. **10**, 82 (1959).
[8] P. J. Bushell, Quart. J. Math. (Oxford) **14**, 106 (1963), Sec. 5.

Nonmetrical Specification of Space-Time Sources

C. Martin Pereira*
Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 10 November 1971; Revised Manuscript Received 17 May 1972)

The role played by the energy-momentum conservation law in general relativity is examined. It is noted that this law can be interpreted in two ways. It may be thought of as a condition determining the evolution of the energy-momentum tensor density in time, or it may be thought of as a condition determining the metric. In the present paper, the second of these ways of thinking about the energy-momentum conservation law is explored. Einstein's nonvacuum gravitational field equations (which imply the conservation law) are examined. It is shown that given any analytic symmetric contravariant energy-momentum tensor density as a function of the space-time coordinates, a solution to the gravitational field equations always exists. Furthermore, this solution is such that the law of conservation of energy-momentum is satisfied. The proof uses a coordinate transformation method to exploit the covariance of the energy-momentum conservation law. Riquier's existence theorem enters as an important part of the proof, and a general discussion of Riquier's existence theorem from a physical point of view is given. Both the geodesic nature of the trajectories of free particles and the unit magnitude of the velocity 4-vector are discussed. An interpretation of the above-described results is given.

1. INTRODUCTION

Recently, a number of new results have been obtained regarding the existence of solutions to Einstein's nonvacuum field equations[1]

$$(-g)^{1/2} G^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu}. \tag{1.1}$$

The novelty of these results does not lie in the existence per se but rather in the variety of the functions $\tilde{T}^{\mu\nu}(x)$ compatible with this existence.

It will be convenient to consider certain unique features of the general relativistic conservation law

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{1.2}$$

Since the special relativistic form of this law

$$\tilde{T}^{\mu\nu}{}_{,\nu} = 0 \tag{1.3}$$

provides a real restriction on $\tilde{T}^{\mu\nu}$ in the sense that not every tensor density $\tilde{T}^{\mu\nu}$ satisfies (1.3), Eq. (1.2) is usually thought of as restricting $\tilde{T}^{\mu\nu}$. However, since (1.2) involves the metric it can also be thought of as restricting the metric alone while leaving $\tilde{T}^{\mu\nu}$ completely unrestricted.

Einstein's ten nonvacuum gravitational field equations

$$(-g)^{1/2} G^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu} \tag{1.4a}$$

imply the generally covariant conservation law

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{1.4b}$$

It is useful to consider the system of 14 equations (1.4a), (1.4b) which is equivalent to Eq. (1.4a) alone.

The basic result described in the present paper is a proof that the combined system (1.4) admits an unrestricted $\tilde{T}^{\mu\nu}(x)$ just as the generally covariant energy momentum law (1.2) does. The discussion below describes this result in somewhat more detail.

Equation (1.3), the special relativistic form of (1.4b), provides a real restriction on $\tilde{T}^{\mu\nu}$. Because of this, the combination of (1.4a) with (1.4b) is usually thought of as providing a restriction on $\tilde{T}^{\mu\nu}$. The basic mathematical result described in this paper shows that because of the structure of (1.4a) and because (1.4b) involves the metric, the combined gravitational system (1.4) can also be thought of as restricting the metric alone while leaving $\tilde{T}^{\mu\nu}$ completely undetermined.

Thus there are three ways of looking at the system (1.4). First, there is the viewpoint enunciated by Schrödinger[2,3] in which Eq. (1.4) is thought of as defining $\tilde{T}^{\mu\nu}$ in terms of the metric $g_{\mu\nu}$. According to this viewpoint, $\tilde{T}^{\mu\nu}$ is taken to be completely determined while $g_{\mu\nu}$ is completely undetermined. Second, there is a viewpoint most closely associated with the existence proof of Lichnerowicz.[4] This viewpoint takes the g_{ij} and the $\tilde{T}^{4\nu}$ to be determined by (1.4), while the $g_{\nu 4}$ and the $\tilde{T}^{ij}$ remain undetermined. Third, there is the viewpoint presented in the present paper. This viewpoint takes the g_{ij} and the $g_{\nu 4}$ to be determined by (1.4), while the $\tilde{T}^{\nu 4}$ and the $\tilde{T}^{ij}$ remain undetermined.

Riquier's existence theorem, the existence theorem used to prove this new result, is not wellknown. Therefore, a general description of the procedures used in applying the theorem to systems of equations is presented in Sec. 2. In addition, diagrams are pre-

sented summarizing the structure of various existence proofs which use Riquier's theorem. The last diagram in Sec. 2 summarizes the existence proof for the gravitational field equations (1.4a).

2. INTEGRABILITY[5]

I wish to discuss here the methods necessary to prove the integrability of systems of partial differential equations. Consider the simple system of equations

$$\psi_{,\mu} = A_\mu, \tag{2.1}$$

where ψ is an unknown function and the A_μ are given functions of the variables x, y, z, t. The integrability conditions for the system (2.1) are

$$A_{[\mu,\nu]} = 0. \tag{2.2}$$

These conditions are sufficient conditions for the existence of the function ψ. Before going further, I should say a few words as to terminology. I will refer to integrability conditions like Eq. (2.2) as an "equation," a "condition," or a "restriction." I will never refer to it as an identity unless it is actually equivalent to

$$0 = 0. \tag{2.3}$$

The reason for this distinction can be seen when one considers the nonvacuum Einstein equations

$$\tilde{G}^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu}, \tag{2.4}$$

where $\tilde{T}^{\mu\nu}$ is a symmetric tensor density. The Bianchi identity (a true identity)

$$\tilde{G}^{\mu\nu}{}_{;\nu} = 0 \tag{2.5}$$

implies the equation

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{2.6}$$

But Eq. (2.6) is not an identity since, for example, when $\tilde{T}^{\mu\nu}(x)$ is given, Eq. (2.6) clearly becomes a restriction on the $g_{\mu\nu}$.

Riquier's existence theorem describes the method of deriving the integrability conditions for a general system of partial differential equations. This method is analogous to the method of deriving Eq. (2.2) from (2.1); but to describe it in detail requires quite an extensive discussion.

Riquier's book on the subject is rather difficult to obtain. One of the more accessible sources for Riquier's theorem is Ritt's book.[6] However, Ritt's presentation is not well suited to the equations of physics since his notation gives no explicit recognition to given functions like A_μ in Eq. (2.1), or $\tilde{T}^{\mu\nu}$ in Eq. (2.4), or $\tilde{J}^\mu$ in Diagram III. And in physics, the restrictions upon these given functions generated by the integrability conditions have physical meaning.[7] For this reason, a simplified and physically oriented discussion of Riquier's theorem is presented below.

In addition to the unknown functions y_m, the original system, hereafter called system S, may contain given functions J^k. We will denote the system S by the following symbolic equation:

$$S\,[y_m, J^k] = 0, \tag{2.7}$$

where the brackets signify that the functions S involve the y_m, the J^k, *and* their derivatives.

A definite procedure is prescribed in the theorem for deriving the integrability conditions of system S. First, one must choose an order for the derivatives and perform those algebraic operations on the system necessary to put the system S in a form which satisfies criteria (a)–(c) of Appendix B. This entire process of ordering and algebraic manipulation will hereafter be called *Procedure O*.

For simplicity and definiteness, we will consider as unknown only the minimum number of functions required to complete the ordering process. Thus, once ordering is completed, all functions whose derivatives do not appear in one or more first members,[8] will be considered to be given functions. In some cases in order to complete the ordering process, it will be necessary to take some of the given functions to be unknown functions. When we wish to emphasize that the ordering process includes such a change in the number of unknown functions, we will call it *Procedure O**. Similarly if we wish to emphasize that the process includes no change in the number of unknown functions, we will call it *Procedure O′*.

If each of the first members is the derivative of a different unknown function, then there are no integrability conditions, and we say that *system S has integrability of the zeroth kind*. If there is at least one unknown function which has derivatives appearing in two or more different first members, then there will be integrability conditions. These are derived according to the following procedure,[9] which for brevity we will call *Procedure D*. One first writes down an integrability condition by differentiating two appropriate equations of the system and eliminating by subtraction the first members of the resulting equations. One then uses the equations of the system and their derivatives to eliminate the principal derivatives occuring in the integrability condition. One does this step by step, starting with the highest principal derivative and proceeding to the lowest.[10] At the end of this procedure one obtains the integrability conditions for the system. We denote the resulting integrability conditions by the equation

$$I_1[y_m, J^k] = 0. \tag{2.8}$$

Both the answer to the question of existence of solutions and also the structure of the system S depend on the nature of Eq. (2.8). We shall next proceed to discuss the various categories. A summary of this discussion in diagrammatic form is given in Diagram I, which will be found at the end of this chapter. We suggest that the reader refer to Diagram I from time to time to clarify the relationship between the various categories.

Integrability Conditions Are the Identity

If Eq. (2.8), hereafter called I_1, is actually the identity $0 = 0$, then a solution to system S exists. Since *one* application of Procedure D was necessary to prove existence of a solution, we say that system S has integrability of the *first* kind. In Diagram II, Einstein's vacuum field equations are shown to be an example of this case.

DIAGRAM I. Summary of general case.

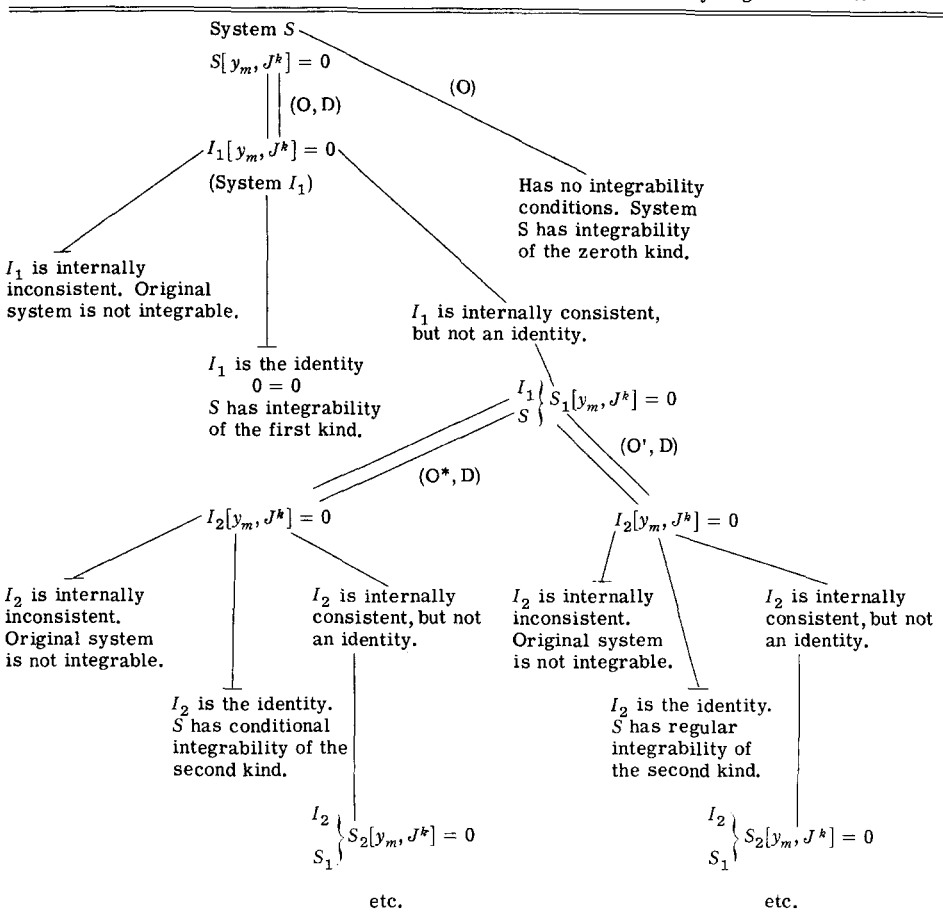
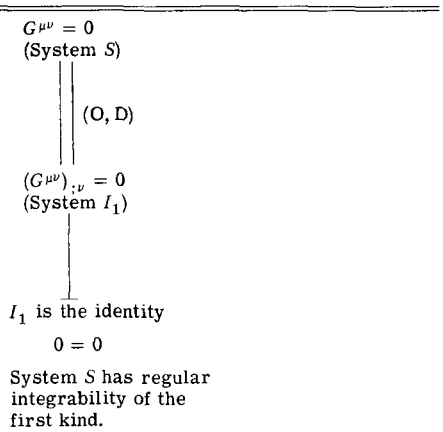


DIAGRAM II. Example of regular integrability of the first kind.



Integrability Conditions Are Internally Inconsistent

If the system I_1 is internally inconsistent, then of course no solution to system S exists.

Integrability Condition I_1 Is Internally Consistent, but Is Not the Identity

In this case one must combine I_1 with the system S, and apply first Procedure O, and then Procedure D. In order to apply Procedure O, it may be necessary to consider certain of the given functions to be unknown functions. And several different choices for these new unknown functions may be possible. Each different choice corresponds to a different mathematical problem. After Procedure O has been applied, one applies Procedure D to derive a new set of integrability conditions I_2. If the new set I_2 is internally inconsistent, of course, the system S is not integrable.

If this new set I_2 is an identity, then one concludes that the original system S has integrability of the second kind. Suppose that in applying the ordering procedure to the combined system S_1, it was necessary to increase the number of unknown functions. Then, if I_2 is the identity we say that the original system S has *conditional* integrability of the second kind. Here the term conditional refers to the fact that integrability is conditioned on an increase in the number of unknown functions. Einstein's and Maxwell's nonvacuum field equations provide examples of this case. See Diagrams III and IV.

If I_2 is internally consistent but not an identity, it must be combined with the system S_1 and Procedures O and D applied all over again. In general, this derivation of new integrability conditions could go on indefinitely in which case Riquier's procedure would give no conclusion as to the integrability of the original system S.

At the end of this section is a general diagram of the various possibilities described above. And on later pages there are similar diagrams of specific cases. One point should be emphasized about these diagrams. The first statement of an integrability condition in each diagram is given in a form which makes clear what operations on the preceding system of equations are used to produce the integrability condition. The operations we refer to are of course the operations dictated by Procedure D of Riquier's theorem.

DIAGRAM III. Example of conditional integrability of the second kind.

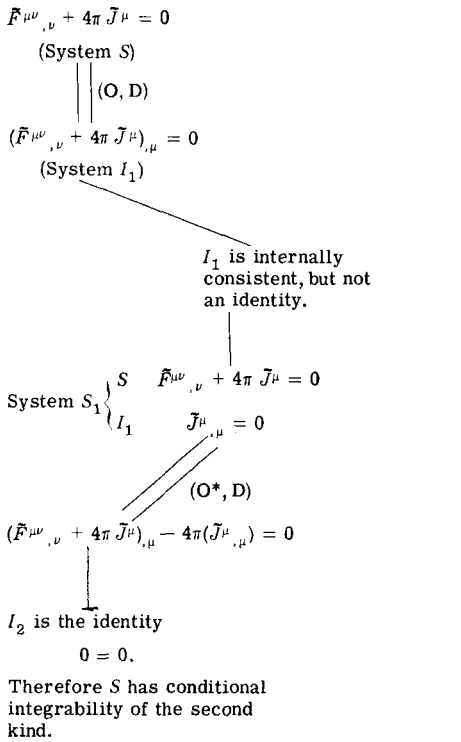
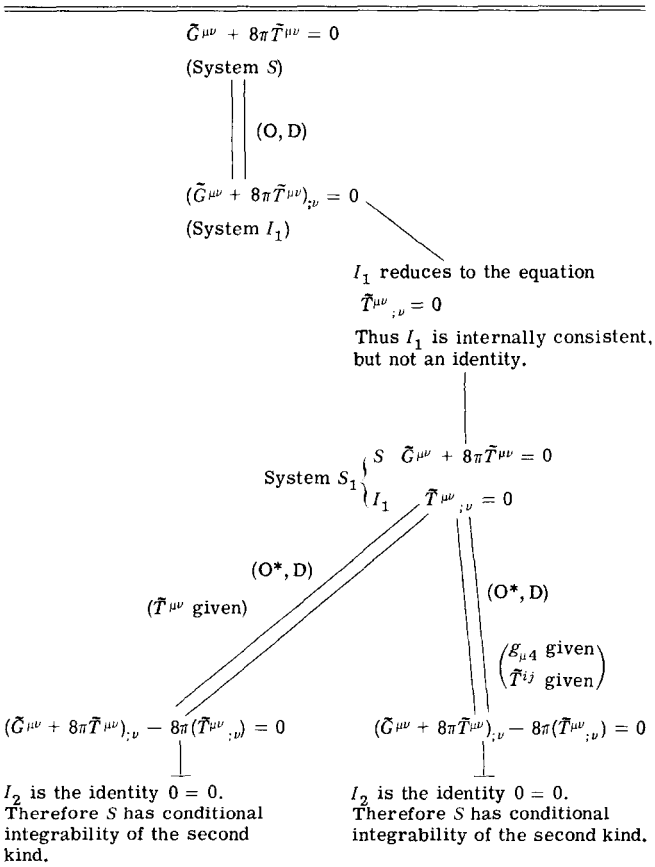


DIAGRAM IV.



Let us give an example. In Diagram III, integrability condition I_1 is first written

$$(\tilde{F}^{\mu\nu}{}_{,\nu} + 4\pi \tilde{J}^{\mu})_{,\mu} = 0. \tag{2.9}$$

This denotes that the effect of following Riquier's procedure is to take the divergence of Maxwell's equations. Then in the same diagram, integrability condition I_2 is written

$$(\tilde{F}^{\mu\nu}{}_{,\nu} + 4\pi \tilde{J}^{\mu})_{,\mu} - 4\pi(\tilde{J}^{\mu}{}_{,\mu}) = 0. \tag{2.10}$$

This denotes that the effect of following Riquier's procedure for system S_1 is to take the divergence of the first component of S_1 and to subtract from it 4π times the second component of S_1. A similar practice is followed in the other diagrams.

3. THE FIELD EQUATIONS

Consider Einstein's field equations

$$(-g)^{1/2}(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = -8\pi\tilde{T}^{\mu\nu}, \tag{3.1}$$

which may be rewritten in the form

$$(-g)^{1/2}R^{\mu\nu} = -8\pi(\tilde{T}^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}\tilde{T}). \tag{3.2}$$

If one is given some symmetric tensor density $\tilde{T}^{\mu\nu}(x)$, what restrictions must it satisfy in order to guarantee the existence of a metric satisfying Eq. (3.1)? To answer this question, one applies Riquier's theorem to (3.2). Now to simplify the necessary calculations, it is convenient to examine Eq. (3.1) in harmonic coordinates.[11] To do this, one considers the combined system

$$(-g)^{1/2}\,(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = -8\pi\tilde{T}^{\mu\nu}, \tag{3.3a}$$

$$[(-g)^{1/2}g^{\alpha\beta}]_{,\beta} = 0. \tag{3.3b}$$

Here the harmonic coordinate condition is expressed by Eq. (3.3b), and the harmonic coordinate system is labeled x.

Furthermore, to avoid any limitations on $\tilde{T}^{\mu\nu}$ that would result from the coordinate conditions (3.3b), one imagines that the $\tilde{T}^{\mu'\nu'}(x')$ are given in some arbitrary coordinate system x', while the metric $g^{\mu\nu}(x)$ is found in the harmonic coordinate system x. With this point of view in mind, the field equations become

$$(-g)^{1/2}\,(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = -8\pi A^{\mu}{}_{\alpha'}[f]A^{\nu}{}_{\tau'}[f]\tilde{T}^{\alpha'\tau'}, \tag{3.4a}$$

$$[1/(-g)^{1/2}][(-g)^{1/2}g^{\mu\beta}]_{,\beta} = 0. \tag{3.4b}$$

Several comments are required to explain the quantities $A^{\mu}{}_{\alpha'}[f]$ appearing in Eqs. (3.4a). Let the transformation from the x coordinate system to the x' coordinate system be

$$x'^{\alpha} = f^{\alpha'}(x), \tag{3.5a}$$

and let the inverse transformation be

$$x^{\tau} = F^{\tau}(x'). \tag{3.5b}$$

Then the quantities $A^{\mu}{}_{\alpha'}$ are defined by

$$A^{\mu}{}_{\alpha'} = \frac{\partial x^{\mu}}{\partial x^{\alpha'}} = \frac{\partial F^{\mu}}{\partial x^{\alpha'}} \tag{3.6a}$$

and

$$A^{\tau'}_{\mu} = \frac{\partial x^{\tau'}}{\partial x^{\mu}} = \frac{\partial f^{\tau'}}{\partial x^{\mu}}. \tag{3.6b}$$

Then the following identity holds

$$A^{\mu}{}_{\alpha'}\frac{\partial f^{\alpha'}}{\partial x^{\nu}} \equiv \delta^{\mu}{}_{\nu}. \tag{3.7}$$

Equation (3.7) can be used to express the quantities $A^{\mu}{}_{\alpha'}$, as algebraic functions of the derivatives $\partial f^{\alpha'}/\partial x^{\nu}$. In Eq. (3.4) the quantities $A^{\tau}{}_{\alpha'}$ are to be considered to be expressed in terms of the $\partial f^{\alpha'}/\partial x^{\nu}$ by means of Eq. (3.7).

One can show[11] that when (3.3b) holds, the Ricci tensor $R^{\mu\nu}$ can be written

$$R^{\mu\nu} = -\tfrac{1}{2}g^{\alpha\beta}g^{\mu\nu}{}_{,\alpha\beta} + g^{\alpha\rho}g^{\beta\sigma}\Gamma^{\mu}{}_{\rho\sigma}\Gamma^{\nu}{}_{\alpha\beta}. \tag{3.8}$$

Using (3.8) Eqs. (3.4) can be written

$$\begin{aligned}(-g)^{1/2}\,&(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)\\ &= -\tfrac{1}{2}g^{\alpha\beta}[(-g)^{1/2}\,(g^{\mu\nu}{}_{,\alpha} - \tfrac{1}{2}g^{\mu\nu}g_{\sigma\tau}g^{\sigma\tau}{}_{,\alpha})]_{,\beta}\\ &\quad + K^{\mu\nu}(g^{\mu\nu}, g^{\mu\nu}{}_{,\alpha}) = -8\pi A^{\mu}{}_{\alpha'}[f]A^{\nu}{}_{\tau'}[f]\tilde{T}^{\alpha'\tau'},\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}K^{\mu\nu}(g^{\mu\nu}, g^{\mu\nu}{}_{,\alpha}) &= (-g)^{1/2}g^{\alpha\rho}g^{\beta\sigma}\Gamma^{\mu}{}_{\rho\sigma}\Gamma^{\nu}{}_{\alpha\beta}\\ &\quad - \tfrac{1}{2}g^{\mu\nu}g^{\alpha\rho}g^{\beta\sigma}\Gamma_{\rho\sigma,\mu}\Gamma^{\mu}{}_{\alpha\beta} + \tfrac{1}{2}g^{\alpha\beta}[(-g)^{1/2}]_{,\beta}\\ &\quad \times (g^{\mu\nu}{}_{,\alpha} - \tfrac{1}{2}g^{\mu\nu}g_{\sigma\tau}g^{\sigma\tau}{}_{,\alpha}) - \tfrac{1}{4}(-g)^{1/2}g^{\alpha\beta}\\ &\quad \times g^{\sigma\tau}{}_{,\alpha}(g^{\mu\nu}g_{\sigma\tau})_{,\beta}.\end{aligned}$$

It will be convenient to replace the metric tensor $g^{\mu\nu}$ by the metric tensor density

$$\mathfrak{g}^{\mu\nu} \equiv (-g)^{1/2}g^{\mu\nu} \tag{3.10}$$

and to express Einstein's equations entirely in terms of the $\mathfrak{g}^{\mu\nu}$. We shall also use the determinant $\mathfrak{g}$ of $\mathfrak{g}^{\mu\nu}$

$$\mathfrak{g} \equiv |\mathfrak{g}^{\mu\nu}| = [(-g)^{1/2}]^4/g = g. \tag{3.11}$$

Equations (3.10) and (3.11) imply that

$$(-g)^{1/2}\,(g^{\mu\nu}{}_{,\alpha} - \tfrac{1}{2}g^{\mu\nu}g_{\omega\tau}g^{\omega\tau}{}_{,\alpha}) = \mathfrak{g}^{\mu\nu}{}_{,\alpha}. \tag{3.12}$$

Note also that

$$g^{\mu\nu} = [1/(-\mathfrak{g})^{1/2}]\,\mathfrak{g}^{\mu\nu} \tag{3.13}$$

and

$$g^{\mu\nu}{}_{,\alpha} = [\mathfrak{g}^{\mu\nu}/(-\mathfrak{g})^{1/2}]_{,\alpha}. \tag{3.14}$$

Using Eqs. (3.12)–(3.14) the field equations (3.9) can be expressed in terms of $\mathfrak{g}^{\mu\nu}$:

$$\begin{aligned}(-g)^{1/2}&[R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R]\\ &= -\tfrac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu\nu}{}_{,\alpha\beta}\\ &\quad + K^{\mu\nu}\left[\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}, \left(\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\right)_{,\tau}\right]\\ &= -8\pi A^{\mu}{}_{\alpha'}[f]\,A^{\nu}{}_{\tau'}[f]\tilde{T}^{\alpha'\tau'}.\end{aligned} \tag{3.15a}$$

Express Eq. (3.4b) in terms of $\mathfrak{g}^{\mu\nu}$ and obtain

$$\mathfrak{g}^{\mu\nu}{}_{,\nu} = 0. \tag{3.15b}$$

To apply Riquier's theorem, first choose an order for the derivatives[12]

$$\mathfrak{g}^{\mu\nu}{}_{,4} > \mathfrak{g}^{\mu\nu}{}_{,3} > \mathfrak{g}^{\mu\nu}{}_{,2} > \mathfrak{g}^{\mu\nu}{}_{,1} \tag{3.16a}$$

$$\mathfrak{g}^{13}{}_{,3} > \mathfrak{g}^{14}{}_{,4} > \mathfrak{g}^{12}{}_{,2} > \mathfrak{g}^{11}{}_{,1} \tag{3.16b}$$

$$\mathfrak{g}^{23}{}_{,3} > \mathfrak{g}^{24}{}_{,4} > \mathfrak{g}^{22}{}_{,2} > \mathfrak{g}^{21}{}_{,1} \tag{3.16c}$$

$$\mathfrak{g}^{33}{}_{,3} > \mathfrak{g}^{34}{}_{,4} > \mathfrak{g}^{32}{}_{,2} > \mathfrak{g}^{31}{}_{,1} \tag{3.16d}$$

$$\mathfrak{g}^{43}{}_{,3} > \mathfrak{g}^{44}{}_{,4} > \mathfrak{g}^{42}{}_{,2} > \mathfrak{g}^{41}{}_{,1} \tag{3.16e}$$

$$\mathfrak{g}^{13} > \mathfrak{g}^{23} > \mathfrak{g}^{33} > \mathfrak{g}^{43} > \begin{pmatrix}\text{all other}\\ \mathfrak{g}^{\mu\nu}\end{pmatrix} \tag{3.16f}$$

The order relations given above are mutually self-consistent. The process of choosing these order relations is by no means trivial. The particular choice given above greatly simplifies the calculation of the integrability conditions via Riquier's procedure. These order relations lead to the following principal derivatives for the system (3.15):

$\mathfrak{g}^{\mu\nu}{}_{,44}$ for Eqs. (3.15a)

$\mathfrak{g}^{\mu 3}{}_{,3}$ for Eqs. (3.15b)

Only four unknowns appear twice in Eq. (3.15), namely the $\mathfrak{g}^{\mu 3}$. And these four functions give rise to four integrability conditions which have the form

$$\mathfrak{g}^{\mu 3}{}_{,344} - \mathfrak{g}^{\mu 3}{}_{,443}. \tag{3.16g}$$

Written out explicitly, the integrability conditions are

$$\begin{aligned}&-\frac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu 3}{}_{,\alpha\beta 3} + \frac{1}{2}\frac{\mathfrak{g}^{44}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu\nu}{}_{,\nu 44}\\ &\qquad + \text{second-order terms} = -8\pi(A^{\mu}{}_{\alpha'}A^{3}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,3}.\end{aligned} \tag{3.17}$$

The terms in $g^{\mu 3}{}_{,443}$ cancel out of Eq. (3.17). The next step in Riquier's procedure is to remove from (3.17) the $g^{\mu 3}{}_{,ij3}$ and $g^{\mu 3}{}_{,i43}$ terms using the ij and $i4$ derivatives of (3.15b), $i = 1, 2, 3$. The result is

$$\begin{aligned}&-\frac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu 3}{}_{,\alpha\beta 3} + \frac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu\nu}{}_{,\nu\alpha\beta}\\ &\qquad + \text{second-order terms} = -8\pi(A^{\mu}{}_{\alpha'}A^{3}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,3}.\end{aligned} \tag{3.18}$$

The only third-order terms remaining in (3.18) are $g^{\mu a}{}_{,a\alpha\beta}(a = 1, 2, 4)$. The next step in Riquier's process removes these terms using the three-divergence of (3.15a). The third order portion of this three-divergence has the form $g^{\mu a}{}_{,a\alpha\beta}(a = 1, 2, 4)$. The result is

$$\begin{aligned}&-\frac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}\mathfrak{g}^{\mu\nu}{}_{,\alpha\beta\nu} + \frac{1}{2}\frac{\mathfrak{g}^{\alpha\beta}}{(-\mathfrak{g})^{1/2}}g^{\mu\nu}{}_{,\nu\alpha\beta}\\ &\qquad + \text{second-order terms} = -8\pi(A^{\mu}{}_{\alpha'}A^{\nu}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,\nu}\end{aligned} \tag{3.19}$$

Note that the third-order terms on the left-hand side of (3.19) completely cancel one another. Also if one expresses the remaining terms on the left-handside of (3.19) in terms of $R^{\mu\nu}$ one obtains

$$\begin{aligned}(-g)^{1/2}\,&(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)_{,\nu} + \tfrac{1}{2}[\mathfrak{g}^{\alpha\beta}/(-\mathfrak{g})^{1/2}]\,g^{\mu\nu}{}_{,\nu\alpha\beta}\\ &= -8\pi(A^{\mu}{}_{\alpha'}A^{\nu}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,\nu}.\end{aligned} \tag{3.20}$$

As can be seen from Eq. (3.19), the left-hand side of (3.20) in reality involves only second-order terms in the $g^{\mu\nu}$. The particular form of these second-order terms can be found most quickly by using the following identity (Bianchi's identity):

$$[(-g)^{1/2}(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)]_{,\nu} + (-g)^{1/2}\Gamma^{\mu}{}_{\alpha\beta}(R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R) = 0. \quad (3.21)$$

Substitute Eq. (3.15b) into (3.20) to remove the second term on the left-hand side of (3.20). Next make similar substitution of Eq. (3.21) in Eq. (3.20) and obtain

$$(-g)^{1/2}\Gamma^{\mu}{}_{\alpha\beta}(R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R) = -8\pi(A^{\mu}{}_{\alpha'}A^{\nu}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,\nu}. \quad (3.22)$$

Having rewritten Eq. (3.19) in a simpler and more explicit form, one can now complete Riquier's procedure. Each of the coefficients of $\Gamma^{\mu}{}_{\alpha\beta}$ in (3.22) contains a principal derivative of the form $g^{\alpha\beta}{}_{,44}$, and the final step in Riquier's procedure is to eliminate these principal derivatives using Eq. (3.15a). The result is

$$-8\pi[(A^{\mu}{}_{\alpha'}A^{\nu}{}_{\tau'}\tilde{T}^{\alpha'\tau'})_{,\nu} + \Gamma^{\mu}{}_{\alpha\beta}\tilde{T}^{\alpha\beta}] = 0 \quad (3.23)$$

which is the integrability condition for the gravitational field equations (3.4).

This can also be written

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \quad (3.24)$$

Thus the energy–momentum conservation law is the integrability condition that results from the application of Riquier's procedure to the field equations (3.15). Now the given quantities are the $\tilde{T}^{\mu'\nu'}(x')$, and the dependent functions of the system are $g_{\mu\nu}(x)$. The next step is to express Eq. (3.24) in terms of $\tilde{T}^{\mu'\nu'}(x')$, $g_{\mu\nu}(x)$, and the transformation functions $f^{\nu'}(x)$. To do this, first rewrite Eq. (3.24) in the x' coordinate system

$$\tilde{T}^{\nu'\mu'}{}_{;\mu'} = \tilde{T}^{\nu'\mu'}{}_{,\mu'} + \Gamma^{\nu'}{}_{\beta'\mu'}\tilde{T}^{\beta'\mu'} = 0. \quad (3.25)$$

Now the relationship between the Christoffel symbols $\Gamma^{\mu'}{}_{\alpha'\beta'}$ in the x' and the $\Gamma^{\mu}{}_{\alpha\beta}$ in the x coordinate system is

$$\Gamma^{\nu'}{}_{\beta'\mu'} = \Gamma^{\tau}{}_{\gamma\phi}A^{\nu'}{}_{\tau}A^{\gamma}{}_{\beta'}A^{\phi}{}_{\mu'} + A^{\nu'}{}_{\tau}(\partial_{\mu'}A^{\tau}{}_{\beta'}). \quad (3.26)$$

The identity

$$\partial_{\mu'}(\delta^{\nu'}{}_{\beta'}) = \partial_{\mu'}(A^{\nu'}{}_{\tau}A^{\tau}{}_{\beta'}) = 0 \quad (3.27)$$

may be written

$$A^{\nu'}{}_{\tau}(\partial_{\mu'}A^{\tau}{}_{\beta'}) = -A^{\tau}{}_{\beta'}(A^{\alpha}{}_{\mu'}\partial_{\alpha}A^{\nu'}{}_{\tau}). \quad (3.28)$$

Using Eq. (3.28), Eq. (3.26) may be written

$$\Gamma^{\nu'}{}_{\beta'\mu'} = \Gamma^{\tau}{}_{\gamma\phi}A^{\nu'}{}_{\tau}A^{\gamma}{}_{\beta'}A^{\phi}{}_{\mu'} - A^{\tau}{}_{\beta'}A^{\alpha}{}_{\mu'}\partial_{\alpha}A^{\nu'}{}_{\tau}. \quad (3.29)$$

Substitute Eq. (3.29) in Eq. (3.25) and obtain

$$[\partial_{\alpha}A^{\nu'}{}_{\tau}(A^{\tau}{}_{\beta'}A^{\alpha}{}_{\mu'}) - \Gamma^{\tau}{}_{\gamma\phi}A^{\nu'}{}_{\tau}A^{\gamma}{}_{\beta'}A^{\phi}{}_{\mu'}]\tilde{T}^{\beta'\mu'} - \tilde{T}^{\nu'\mu'}{}_{,\mu'} = 0 \quad (3.30)$$

as the final form of the integrability condition for the gravitational field equations (3.4).

Since the integrability condition equation (3.30) is not an identity, Riquier's procedure has so far given no answer to the question of integrability[13] of the field equations (3.4). To obtain an answer one must affix equation (3.30) to the system of 14 equations (3.4) and apply Riquier's procedure to the combined system of 18 equations given below:

$$(-g)^{1/2}(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = -8\pi A^{\mu}{}_{\alpha'}[f']A^{\nu}{}_{\tau'}[f']\tilde{T}^{\alpha'\tau'}, \quad (3.31a)$$

$$g^{\mu\beta}{}_{,\beta} = 0, \quad (3.31b)$$

$$\left(\frac{\partial^2 f^{\nu'}}{\partial x^{\alpha}\partial x^{\tau}}(A^{\tau}{}_{\beta'}A^{\alpha}{}_{\mu'}) - \Gamma^{\tau}{}_{\gamma\phi}A^{\nu'}{}_{\tau}A^{\gamma}{}_{\beta'}A^{\phi}{}_{\mu'}\right)\tilde{T}^{\beta'\mu'} - \tilde{T}^{\nu'\mu'}{}_{,\mu'} = 0. \quad (3.31c)$$

The principal derivatives of the system (3.31) are

$g^{\mu\nu}{}_{,44}$ for Eqs. (3.31a),

$g^{\mu 3}{}_{,3}$ for Eqs. (3.31b),

$f^{\nu'}{}_{,44} \equiv \dfrac{\partial}{\partial x^4}\left(\dfrac{\partial f^{\nu'}}{\partial x^4}\right)$ for Eqs. (3.31c).

Thus the number of unknown functions has been increased from 10 to 14. $\tilde{T}^{\nu'\mu'}[f']$ is given and the system of field equations (3.31) is thought of as determining the $g^{\mu\nu}(x)$ and the transformation functions $f^{\nu'}(x)$.

Since the principal derivatives for Eqs. (3.31a) and (3.31b) are the same as for (3.15a) and (3.15b), Riquier's procedure leads as before to Eq. (3.30). This time, however, the derivatives $f^{\nu'}{}_{,44}$ which equation (3.30) contains are principal derivatives. Riquier's procedure now requires that these derivatives be eliminated by substituting for the $f^{\nu'}{}_{,44}$ as defined by Eqs. (3.31c). Since Eqs. (3.30) and (3.31c) are identical, this last step in Riquier's procedure leads to the identity

$$0 = 0, \quad (3.32)$$

which proves that the system of Eqs. (3.31) has an analytic solution. And an analytic solution exists regardless of the form one is given for the ten analytic functions $\tilde{T}^{\mu'\nu'}(f')$. In Appendix B, it is shown in addition that the initial conditions can always be chosen so that both the coordinate transformation (3.5a) and the corresponding inverse transformation (3.5b) are nonsingular. Once such a solution to the system (3.31) corresponding to nonsingular transformation functions $f^{\nu'}$ and F^{α} has been found giving $g^{\mu\nu}(x)$ and $f^{\nu'}(x)$, one can transform that solution into the x' coordinate system to obtain $g^{\mu'\nu'}(x')$. From this using Eq. (3.10) one can obtain the $g^{\mu'\nu'}(x')$ corresponding to the given functions $\tilde{T}^{\mu'\nu'}(x')$. This completes the proof of the following theorem.

Theorem 1: Given any symmetric[14–16] analytic tensor density $\tilde{T}^{\mu'\nu'}(x')$, there always exists a corresponding metric $g^{\mu'\nu'}(x')$ which satisfies the field equations (3.1).

4. CONCLUSION

In the introduction to this paper, three mathematical points of view concerning Einstein's field equations (1.4a) and their consequence, the law of conservation of energy, Eqs. (1.4b) were discussed. According to the first viewpoint, Eqs. (1.4) are considered to be restrictions on the $\tilde{T}^{\mu\nu}$. According to the second viewpoint, Eqs. (1.4) are considered to be restrictions on the g_{ij} and the $\tilde{T}^{\nu 4}$. According to the third viewpoint, Eqs. (1.4) are considered to be restrictions on the $g_{\mu\nu}$.

If one adopts the first mathematical viewpoint, one chooses the metric arbitrarily beforehand, and then one proceeds to calculate the corresponding energy–momentum tensor density. If one adopts the third mathematical viewpoint, one arbitrarily chooses the symmetric energy–momentum tensor density $\tilde{T}^{\mu\nu}$ beforehand, and then one proceeds to calculate the corresponding metric. Theorem 1 states that such a calculation of the metric is always possible. (See Section 3).

To the first mathematical point of view there corresponds the physical viewpoint according to which Eqs. (1.4) are thought of as a definition of the matter tensor density $\tilde{T}^{\mu\nu}$. According to this view, regions where matter is present are defined to be regions where the left-hand side of (1.4a) is nonzero. The presence of matter is no more than an interpretation assigned to the nonzero values of $(-g)^{1/2}(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R)$. (See Ref. 2.)

To the second mathematical point of view there corresponds a widely accepted physical viewpoint according to which the conservation law (1.4b) is thought of as expressing a property of the energy–momentum tensor density $\tilde{T}^{\mu\nu}$, while the field equations (1.4a) are thought of as expressing a property of the gravitational field $g_{\mu\nu}$.

To the third mathematical point of view there coresponds the physical viewpoint according to which Eqs. (1.4) are thought of as a definition of the metric $g_{\mu\nu}$. The metric is thought of as no more than an interpretation of the properties of the symmetric energy–momentum tensor density $\tilde{T}^{\mu\nu}$. Theorem 1 implies that such a physical viewpoint is mathematically self-consistent, provided that one is considering only Eqs. (1.4). However, to establish the physical self-consistency of this third physical viewpoint, one must show the quantities $\tilde{T}^{\mu\nu}(x)$ to be obtainable by physical measurements which do not involve the metric. Such a complete analysis of the physical meaning of $\tilde{T}^{\mu\nu}(x)$ is not given here. The present paper does include, however, a discussion of the special case when $\tilde{T}^{\mu\nu}$ corresponds to a pressure-free fluid. (See Appendix C).

Also, note that Eqs. (1.4) are not the only equations of classical physics in which the metric appears. The metric also appears in the covariant form of Maxwell's equations and in the Lorentz force equation. One may ask whether this third physical viewpoint according to which the field equations (1.4) are thought of as a definition of the metric remains mathematically self-consistent when the combined Maxwell–Einstein equations are considered. This question will be discussed in a future paper.

In conclusion, much remains to be done to fully establish the mathematical and physical self-consistency of the third physical viewpoint. The results presented in the present paper do suggest, however, that the definition of the metric (1.4) may not be the only possible one if the rules of measurement are appropriately changed. Indeed, these results suggest that certain features of the rules of measurement, namely those features which determine the metric, may be arbitrary in character.[17–21] These results also suggest that the law of conservation of energy-momentum may be one of these arbitrary features of the rules of measurement.

APPENDIX A

To complete the proof of integrability of the system (3.31), it must be shown that the initial conditions for the system can always be chosen so that both the coordinate transformation (3.5a) and the corresponding inverse transformation (3.5b) are nonsingular.

Since the solution is analytic one can conclude that the functions $f^{\alpha'}(x)$ are finite and that the Jacobian

$$|\partial_\mu f^{\alpha'}|$$

is finite. From this fact one concludes that the Jacobian of the inverse transformation is nonzero:

$$|\partial_{\alpha'} F^{\mu}| \neq 0, \tag{A1}$$

which means that the coordinate transformation (3.5b) is nonsingular.

Next consider the transformation (3.5a). Suppose that the initial conditions cannot be chosen so as to make this transformation nonsingular. This would be the case only if Eqs. (3.31c) imply that

$$|\partial_\alpha f^{\nu'}| = 0. \tag{A2}$$

But the system (3.31c) is a system of second-order partial differential equations, and such a system can only imply a first-order differential equation like Eq. (A2) if it has at least one of two properties. Either there are more equations than unknown functions (so that one of the second-order derivatives can be eliminated algebraically), or the system has integrability conditions (so that derivatives can be eliminated after appropriate differentiation of the equations of the system). The system (3.31) has neither of these properties. Therefore, it can lead to a first-order differential equation only by integration. Such an integral would involve at least one arbitrary constant of integration. Equation (A2) contains no arbitrary constants. Therefore it cannot be a consequence of the system (3.31c). Thus, one may conclude that it is always possible to choose the initial conditions in the system (3.31) so as to obtain

$$|\partial_\alpha f^{\nu'}| \neq 0. \tag{A3}$$

APPENDIX B: RIQUIER'S ORDERING PROCEDURE

To apply Riquier's existence theorem to a system of equations one must first assign an order to the various partial derivatives which appear in the system. This order determines the order in which those derivatives will be eliminated in Riquier's procedure

for calculating the integrability conditions. One writes

$$A > B \tag{B1}$$

(read "A is greater than B") to state that the derivative A is to be eliminated before the derivative B. We will refer to equations such as (B1) as "ordering inequalities." The order of the derivatives may be chosen at will except for the restrictions given below.

Let the indices i and j label the various dependent functions which appear in the system and the Greek letters μ, ν, etc. label the independent variables x^μ. Then we define f^i as the dependent functions, and we write

$$f^i{}_\mu = \frac{\partial f^i}{\partial x^\mu}, \qquad f^i{}_{\mu\nu} = \frac{\partial f^i}{\partial x^\nu \partial x^\mu}, \tag{B2}$$

and so on. Also note that in the following two equations, Eq. (B3) and (B4), the symbol " = " will mean "implies and is implied by." Then the restrictions on the ordering may be expressed as follows:

$$\begin{aligned} &(f^i > f^j) = (f^i_\mu > f^j_\mu) = (f^i_\nu > f^j_\nu), \\ &(f^i_\mu > f^i_\nu) = (f^j_\mu > f^j_\nu), \\ &f^i_{\mu\nu} > f^i_\mu > f^i. \end{aligned} \tag{B3}$$

Also each ordering inequality must remain true no matter how many times it is differentiated. For example,

$$(f^i_\mu > f^j_\mu) = (f^i_{\mu\nu\tau} > f^j_{\mu\nu\tau}). \tag{B4}$$

With the above definitions and restrictions on the ordering in mind, we are now in a position to describe Riquier's procedure for calculating the integrability conditions for a system of partial differential equations. We consider a finite system of n differential equations

$$\frac{\partial^{i_1+\cdots+i_m} y_j}{\partial x_1{}^{i_1} \cdots \partial x_m^{i_m}} = h_{i_1 \cdots i_m, j}, \tag{B5}$$

where the y_j are the unknown functions, the x_m are the independent variables, and the indicies $i_1 \cdots i_m$, j, m are not summed and may run over different ranges of numbers; where

(a) in each equation h is a function of the given functions J^k and their derivatives, and also of certain of the derivatives of the unknown functions y_i, every such y_i derivative in h being lower than the left-hand member of the equation.

(b) The left-hand side of each equation is different from that of every other.

(c) If w is the left-hand member of some equation, no derivative of w appears on the right-hand side of any equation.

Note that once criteria (a)–(c) are satisfied, those y_i which do not appear on the left-hand side of some equation can be considered to be given functions. Furthermore, the n highest derivatives appearing on the left are not necessarily the n highest derivatives appearing in the system.

Riquier's existence theorem applies to all systems of equations which, by algebraic manipulation, and proper choice of the ordering of the derivatives, can be written in the form (B5). The theorem applies in any region R in which the functions h in Eq. (B5) are analytic functions of their arguments. The theorem shows how to construct a power series solution to the system (B5). Furthermore, the theorem shows that if the power series is centered about any point in the region R, it converges within a small (but not infinitesimal) region about that point.

We call the derivatives on the left in Eq. (B5) the "first members" of Eqs. (B5). Derivatives which are first members or are derivatives of first members are called principal derivatives.

Next we shall describe what happens to the integrability conditions of a system of differential equations T if several derivatives of its component equations are added to it to form a new system T'. We consider a simple example. Let the system T consist of the single equation

$$U_{,1} + U^2 + F(x^1, x^2) = 0. \tag{B6}$$

Differentiate Eq. (B6) with respect to x^2, and consider the combined system T' given below:

$$U_{,1} + U^2 + F(x^1, x^2) = 0, \tag{B7a}$$

$$U_{,12} + 2UU_{,2} + F_{,2} = 0. \tag{B7b}$$

Now Eq. (B6) has no integrability conditions; but Eqs. (B7) do have an integrability condition in the sense of Riquier. This condition is

$$[U_{,1} + U^2 + F]_{,2} - [U_{,12} + 2UU_{,2} + F_{,2}] = 0, \tag{B8}$$

which immediately reduces to the identity $0 = 0$.

Thus the addition of the derivative equation (B7b) to the system does increase the number of integrability conditions. But the new integrability condition is identically satisfied. Integrability conditions like Eq. (B8) which (a) are identically satisfied and (b) arise from the combination of an equation with one of its derivatives will be called *trivial* integrability conditions. Integrability conditions which are not trivial will be called *nontrivial*. The property we have demonstrated in a simple case is true for any system of equations. That is, given any system of equations T in the form (B5), suppose one adds to that system additional equations which are derivatives of its component equations to form a new system T'. Then the system T' will have the same nontrivial integrability conditions as the system T.

In the proof of Riquier's theorem, it is necessary to extend the system (B5) by adding to it certain equations which are derivatives of some of the equations of the system. The additional equations are chosen so that the first members of the new system form a complete set in the sense of Riquier. (See Ref. 6, pp. 148–151.) Such a process of completion is always possible for a system of the form (B5).

From the discussion above of Eqs. (B6) and (B7), it is clear that the process of completion will not change or add to the nontrivial integrability conditions of the general system (B5). Thus the process of completion

can be ignored when one is calculating the nontrivial integrability conditions of a system. For this reason, no mention is made of the process of completion in Sec. 2, and no use of this process is made in Sec. 3. All the integrability conditions discussed in Sec. 2 and 3 and in Diagrams I–IV are nontrivial integrability conditions.

APPENDIX C: COMMENTS ON THE MEANING[22] OF $\tilde{T}^{\mu\nu}(x)$

In Theorem 1 the energy-momentum tensor density is given beforehand as a function of the coordinates and independent of the metric. Several comments as to the possible physical meaning corresponding to the giving of the energy-momentum tensor density in this way are given below. Since the metric has not been given, all concepts used in this section in defining $\tilde{T}^{\mu\nu}$ [i.e., all those concepts leading up to Eq. (C3)] are nonmetrical in nature.[23]

First, consider a field of trajectories, with a mass assigned to each trajectory. To such a field there corresponds a unique conserved mass-current density $\tilde{J}^{\mu}$ and a parallel tangent velocity vector v^{μ} unique up to a scaling factor $\alpha(x)$

$$\tilde{J}^{\mu}{}_{,\mu} = 0, \tag{C1a}$$

$$\tilde{J}^{[\mu}v^{\nu]} = 0. \tag{C1b}$$

Note that since $\tilde{J}^{\mu}$ is a current density, the divergence in Eq. (C1a) is equivalent to a covariant divergence so Eq. (C1a) may be rewritten

$$\tilde{J}^{\mu}{}_{;\mu} = 0. \tag{C2}$$

From such a trajectory field (which describes pressure-free dust) one can define the symmetric energy-momentum tensor density

$$\tilde{T}^{\mu\nu} = v^{\mu}\tilde{J}^{\nu}. \tag{C3}$$

Apply *Theorem 1* and conclude that there exists a metric $g^{\mu\nu}$ which satisfies the field equations (1.4a) and the conservation law (1.4b). Substitute Eq. (C3) into the conservation law (1.4b) written in the x' coordinate system appearing in Eq. (3.31c):

$$v^{\mu'}\tilde{J}^{\nu'}{}_{;\nu'} + v^{\mu'}{}_{;\nu'}\tilde{J}^{\nu'} = 0. \tag{C4}$$

Substitute Eq. (C2) in Eq. (C4) and obtain the geodesic equation

$$v^{\mu'}{}_{;\nu'}\tilde{J}^{\nu'} = 0. \tag{C5}$$

Thus the metric satisfying (1.4a) makes the trajectories geodesics. Transvect Eq. (C5) on $v_{\mu'}$ and obtain after minor simplification

$$(v_{\mu'}v^{\mu'})_{,\nu'}\tilde{J}^{\nu'} = 0. \tag{C6}$$

Note from Eq. (C6) that the length of $v^{\mu'}$ remains unchanged along a trajectory. Thus if $v^{\mu'}$ has unit length on a spacelike hypersurface, it retains that length throughout time. To obtain this unit length, one rescales $v^{\mu'}$ replacing $v^{\mu'}$ by $u^{\mu'}$ with

$$u^{\mu'} = \alpha v^{\mu'}. \tag{C7}$$

This does not change the trajectories. The requirement that the geodesic equation (C5) be maintained during the rescaling process implies that

$$\alpha_{,\mu'}\tilde{J}^{\mu'} = 0. \tag{C8}$$

Equation (C8) leaves one free to choose $\alpha(x')$ on an x^4 = const hypersurface. One requires that

$$(\alpha^2)_{x^4=0} = \mp\left(\frac{1}{g_{\mu\nu}v^{\mu}v^{\nu}}\right)_{x^4=0} = \mp\left(\frac{1}{g_{\mu\nu}F^{\mu}{}_{\alpha'}F^{\nu}{}_{\tau'}v^{\alpha'}v^{\tau'}}\right)_{x^4=0}. \tag{C9}$$

This gives[24]

$$(g_{\mu\nu}u^{\mu}u^{\nu})_{x^4=0} = \mp 1. \tag{C10}$$

Combine the initial condition (C10) with equation (C6) to obtain the result

$$u_{\mu}u^{\mu} = \mp 1 \tag{C11}$$

throughout all space–time.

This completes the proof of the following theorem.

Theorem 2: Given any analytic field of mass-carrying trajectories in which the trajectories remain distinct from one another, it is always possible to find a metric which satisfies Einstein's equations corresponding to the energy-momentum tensor density arising from these trajectories. Furthermore, this metric makes the trajectories geodesics, and a rescaling of the tangent vector field $v^{\mu}(x)$ corresponding to the trajectories can make $v^{\mu}(x)$ into a unit vector while retaining the geodesic equation.

ACKNOWLEDGMENTS

I wish to express my grateful appreciation to my thesis advisor, Professor J. W. Weinberg for his patient, relevant, useful and thought-provoking criticism, and for important help in clarifying the presentation of my results. I also wish to express my thanks to Professor G. E. Tauber, who served as advisor during an early portion of the research.

---

* Work supported in part by the Ohio State University Research Foundation under Air Force-Ohio State University Research Foundation Contract F33615-67-C-1758 during a six-month visit at Aerospace Research Laboratories, Wright–Patterson AFB, Ohio.

1 This paper presents with certain major simplifications several of the results obtained in the author's Ph.D. thesis (Case Western Reserve University, 1967). These results include a proof of the theorem referred to by J. Stachel in Phys Rev. **180**, 1256 (1969).

2 See E. Schrödinger, *Space–Time Structure* (Cambridge U. P., New York, 1969), p. 99.

3 This viewpoint is also closely related to (but not precisely equivalent to) the viewpoint presented in the Rainich–Misner–Wheeler theory of electromagnetism and gravitation. See J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962), especially p. 225.

4 See A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson, Paris, 1955). I wish to thank J. N. Goldberg for several enlightening comments which helped to clarify the relationship between the second and third view-

points.

[5] I wish to thank G. Kuerti and J. R. Stachel for raising questions which helped to clarify the discussion in this section.

[6] J. F. Ritt, *Differential Algebra*, Amer. Math. Soc. Colloq. Publ., Vol. 33, (Dover, New York, 1966), Chap. VIII. Also see J. F. Ritt, *Differential Equations from the Algebraic Standpoint*, AMS Colloq. Publ. (Amer. Math. Soc., Providence, R.I., 1932) Vol. 14, Chap. IX. For a list of references on Riquier's existence theorem see J. A. Schouten and W. v. d. Kulk, *Pfaff's Problem* (Oxford U. P. New York, 1949), reprinted (Chelsea New York, 1969), Chap. X, p. 506. Also in Chap. X is a discussion of the closely related existence theorem of Cartan and a comparison of Cartan's theorem with Riquier's theorem.

[7] For example, consider system S in Diagram III. In Diagram III the current is the given function, and the integrability condition is the conservation of charge law. [See Eq. (2.9).]

[8] Defined in Appendix B.

[9] The main thing about Procedure D which distinguishes it from Procedure O is that it involves differentiation.

[10] See Appendix B for the definition of highest and lowest, and also for the definition of principal derivative. See Sec. 3 for an example of how highest and lowest derivatives are chosen.

[11] V. Fock, *The Theory of Space, Time and Gravitation* (Macmillan, New York, 1964), p. 193.

[12] See Sec. 2 and Appendix B for the meaning of "ordering."

[13] See Sec. 2.

[14] I am grateful to G. E. Tauber for posing the question: Do any solutions to the field equations corresponding to the interior and exterior of a steadily rotating ideal fluid with axial symmetry exist? It was in an attempt to answer this question that Theorem 1 was first proved (in the special case $g_{Aa} = 0, A = 1, 2; a = 3, 4$; $g_{\mu\nu,a} = 0$. I also wish to thank R. Trümper for certain helpful remarks made during an early portion of the research).

[15] See Appendix B for details as to the range of validity of this theorem.

[16] In the present proof, $\bar{T}^{\mu\nu}(x)$ is for convenience assumed to be symmetric. If this assumption had not been made, the restrictions $\bar{T}^{[\mu\nu]} = 0$ would have appeared as integrability conditions. (See Ref. 1, p. 5). It is of interest to note that in special relativity, the symmetric nature of the energy–momentum tensor is necessary for the conservation of angular momentum. See, for example, C. Møller, *The Theory of Relativity* (Oxford U.P., London, 1952), 1st ed., p. 169.

[17] Similar ideas are presented by Poincaré under the title "Relativity of Space." (See Ref. 18–21.) Poincaré's comments form a helpful background for the understanding of the third physical viewpoint. Indeed the present paper can, in a certain sense, be considered an extension of these ideas of Poincaré (which were presented before the advent of general relativity) to curved space-time. I am grateful to J. W. Weinberg for pointing out the similarity between my own ideas and those of Poincaré.

[18] H. Poincaré, *Science and Hypothesis* (Dover, New York, 1952), (originally published in French in 1902), pp. 70–1.

[19] H. Poincaré, *Science and Method* (Dover, New York, 1952) (originally published in French in 1908), pp. 93–109.

[20] H. Poincaré, *Value of Science* (Dover, New York, 1958) (originally published in French in 1905), pp. 26–40.

[21] H. Poincaré, *Last Essays* (Dover, New York, 1963) (originally published in French in 1913), p. 27.

[22] I am grateful to J. W. Weinberg and P. G. Bergmann for important and useful comments on the questions discussed in this section.

[23] Nonmetrical discussions of contravariant vectors and vector densities may be found, for example, in the following: J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed., Chap. I and II, especially Chap. II, Secs. 8 and 6; J. L. Synge and A. Schild, *Tensor Calculus* (U. of Toronto Press, Toronto, 1949), Sec. 7.3; and *Astrophysics and General Relativity*, edited by M. Chrétien, S. Deser, and J. Goldstein (Gordon and Breach, New York, 1969), Vol. I., pp. 120-27.

[24] *Note added in proof:* The use of the plus-or-minus signs here is necessitated by the requirement that α^2 be positive.

Properties of Lorentz Covariant Analytic Functions*

Henry P. Stapp
Lawrence Radiation Laboratory, University of California, Berkeley, California 94702
(Received 20 September 1965; Revised Manuscript Received 29 March 1972)

A theorem is proved that asserts, roughly, that a function that is real Lorentz covariant anywhere is complex Lorentz covariant everywhere in its domain of regularity. It is also shown that the analytic continuation of a scattering function from a regularity domain in the physical region of a given process along all paths generated by complex Lorentz transformations leads to a function that is single-valued in the neighborhood of all these paths. Applications are discussed. The results derived constitute necessary preliminaries to a discussion of the analytic structure of scattering functions given in other papers.

The requirement that transition probabilities be invariant under physical Lorentz transformations implies[1] that the scattering functions $M(K)$ satisfy the Lorentz covariance condition[2,3]

$$M(K) = \Lambda_s M(\Lambda^{-1}K)$$

for all real K corresponding to physical points and for Λ any element of the real proper orthochronous homogeneous Lorentz group. Here K is the set of variables

$$K = \{k_i, m_i, t_i\},$$

where k_i, m_i, and t_i are the momentum–energy, spin quantum number, and particle type of particle i, and Λ_s is an operator that applies to each spin index m_i a matrix transformation corresponding to Λ. The specific form of Λ_s is given in the Appendix.

In this paper some consequences of assuming that $M(K)$ is also regular analytic at some physical point will be examined. The main result to be established is that if an M function is regular at some physical point, then the complete analytic extension of the function is defined over a multisheeted manifold, each sheet of which maps onto itself under any proper complex Lorentz transformation. Furthermore, the function defined (single-valuedly) and regular over any sheet is covariant under proper complex Lorentz transformations. Finally, if M is regular at each point of some real domain containing only physical points, then the sheets described above can be chosen so that all the points of any closed bounded subset of this domain lie in a single sheet. These results have some important consequences, which will be mentioned at the end of the paper.

The initial considerations will refer to a function $F(K)$ whose domain of definition is not restricted by the mass shell and conservation-law constraints. Also the type variables $T = \{t_i\}$ will be considered fixed. Thus the argument of $F(K)$ will be a set of the type introduced above but with the mass constraints and type variables removed.
Let the following definitions be made.

Definition: L will denote the real proper orthochronous homogeneous Lorentz group. It is continuously connected to the identity.

Definition: $\mathcal{L}$ will denote the complex proper homogeneous Lorentz group. It is continuously connected to the identity.

points.

[5] I wish to thank G. Kuerti and J. R. Stachel for raising questions which helped to clarify the discussion in this section.

[6] J. F. Ritt, *Differential Algebra*, Amer. Math. Soc. Colloq. Publ., Vol. 33, (Dover, New York, 1966), Chap. VIII. Also see J. F. Ritt, *Differential Equations from the Algebraic Standpoint*, AMS Colloq. Publ. (Amer. Math. Soc., Providence, R.I., 1932) Vol. 14, Chap. IX. For a list of references on Riquier's existence theorem see J. A. Schouten and W. v. d. Kulk, *Pfaff's Problem* (Oxford U. P. New York, 1949), reprinted (Chelsea New York, 1969), Chap. X, p. 506. Also in Chap. X is a discussion of the closely related existence theorem of Cartan and a comparison of Cartan's theorem with Riquier's theorem.

[7] For example, consider system S in Diagram III. In Diagram III the current is the given function, and the integrability condition is the conservation of charge law. [See Eq. (2.9).]

[8] Defined in Appendix B.

[9] The main thing about Procedure D which distinguishes it from Procedure O is that it involves differentiation.

[10] See Appendix B for the definition of highest and lowest, and also for the definition of principal derivative. See Sec. 3 for an example of how highest and lowest derivatives are chosen.

[11] V. Fock, *The Theory of Space, Time and Gravitation* (Macmillan, New York, 1964), p. 193.

[12] See Sec. 2 and Appendix B for the meaning of "ordering."

[13] See Sec. 2.

[14] I am grateful to G. E. Tauber for posing the question: Do any solutions to the field equations corresponding to the interior and exterior of a steadily rotating ideal fluid with axial symmetry exist? It was in an attempt to answer this question that Theorem 1 was first proved (in the special case $g_{Aa} = 0, A = 1, 2; a = 3, 4$; $g_{\mu\nu,a} = 0$. I also wish to thank R. Trümper for certain helpful remarks made during an early portion of the research).

[15] See Appendix B for details as to the range of validity of this theorem.

[16] In the present proof, $\bar{T}^{\mu\nu}(x)$ is for convenience assumed to be symmetric. If this assumption had not been made, the restrictions $\bar{T}^{[\mu\nu]} = 0$ would have appeared as integrability conditions. (See Ref. 1, p. 5). It is of interest to note that in special relativity, the symmetric nature of the energy–momentum tensor is necessary for the conservation of angular momentum. See, for example, C. Møller, *The Theory of Relativity* (Oxford U.P., London, 1952), 1st ed., p. 169.

[17] Similar ideas are presented by Poincaré under the title "Relativity of Space." (See Ref. 18–21.) Poincaré's comments form a helpful background for the understanding of the third physical viewpoint. Indeed the present paper can, in a certain sense, be considered an extension of these ideas of Poincaré (which were presented before the advent of general relativity) to curved space-time. I am grateful to J. W. Weinberg for pointing out the similarity between my own ideas and those of Poincaré.

[18] H. Poincaré, *Science and Hypothesis* (Dover, New York, 1952), (originally published in French in 1902), pp. 70–1.

[19] H. Poincaré, *Science and Method* (Dover, New York, 1952) (originally published in French in 1908), pp. 93–109.

[20] H. Poincaré, *Value of Science* (Dover, New York, 1958) (originally published in French in 1905), pp. 26–40.

[21] H. Poincaré, *Last Essays* (Dover, New York, 1963) (originally published in French in 1913), p. 27.

[22] I am grateful to J. W. Weinberg and P. G. Bergmann for important and useful comments on the questions discussed in this section.

[23] Nonmetrical discussions of contravariant vectors and vector densities may be found, for example, in the following: J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed., Chap. I and II, especially Chap. II, Secs. 8 and 6; J. L. Synge and A. Schild, *Tensor Calculus* (U. of Toronto Press, Toronto, 1949), Sec. 7.3; and *Astrophysics and General Relativity*, edited by M. Chrétien, S. Deser, and J. Goldstein (Gordon and Breach, New York, 1969), Vol. I., pp. 120-27.

[24] *Note added in proof:* The use of the plus-or-minus signs here is necessitated by the requirement that α^2 be positive.

Properties of Lorentz Covariant Analytic Functions*

Henry P. Stapp
Lawrence Radiation Laboratory, University of California, Berkeley, California 94702
(Received 20 September 1965; Revised Manuscript Received 29 March 1972)

A theorem is proved that asserts, roughly, that a function that is real Lorentz covariant anywhere is complex Lorentz covariant everywhere in its domain of regularity. It is also shown that the analytic continuation of a scattering function from a regularity domain in the physical region of a given process along all paths generated by complex Lorentz transformations leads to a function that is single-valued in the neighborhood of all these paths. Applications are discussed. The results derived constitute necessary preliminaries to a discussion of the analytic structure of scattering functions given in other papers.

The requirement that transition probabilities be invariant under physical Lorentz transformations implies[1] that the scattering functions $M(K)$ satisfy the Lorentz covariance condition[2,3]

$$M(K) = \Lambda_s M(\Lambda^{-1}K)$$

for all real K corresponding to physical points and for Λ any element of the real proper orthochronous homogeneous Lorentz group. Here K is the set of variables

$$K = \{k_i, m_i, t_i\},$$

where k_i, m_i, and t_i are the momentum–energy, spin quantum number, and particle type of particle i, and Λ_s is an operator that applies to each spin index m_i a matrix transformation corresponding to Λ. The specific form of Λ_s is given in the Appendix.

In this paper some consequences of assuming that $M(K)$ is also regular analytic at some physical point will be examined. The main result to be established is that if an M function is regular at some physical point, then the complete analytic extension of the function is defined over a multisheeted manifold, each sheet of which maps onto itself under any proper complex Lorentz transformation. Furthermore, the function defined (single-valuedly) and regular over any sheet is covariant under proper complex Lorentz transformations. Finally, if M is regular at each point of some real domain containing only physical points, then the sheets described above can be chosen so that all the points of any closed bounded subset of this domain lie in a single sheet. These results have some important consequences, which will be mentioned at the end of the paper.

The initial considerations will refer to a function $F(K)$ whose domain of definition is not restricted by the mass shell and conservation-law constraints. Also the type variables $T = \{t_i\}$ will be considered fixed. Thus the argument of $F(K)$ will be a set of the type introduced above but with the mass constraints and type variables removed.
Let the following definitions be made.

Definition: L will denote the real proper orthochronous homogeneous Lorentz group. It is continuously connected to the identity.

Definition: $\mathcal{L}$ will denote the complex proper homogeneous Lorentz group. It is continuously connected to the identity.

Definition: Λ will represent a Lorentz transformation and

$$(\Lambda K) \equiv \{\Lambda k_i, m_i\}. \tag{1}$$

Definition: Points K_1 and K_2 related by $K_1 = \Lambda K_2$ will be said to be *connected by* Λ.

Definition: The set of points connected to K by some $\Lambda \in \mathcal{L}$ (or L) will be denoted by $\mathcal{L}K$ (or LK).

Definition: The set of points connected to some element of the set D by some $\Lambda \in \mathcal{L}$ (or L) will be denoted by $\mathcal{L}D$ (or LD).

Definition: A point K is *real* if and only if the four vectors $\{k_i\}$ are real.

Definition: A *real set* is a set of real points.

Definition: A *function* $F(K)$ is a (*single-valued*) mapping to the complex numbers.

Definition: The spin indices of (K) will be presumed to have some spinor index-type label, and $\Lambda_s F(K)$ will represent the result of the action upon $F(K)$ of the corresponding spinor transformations associated with Λ, as discussed in the Appendix.

Lemma 1: If $F(K)$ is defined (single-valuedly) over a real set D and satisfies for all $\Lambda \in L$ and all K such that K and ΛK are elements of D the covariance condition

$$F(K) = \Lambda_s F(\Lambda^{-1}K), \tag{2}$$

then (2) with $\Lambda^{-1}K \in D$ and $\Lambda \in \mathcal{L}$ defines a (single-valued) function over $\mathcal{L}D$, provided any two points of D connected by a real element of $\mathcal{L}$ are also connected by an element of L.

Proof: The prescription will uniquely define $F(K)$ at K' of $\mathcal{L}D$ if for any two points K_1 and K_2 of D for which $K' = \Lambda_1 K_1 = \Lambda_2 K_2$, with Λ_1 and $\Lambda_2 \in \mathcal{L}$, one has

$$\Lambda_{1s} F(K_1) = \Lambda_{2s} F(K_2). \tag{3}$$

But by the group property $K_2 = \Lambda_2{}^{-1}\Lambda_1 K_1 \equiv \Lambda K_1$. Thus (2) gives

$$F(K_2) = \Lambda_{2s}^{-1}\Lambda_{1s} F(K_1) \tag{4}$$

provided $\Lambda \equiv (\Lambda_2^{-1}\Lambda_1) \in L$. Hence it is sufficient to show that Λ is an element of L. If the rank $r(K_1)$ of the Gram determinant $G(G_{ij} = k_i \cdot k_j)$ at the point K_1 is 4, or equivalently[4] if there are four linearly independent vectors among the vectors of K_1, then the rank is also 4 at K_2, since inner products are unchanged, and the same four vectors are also linearly independent at K_2. In this case the linear transformation Λ is unique. Since K_1 and K_2 are real, Λ is a real element of $\mathcal{L}$. By hypothesis it is then, by virtue of its uniqueness, an element of L. This completes the proof for the case $r(K_1) = 4$. For $r(K_1) = 3$ the transformation Λ is still unique[4] and the same argument holds.

If $r(K_1) < 3$, then the transformation Λ is not always uniquely defined by the equation $K_2 = \Lambda K_1$ and it may not be real, as required for the above argument. There are several cases. If the rank $r(K_1)$ is equal to $n(K_1)$, the number of linearly independent vectors of K_1, then the space separates into a manifold $M(K_1)$ of dimension $n(K_1) = r(K_1)$ spanned by the set K_1 and the orthogonal manifold $M^{\perp}(K_1)$. One can construct a set of real orthogonal basis vectors $e_\rho(K_1)$, each of length ± 1, such that the first n span $m(K_1)$ and the last $(4 - n)$ span $M^{\perp}(K_1)$. To construct such a basis, one first takes $n(K_1)$ linearly independent real vectors from the set K_1. This set is augmented by $[4 - n(K_1)]$ real vectors to give a complete set of real linearly independent vectors. Because the rank $r(K_1)$ equals $n(K_1)$, the linear equations arising in the construction of $e_\rho(K_1)$ are soluble. The details have been given by Hall and Wightman.[4] Since the original vectors are, for us, real, the coefficients in the linear equations are real, and hence the solutions can be taken to be real.[5] A similar real basis, $e_\rho(K_2)$, can be constructed for K_2.

Our interest is in the various Lorentz transformations Λ' satisfying $K_2 = \Lambda' K_1$, the K_1 and K_2 being the fixed points of D connected by $\Lambda \in \mathcal{L}; K_2 = \Lambda K_1$. The transformations Λ' can be represented by the matrices $\Lambda''_{\rho\sigma}$ defined by

$$\Lambda' e_\rho(K_1) \equiv e^\sigma(K_2)\Lambda''_{\sigma\rho} \equiv e_\sigma(K_2) G^{\sigma\tau}(K_2)\Lambda''_{\tau\rho}, \tag{5}$$

where a summation convention is used. The labels ρ, σ, and τ specify the basis vectors, not components, and

$$G^{\sigma\rho}(K_j) \equiv e^\sigma(K_j)\cdot e^\rho(K_j) = \pm\,\delta_{\sigma\rho} \quad \text{for } j = 1, 2. \tag{6}$$

For either value of j three of the vectors $e^\sigma(K_j)$ have length -1 and the other has length $+1$.[6] That all four have length -1 is impossible because any vector v can be expanded as

$$v = v_\sigma(K_j) e^\sigma(K_j) \tag{7}$$

with

$$v_\sigma(K_j) = e_\sigma(K_j)\cdot v = e_\sigma{}^\mu(K_j) v_\mu, \tag{8}$$

where μ labels the component of the vector. Then

$$v\cdot v \equiv v_\mu G^{\mu\nu} v_\mu = v_\rho(K_j) G^{\rho\sigma}(K_j) v_\sigma(K_j). \tag{9}$$

If the negative sign were always to occur in (6), then all vectors represented by real $v_\rho(K_j)$ would have negative length. But the vector v with components $v_\mu = \delta_{\mu 0}$ has real v_ρ and positive length $[G^{\mu\nu} = (1, -1, -1, -1)]$, which is a contradiction. On the other hand, if there were two real orthogonal vectors v^1 and v^2 of length $+1$, then

$$(v_0^1)^2 - |\mathbf{v}^1|^2 = 1, \quad (v_0^2)^2 - |\mathbf{v}^2|^2 = 1, \tag{10}$$

and

$$v_0^1 v_0^2 = \mathbf{v}^1\cdot\mathbf{v}^2. \tag{11}$$

From these it would follow that

$$(\mathbf{v}^1\cdot\mathbf{v}^2)^2 = (1 + |\mathbf{v}^1|^2)(1 + |\mathbf{v}^2|^2), \tag{12}$$

and hence that

$$(\mathbf{v}^1\cdot\mathbf{v}^2) > |\mathbf{v}^1|^2 |\mathbf{v}^2|^2, \tag{13}$$

which is not possible for real vectors. Thus there is, for each j, precisely one vector $e^\sigma(K_j)$ of length $+1$.

Because of this the vectors $e^\sigma(K_j)$ can be generated from the original set of basis vectors by real Lorentz transformations. The transformation Λ_b connecting the two sets

$$\Lambda_b e^\sigma(K_1) = e^\sigma(K_2) \tag{14}$$

will then also be a real Lorentz transformation.

The basis set $e^\sigma(K_2)$ is not completely specified by this construction. It is possible to take the first n vectors (which may or may not include the one of positive length) to be given by

$$e^\sigma(K_2) = \Lambda e^\sigma(K_1), \qquad \sigma = (1, 2, \ldots, n). \tag{15}$$

For, since $K_2 = \Lambda K_1$, these vectors span the space $M(K_2)$. They are orthogonal, since the $e^\sigma(K_1)$ are and $\Lambda \in \mathcal{L}$. And they are real, since Λ takes all the real vectors of K_1 into the real vectors of K_2, and hence by linearity all real vectors of $M(K_1)$ into real vectors of $M(K_2)$. Because $n(K_1) < 4$, one can by proper choice of the sense of the vectors $e^\sigma(K_1)$ with $\sigma > n(K_1)$ make Λ_b a real element of $\mathcal{L}$.

With the basis vectors fixed in this way, it is clear that the basis vector of positive length occurs either in the first n vectors of both sets $e^\sigma(K_2)$ and $e^\sigma(K_1)$ or in the last $(4-n)$ vectors of both sets. Also, with this choice the first $n \times n$ submatrix of $(\Lambda'')^\sigma_\rho$ is the $n \times n$ unit matrix. Since Λ' takes all vectors of $M(K_1)$ into vectors of $M(K_2)$, the first n columns of Λ'' have zeros except in the diagonal positions. The same property holds also for the first n rows as a consequence of the relations $\Lambda'^{-1}K_2 = K_1$ and

$$(G\Lambda''G)_{\rho\sigma} = (\Lambda''^{-1})_{\sigma\rho}, \tag{16}$$

which is the characteristic property of Lorentz transformations. That Λ'' is a Lorentz transformation follows from (5) and (14); one obtains

$$\Lambda_b^{-1}\Lambda' e_\mu(K_1) = e_\sigma(K_1)\Lambda''^\sigma_\rho \Lambda' e_\rho(K_1), \tag{17}$$

which shows that $\Lambda'' \equiv \Lambda_b^{-1}\Lambda'$. Since Λ_b is real, the transformation Λ'' will be real if Λ' is.

The conclusion from the above remarks is that for the case $n(K_1) = r(K_1)$ all Lorentz transformations $\Lambda' \in \mathcal{L}$ satisfying $K_2 = \Lambda' K_1$ and $K_2 \in D$, and with $K_2 = \Lambda K_1$ for some $\Lambda \in \mathcal{L}$, can be represented in the form

$$\Lambda' = \Lambda_b \Lambda'' \tag{18}$$

with a fixed real $\Lambda_b \in \mathcal{L}$ and a $\Lambda'' \in \mathcal{L}$ differing from the identity only in the $(4-n) \times (4-n)$ subspace corresponding to $M^\perp(K_1)$. And conversely, for all $\Lambda'' \in \mathcal{L}$ satisfying this property, which we call P, the transformation $\Lambda' \equiv \Lambda_b \Lambda''$ is an element of $\mathcal{L}$ satisfying $\Lambda' K_1 = K_2$.

This result is used in the following way: The transformations $\Lambda'' \in \mathcal{L}$ satisfying P can be parametrized in such a way that the matrix elements $(\Lambda''_s)^\sigma_\rho$ are analytic functions of these parameters regular in a neighborhood N of the identity, and such that real parameters give there real $\Lambda'' \in L$. Such a parameterization has been given by Jost,[7] for the case with no constraint P. The restriction to a submatrix is accomplished by setting some of his parameters to zero. Now suppose first that $\Lambda_b \in L$. Then the hypothesis of the lemma gives

$$F(K_2) = \Lambda'_s F(K_1) = \Lambda_{bs}^{-1}\Lambda''_s F(K_1), \tag{19}$$

for all $\Lambda'' \in L$ satisfying P. For then $K_2 = \Lambda' K_1$, with $\Lambda' \in L$, the K_1 and K_2 being the fixed points of D connected by $K_2 = \Lambda K_1$, with $\Lambda \in \mathcal{L}$. But the validity of this equation for real values of the parameters of Λ'', together with regularity in N, implies its validity throughout N. Thus (19) is true for $\Lambda'' \in \mathcal{L}$ satisfying P, in a neighborhood of the identity. The restriction P does not destroy the group property, since products of matrices having this property will also have it, and inverses of matrices having this property must also have it. Using the fact that the subgroups of $\mathcal{L}$ specified by the constraints P are connected, or more specifically, that any element of $\mathcal{L}$ satisfying P can be expressed as a product of a finite number of elements of $\mathcal{L}$ satisfying P from any fixed neighborhood of the origin, one obtains the result that (19) is true for all $\Lambda' \in \mathcal{L}$ satisfying $\Lambda' K_1 = K_2$. This ensures the validity of (4), from which the lemma follows, for the case $n(K_1) = r(K_1)$, provided Λ_b is an element of L.

In the above argument it was supposed that Λ_b was an element of L; then for $\Lambda'' \in L$ it followed that $\Lambda' \in L$, and (2) was immediately applicable. Now Λ_b is by construction a real element of $\mathcal{L}$ satisfying $\Lambda_b K_1 = K_2$. Thus, by virtue of the hypothesis of the lemma, there exists some $\Lambda' \in L$ such that $\Lambda' K_1 = K_2$. For this Λ' the transformation $\Lambda'' = \Lambda_b^{-1}\Lambda'$ must be a real element of $\mathcal{L}$. Thus it is either an element of L or it can be written in the form $\Lambda'' = \Lambda_0 \Lambda''_1$, where Λ''_1 is an element of L and Λ_0 is the (CPT) transformation $k \to -k$, which is a real element of $\mathcal{L}$. Parametrizing $\Lambda''_1 \in L$ instead of Λ'' one can develop the same argument as before and prove, from the validity of (14) for the $\Lambda' \in L$ just introduced, its validity for all $\Lambda' \in \mathcal{L}$ satisfying $\Lambda' K_1 = K_2$. This again validates (4), and completes the proof of the lemma for this case $n(K^1) = r(K^1)$.

The remaining possibility is $n(K_1) > r(K_1) < 3$. For these cases the vectors of K_1 are linear combinations of $r(K_1)$ orthogonal vectors of nonzero length and a single vector of zero length orthogonal to these. The $r(K_1)$ vectors of nonzero length are obtained by first picking $r(K_1)$ of the vectors of K_1 such that the Gram determinant of these $r(K_1)$ vectors is nonvanishing. This is always possible.[4] If any one of these vectors has nonzero length, then normalize it to ± 1, by multiplying by a real scalar, and let it be the first vector of a real basis. If on the other hand all these vectors have zero length, then some real multiple of a combination of the form $(k_i + k_j)$ must have length ± 1, since the (Gram) determinant of the matrix $(G_{ij}) = (k_i \cdot k_j)$ is nonvanishing. Subtracting a real multiple of this normalized vector from the other vectors, in the usual way, one gets a set of $[r(K_1) - 1]$ vectors orthogonal to it. Since the Gram determinant is still nonvanishing, the process can be repeated to give a real orthonormalized (i.e., to ± 1) set of $r(K_1)$ vectors $(e^1(K_1), \ldots, e^{r(K_1)}(K_1))$. This same construction was used (though not described) in the case $r(K_1) = n(K_1)$.

Since in the present case $n(K_1) > r(K_1)$, there must be a vector of K_1 that is linearly independent of these first $r(K_1)$ vectors. By subtracting from it multiples of the $e^\sigma(K_1)$, $\sigma = 1, \ldots, r(K_1)$, a linearly independent vector w orthogonal to them can be obtained. Since the value of the Gram determinant is unaltered by adding linear combinations of certain of the vectors

to others, the Gram determinant of the first $r(K_1)$ vectors together with w must vanish. But then w must have zero length. The next step is to augment the set K_1 by adding $[4 - r(K_1)]$ real vectors that together with the first $r(K_1)$ basis vectors give four linearly independent vectors. Since $n = 4$ implies[4] $r = 4$, one can complete the construction of a complete set of real orthonormalized basis vectors $e^\sigma(K_1)$, using the procedure just described.

The vector w is orthogonal to the first $r(K_1)$ of the $e^\sigma(K_1)$, and hence it is a linear combination of the remaining ones. Since it is real and of zero length it must, for the case $r(K_1) = 2$, be of the form

$$w = a[e^0(K_1) \pm e^3(K_1)], \tag{20}$$

where $a \neq 0$ is real and $e^0(K_1)$ is the basis vector having positive length. That the coefficients of the $e^\sigma(K_1)$ are real for real w follows from the existence of the real inverse of the real Lorentz transformation generating the $e^\sigma(K_1)$ from the original basis vectors.

The sign of $\pm e^3(K_1)$ in (20) depends on the sense of the vector $e^3(K_1)$. However, only one sign is possible; if different vectors of K_1 were to give w's having different signs in (20), then one would have $n(K_1) = r(K_1) + 2 = 4$, which is impossible since $n(K_1) = 4$ implies $r(K_1) = 4$.

For the case $r(K_1) = 1$ the vector w must be of the form

$$w = a[e^0(K_1) + \sin\theta e^2(K_1) + \cos\theta e^3(K_1)], \tag{21}$$

with a and θ real and $a \neq 0$. Moreover, for this case all vectors of K_1 must, when the part along $e^1(K_1)$ is removed, give multiples of this same vector w. To see this, note that the Gram determinant of two vectors w and w' of the form (21) is

$$G(w,w') = aa'[1 - \cos(\theta - \theta')]^2, \tag{22}$$

which is different from zero unless w' is a multiple of w. Thus if two vectors w and w' of the form (21) can be obtained as linear combinations of the vectors of (K_1), then either w' is a multiple of w or $r(K_1) \geq 2$. The second possibility contradicts the assumption $r(K_1) = 1$. The form (21) can be brought to the form (20) by a redefinition of the basis vectors that leaves them real and orthonormalized.

In the case $r(K_1) = 0$ all the vectors of K_1 are of zero length and they are mutually orthogonal. Expanding them in terms of an arbitrary real orthonormalized basis $e^\sigma(K_1)$, each one has the form

$$w = a[e^0(K_1) + \alpha e^1(K_1) + \beta e^2(K_2) + \gamma e^3(K_3)], \tag{23}$$

where a and (α, β, γ) are real and

$$\alpha^2 + \beta^2 + \gamma^2 = 1. \tag{24}$$

If $w \neq 0$, then any vector w' of the same form for which

$$G(w,w') = 0 \tag{25}$$

is, as before, a multiple of w. Thus for all the cases $n(K_k) > r(K_1)$ one can construct a real orthonormalized basis $e^\sigma(K_1)$ such that the vectors of K_1 are real linear combinations of a zero-length vector

$$w = e^0(K_1) + e^3(K_1) \tag{26}$$

and the vectors

$$e^\sigma(K_1), \quad \sigma = 1, \ldots, r(K_1) < 3. \tag{27}$$

A similar basis can be constructed for K_2. The set K_2 is related to the set K_1 by the relation $K_2 = \Lambda K_1$, $\Lambda \in \mathcal{L}$. Since Λ is not necessarily real, the vectors $\Lambda e^\sigma(K_1)$ need not be real. However, for $\sigma = 1, \ldots, r(K_1)$ these vectors must be real; the basis vectors $e^\sigma(K_1)$ can be expressed as real linear combinations of vectors of K_1, and hence the $\Lambda\sigma^\sigma(K_1)$ will be the same linear combination of the corresponding vectors ΛK_1 or ΛK_2, and hence also real. They have a Gram determinant of rank $r(K_2) = r(K_1)$ and are orthogonal and of length -1, and hence they can be chosen to be the corresponding $e^\sigma(K_2)$:

$$e^\sigma(K_2) \equiv \Lambda e^\sigma(K_1), \quad \sigma = 1, \ldots, r(K_1). \tag{28}$$

The entire set of real vectors $e^\sigma(K_2)$, constructed in the same manner as the $e^\sigma(K_1)$, and using (28) for $\sigma = (1, \ldots, r(K_1))$, can be related to the set $e^\sigma(K_1)$ by the equation

$$e^\sigma(K_2) = \Lambda_b e^\sigma(K_1), \tag{29}$$

where Λ_b is a real Lorentz transformation uniquely defined by this equation, once $e^\sigma(K_1)$ and $e^\sigma(K_2)$ are picked.

All real vectors of zero length in $M(K_1)$, the manifold spanned by the vectors of K_1, are multiples of the single vector

$$w(K_1) = e^0(K_1) + e^3(K_1), \tag{30}$$

since any real linear combination of the vectors of (27) is orthogonal to $w(K_1)$ and of nonzero length unless zero. Similarly all real zero-length vectors of $M(K_2)$ are multiples of

$$w(K_2) = e^0(K_2) + e^3(K_2). \tag{31}$$

Since $w(K_1)$ is a linear combination of the vectors of K_1, the vector $\Lambda w(K_1)$ is in $M(K_2)$, the manifold spanned by the vectors of K_2. But then $\Lambda w(K_1)$ is a real nonzero vector of zero length in $M(K_2)$. Hence it is a multiple $w(K_2)$:

$$w(K_2) = c\,\Lambda w(K_1) \neq 0. \tag{32}$$

The factor c can be taken to be unity. This follows from the fact that a real Lorentz transformation in the (0, 3) subspace gives simply a scale transformation of a vector of the form (31):

$$\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cosh\alpha + \sinh\alpha \\ \sinh\alpha + \cosh\alpha \end{pmatrix}. \tag{33}$$

This transformation preserves the reality and orthonormality properties of the $e^\sigma(K_1)$. Thus it can, and will, be assumed that the basis $e^\sigma(K_2)$ is chosen so that

$$c = 1 \tag{32'a}$$

or, equivalently, that

$$e^0(K_2) + e^3(K_2) = \Lambda e^0(K_1) + \Lambda e^3(K_1). \tag{32'b}$$

Using (29), one obtains, then,

$$e^0(K_1) + e^3(K_1) = \Lambda_b^{-1}\Lambda[e^0(K_1) + e^3(K_1)]. \tag{34}$$

The general form of the Lorentz transformation $\Lambda_b^{-1}\Lambda \in \mathcal{L}$ satisfying (34) is readily computed. If the rows and columns are placed in the order (0, 3, 2, 1), the general transformation matrix $(\Lambda'')^\sigma_\rho$ defined by

$$\Lambda'' e^\sigma(K_1) = e^\rho(K_1)(\Lambda'')^\sigma_\rho, \tag{35}$$

and consistent with (34), with $\Lambda'' \in \mathcal{L}$ in place of the fixed $\Lambda_b^{-1}\Lambda$, can be written

$1 + a$	$-a$	c	f
a	$1 - a$	c	f
$(c\cos\theta + f\sin\theta)$	$-(c\cos\theta + f\sin\theta)$	$\cos\theta$	$\sin\theta$
$(f\cos\theta - c\sin\theta)$	$-(f\cos\theta + c\sin\theta)$	$-\sin\theta$	$\cos\theta$

(36a)

where c, f, and Θ are arbitrary complex numbers and

$$2a = c^2 + f^2. \tag{36b}$$

The condition (34) imposes the constraint that the first two columns are the negatives of each other, aside from the unit contributions on the diagonal. This gives four conditions, only three of which are independent of Lorentz transformation condition (16).

Since the relations (28) and (34) are maintained if Λ is replaced by any Λ' satisfying $\Lambda' K_1 = K_2$, of which one is Λ_b, the general form of $(\Lambda'')^\sigma_\rho \equiv (\Lambda_b^{-1}\Lambda')^\sigma_\rho$ defined by (35), with $\Lambda' K_1 = K_2$, is given by (36) with the last $r(K_1)$ rows and columns having unity in the diagonal position and zeros elsewhere, provided $\Lambda'' \in \mathcal{L}$.

It can be assumed that $\Lambda_b \in \mathcal{L}$. If $r(K_1) < 2$, then there is freedom in the sign of at least one $e^\sigma(K_2)$, and Λ_b can be made a proper transformation. Then Λ_b will be a real element of $\mathcal{L}$. For the other case, $r(K_2) = 2$, the basis $e^\sigma(K_2)$ is uniquely specified by the conditions that have been imposed, and one cannot adjust Λ_b. However, in this case the conditions on $(\Lambda'')^\sigma_\rho$ require it to be unity even without the condition $\Lambda'' \in \mathcal{L}$, for one then has $c = f = \theta = 0$ from the conditions on $e^\sigma(K_1)$ for $\sigma = (1, 2)$, and condition (36) then gives the unique solution $\Lambda_b = \Lambda \in \mathcal{L}$.

To complete the argument for the case $n(K_1) > r(K_1)$, one first notes that Λ_b is a real element of $\mathcal{L}$ satisfying $\Lambda_b K_1 = K_2$. Thus there must, by hypothesis, exist some $\Lambda' \in L$ satisfying $\Lambda' K_1 = K_2$. But then

$$F(K_2) = (\Lambda_b\Lambda'')_s F(K_1) \tag{37}$$

is valid when $\Lambda'' \equiv \Lambda_b^{-1}\Lambda'$ corresponds to this $\Lambda' \in L$. Since Λ_b is a real element of $\mathcal{L}$, either Λ'' is an element of L or $\Lambda_0^{-1}\Lambda'' = \Lambda''_1$ is, where Λ_0 is the CPT transformation. Then Λ'' or Λ''_1, whichever is in L, can be parametrized as in (36), with the appropriate constraints if $r(K_1) > 0$. For a neighborhood of real values of the parameters, subject to these constraints, one still has $\Lambda' K_1 = K_2$ with $\Lambda' \in L$. But the spinor transformation

$$\Lambda'_s \equiv \Lambda_{bs}\Lambda''_s \quad \text{or} \quad (\Lambda_b\Lambda_0)_s(\Lambda''_1)_s \tag{38}$$

is an analytic function of these parameters, regular in a neighborhood of the origin of the free variables of (c, f, θ). Since (37) is true for real values of these variables, it is also valid for complex values in this neighborhood. One sees by inspection of (36) that the set of Λ'' satisfying the conditions corresponding to $\Lambda' K_1 = K_2$, $\Lambda' \in \mathcal{L}$, is a connected set of transformations in $\mathcal{L}$. From this it follows that any element of the set can be expressed as a product of a finite number of elements of the set lying within any neighborhood of the identity, and hence that (37) is valid for all $\Lambda' \in \mathcal{L}$ satisfying $\Lambda' K_1 = K_2$. This validates (4) for this last case and completes the proof of Lemma 1.

Lemma 1A: Real points connected by a Lorentz transformation $\Lambda \in \mathcal{L}$ are connected by some real $\Lambda \in \mathcal{L}$.

Proof: The transformation Λ_b constructed in the course of the proof of Lemma 1 is the required real $\Lambda \in \mathcal{L}$.

Lemma 2: Let K_0 be a set of n linearly independent vectors. For any neighborhood N of the identity in $\mathcal{L}$ there is a neighborhood $D(N, K_0)$ of K_0 such that any two points in $D(N, K_0)$ connected by a Lorentz transformation are connected by a Lorentz transformation $\Lambda \in N$.

Proof: Suppose the rank of the Gram determinant of the vectors of the set $K_0 \equiv \{k_i^0\}$ is $r(K_0) = r$. One can arrange the vectors of K_0 such that the rank of the Gram determinant of the first r vectors of the set is r. By using the procedure discussed in Lemma 1, but without the reality condition, a set of r orthonormal basis vectors $e_1(K_0), \ldots, e_r(K_0)$ can be constructed as linear combinations of the first r vectors of K_0. Completing the set K_0 to a set of four linearly independent vectors by the addition of $(4 - n)$ new vectors, one can construct $(4 - n)$ more vectors $e_{n+1}(K_0)$, $\ldots, e_4(K_0)$ that are orthonormal and orthogonal to the first r of the basis vectors. For the case $r = n$ this gives a complete set of basis vectors $e_\sigma(K_0)$.

For the case $n = r + 1$ the subtraction from k^0_{r+1} of its components along $e_1(K_0), \ldots, e_r(K_0)$ leaves a vector $w_0 = w \neq 0$, which must be of zero length, since otherwise the rank r would be n. For some $\sigma > n$ one must have $e_\sigma(K_0)\cdot w \neq 0$, since otherwise w_0 would be a zero-length vector orthogonal to three orthonormal vectors in a four-dimensional (nondegenerate[4]) space and hence zero. Take this vector $e_\sigma(K_0)$ to be $e_4(K_0)$. Then $i\{e_4(K_0) - w[e_4(K_0)\cdot w]^{-1}\}$ is a vector of unit length orthogonal to $e_4(K_0)$ and to $e_4(K_0), \ldots, e_r(K_0)$. Take this to be the final basis vector $e_{r+1}(K_0)$, and reorthogonalize $e_{r+2}(K_0), \ldots, e_3(K_0)$ following the standard procedure.

For the case $r = n - 2$ the subtraction of components along $e_1(K_0), \ldots, e_r(K_0)$ from the vectors k^0_{r+1}, k^0_{r+2} must leave two linearly independent orthogonal vectors w_{r+1} and w_{r+2} having zero length. Otherwise there would be fewer than n linearly independent vectors, or the rank of the vectors of K_0 would be greater than r. The vectors $w \equiv w_{r+1}$ and $\bar{w}' \equiv w_{r+2}$ cannot both be orthogonal to $e_\sigma(K_0)$ for all $\sigma > n$, for then they would be orthogonal to two orthonormal vectors. This would provide two linearly independent zero-length vectors in a two-dimensional space, which is impossible.[4] One can order the vectors of K_0 and of the $e_\sigma(K_0)$, $\sigma > n$, so that $w\cdot e_4(K_0) \neq 0$. Then the vector $i\{e_4(K_0) - w[e_4(K_0)\cdot w]^{-1}\}$ is a vector of unit length

orthogonal to the vectors $e_1(K_0), \ldots, e_r(K_0)$, and to $e_4(K_0)$. Let it be called $e_{r+1}(K_0)$. The vectors $e_\sigma(K_0)$, $4 > \sigma > n$, can then be reorthonormalized following the standard procedure so that the $e_\sigma(K_0)$ for $\sigma \leq r + 1$ and $\sigma > n$ become an orthonormal set. If the original $e_\sigma(K_0)$, $\sigma > n$, are appropriately chosen the subtractions of the required vectors will not give any zero-length vectors.

From the relation $\overline{w}' \cdot e_{r+1}(K_0) = i[\overline{w}' \cdot e_4(K_0)]$ it follows that $\{\overline{w}' - e_4(K_0)[e_4(K_0) \cdot \overline{w}'] - e_{r+1}(K_0)[e_{r+1}(K_0) \cdot \overline{w}']\} \equiv \overline{w}$ is a zero-length vector orthogonal to $e_4(K_0), e_1(K_0), \ldots, e_{r+1}(K_0)$. It cannot vanish since $\overline{w}'$ is linearly independent of w whereas $e_4(K_0)[e_4(K_0) \cdot \overline{w}^1] + e_{r+1}(K_0) \cdot \overline{w}']$ one must have $e_\sigma(K_0) \cdot \overline{w} \neq 0$. Otherwise $\overline{w}$ would be a zero-length vector orthogonal to the first $r + 1$ basis vectors and the last $4 - n$ basis vectors and hence orthogonal to $4 - n + r + 1 = 3$ orthonormal basis vectors. Let this $e_\sigma(K_0)$ be $e_3(K_0)$, since it is not vectors. Let this $e_\sigma(K_0)$ be $e_3(K_0)$, since it is not $e_4(K_0)$. Then the vector $i\{e_3(K_0) - \overline{w}[e_3(K_0) \cdot \overline{w}]^{-1}\} \equiv e$ is a vector of unit length orthogonal to all $e_\sigma(K_0)$ with $\sigma \leq r + 1$ or $\sigma \geq 3$, where these vectors are all orthonormal. This is impossible unless $r = 0$, since a vector orthogonal to four orthonormal vectors is zero. Thus one can set $e_2(K_0) = e$. This completes the construction of the orthonormal basis $e_\sigma(K_0)$ for the case $n = r + 2$. The case $n > r + 2$ is not possible.

For K in a sufficiently small neighborhood of K_0 one can construct a basis $e_\sigma(K)$ following the procedure just described, except for the following changes: The $(4 - n)$ vectors that are added to the set K to make a linearly independent set will, for all K, be taken to be the fixed vectors $e_\sigma(K_0)$ for $\sigma > n$, constructed above. For K in a sufficiently small neighborhood $D'(K_0)$ of K_0 the augmented set will continue to have four linearly independent vectors, and one can proceed with the construction; one constructs a set $e_\sigma(K)$, $\sigma > n$, by subtracting in the standard way the components along $e_\sigma(K)$, $\sigma \leq r$, etc., and normalizing. For $K \in D''(K_0) \subset D'(K_0)$ the vectors arising in this procedure will have nonzero length, so that a uniform procedure can be followed for all $K \in D''(K_0)$. At the next stage the vectors $e_{r+1}(K)$ [and $e_{r+2}(K)$] can be defined in the same way as above except that additional normalization factors η (and $\overline{\eta}$) must be supplied. For K in a sufficiently small neighborhood $D'''(K_1) \subset D''(K_1)$ the various factors that are required to be nonvanishing will continue to be nonvanishing, since they will depend continuously on the vectors of K. The only ambiguity in the procedure is in the choice of sign for the normalization factors. This sign can be fixed by requiring the normalization factors to be continuous functions of K. Thus in a sufficiently small neighborhood $D(K_0)$ of K_0 a basis $e_\sigma(K)$ can be defined so that these basis vectors depend continuously on the vector K. Also, for the case $r = n - 1$ the vector w obtained by subtracting from $k_n{}^0 \equiv k_{r+1}^0$ its components along $e_\sigma(K)$, $\sigma = 1, \ldots, r$, will always have the standard form $w = [e_4(K) + i\eta e_{r+1}(K)][w \cdot e_4(K)]$. For the case $r = n - 2 = 0$ this vector will have the form $w = [e_4(K) + i\eta e_1(K)][w \cdot e_4(K)]$, and the other vector, $\overline{w}$, will have the standard form $\overline{w} = [e_3(K) + i\overline{\eta} e_2(K)] \cdot [\overline{w} \cdot e_3(K)]$.

For *any* two vectors K_1 and K_2 in $D(K_0)$ a Lorentz transformation $\Lambda(K_1, K_2) \in \mathcal{L}$ is defined by the equation

$$e_\sigma(K_1) = \Lambda(K_1, K_2) e_\sigma(K_2). \tag{39}$$

If K_1 and K_2 are connected by a Lorentz transformation, then $K_1^r = \Lambda(K_1, K_2) K_2^r$, where K^r is the set consisting of the first r vectors of K. This is because the vectors $e_\sigma(K)$ are constructed, following a standardized procedure, as a linear combination of the vectors of K^r, and the coefficients are given as functions only of the inner products of the vectors of K^r. For the case $r = n$, $K^r = K^n$ and this transformation connects K_1 to K_2. Since the transformation $\Lambda(K_1, K_2)$ is a continuous function of (K_1, K_2) the inverse image of any open set in N containing the identity contains a neighborhood of the point (K_0, K_0). This neighborhood must contain a neighborhood of the form $K_1 \in D(N, K_0)$, $K_2 \in D(N, K_0)$, with $D(N, K_0) \subset D(K_0)$. This $D(N, K_0)$ satisfies the requirements of the lemma for the case $r = n$.

For the cases $r < n$ any points K_1 and $K_2 \in D(K_0)$ connected by a Lorentz transformation are connected by a Lorentz transformation of the form

$$K_1 = \Lambda^w \Lambda(K_1, K_2) K_2 \equiv \Lambda^w K_2', \tag{40}$$

where $\Lambda^w K_1^r = K_1^r$. For the subcase $r = n - 1$ the K_1 and K_2' differ only in the value of the vector w, and both values, w_1 and w_2', lie in the $(e_3(K_1); e_4(K_1))$ subspace. But two vectors in a subspace connected by a Lorentz transformation are connected by a Lorentz transformation in the subspace. This is a consequence of Lemma 2 of Hall and Wightman.

The Lorentz transformations in a two-dimensional subspace can be expressed as a product of possible inversions about the space or time axis times a transformation

$$a_\pm \to [\exp(\pm\Gamma)] a_\pm = \Lambda(\Gamma) a_\pm, \tag{41}$$

where Γ is a complex number and the $a_\pm$ are components along two orthogonal light-cone vectors. If two points are connected by a transformation of the form $\Lambda(\Gamma)$ then this transformation is unique.

If two points are in a neighborhood of the point $(a_+, a_-) = (1, 0)$ that contains no point with $a_+ = 0$, then, if they are connected by any Lorentz transformation, they are also connected by a $\Lambda(\Gamma)$. This is because for the case $a_-(0) \neq 0$ one can transform—using a $\Lambda(\Gamma)$—to a point where $a_+ = \pm a_-$. At such a point the reflections are equivalent either to the identity or to the particular $\Lambda(\Gamma)$ given by $\exp\Gamma = \exp(-\Gamma) = -1$. As a consequence of this, any sequence of reflections and proper transformations can be reduced to a single transformation $\Lambda(\Gamma)$, for this case, by the elimination of reflections in pairs. On the other hand, if $a_- = 0$, any product of reflections and $\Lambda(\Gamma)$ takes the point to a point with $a_- = 0$, which can be reached by $\Lambda(\Gamma)$ alone, or to a point with $a_+ = 0$, which by assumption is not in the original domain. Thus, with the neighborhood taken small enough so that points $a_+ = 0$ are not included, all points in the neighborhood connected by a Lorentz transformation are connected by a unique transformation of the form $\Lambda(\Gamma)$. One can therefore define a unique $\Lambda_1(K_1, K_2) = \Lambda(\Gamma)\Lambda(K_1, K_2)$ that satisfies

$$K_1 = \Lambda_1(K_1, K_2) K_2.$$

This transformation is a uniquely defined and contin-

uous function of the K_1 and K_2, provided (K_1, K_2) is restricted to a sufficiently small neighborhood of (K_0, K_0).

In case K_1 and K_2 are not connected by a Lorentz transformation Eq. (41) can be modified by the inclusion of a scale factor λ defined by

$$a_\pm \to \lambda[\exp(\pm\Gamma)]a_\pm \equiv \lambda\Lambda(\Gamma)a_\pm.$$

The Λ_1 is still defined to be $\Lambda(\Gamma)\Lambda(K_1, K_2)$. This Λ_1 is again continuous in K_1 and K_2.

Since $\Lambda_1(K_1, K_2)$ is continuous, one can proceed just as before, and $D(N, K_0)$ can be taken to be any neighborhood of K_0 such that $D(N, K_0) \otimes D(N, K_0)$ is in the inverse image of any neighborhood of the identity contained in N. Such a $D(N, K_0)$ must exist since the inverse image contains a neighborhood of (K_0, K_0). The neighborhood $D(N, K_0)$ is to be restricted also by the condition that the vectors w do not have a zero component along the w + axis. This is possible since for K_0 this condition is satisfied (for this case $r = n - 1$).

For the remaining case $n = r + 2 = 2$ similar arguments apply. The vectors of K are specified by the two vectors w and $\overline{w}$. The vectors w_1 and w_2' both lie in the $(e_4(K_1); e_2(K_1))$ subspace and the vectors $\overline{w}_1$ and $\overline{w}_1'$ both lie in the $(e_3(K_1); e_1(K_0))$ subspace. Thus the transformation Λ^w will be a product of transformations in two orthogonal subspaces. The problem separates then into two disconnected parts each of which is treated in the same way as Λ^w for the $r = n - 1$ case.

Lemma 3: Let K_0 be an arbitrary set of vectors. Let the first n vectors of K_0 be linearly independent. For any neighborhood N of the identity in $\mathcal{L}$ there is a neighborhood $D(N, K_0)$ of K_0 such that if any two points K_1 and K_2 in $D(N, K_0)$ are connected by a Lorentz transformation then $K_1^n = \Lambda K_2^n$ with $\Lambda \in N$, where K^n is the set consisting of the first n vectors of K.

Proof: This is a trivial extension of the preceding lemma. The neighborhood $D(N, K_0)$ can be the intersection of any (full) neighborhood of K_0 with $D^n(N, K_0^n)$, the neighborhood in the subspace associated with the K^n specified by Lemma 2.

Definition: A *simple point* K is a point where $n = 3, 4,$ or r.

Lemma 4: Let K_0 be any simple point and $D(K_0)$ be any neighborhood of K_0. Then there is a neighborhood $D_0(K_0)$ of K_0, contained in $D(K_0)$, such that any two points K_1 and K_2 in $D_0(K_0)$ connected by a $\Lambda \in \mathcal{L}$ are connected by a continuous path $K(t) = \Lambda(t)K_2$, with $K(0) = K_2$ and $K(1) = K_1$, such that $\Lambda(t) \in \mathcal{L}$ and $K(t) \in D(K_0)$ for $0 \leq t \leq 1$.

Proof: Let n be the number of linearly independent vectors of K_0 and r the rank of their Gram determinant. Arrange the vectors of K_0 so that the first n are linearly independent and the rank of the Gram determinant of the first r is r. Then, according to Lemma 3 there is, for any arbitrary neighborhood N of the identity in $\mathcal{L}$, a neighborhood $D(N, K_0)$ of K_0 small enough so that if K_1 and K_2 are in $D(N, K_0)$ and are connected by a Lorentz transformation $\Lambda \in \mathcal{L}$, then there is a $\Lambda_1 \in N$ such that $K_1^n = \Lambda_1 K_2^n$, where K_1^n and K_2^n are the subsets of K_1 and K_2 consisting of their first n vectors. The neighborhood N can be taken to be a domain (i.e., connected), and hence a path $\Lambda(t)$ in N can be constructed with $\Lambda(0) = 1, \Lambda(\frac{1}{2}) = \Lambda_1$, and $\Lambda(t) \in N$ for $0 \leq t \leq \frac{1}{2}$. The $D_0(K_0) \subset D(N, K_0)$ and N can evidently be chosen small enough so that all points $\Lambda' K_2$ with $\Lambda' \in N$ and $K_2 \in D_0(K_0)$ are in any preassigned neighborhood of K_0, say $D_1(K_0) \subset D(K_0)$.

Consider first the case $r = n$. The neighborhood $D_1(K_0)$ will be taken small enough so that for all $K \in D_1(K_0)$, the rank $r(K)$ of the Gram determinant of the first r vectors of K remains equal to r. Then any $K \in D_1(K_0)$ can be uniquely decomposed into a sum of two terms, $K \equiv K^r + V$, where the vectors of K^r are in the subspace spanned by the first r vectors of K, and the vectors of V lie in the subspace orthogonal to those r vectors. (Note that K^r is not the same as in Lemma 2.)

The neighborhood $D_1(K_0)$ can be specified by conditions of the form $\|K^r - K_0^r\| < \rho_r$ and $\|V\| < \rho$, with ρ and $\rho_r > 0$, since this is an arbitrarily small open set containing $K_0 = K_0^r$. One can use here for instance the Euclidian norms; e.g.,

$$\|V\| = \sum_i |v_i|^2 = \sum_{i,\mu} |v_i^\mu|^2 \tag{42}$$

The proof will be completed, for this case, if a continuous $\Lambda(t)$ for $\frac{1}{2} \leq t \leq 1$, with $\Lambda(\frac{1}{2}) = \Lambda_1$ and $\Lambda(1)K_2 = K_1$, can be found that acts only in the space orthogonal to the space spanned by the set K_1^n and keeps $\|v\| < \rho$.

The Lorentz transformation $\Lambda = \Lambda(1)\Lambda_1^{-1} \in \mathcal{L}$, which takes the point $\Lambda(\frac{1}{2})K_2 \equiv \Lambda_1 K_2$ to $\Lambda(1)K_2 \equiv K_1$, can, as any $\Lambda \in \mathcal{L}$, be expressed in the form[8]

$$\Lambda = R \exp A, \tag{43}$$

where R is a unimodular real orthogonal (hence unitary) transformation and A is Hermitian and imaginary:

$$A = -A^* = A^\dagger. \tag{44}$$

(The metric tensor G has been converted to the unit matrix by the introduction of the appropriate imaginary units.) The required transformation $\Lambda(t)$ for $\frac{1}{2} \leq t \leq 1$ can be taken to be defined by

$$\Lambda'(t) \equiv \Lambda(t)\Lambda_1^{-1} = \begin{cases} \exp[4(t - \frac{1}{2})A] & \text{for } \frac{1}{2} \leq t \leq \frac{3}{4} \quad (45) \\ R(t) \exp A & \text{for } \frac{3}{4} \leq t \leq 1, \quad (46) \end{cases}$$

where $R(t)$ for $\frac{3}{4} \leq t \leq 1$ is any continuous curve from the identity E to R in the connected space of real unimodular orthogonal matrices.

The Euclidean norm $\|V(t)\|$ of $V(t) \equiv \{\Lambda(t)v_i\}$ is the square root of

$$D^2(t) = \sum_i |\Lambda(t)v_i|^2 = \sum_i v_i^* \Lambda^\dagger(t)\Lambda(t)v_i \equiv \langle \Lambda^\dagger(t)\Lambda(t) \rangle_V. \tag{47}$$

In the interval $\frac{3}{4} \leq t \leq 1$ the $\|V(t)\|$ is constant, because of the unitarity of $R(t)$:

$$R^\dagger(t)R(t) = \tilde{R}(t)R(t) = E. \tag{48}$$

On the other hand, in the interval $\frac{1}{2} < t < \frac{3}{4}$ one has, since $A = A^\dagger$,

$$\frac{d^2}{dt^2} D^2(t) = \frac{d^2}{dt^2} \langle \Lambda^\dagger(t) \Lambda(t) \rangle_V$$
$$= 64 \langle \Lambda^\dagger(t) A^\dagger A \Lambda(t) \rangle_V \geq 0. \quad (49)$$

Because the second derivative of $\| V(t) \|^2$ is nonnegative, its maximum value must be assumed at an end point. As the end points are in $D_1(K_0)$ they satisfy $\| V(t) \| < \rho$. Thus for all $0 \leq t \leq 1$ this condition is satisfied. Consequently, all points $K(t) = \Lambda(t) K_2$ are in $D_1(K_0) \subset D(K_0)$. This completes the proof for the case that r, the rank of the Gram determinant of K_0, is equal to n, the number of linearly independent vectors of K_0.

In the case $r < n$ the first part of the transformation, $0 \leq t \leq \frac{1}{2}$, can be performed as before. For $n \geq 3$ this already completes the proof, since the coincidence of three linearly independent vectors ensures the coincidence of all vectors. The special form of $D_1(K_0)$ is not needed for this case. This completes the proof.

Definition: A function $F(K)$ will be said to be *regular at a point* K if and only if the various functions of K corresponding to the various combinations of the spin indices are all regular analytic functions of the components of the four vectors $\{k_i\}$ at the point K.

Lemma 5: Let Λ be a fixed Lorentz transformation. Let $F_\Lambda(K)$ be defined by

$$F_\Lambda(K') = \Lambda_s F(\Lambda^{-1} K'). \quad (50)$$

If $F(K)$ is regular at the point $K = \Lambda^{-1} K'$, then $F_\Lambda(K)$ is regular at the point $K = K'$.

Proof: This is an immediate consequence of the theorem in several complex variables that an analytic function of an analytic function is analytic. This well-known theorem is easily proved by using the Cauchy–Riemann equations.

Corollary A: Let $F_\Lambda(K)$ be defined by (50), where Λ is fixed. Then $F_\Lambda(K)$ is regular at $K = K'$ if and only if $F(\Lambda^{-1} K')$ is regular at $K = \Lambda^{-1} K'$.

Proof: The first part of the corollary is just the lemma. To prove the converse, apply the lemma to the function

$$F''(K) \equiv \Lambda_s^{-1} F_\Lambda(\Lambda K) \quad (51)$$

to show that $F''(K)$ is regular at K if $F_\Lambda(\Lambda K)$ is regular at ΛK. But $F''(K)$ is just $F(K)$. The substitution $K = \Lambda^{-1} K'$ gives the desired results. The fact that the inverses Λ^{-1} and Λ_s^{-1} exist is essential to the proof.

Corollary B: The property of being regular at a point does not depend on the choice of coordinate system relative to which the components of the vectors k are measured, provided the components in the two systems are related by a Lorentz transformation.

Proof: The proof is the same as for the lemma.

Definition: A domain is an arcwise connected open set.

Definition: A real domain is an arcwise connected real set open with respect to the set of real points.

Lemma 6: Suppose $F(K)$ is defined (single-valuedly) in a domain $D(K_0)$ containing K_0 and is regular at all points of $D(K_0)$. And suppose $F(k)$ satisfies the covariance condition

$$F(K) = \Lambda_s F(\Lambda^{-1} K), \quad (52)$$

for $\Lambda \in L$, and K and $\Lambda^{-1} K$ in a real domain D containing the point K_0. Then for each point K in $D(K_0)$ Eq. (52) is satisfied for $\Lambda \in N_r(K)$, where $N_r(K)$ is some neighborhood of the identity in L.

Proof: Let K_1 be a fixed arbitrary point of $D(K_0)$. Since $D(K_0)$ is a domain there exists a continuous curve $K(t)$, $0 \leq t \leq 1$, from K_0 to K_1, all points of which are in $D(K_0)$. Let the distance between two points be defined as maximum of the absolute values of the differences of the components of the vectors $\{k_i\}$. Then the distance of a point K in $D(K_0)$ to the boundary of $D(K_0)$ will be defined as the maximum (real) number $\Delta(K)$ such that every point whose distance from K is less than $\Delta(K)$ is inside $D(K_0)$. Since $D(K_0)$ is a domain $\Delta(K) > 0$ for all $K \in D(K_0)$. Moreover,

$$\Delta(K(t)) \geq a > 0, \quad \text{for } 0 \leq t < 1, \quad (53)$$

for if there were no positive lower bound $a > 0$ of $\Delta(K(t))$, one could find a sequence t_n, $0 \leq t_n \leq 1$, with $\Delta(K(t_n)) < 2^{-n}$. These t_n would have to have an accumulation point $\bar{t}$, $0 \leq t \leq 1$. But $\Delta(K(\bar{t})) \equiv b > 0$. Hence, for all t such that the distance between $K(t)$ and $K(\bar{t})$ is less than $b/2$, one would have $\Delta(K(t)) \geq b/2$, by the triangle inequality. Since $K(t)$ is a continuous curve, the inverse map of the open set $\| K(t) - K(\bar{t}) \| < b/2$ contains an open interval Δt about $\bar{t}$. But, since $\Delta(K(t)) > b/2$ for $t \in \Delta t$, only a finite number of the t_n can be in Δt. Hence $\bar{t}$ cannot be an accumulation point. This is a contradiction. Thus there is a positive lower bound a.

Let the maximum value of $\| K(t) \|$ for $0 \leq t \leq 1$ be A. Let $N(K_1)$ be a neighborhood of the identity in $\mathcal{L}$ such that if $\Lambda^{-1} \in N(K_1)$, then $\| (\Lambda^{-1})_\mu^\nu - \delta_\mu^\nu \| < (a/4A)$. Then, for $\Lambda^{-1} \in N(K_1)$, it follows that $\| \Lambda^{-1} K(t) - K(t) \| < a$, and the (continuous) curve $K_\Lambda(t) \equiv \Lambda^{-1} K(t)$ remains inside of $D(K_0)$ for all $0 \leq t \leq 1$.

Let N_r be a neighborhood of the identity in L such that $\Lambda^{-1} K_0 \in D$ for $\Lambda^{-1} \in N_r$. The existence of such a neighborhood follows immediately from the continuity of $\Lambda^{-1} K_0$ in Λ at the identity. For any fixed $\Lambda^{-1} \in N_r \cap N(K_1) \equiv N_r(K_1)$ there is a real domain $D(\Lambda, K_1) \subset D$, with $K_0 \in D(\Lambda, K_1)$, such that for all $K \in D(\Lambda, K_1)$ the points K and $\Lambda^{-1} K$ are in $D \cap D(K_0)$. The existence of such a $D(\Lambda, K_1)$ follows from the fact that K_0 and $\Lambda^{-1} K_0$ are in $D \cap D(K_0)$, in conjunction with the continuity of $\Lambda^{-1} K$ as a function of K. Thus (52) is valid for any fixed $\Lambda^{-1} \in N_r \cap N(K_1)$ for all $K \in D(\Lambda, K_1)$. The validity of (52) for fixed $\Lambda^{-1} \in N_r \cap N(K_1)$, for all K in the real domain $D(\Lambda, K_1)$, together with the analyticity of both sides of the equation, as functions of K (Lemma 5) implies the validity also at the point K_1, since one can analytically continue along $K(t)$ with the argument of the function on the right tracing simultaneously the curve $K_\Lambda(t)$, which remains inside the domain of regularity $D(K_0)$.

Lemma 6A: Lemma 6 modified by the substitution of $\mathcal{L}$ for L and of a (full complex) domain D_c for the real domain D is also valid.

Proof: Make these substitutions throughout the proof of Lemma 6.

Lemma 7: Suppose $F(K)$ is defined (single-valuedly) in a domain $D(K_0)$ containing K_0 and is regular at all points of $D(K_0)$. And suppose $F(K)$ satisfies the covariance condition (52) for $\Lambda \in L$, and K and $\Lambda^{-1}K$ in a real domain D containing the point K_0. Then (52) is also valid for all $K \in D(K_0)$ and $\Lambda \in \mathcal{L}$ such that there is a continuous path $\Lambda(t) \in \mathcal{L}$, $0 \leq t \leq 1$, with $\Lambda(0) = E$ and $\Lambda(1) = \Lambda$, such that $K(t) \equiv \Lambda^{-1}(t)K \in D(K_0)$ for $0 \leq t \leq 1$.

Proof: The assumptions of the lemma are the same as those of Lemma 6. Thus the conclusions of Lemma 6 may be used; (52) is valid for every point $K \in D(K_0)$ for $\Lambda^{-1} \in N_r(K)$, a neighborhood of the identity in L. Following Hall and Wightman[4] and Jost,[7] the Lorentz transformations Λ in a neighborhood N of the identity in $\mathcal{L}$ can be parametrized by a continuous one-to-one mapping $\Lambda(\lambda_j)$ in such a way that the representations of Λ^{-1} and Λ_s are regular analytic functions of the λ_j for $\Lambda^{-1} \in N$; and such that for $\Lambda^{-1} \in N \cap L$ the λ_j are real; and such that the origin in λ_j maps into the identity in Λ. Such a parametrization has been given by Jost.[7]

Considered as a function of the λ_j the right-hand side of (52) is an analytic function regular at all points for which $\Lambda^{-1} \in N$ and $\Lambda^{-1}K \in D(K_0)$. But for Λ^{-1} in the real neighborhood of the origin $N_r(K)$ the right-hand side of the equation is, by Lemma 6, equal to the left-hand side, which is independent of λ_i. Thus the right side must be equal to the left for all $\Lambda = \Lambda(t)$ such that $\Lambda^{-1}(t') \in N$ and $\Lambda(t')K \in D(K_0)$ for $0 \leq t' \leq t$, since one can analytically continue to this point, the right-hand side remaining regular. If for all $0 \leq t \leq 1$ the $\Lambda^{-1}(t)$ are not contained in N, then the continuation can be carried out stepwise by expanding $\Lambda^{-1}(t)$, in the manner specified above, about a finite sequence of intermediate points, t_n, and by using the group properties. The covariance equation is in this way validated for all points $K, \Lambda^{-1}K$ connected by a continuous path $\Lambda(t)K$ that remains always inside the domain of regularity $D(K_0)$. That only a finite number of t_n are required follows from the Heine–Borel covering theorem.

Lemma 7A: Lemma 7 is also true if the real D and L are replaced by complex D_c and $\mathcal{L}$.

Proof: Make these substitutions throughout the proof of Lemma 7.

Lemma 8: Let $F(K)$ be defined (single-valuedly) and regular for points in a domain $D(K_0)$ containing K_0. And suppose

$$F(K) = \Lambda_s F(\Lambda^{-1}K) \tag{54}$$

for $\Lambda \in L$ and Λ and $\Lambda^{-1}K$ in a real domain D containing K_0. Then for every simple point $K_1 \in D(K_0)$ there is a domain $D_0(K_1)$ containing K_1 such that the equation

$$F(K; D_0(K_1)) \equiv \Lambda_s F(\Lambda^{-1}K) \tag{55}$$

with $\Lambda^{-1}K \in D_0(K_1)$ and $\Lambda \in \mathcal{L}$ defines a (single-valued) function $F(K; D_0(K))$ over the points $K \in \mathcal{L}D_0(K_1)$. This function is regular throughout its domain of definition and coincides with $F(K)$ in the domain $D_0(K_1) \subset D(K_0)$.

Proof: The assumptions are the same as those of Lemma 7. Thus the covariance equation (54) holds for all K and $\Lambda^{-1}K$ connected by a path $\Lambda(t)K$, $0 \leq t \leq 1$, that is everywhere in $D(K_0)$. Consider an arbitrary point $K_1 \in D(K_0)$. According to Lemma 4, there is a domain $D_0(K_1)$ containing K_1 such that the points of every pair of points in $D_0(K_1)$ connected by a Lorentz transformation are connected by a continuous path $\Lambda(t)K$, $0 \leq t \leq 1$, that is everywhere in $D(K_0)$. Lemma 7 then ensures that the covariance equation (54) is valid for all $K, \Lambda^{-1}K \in D_0(K_1)$. This in turn ensures that (55) defines a (single-valued) function $F(K; D_0(K_1))$. To show this, suppose for some $K \in \mathcal{L}D_0(K_1)$ the points $\Lambda_1^{-1}K$ and $\Lambda_2^{-1}K$ are both in $D_0(K_1)$. Then one can write

$$F_1(K; D_0(K_1)) = \Lambda_{1s} F(\Lambda_1^{-1}K) \tag{56}$$

and

$$F_2(K; D_0(K_1)) = \Lambda_{2s} F(\Lambda_2^{-1}K). \tag{57}$$

That these are equal follows from Eq. (54) expressed in the form

$$F(\Lambda_1^{-1}K) = \Lambda_{1s}^{-1}\Lambda_{2s} F((\Lambda_2^{-1}\Lambda_1)\Lambda_1^{-1}K), \tag{58}$$

which is true because both arguments are in $D_0(K_1)$.

Since $F(K; D_0(K_1))$ is independent of the particular Λ used on the right of (55), so long as $\Lambda^{-1}K \in D_0(K_1)$, the values of $F(K; D_0(K_1))$ in some neighborhood of any $K \in \mathcal{L}D_0(K_1)$ can be generated from a fixed Λ, as a consequence of the continuity of $\Lambda^{-1}K$ as a function of K, for fixed Λ. That is, the inverse map of the open set $D_0(K_1)$ of $\Lambda^{-1}K$'s is an open set $D_\Lambda(K_1)$ of K's. But for fixed Λ the regularity of the left-hand side of (55) is ensured by Lemma 5, since $\Lambda^{-1}K \in D_0(K_1) \subset D(K_0)$. Finally, that $F(K; D_0(K_1))$ coincides with $F(K)$ for $K \in D_0(K_1)$ is true by virtue of (55) with $\Lambda = I$.

Remark: Minkowski and Williams[9] have shown that Lemma 8 can be proved without the restriction to simple points. This restriction will therefore be henceforth omitted. Lemma 4, on the other hand, is not true for nonsimple points, as shown by a counter example of Jost generalized by Seiler.[9]

Lemma 8A. The lemma remains valid if the real D and L are replaced by a complex D and $\mathcal{L}$.

Some concepts from the theory of functions of several complex variables will now be introduced.[10]

Definition: A regular function element e is a triple $[K_e; D_e; F_e(K)]$ consisting of a base point K_e, a domain D_e containing K_e, and an associated function $F_e(K)$ defined (single-valuedly) and regular in D_e.

Definition: Two regular function elements will be called *equivalent* if and only if they have the same base point and their functions coincide in some neighborhood of this point.

Definition: A *germ* is a set of regular function elements such that

(1) any two elements of the set are equivalent, and

(2) any regular function element equivalent to an element of the set is also in the set.

Definition: A germ neighborhood $N(D_N, F_N(K))$ is the set of all germs containing a regular functions element $[K; D_N; F_N(K)]$. The domain D_N and the function $F_N(K)$ are called the *base domain* and the *characteristic function* of the germ neighborhood, respectively.

Definition: The topological (Hausdorff) space with germs as points and germ neighborhoods as neighborhoods will be called the *germ space.*

Definition: The *domain of regularity* of a function $F(K)$ defined (single-valuedly) and regular in a domain D is the set of all germs connected to any germ of $N(D, F(K))$ by a continuous curve in the germ space.

Definition: The unique germ $g[e]$ containing e is called the germ *specified* by e. (Uniqueness is easily proved.)

Definition: The *base point* $K(g)$ of a germ g is the common base point of the $e \in g$.

Definition: $\overline{F}(g) \equiv F_e(K(g))$, with $e \in g$. $[\overline{F}(g)$ is independent of the choice of $e \in g.]$

Definition: Let $N \equiv N(D_N, F_N(K))$ be a germ neighborhood. Then, for $K \in D_N$, define $g_N(K) \equiv g[e]$, where $e = [K; D_N; F_N(K)]$.

Remark: $g_N(K)$ is the unique $g \in N$ such that $K(g_N(K')) = K'$. Restated, $g_N(K)$ is the unique inverse of $K(g)$, subject to the condition that $g \in N$.

Lemma 9: If the characteristic functions of two germ neighborhoods N and N' coincide in a domain $D \subset (D_N \cap D_{N'})$, then $g_N(K) = g_{N'}(K)$, for $K \in D$.

Proof: The associated function of any element e of $g_N(K)$ coincides with $F_N(K)$ for K in some neighborhood $N(K)$ of $K \in D$. Thus it must coincide with $F_{N'}(K)$ in $N(K) \cap D$ and hence in some neighborhood of K. Thus e is in $g_{N'}(K)$. Conversely every $e \in g_{N'}(K)$ is in $g_N(K)$.

Some terminology associated with Lorentz covariant analytic functions will now be introduced.

Definition: A function will be called $\mathfrak{L}$ *(or* L*)-covariant* over a set of points S if and only if it satisfies

$$F(K) = \Lambda_s F(\Lambda^{-1}K)$$

for any K and Λ such that Λ is in $\mathfrak{L}$ (or L) and both K and $\Lambda^{-1}K$ are in S.

Definition: An *orbit* is a set of points K all connected to a single point by Lorentz transformations $\Lambda \in \mathfrak{L}$.

Definition: A *regular orbit* is a set of germs whose base points cover exactly once the points of an orbit, and such that the image in the germ space of any continuous curve in the orbit is a continuous curve in the germ space.

Definition: Let $g(K)$ for $K \in \mathfrak{L}K_0, K(g(K')) = K'$, be the germs of a regular orbit. This regular orbit will be called $\mathfrak{L}$*-covariant* if and only if the function $F(K) \equiv \overline{F}(g(K))$ is $\mathfrak{L}$*-covariant* over the orbit $\mathfrak{L}K_0$.

Definition: A domain of regularity will be called $\mathfrak{L}$*-covariant* if and only if it is a union of $\mathfrak{L}$-covariant regular orbits.

Theorem 1: A function defined (single-valuedly) and regular in a domain containing a point and L-covariant over a real domain containing the point has an $\mathfrak{L}$-covariant domain of regularity.

Proof: Let K_0 be the point in the real domain and let the function be called $F(K)$. There is a domain $D(K_0)$ containing K_0 on which $F(K)$ is defined and regular. Thus the set $e_0 \equiv [K_0; D(K_0); F(K)]$ constitutes a regular function element. Let g_0 be the germ specified by e_0. This g_0 is in $N \equiv N(D(K_0), F(K))$. Let g_1 and g_2 be any two germs in N. Then there is a continuous curve in the germ space connecting g_1 and g_2. In particular, if $K(t)$ is a continuous curve in $D(K_0)$ connecting $K(g_1)$ and $K(g_2)$, then $g_N(K(t))$ will be a continuous curve in the germ space connecting g_1 and g_2. For consider any germ neighborhood $N' \equiv N(D', F'(K))$ that contains a germ $g_N(K(t_0))$, where t_0 is some fixed value of t, $0 \leq t \leq 1$. Let D'' be a domain in $D' \cap D(K_0)$ containing $K(t_0)$. Any germ of N' with base point in D'' is identical to the germ of N with the same base point, for D'' is a domain and hence the function $F'(K)$ must be identical with $F(K)$ for $K \in D''$. This is true because $F(K)$ and $F'(K)$ are both regular over the domain D'' and they coincide over some neighborhood of $K(t_0) \in D''$, since $g(K(t_0))$ contains both $[K(t_0); D; F]$ and $[K(t_0); D'; F']$. Since the functions $F'(K)$ and $F(K)$ are identical for $K \in D''$, the germs of N' and N with base points in D'' must be identical, by virtue of Lemma 9. Because $K(t_0)$ is in the domain D'', and $K(t)$ is a continuous curve, the inverse image of the points $K(t) \in D''$ contains an interval Δt that contains t_0 and is open with respect to the set $0 \leq t \leq 1$. The germs $g_N(K(t))$ with t in the interval Δt are all in the arbitrary neighborhood N' containing $g_N(K(t_0))$. Thus, this curve $g_N(K(t)) \equiv g(t)$ is continuous. Hence any two germs in N can be connected by a continuous curve. This means that the word "any" in the definition of domain of regularity can be replaced by "every" with no change in the meaning. (That two continuous curves joined at their end points give a continuous curve follows easily.)

Consider now an arbitrary germ g in the domain of regularity of $F(K)$. It is connected to g_0 by a continuous curve $g(t)$ in the germ space. Since $g(t)$ is continuous the inverse image of any germ neighborhood containing a germ $g(t_0)$ contains an interval Δt containing t_0 that is open with respect to the set $0 \leq t \leq 1$. By the Heine–Borel theorem, the closed bounded set $0 \leq t \leq 1$ is covered by a finite number of these intervals, Δ_i, with $i = 1, 2, \ldots, n$. Associated with these intervals are corresponding germ neighborhoods N_i, with $i = 1, 2, \ldots, n$, such that, for $t \in \Delta_i$, $g(t) \in N_i$. And there is then a sequence $\{t_i\}$ so that $g(t_i)$ is in both N_i and N_{i+1}.

The assumptions of the theorem are a paraphrasing of the assumptions of Lemma 8. Thus for each point K_1 of $D(K_0)$ there is a domain $D_0(K_1) \subset D(K_0)$ containing K_1 such that $F(K)$ is $\mathcal{L}$-covariant in $D_0(K_1)$. The first N_i can be taken to be $N_1 = N$. Take $K_1 = K(t_1)$. Then K_1 will also lie in the domain D_2, in which lie the base points of the germs of N_2. The germ neighborhood N_2 is characterized by the requirement that each of its germs has an element having the domain D_2 and the function $F_2(K)$. Also, N_2 contains the germ $g(t_1)$, which is also in $N_1 = N$, and which therefore has the element $[K(t_1); D(K_0); F(K)]$. But then $F_2(K)$ and $F(K) \equiv F_1(K)$ must coincide with each other in some neighborhood of K_1. But since $F(K)$ is $\mathcal{L}$-covariant in $D_0(K_1)$ the function $F_2(K)$ is $\mathcal{L}$-covariant in some domain containing K_1. Thus the conditions for Lemma 8A are satisfied for $F_2(K)$. Hence for any point K_2 in D_2 there is a domain containing K_2 such that $F_2(K)$ is $\mathcal{L}$-covariant in this domain. Take $K_2 = K(t_2)$. The argument may then be repeated to give $\mathcal{L}$-covariance in a domain about $K_i = K(t_i)$ for $i = 3$, and by iteration, for $i = n - 1$. In particular, there is a point K_{n-1} of the domain D_n, in which lie the base points of N_n, such that $F_n(K)$ is $\mathcal{L}$-covariant in some domain containing K_{n-1}. Lemma 8A now shows that there is a domain $D_n(K_g)$ containing K_g, the base point of the germ g, such that there is a function $F_g(K)$ defined (single-valuedly) over $\mathcal{L}D_n(K_g)$, where it is regular and $\mathcal{L}$-covariant, and which coincides with $F_n(K)$ in $D_n \cap \mathcal{L}D_n(K_g)$, which contains K_g. The germ neighborhood $N_g \equiv N(\mathcal{L}D_n(K_g), F_g(K))$ contains g, by virtue of Lemma 9, since $F_g(K)$ coincides with $F_n(K)$ in a neighborhood of K_g.

The set of germs $g' \in N_g$ with $K(g') \in \mathcal{L}K_g$ constitute an $\mathcal{L}$-covariant regular orbit containing g. Let $g(K(g')) = g'$ for $g' \in N_g$. That any continuous $K(t) \in \mathcal{L}K_g$ has a continuous image $g(K(t))$ follows from the argument given earlier, since $g(K(t)) \in N_g$ (see Lemma 10). The $\mathcal{L}$-covariance of the set $g \in N_g$ with $K(g) \in \mathcal{L}K_g$ follows from the $\mathcal{L}$-covariance of $F_g(K)$ over $\mathcal{L}D_n(K_g) \supset \mathcal{L}K_g$. Thus each germ g in the domain of regularity of $F(K)$ is on an $\mathcal{L}$-covariant regular orbit. Since all points of this orbit are connected to g by a continuous path, they are also contained in the domain of regularity of $F(K)$. Thus each germ g in the domain of regularity of $F(K)$ is a member of an $\mathcal{L}$-covariant regular orbit each of whose members is also in the domain of regularity of $F(K)$. This is what was to be proved.

Theorem 1A: Theorem 1 is also true if "L-covariant" is replaced by "$\mathcal{L}$-covariant", and the real domain is replaced by a (complex) domain.

Definition: A germ neighborhood will be said to be *$\mathcal{L}$-covariant* if only if its base domain is of the form $\mathcal{L}D$ and its characteristic function is $\mathcal{L}$-covariant over $\mathcal{L}D$.

Theorem 1': The domain of regularity of a function satisfying the conditions of Theorem 1 is a union of $\mathcal{L}$-covariant germ neighborhoods.

Proof: In the course of proving Theorem 1 it was shown that each g in the domain of regularity of such a function is in an $\mathcal{L}$-covariant germ neighborhood N_g. All the points of this neighborhood are in the domain of regularity since one is, by virtue of the following lemma, which was also proved in the course of proving Theorem 1.

Lemma 10: The image in a germ neighborhood of a continuous curve in its base domain is a continuous curve in the germ space.

The converse of this lemma is:

Lemma 10': The image $K(g(t))$ of a continuous curve $g(t)$ in the germ space is continuous.

Proof: A continuous function of a continuous function is continuous. But $K(g)$ is continuous, since given any domain D containing $K(g)$ one can take a germ neighborhood N_g containing g specified by a function element whose domain D', which contains $K(g)$, is contained in D. Then, for all $g' \in N_g, K(g) \in D$.

Lemma 11: Let D be a real domain satisfying the condition of Lemma 1 that points of D connected by a real $\Lambda \in \mathcal{L}$ are connected by a $\Lambda \in L$. Let there be two converging sequences $K_i \to K_0$ and $\overline{K}_i \to \overline{K}_0$ whose limit points K_0 and $\overline{K}_0$ are in D. And suppose $\overline{K}_i \in \mathcal{L}K_i$. Then $\overline{K}_0 \in \mathcal{L}K_0$.

Proof: The scalar and pseudoscalar invariants formed from corresponding vectors of K_i and $\overline{K}_i = \overline{\Lambda}_i K_i$ are equal. Thus these points map into the same points in the space of scalar and pseudoscalar invariants. As the mapping from K to the space of invariants is continuous, the converging sequences $K_i \to K_0$ and $\overline{K}_i \to \overline{K}_0$ map into converging sequences in the space of the invariants. Thus K_0 and $\overline{K}_0$ have the same scalar and pseudoscalar invariants.

In case r, the rank of the Gram determinant of K_0 or $\overline{K}_0$, is greater than two, it follows from a trivial generalization of Lemma 2 of Hall and Wightman that K_0 and $\overline{K}_0$ are connected by a Lorentz transformation $\Lambda \in \mathcal{L}$; that the transformation is *proper* in the case $r = 4$ follows from the invariance of the pseudoscalar invariants, and for $r = 3$ there is sufficient freedom to allow Λ to be made proper. Thus the lemma is proved for the case $r > 2$.

Let $n(K)$ be the number of linearly independent vectors in the set K. And let $n = \max(n(K_0), n(\overline{K}_0))$. The above argument works equally well for all the cases $r = n$. One constructs the orthonormalized basis vectors $e_\sigma(K_0)$ and $e_\sigma(\overline{K}_0)$ in the manner specified in Lemma 1 above and obtains $\overline{K}_0 = \Lambda_b K_0$, where Λ_b is the real $\Lambda \in L$ defined by $e_\sigma(\overline{K}_0) = \Lambda_b e_\sigma(K_0)$. Thus $\overline{K}_0$ and K_0 are connected by an element of $\mathcal{L}$. This completes the proof for the case $r = n$.

Because K_0 and $\overline{K}_0$ arereal, the only other cases are $n = r + 1 < 4$. Suppose $n(K_0) = r + 1 < 4$. Then, as in Lemma 1, one can construct a set $e_1(K_0), \ldots, e_r(K_0), e_0(K_0) + e_3(K_0)$ which spans the space of the vectors of K_0. The combination $e_0(K_0) + e_3(K_0)$ is chosen to be equal to some vector ω of zero length formed as a linear combination of vectors of K_0. Such a vector must exist in this case. If $\overline{\omega}$, the same linear combination of the corresponding vectors of $\overline{K}_0$, is not zero, then one can construct a set $e_1(\overline{K}_0), \ldots, e_r(\overline{K}_0), \pm e_0(\overline{K}_0) \pm e_3(\overline{K}_0)$, by means of the same operations as before, but with the corresponding vectors of $\overline{K}_0$. The two $\pm$ signs are independent and will be specified by the condition that the Λ_b defined by $e_\sigma(\overline{K}_0) = \Lambda_b e_\sigma(K_0)$ is in L. For $r(K_0) < 2$ the

sign of $e_3(\bar{K}_0)$ is not determined by this condition and it can, and will, be taken positive.

The points $\bar{K}_0$, $\bar{K}_i$, and $\bar{\omega}$ can be represented by the transformed quantities $K_0' = \Lambda_b^{-1}\bar{K}_0$, $K_i' = \Lambda_b^{-1}\bar{K}_i$, and $\omega' = \Lambda_b^{-1}\bar{\omega}$. This, in effect, refers the barred points to the same coordinate system, $e_\sigma(K_0)$, used for the unbarred points K_0, K_i, and ω. In particular $\omega' = \pm e_0(K_0) \pm e_3(K_0)$, where the $\pm$ signs are the same as the corresponding ones in $\bar{\omega}$. The vectors ω (or ω') are what is left after removing from some vector of K_0 (or the corresponding vector of K_0') the parts along $e_1(K_0), \ldots, e_r(K_0)$. In this same way one constructs from the sets $K_0 \equiv \{k_{0\alpha}\}$ and $K_0' \equiv \{k'_{0\alpha}\}$ the sets of light-zone vectors $\{\omega_\alpha\} \equiv \{a_\alpha\omega\}$ and $\{\omega'_\alpha\} = \{a'_\alpha\omega'\}$ by removing the parts along $e_1(K_0), \ldots, e_r(K_0)$.

That the vectors of these sets are colinear follows from the condition $n = r + 1 < 4$. In the special case that $\omega' = \omega$ and $a_\alpha = a'_\alpha$, one has again $\bar{K}_0 = \Lambda_b K_0$ with $\Lambda_b \in L$. But if $\omega' \neq \omega$ or $a_\alpha \neq a'_\alpha$, for some α, then $\bar{K}_0$ and K_0 are not connected by a $\Lambda \in L$. However, these cases cannot occur. This will now be shown by an examination of points in D near K_0 and $\bar{K}_0$.

In the real 0–3 plane consider a set of small circles $\{C(\omega_\alpha)\}$ drawn around the points $\{\omega_\alpha\}$ and a set of small circles $\{C(\omega'_\alpha)\}$ drawn around the points $\{\omega'_\alpha\}$. A set of points with one in each $C(\omega_\alpha)$ corresponds to a real point near K_0. And a set of points with one in each $C(\omega'_\alpha)$ corresponds to a real point near $\bar{K}_0$. By taking the circles sufficiently small, these two points near K_0 and $\bar{K}_0$, respectively, will be constrained to lie in arbitrarily small real neighborhoods about K_0 and $\bar{K}_0$, and hence in D.

The plan is to show that there is a real point arbitrarily close to K_0 connected to a point arbitrarily close to $\bar{K}_0$ by a real $\Lambda \in \mathcal{L}$, but not a $\Lambda \in L$. The sets of points in the real 0–3 plane connected by $\Lambda \in L$ lie on the various hyperbolas having the light-cone lines as asymptotes. The circles are centered on these light-cone lines, the $C(\omega_\alpha)$ lying on the line with positive slope and the $C(\omega'_\alpha)$ lying either on this line or on the other one, depending on the signs in $\omega' = \pm e_0(K_0) \pm e_3(K_0)$.

If $C(\omega_\alpha)$ and $C(\omega'_\alpha)$ lie on the positively and negatively sloped light-cone lines, respectively, then there is always a $\Lambda \in L$ connecting some point of $C(\omega_\alpha)$ to some points of $C(\omega'_\alpha)$. Moreover, there are then also points in these circles connected by any still "larger" $\Lambda \in L$. The magnitude of the Lorentz transformation is measured by the quotient of the initial over the final (Euclidean) distances of the point from the negatively sloped light-cone line. From these facts it follows that some set of points, one in each of a given set of circles along the positively sloped light-cone line, can be taken into some set of points, one in each of any given set of corresponding circles along the negatively sloped light-cone line, by a single Lorentz transformation $\Lambda \in L$. Thus for the cases $\omega' = \pm (e_0(K_0) - e_3(K_0))$ one can find a $\Lambda \in L$ connecting some real point in any real neighborhood of K_0 to some real point in any neighborhood of $\bar{K}_0$, even though the points themselves cannot be so connected.

The same conclusion holds if one uses instead of $\Lambda \in L$ the real $\Lambda \in \mathcal{L}$ obtained by multiplying the $\Lambda \in L$ by a reflection through the origin in the 0–3 plane. However, as will soon be shown, the points connected in this way cannot be connected by any $\Lambda \in L$. Since by taking the neighborhoods of K_0 and $\bar{K}_0$ small enough the points will be in D, one obtains a contradiction with the assumed property of D. Thus this case $\omega' = \pm (e_0(K_0) - e_3(K_0))$ can, in fact, not occur.

To see that there would be points in D connected by real $\Lambda \in \mathcal{L}$ but not by $\Lambda \in L$, consider first the case $\omega' = -e_0(K_0) + e_3(K_0)$. A timelike point in the circle $C(\omega)$ will be carried to a timelike point in the corresponding circle $C(\omega')$ by the real $\Lambda \in \mathcal{L}$. Since these two points are in the forward and backward light cones, respectively, they cannot be connected by an $\Lambda \in L$. The other case, $\omega' = e_0(K_0) - e_3(K_0)$, occurs only if $r = 2$, as previously mentioned. But now a spacelike point in $C(\omega)$ is taken to a spacelike point in $C(\omega')$ by the real $\Lambda \in \mathcal{L}$. However, transformations involving the first two vectors, $e_1(K_0)$ and $e_2(K_0)$, are not allowed, because the components of vectors of K_0 and K_0' in these subspaces are fixed and equal, and hence these two spacelike vectors, which lie in the right and left space cones, respectively, cannot be connected by a $\Lambda \in L$.

The remaining cases are $\omega' = \pm \omega$, or zero. If $\omega' = \pm \omega$ and $r < 2$, then the construction used above again allows certain points near K_0 to be connected to corresponding points near K_0'. One first uses a $\Lambda \in L$ in the 0-3 plane to take the points of the $C(\omega_\alpha)$ to points near the negatively sloped light-cone line, and then uses a rotation through π in the 2–3 plane to bring the points to the desired positions in the 0–3 plane. In particular, if ω_α and ω'_α have the same sense, certain timelike vectors near ω_α can be taken to timelike vectors near ω'_α. If ω_α and ω'_α have opposite senses, then spacelike vectors can be connected. However, if ω_α and ω'_α have the same (opposite) sense a spacelike (timelike) point near ω_α can be carried to a spacelike (timelike) point near ω'_α by a real $\Lambda \in \mathcal{L}$. But these points cannot be connected by a $\Lambda \in L$ unless $\omega = \omega'$ and $a_\alpha = a'_\alpha$. In that case $K_0' = K_0$ and $\bar{K}_0 = \Lambda_b K_0$, as asserted by the lemma.

The next case is $\omega' = \omega$ and $r = 2$. If $\omega' = \omega$ and $a_\alpha = a'_\alpha$ for all α, then $K_0' = K_0$ and $\bar{K}_0 = \Lambda_b K_0$, which proves the lemma. If $a_\alpha \neq a'_\alpha$ for some α, then $\bar{K}_0$ and K_0 are, in fact, not connected by a $\Lambda \in L$. In any event it is sufficient to show that $\omega' = \omega$ and $r = 2$ imply $a_\alpha = a'_\alpha$ for all α.

The conditions $K_i \to K_0$ and $K_i' = \Lambda_i K_i \to K_0'$ are now involved, for the first time. Let $e_1(K)$, $e_2(K)$, and $\omega(K)$ be the linear combinations of the vectors of K that become $e_1(K_0)$, $e_2(K_0)$, and $\omega(K_0) \equiv \omega$ when K becomes K_0. The $e_i(K)$ are then generally not orthonormalized and $\omega(K)$ is not a null vector. The Λ_i are specified by the conditions $\Lambda_i \in \mathcal{L}$ and by the quantities $e_1(K_i) \equiv e_{i1}$, $e_2(K_i) = e_{i2}$, and $\omega(K_i) \equiv \omega_i$; and $e'_{i1} \equiv e_1(K_i') \equiv e_1(\Lambda_i K_i) = \Lambda_i e_1(K_i) \equiv \Lambda_i e_{i1}$, $e'_{i2} = \Lambda_i e_{i2}$, and $\omega_i' = \Lambda_i \omega_i$, at least for sufficiently large i, where the e_{i1}, e_{i2}, and ω_i are linearly independent. For these quantities give the effect of Λ_i on three linearly independent vectors. But since $e'_{i1} \to e_{i1}$, $e'_{i2} \to e_{i2}$, and $\omega_i' \to \omega_i$, it follows from Lemma 3 that $\Lambda_i \to 1$, for Lemma 3 says that, given any neighborhood N of the identity in $\mathcal{L}$, one can find a neighborhood N' of $(e_1(K_0), e_2(K_0), \omega)$ such that any points in N' connected by a $\Lambda \in \mathcal{L}$ are connected by a $\Lambda \in N$. Since for the case of three linearly inde-

pendent vectors the $\Lambda \in \mathcal{L}$ is uniquely defined by these points, one concludes that since the sets $(e_{i1}, e_{i2}, \omega_i)$ and $(e'_{i1}, e'_{i2}, \omega'_i)$ both converge to $(e_1(K_0), e_2(K_0), \omega)$, the $\Lambda_i \in \mathcal{L}$ connecting them must approach the identity. But if $\Lambda_i \to 1$ and $K_i \to K_0$, then $\Lambda_i K_i \to K_0$. Thus $\bar{K}_0 = \Lambda_b K_0$, which proves the lemma for this case.

If $\omega' = -\omega$, a reflection through the origin in the 0–3 plane takes one to the previous case $\omega' = \omega$. Because of the condition on D this case is then ruled out, since $\bar{K}_0$ is connected to K_0 by a real $\Lambda \in \mathcal{L}$ but not by a $\Lambda \in L$.

Next there is the case $\omega' = 0$. If all of the $\omega'_\alpha = 0$ [i.e., if $n(\bar{K}_0) = r$], then this case is ruled out by the same argument that was used in the case $\omega' = -e_0(K_0) + e_3(K_0)$; there are points of D connected by real $\Lambda \in \mathcal{L}$ but not by $\Lambda \in L$. [The possibility $n(K_0) = n(\bar{K}_0) = r$ with $e_1(K_0), \ldots, e_r(K_0)$ all spacelike is also ruled out in this way, it might be added.] If $\omega' = 0$ but some ω'_α is a nonzero vector lying along the negatively sloped light-cone line, one may again use the same argument as was used for the case $\omega' = -e_0(K_0) + e_3(K_0)$ case; the $C(\omega')$ is simply centered at the origin instead of at its former position.

For the case $r < 2, \omega' = 0$, and $\omega'_\alpha = a'_\alpha \omega \neq 0$ for some α, the argument used in the case $r < 2, \omega' = \pm\omega$, goes through without any change.

Finally there is the same case but with $r = 2$. Every ω_α and ω'_α is either zero or on the positively sloped light-cone line. For every α either ω_α or ω'_α is zero; otherwise it can be made into the case $\omega' = \pm\omega$. And not every ω'_α is zero; otherwise it is the previously considered case $n(\bar{K}_0) = r$. This means that the Λ_i are such that the following conditions can be satisfied:

$$(e_{i1}, e_{i2}, \omega_i) \to (e_1(K_0), e_2(K_0), \omega),$$

$$\Lambda_i \quad (e_{i1}, e_{i2}, \omega_i) \to (e_1(K_0), e_2(K_0), 0),$$

$$(e''_{i1}, e''_{i2}, \omega''_i) \to (e_1(K_0), e_2(K_0), \omega_1),$$

and

$$\Lambda_i^{-1} \ (e''_{i1}, e''_{i2}, \omega''_i) \to (e_1(K_0), e_2(K_0), 0).$$

Here the double-primed quantities are a particular set of primed quantities, the ω''_i being an $\omega'_\alpha(K_i)$ whose limit is $\omega_1 \neq 0$.

These two conditions on the set Λ_i are incompatible. The first two equations imply that, for sufficiently large i, the points $\Lambda_i\omega$ must lie in a narrow conelike region about the negatively sloped light-cone line, whereas the second two imply that $\Lambda_i\omega$ must lie far from the origin in some narrow cone-like region about the positively sloped light-coned line. The incompatibility of these conditions rules out this last possibility.

The consequences for the $\Lambda_i\omega$ asserted above follow from a detailed examination of the converging sequences. A general description of the argument should be sufficient. Since $(e_{i1}, e_{i2}) \to (e_1(K_0), e_2(K_0))$ one can choose basis vectors e_{0i} and e_{3i} in the subspace orthogonal to the one spanned by the (e_{i1}, e_{i2}) in such a way that $(e_{0i}, e_{3i}) \to (e_0(K_0)e_3(K_0))$. The (e_{0i}, e_{3i}), unlike the (e_{i1}, e_{i2}) are to be parts of an orthonormal basis. A set (e'_{0i}, e'_{3i}) similarly related to the $(e'_{i1}, e'_{i2}) \equiv \Lambda_i(e_{i1}, e_{i2})$ is constructed. Then $\Lambda_i{}^r$ is defined by the conditions $\Lambda_i{}^r(e_{i1}, e_{i2}) = \Lambda_i(e_{i1}, e_{i2})$ and $\Lambda_i{}^r(e_{0i}, e_{3i}) = (e'_{0i}, e'_{3i})$. Since $(e_{i1}, e_{i2}, e_{3i}, e_{0i})$ and $(e'_{i1}, e'_{i2}, e'_{3i}, e'_{0i})$ both converge to $(e_1(K_0), e_2(K_0), e_3(K_0), e_0(K_0))$, it follows that $\Lambda_i{}^r \to 1$, by Lemma 3.

Since $\Lambda_i{}^r(e_{i1}, e_{i2}) = \Lambda_i(e_{i1}, e_{i2})$, it follows that $\Lambda_i{}^\omega \equiv (\Lambda_i{}^r)^{-1}\Lambda_i$ acts only in the $e_{0i}, e_{3i})$ subspace. Also since $(e_{0i}, e_{3i}, \omega_i) \to (e_0(K_0), e_3(K_0), \omega)$, with $\omega = e_0(K_0) + e_3(K_0)$, one has $\omega_i \to (e_{0i} + e_{3i})$.

Since $\Lambda_i{}^r \to 1$ and $\Lambda_i = \Lambda_i{}^r\Lambda_i{}^\omega$ the condition $\Lambda_i\omega_i \to 0$ implies $\Lambda_i{}^\omega\omega_i \to 0$. Since $\Lambda_i{}^\omega$ acts only in the (e_{0i}, e_{3i}) subspace, the problem is reduced now to a problem in this two-dimensional space. The two conditions $\Lambda_i{}^\omega\omega_i \to 0$ and $\omega_i \to e_{0i} + e_{3i}$ imply that $\Lambda_i{}^\omega(e_{0i} + e_{3i}) \to 0$; the general Lorentz transformation in this two-dimensional space is represented by

$$(e_{0i} + e_{3i}) \to (\exp \Gamma_i)(e_{0i} + e_{3i})$$

and

$$(e_{0i} - e_{3i}) \to [\exp(-\Gamma_i)](e_{0i} - e_{3i}),$$

and hence one cannot transform a point near $(e_{0i} + e_{3i})$ to a point near the origin unless $\mathrm{Re}\ \Gamma_i \gg 0$. But in this case the point $(e_{0i} + e_{3i})$ is also brought close to the origin. Moreover, any point is brought closer to the line $\lambda(e_{0i} - e_{3i})$. Thus the point ω will be brought closer to the line $\lambda(e_{0i} - e_{3i})$. As i increases, the lines $\lambda(e_{0i} - e_{3i})$ are constrained to lie in smaller and smaller cones about the line $\lambda(e_0(K_0) - e_3(K_0))$. Thus for sufficiently large i the point ω must be taken by Λ_i closer to a point near some small cone about $\lambda(e_0(K_0) - e_3(K_0))$, the cone becoming narrower with increasing i. Thus for sufficiently large i the $\Lambda_i\omega$ are constrained to lie in a conelike region about the $e_0(K_0) - e_3(K_0)$axis.

If, on the other hand, Λ_i^{-1} takes a point near the $(e''_{0i} + e''_{3i})$ axis to a point near the origin, then $\mathrm{Re}\ \Gamma_i \gg 0$. But then under Λ_i all points are moved further from the line $\lambda(e''_{0i} - e''_{3i})$ and closer to the line $\lambda(e''_{0i} + e''_{3i})$. Thus $\Lambda_i\omega$ must for sufficiently large i be far from the origin in a narrow cone-like region about the line $\lambda(e_0(E_0) + e_3(K_0))$. By taking i large enough, these two cones can be made arbitrarily narrow. Hence the allowed regions will not overlap. This gives the contradiction.

Theorem 2: Let D be a real domain satisfying the conditions of Lemma 1. Let $F(K)$ be defined (single-valuedly) and L-covariant over D, and be regular at points of D, in the (weak) sense that for any point $K' \in D$ there is a domain $D(K')$ containing K', and a function $F(K, K')$ that is regular at points $K \in D(K')$ and which coincides with $F(K)$ at points $D_r(K')$, some real domain contained in $D \cap D(K')$ and containing K'. Let C be any closed, bounded subset of D. Then there is an $\mathcal{L}$-covariant germ neighborhood whose base domain $B = \mathcal{L}B$ contains C and whose characteristic function coincides with $F(K)$ for $K \in C$.

Proof: Let K_0 be any point of C. Let $C(K_0, \rho)$ be a polysphere of radius ρ centered at K_0. Let $\rho_i \to 0$ be a monotonically decreasing set of radii converging to zero, and let the first ρ_i be small enough so that $C(K_0, \rho_i) \subset D(K_0)$, for all i. Suppose K_i is an infinite sequence of points in $\mathcal{L}C$ such that $K_i \in C(K_0, \rho_i)$ and

such that $F(K_i, K_0) \neq F'(K_i)$, where $F'(K)$ is the (single-valued) $\mathcal{L}$-covariant extension of $F(K)$ to $\mathcal{L}D$, which according to Lemma 1 exists. For each point $K_i \in \mathcal{L}C$ there is a point $\overline{K}_i \in C \cap \mathcal{L}K_i$. Since C is closed and bounded, the $\overline{K}_i$ have an accumulation point $\overline{K}_0 \in C$, and one can find a subsequence $\overline{K}_i \to \overline{K}_0 \in C$.

The point $\overline{K}_0$ cannot be on $\mathcal{L}K_0$. If it were, there would, according to Lemma 1A and the property of D, be a $\Lambda \in L$ such that $\overline{K}_0 \in \Lambda K_0$. This Λ would map the real domain $D_r(K_0)$ containing K_0 into some real domain containing $\overline{K}_0 \in C$. The intersection of this domain $\Lambda D_r(K_0)$ with D contains a real domain $D'_r(\overline{K}_0)$ containing $\overline{K}_0$. At points of $D'_r(\overline{K}_0)$ the value of $F(K)$ is given in terms of $F(K)$ at points of $D_r(K_0)$ by the L-covariance condition. Now according to Lemma 8 there is an $\mathcal{L}$-covariant germ neighborhood, with a base domain $\mathcal{L}D_0(K_0)$, having a characteristic function that coincides with $F(K; K_0)$ for $K \in D_0(K_0) \subset D(K_0)$. The value of $F(K)$ at points of $D'_r(\overline{K}_0) \cap \mathcal{L}D_0(K_0)$ must coincide with the value of the characteristic function at these points, since both are given in terms of $F(K)$ at $K \in D_r(K_0)$ by the L-covariance condition. But then $F'(K)$ must coincide with this characteristic function for all points of $\mathcal{L}D'_r(K_0) \cap \mathcal{L}D_0(K_0)$. Therefore $F'(K_i) = F(K_i, K_0)$ for all $K_i \in \mathcal{L}D'_r(K_0) \cap D_0(K_0)$. This precludes the possibility that a subsequence of the $\overline{K}_i \in C$ converge to $\overline{K}_0$. Thus the limit point $\overline{K}_0$ cannot lie on $\mathcal{L}K_0$.

But according to Lemma 11 the point $\overline{K}_0$ must lie on $\mathcal{L}K_0$, since $K_i \to K_0, \overline{K}_i \to \overline{K}_0, K_0$ and $\overline{K}_0$ are in D, and $\overline{K}_i \in \mathcal{L}K_i$. Thus there can be no infinite sequence of K_i with the specified properties. In particular for some $\rho_0 > 0$ there can be no points $K_i \in (C(K_0, \rho_0) \cap \mathcal{L}C)$ with $F'(K_i) \neq F(K_i, K_0)$.

Take some ρ'_0 with $\rho_0 > \rho'_0 > 0$ such that $C(K_0, \rho'_0) \subset D_0(K_0)$. Then the restriction of the $\mathcal{L}$-covariant germ neighborhood over $\mathcal{L}D_0(K_0)$ to the $\mathcal{L}$-covariant germ neighborhood over $\mathcal{L}C(K_0, \rho'_0)$ is an $\mathcal{L}$-covariant germ neighborhood whose characteristic function coincides with $F'(K)$ for $K \in (\mathcal{L}C \cap \mathcal{L}C(K_0, \rho'_0))$.

The point K_0 was an arbitrary point of C. This construction can be carried though for every point $K' \in C$. Let the radius corresponding to ρ'_0, but for the general $K' \in C$, be denoted by $\rho(K')$. One can take $\rho(K') < A$, some positive upper bound.

Let $r_i \to 0$ be an infinite sequence of positive numbers that decrease monotonically to zero. Let K_0 be an arbitrary point of C and let $C(K_0, r(K_0))$ be a polysphere of radius $r(K_0)$ about the point K_0. Let $r_i(K) > 0$ be less than $\rho(K)$ and less than r_i. Let K_i be a new set of points such that for each K_i there is a $K'_i \in C$ such that $K_i \in C(K_0, r_i(K_0)) \cap \mathcal{L}C(K'_i, r_i(K'_i))$ and such that the characteristic functions constructed above for K_0 and K'_i fail to coincide at $K = K_i$. Either an infinite sequence of K_i can be found or there is some $a(K_0)$ such that for $r_i < a(K_0)$ no such K_i exists. Suppose there is an infinite sequence of K_i. For each K_i there is a $\overline{K}_i \in \mathcal{L}K_i$ that is in $C(K'_i, r_i(K'_i))$. Since the union of the $C(K', \rho(K'))$, $K' \in C$, is a-bounded set the $\overline{K}_i$ must have an accumulation point K_0. This point must be in C, since the $r_i(K'_i) \to 0$. This point $\overline{K}_0$ is a limit point for a subsequence of the $\overline{K}_i$. The other $\overline{K}_i$ can be omitted. This limit point must, according to Lemma 11, lie on $\mathcal{L}K_0$. By virtue of the property of D there must then be a $\Lambda \in L$ such that $\overline{K}_0 = \Lambda K_0$.

Thus $\overline{K}_0$ is in $\mathcal{L}C(K_0, \rho(K_0))$. But since $\overline{K}_i \to \overline{K}_0$ and $K'_i \to \overline{K}_i$, also $K'_i \to \overline{K}_0$, and the $K'_i \in C$ must be in $\mathcal{L}C(K_0, \rho(K_0))$, except for a finite few which can be omitted. Then also the $C(K'_i, r_i(K'_i))$ will be completely inside $\mathcal{L}C(K_0, \rho(K_0))$, except for a finite few, which can be omitted. But then the characteristic functions over $\mathcal{L}C(K_0, \rho(K_0))$ and $\mathcal{L}C(K'_i, \rho(K'_i))$ must coincide at the points in $C(K'_i, r_i(K'_i))$ since they coincide over points of C contained in this polysphere, whose intersection with $\mathcal{L}C(K_0, \rho(K_0))$ is a domain, $C(K'_i, r_i(K'_i))$. But then the two characteristic functions must coincide at $\overline{K}_i$, and hence also at points of $\mathcal{L}\overline{K}_i$, and hence at K_i. This contradicts the assumption concerning the K_i. Thus there cannot be an infinite sequence of K_i satisfying those conditions, and hence there is an $a(K_0)$ such that for $r_i < a(K_0)$ the characteristic function over $\mathcal{L}C(K_0, \rho(K_0))$ coincides with the characteristic function over $\mathcal{L}C(K', \rho(K'))$ for all $K' \in C$, at all points $K \in C(K_0, r_i(K_0)) \cap \mathcal{L}C(K', r_i(K'))$ and hence at all points $K \in \mathcal{L}C(K_0, r_i(K_0)) \cap \mathcal{L}C(K', r_i(K'))$, where $r_i(K) < \min(r_i, \rho(K))$.

The point K_0 was an arbitrary point of C. Thus there is for every $K' \in C$ a characteristic radius $a(K') > 0$. If there is no lower bound $\bar{a} > 0$ such that $a(K') \geq \bar{a} > 0$ for all $K' \in C$, then one can find a sequence of $K_i \in C$ such that $a(K_i) \to 0$. These K_i must have an accumulation point $\overline{K} \in C$, though $a(\overline{K}) > 0$. But such an abrupt jump in $a(K)$ at $K = \overline{K}$ is not possible, for if $b(\overline{K}) = \min\{a(\overline{K}), \rho(\overline{K})\}$ then certainly $a(K) \geq \frac{1}{3}b(\overline{K}) > 0$ for $K \in C(K, \frac{1}{2}b(\overline{K})) \cap C$, since for these K all points of $C(K, \frac{1}{3}b(\overline{K}))$ are in $C(\overline{K}, b(\overline{K}))$, where the various characteristic functions coincide even with the weaker limit $a(\overline{K})$ on the r_i, and hence certainly for $r_i < \frac{1}{3}b(\overline{K})$. Thus there must be an $\bar{a} > 0$ such that $a(K') > \bar{a}$ for all $K' \in C$. Thus the union of the $\mathcal{L}$-covariant germ neighborhoods over the base domains $\mathcal{L}C(K', b'(K'))$, with $K' \in C$ and $b'(K') = \min(\bar{a}, \rho(K'))$, satisfies the required conditions; its base domain contains all points $K' \in C$, it has an $\mathcal{L}$-covariant characteristic function defined (single-valuedly) over its base domain $B = \mathcal{L}B$, and this characteristic function coincides with $F'(K)$ for $K \in B \cap \mathcal{L}C$.

Definition: An *enlargement* of a germ neighborhood N is a germ neighborhood containing N but not contained in N.

Definition: A germ neighborhood N will be called *maximal* if and only if no enlargement of N exists.

Lemma 12: Every germ neighborhood is contained in a maximal germ neighborhood.

Proof: Let N be an arbitrary germ neighborhood. A maximal germ neighborhood $N_M \supset N$ can be constructed as follows: Let $\{K_i\}$ be a denumerable sequence of points that is everywhere dense in the space in which lie the base points of the germs of the germ space. Let the K_i be enumerated. If a point K_i is reached that is in the base domain of an enlargement of N, then replace N by this enlargement (probably one of many possible enlargements) and proceed iteratively with the enumeration of the points of the sequence $\{K_i\}$. Because the union of a (finite or infinite) set of open sets is an open set, the result of this denumerable sequence of operations is a germ neighborhood N_a, since the base domain D_a is certainly con-

nected and the function $F_a(K)$ is defined (single-valuedly) over D_a and is regular at any point in D_a.

Let $\overline{D}_a$ be the set of accumulation points of the points $K_i \in D_a$. No enlargement of N_a can contain a point whose base point $\overline{K}$ is not in $\overline{D}_a$. For any such point $\overline{K}$ must be an accumulation point of points K_i not in D_a. Hence any enlargement containing a point with such a base point $\overline{K}$ would also contain a point with base point K_i not in D_a. This is impossible, for if there were such a K_i, then, when this K_i was reached in the enumeration, it could have been included in the base domain of an enlargement of the then current germ neighborhood, since enlargements of enlargements are themselves also enlargements. But the construction was such that if any K_i can be included in the base domain of any enlargement of the then-current germ neighborhood then it is in fact included in the enlargement associated with this K_i. Thus this K_i would be in D_a. Thus no K_i not in D_a and no accumulation point $\overline{K}$ of these K_i can be the base point of a point in any enlargement of N_a; the base points of all points of every enlargement of N_a are in $\overline{D}_a$.

If a point with base point $\overline{K} \in \overline{D}_a$ is in an enlargement of N_a, then the value of the characteristic function of the enlargement at $K = \overline{K}$ is unique; it is the same for any enlargement. For in order that a point with base point $\overline{K} \in \overline{D}_a$ be in an enlargement of N_a the corresponding characteristic function must be defined (single-valuedly) and regular in a neighborhood $N(\overline{K})$ of $\overline{K}$, and it must coincide with $F_a(K)$ for $K \in D_a \cap N(\overline{K})$. Thus it must coincide with $F_a(K)$ at the points $K_i \in D_a \cap N(\overline{K})$, which are dense in a neighborhood of $\overline{K}$. But the value of $F_a(K)$ at these points then determines the function at $K = \overline{K}$ by virtue of the continuity requirement implied by the regularity at $\overline{K}$ of the characteristic function of the enlargement.

Let D_M be the subset of $\overline{D}_a$ consisting of all the points of $D_a \subset \overline{D}_a$ and of all the base points of the points of any enlargement of N_a. Since the D_M is a union of domains each of which has a point in common with D_a the set D_M is a domain. Since the value of the characteristic function of any enlargement of N_a is uniquely defined for every $K \in D_M$ one may denote it by $F_M(K)$. This function is regular at every $K \in D_M$ because it is defined for $K \in D_M$ by an enlargement of N_a. Thus one may define a germ neighborhood $N_M \equiv N(D_M, F_M)$. This germ neighborhood contains N_a and hence N. Moreover, this germ neighborhood N_M is maximal. For any enlargement of N_M would also be an enlargement of N_a. But no enlargement of N_a exists that is also an enlargement of N_M because N_M contains every point of every enlargement of N_a.

Lemma 12A: Every $\mathfrak{L}$-covariant germ neighborhood is contained in a maximal germ neighborhood that is $\mathfrak{L}$-covariant.

Proof: Let $N = N(D, F)$ be an $\mathfrak{L}$-covariant germ neighborhood. If an enlargement of N exists, then an $\mathfrak{L}$-covariant enlargement also exists. To prove this, note first that any enlargement of N is a domain containing a point of N and some point not in N. By connecting these with a continuous curve one can, by a simple construction, find, in the enlargement, a point P_0 not in N such that any neighborhood of P_0 contains a point of N. Let the base point of P_0 be K_0. According to the Corollary to Lemma 8 there is a domain $D_0(K_0)$ containing K_0 such that the function defined in $D_0(K_0)$ as the characteristic function of the enlargement of N can be extended to a function $F'(K)$ that is $\mathfrak{L}$-covariant throughout $\mathfrak{L}D_0(K_0)$ and regular there. It must coincide with the characteristic function of the original $\mathfrak{L}$-covariant germ neighborhood, wherever both are defined, since both functions are $\mathfrak{L}$-covariant over their domains of definition and they coincide in $D_0(K_0) \cap D$, which contains a point of every orbit common to both domains. Thus the union of the original $\mathfrak{L}$-coinvariant given neighborhood N with the $\mathfrak{L}$-covariant germ neighborhood $N' \equiv N(\mathfrak{L}D_0, F')$ constitutes an enlargement of the original one, and this enlargement is $\mathfrak{L}$-covariant. Thus if an $\mathfrak{L}$-covariant germ neighborhood has an enlargement it has an $\mathfrak{L}$-covariant enlargement.

By virtue of this, one may proceed just as in Lemma 12, using however only $\mathfrak{L}$-covariant enlargements. After running through the denumerable set K_i one has an $\mathfrak{L}$-covariant germ neighborhood $N_a = N(D_a, F_a)$. Now, no point not in $\overline{D}_a$ can be the base point of an $\mathfrak{L}$-covariant enlargement. The set $D_M \subset \overline{D}_a$ is defined by using only $\mathfrak{L}$-covariant enlargements. Thus $N_M = N(D_M, F_M)$ is a germ domain that is maximal with respect of $\mathfrak{L}$-covariant enlargements. But then according to the first paragraph N_M is also maximal. Thus it is a maximal germ neighborhood that is $\mathfrak{L}$-covariant.

Definition: The base domain of a maximal germ neighborhood will be called a *sheet*.

Theorems 1A and 2, in conjunction with Lemma 12A, are summarized in

Theorem 3: Let $F(K)$ be a function defined (single-valuedly) over a real domain D. For every Λ in the real proper orthochronous homogeneous Lorentz group L and every K such that K and ΛK are in D, let $F(K)$ satisfy the Lorentz covariance condition

$$F(K) = \Lambda_s^{-1} F(\Lambda K).$$

If $F(K)$ is regular at some point $K \in D$, then the analytic continuation of $F(K)$ from the neighborhood of this point is defined over a manifold covered by a set of sheets each of which maps onto itself under any element of the proper homogeneous complex Lorentz group $\mathfrak{L}$. And for any sheet the associated function defined (single-valuedly) and regular at all points of this sheet satisfies the Lorentz covariance condition for all $\Lambda \in \mathfrak{L}$.

Moreover, if every point of D is a regular point of $F(K)$ and D has the property, specified in Lemma 1, that any points of D connected by a real $\Lambda \in \mathfrak{L}$ are connected by a $\Lambda \in L$, then any closed bounded subset C of D can be completely contained in a single $\mathfrak{L}$-covariant sheet, with $F(K)$ coinciding with the function defined over that sheet for $K \in C$.

Definition: The *restricted mass shell* is the subset W in the space of points $K \equiv \{k_1, \ldots, k_n\}$ that satisfy the n mass constraints

$$k_i^2 \equiv \sum_\mu (k_i^\mu)^2 = m_i^2, \qquad i = 1, \ldots, n,$$

the four conservation laws

$$\sum_i k_i^\mu = 0, \quad \mu = 0, 1, 2, 3,$$

and the condition that at least one pair of vectors k_i from the set $K = \{k_1, \ldots, k_n\}$ be linearly independent. The m_i are fixed positive numbers and $n \geq 4$.

Lemma 13: The restricted mass shell W is a $(3n - 4)$ complex-dimensional manifold.

Proof: Consider any point $\overline{K} \in W$. Let the k_i be ordered so that the last two are linearly independent. Let $\Lambda(\overline{K})$ be a Lorentz transformation that is such that the energy components of the vectors of $\overline{K}'(\overline{K}) \equiv \Lambda(\overline{K})\overline{K}$ are all nonzero. Such a $\Lambda(\overline{K})$ surely exists since the $\overline{K}$ are a finite set of nonzero vectors. Let the components 1, 2, 3, be numbered so that $\bar{k}'^0_{n-1}/\bar{k}'^0_n \neq \bar{k}'^3_{n-1}/\bar{k}'^3_n$. This is possible because k_{n-1} and k_n are linearly independent. By a small change in $\Lambda(\overline{K})$ that does not upset the above inequalities, one can also ensure that $(\bar{k}'^0_{n-1} + \bar{k}'^0_n)^2 \neq (k'^3_{n-1} + k'^3_n)^2$, since $k_{n-1} + k_n \neq 0$.

With $\Lambda(\overline{K})$ fixed in this way the set of vectors $K'(\overline{K}, K)$ is defined by $K'(\overline{K}, K) \equiv \Lambda(\overline{K})K$. The set $Z(\overline{K}, K)$ is then defined as the set of $(3n - 4)$ complex variables consisting of the three space components of the first $(n - 2)$ vectors of $K'(\overline{K}, K)$ and the first two components of the $(n - 1)$th vector of $K'(\overline{K}, K)$. The set of functions $Z(\overline{K}, K)$ are analytic functions (in fact linear functions) of the vectors of K. They define a set of mappings of K space onto Z space.

By virtue of the conditions that have been imposed on the vectors of K' the inverse transformation $K'(\overline{K}; Z)$ that maps Z back into $K' \in W$ is uniquely defined for $Z \in U(\overline{K})$, a domain containing $Z \equiv Z(\overline{K}, \overline{K})$, and is an analytic function of Z there. This follows from simple algebra or from the implicit function theorem,[11] the conditions of which are easily verified.

The set W can be made into a topological (Hausdorff) space by defining the open sets in W to be the restriction of open sets in K space to W. The topology in K space and Z space will be taken as the usual one induced by the Euclidian norm. With the topology of W defined in this way, the continuity of the functions $K(\overline{K}; Z)$ and $Z(\overline{K}; K)$, considered as mappings between K space and Z space, which follows from their analyticity, implies that these mappings are continuous mappings between $U(\overline{K})$ and its image $U_W(\overline{K}) \subset W$. For if a neighborhood of a point $Z \in U(\overline{K})$ maps into a K-space neighborhood of its image $K = K(\overline{K}; Z)$, then it must also map into a W-space neighborhood of $K = K(K; Z)$, since it maps into W. And, conversely, if a neighborhood of $K \in W$ in K space maps into a neighborhood in Z space, then its restriction to W also maps into this neighborhood. Thus the transformation $K(\overline{K}; Z)$ defines a one-to-one continuous mapping of neighborhoods of $\overline{K} \in W$ contained in $U_W(\overline{K})$ onto neighborhoods of $\overline{Z}$ contained in $U(\overline{K})$. Since the inverse is also continuous, the transformation is, by definition, a homeomorphism and the open sets in $U_W(\overline{K})$ and $U(\overline{K})$ are homeomorphic images of each other. Since $\overline{K}$ was an arbitrary point of W, the set W has an open covering by sets homeomorphic with open sets of $C^{(3n-4)}$, and hence W is a $(3n - 4)$ (complex)-dimensional manifold.[12]

Definition: The functions $K(\overline{K}; Z)$ and $Z(\overline{K}; K)$ will denote the functions introduced in the proof of Lemma 13. The function $Z(\overline{K}; K)$ is defined for $\overline{K} \in W$ and for all K, and for each $\overline{K} \in W$ it is an analytic function of K. The function $K(\overline{K}; Z)$ is defined for $\overline{K} \in W$ and $Z \in U(\overline{K})$, a domain containing $Z = Z(\overline{K}, \overline{K})$, and for each $\overline{K} \in W$ it is an analytic function of Z for $Z \in U(\overline{K})$. The function $K(\overline{K}; Z)$ maps points $Z \in U(\overline{K})$ into $U_W(\overline{K}) \subset W$. Its reciprocal is $Z(\overline{K}, K)$ in the sense that $Z(\overline{K}; K(\overline{K}; Z')) = Z'$ for $Z' \in U(\overline{K})$ and $K(\overline{K}; Z(\overline{K}; K')) = K'$ for $K' \in U_W(\overline{K}) \subset W$.

Remark: The set $U_W(\overline{K})$, as a homeomorphic image of the domain $U(\overline{K})$, is a domain.

Definition: The mapping $\phi(\overline{K})$ is a mapping of $K \in U_W(\overline{K})$ to $Z \in U(\overline{K})$ defined by $\phi(\overline{K})K = Z(\overline{K}; K)$ for $\overline{K} \in W$ and $K \in U_W(\overline{K})$.

Definition: The restricted mass shell W together with the complex structure induced by the collection $\{U_W(\overline{K}), \phi(\overline{K})\}$, $\overline{K} \in W$, is called the *complex analytic manifold* W.

Definition A: A function $M(K)$ defined on a restricted mass shell W will be called *regular* at $\overline{K} \in W$ if and only if $M(\phi^{-1}(\overline{K})Z) \equiv (M \circ \phi^{-1}(\overline{K}))Z$ is a regular function of Z at $Z = \phi(\overline{K})\overline{K}$.

Definition A': A function $M(K)$ defined on a restricted mass shell W will be called *regular* at $\overline{K} \in W$ if and only if $M \circ \phi^{-1}$ is regular at $Z = \phi\overline{K}$ for every one-to-one mapping ϕ, such that $\phi^{-1}Z \equiv K(Z) \in W$ is an analytic function at $Z = \phi\overline{K}$.

Lemma 14: Definitions A and A' are equivalent.

Proof: If $M(K)$ is regular (A') at $\overline{K} \in W$, it is certainly regular (A) at $\overline{K} \in W$ since $\phi(\overline{K})$ is a particular ϕ. If $M(K)$ is regular (A) at $\overline{K} \in W$ and ϕ is a one-to-one mapping such that $\phi^{-1}Z \equiv K(Z) \in W$ is an analytic function at $Z = \phi\overline{K}$, then $(M \circ \phi^{-1}(\overline{K}) \equiv M(K(Z)) \equiv M(\phi^{-1}(\overline{K})Z(\overline{K}; K(Z)))$. But $M \circ \phi^{-1}(\overline{K})$ is an analytic function of its argument Z for $Z = Z(\overline{K}; \overline{K})$, and $Z(\overline{K}; K)$ is an analytic function of K for $K = \overline{K}$, and $K(Z)$ is an analytic function at $Z = \phi\overline{K}$. Thus $M \circ \phi^{-1}$ is an analytic function of Z at $\phi\overline{K}$, since it is an analytic function of an analytic function of an analytic function.

Theorem 4: The preceding theorems and lemmas remain valid if $F(K)$ is replaced by $M(K)$ defined on a restricted mass shell W, and all domains are taken to be domains relative to W.

Proof: The mass shell contains all points having the same scalar invariants as any point on it, and in particular all points on any orbit intersecting it. This is the only global property of the K space that was used in any of the above proofs. For local properties one replaces the topology of K space by the topology of W space. Some of the proofs become vastly simplified because for real $K \in W$ one has $n = r$.

Remark 1: Any real domain of W satisfies the condition of Lemma 1; two real points of W connected by a real $\Lambda \in \mathcal{L}$ that is not a $\Lambda \in L$ must have opposite energy components, and hence they cannot both be in a real domain in W. The M_c functions have

been shown to satisfy the L-covariance condition at regular physical points. Thus if D_γ is a real (physical) domain of regularity of M_c (defined over W), then, by Theorem 3, any closed bounded set $C \subset D_\gamma$ is contained in a sheet S that maps onto itself under any $\Lambda \in \mathfrak{L}$, and the function M_c has a single-valued analytic continuation throughout S, and is $\mathfrak{L}$ covariant there.

Remark 2: One consequence of the above remark is a slight weakening of the assumptions needed for the S-matrix proof of CPT invariance. In the original proof[3] the postulate of minimal analyticity required the existence of a physical sheet that was bounded by cuts defined by equations involving only scalar invariants. This condition on the boundary was imposed specifically to eliminate problems associated with a possible multivaluedness in the continuation to the CPT image point. However, a consequence of Theorem 3 drawn in the above remark is the existences of the single-valued $\mathfrak{L}$-covariant continuation to the CPT-image point.[13] The proof of CPT invariance in this way is similar to the field-theoretic proof of Jost[7]; that proof rested heavily on Lemma 1 of Hall and Wightman, which is rather analogous to Theorem 3.

Remark 3: In the construction of the decomposition of the analytic M_c functions into analytic functions of scalar invariants times standard (polynomial) covariants,[15,16] the $\mathfrak{L}$-covariance of the domains of regularity is a basic ingredient. A fundamental result that can be drawn from this paper (Theorems 1 and 3, and the L-covariance at physical points established in Ref. 1) is that any domain of regularity of M_c containing a physical point is $\mathfrak{L}$-covariant. Since M_c is defined by analytic continuation from physical points, any domain of regularity of M_c is $\mathfrak{L}$-covariant.

Remark 4: Hepp[15] has shown that the complex mass shell is a "normal analytic set." It follows that Theorem 4 can be proved on the entire complex mass shell: The restriction to the restricted mass shell W is unnecessary.

Remark 5: The results of this paper have been extended to the orthogonal and symplectic groups by Seiler in Ref. 17.

ACKNOWLEDGMENTS

The author is grateful to Professor R. Jost, Dr. David Williams, Dr. P. Minkowski, and Dr. Seiler for important communications regarding this work.

APPENDIX: GENERALIZED SPINOR CALCULUS

The Lorentz transformations $\Lambda^\mu_\nu(A,B)$ are defined by the equation

$$A\sigma_\nu B = \sigma_\mu \Lambda^\mu_\nu(A,B), \tag{A1}$$

where $\sigma_\mu = (\sigma_0, \boldsymbol{\sigma})$ are the usual Pauli matrices, and A and B are unimodular 2×2 matrices. The unimodular 2×2 matrices form a group. The canonical irreducible representations of this group of dimension $(2a + 1)$ are generated by the recursion relation

$$A^{(a)\alpha'}{}_{\alpha} = C_{bc}(a,\alpha;\beta,\gamma)C_{bc}(a,\alpha';\beta',\gamma')A^{(b)\beta'}{}_{\beta}A^{(c)\gamma'}{}_{\gamma}, \tag{A2}$$

where the coefficients C are the usual Clebsch–Gordan coefficients. The $A^{(1/2)}$ identified with A.

Generalized spinor indices of order $(2a + 1)$ are introduced. They can be either upper or lower and either dotted of undotted. The distinction between indices of these various types is with respect to the effect upon them of the operator Λ_s. The action of this operator is defined as follows:

$$\begin{aligned} &\Lambda_s\xi_\alpha = A^{(a)}{}_{\alpha}{}^{\alpha'}\xi_{\alpha'}, \qquad \Lambda_s\xi_{\dot\alpha} = \xi_{\dot\alpha'}B^{(a)\dot\alpha'}{}_{\dot\alpha}, \\ &\Lambda_s\xi^\alpha = \xi^{\alpha'}(A^{(a)})^{-1}{}_{\alpha'}{}^{\alpha}, \qquad \Lambda_s\xi^{\dot\alpha} = (B^{(a)})^{-1}{}^{\dot\alpha}{}_{\dot\alpha'}\xi^{\dot\alpha'} \end{aligned} \tag{A3}$$

Here $B^{(a)}$ is defined by the analog of (A2) with B's in place of A's. If a function has several spinor indices, then Λ_s acts individually on each in the manner given by (A3).

Let $f(V)$ be a function of a set $V = \{v \cdots v_n\}$ of 4-vectors. Let $\Lambda V = \{\Lambda v_1, \ldots, \Lambda v_n\}$, where

$$(\Lambda v)^\mu = \Lambda^\mu{}_\nu(A,B)v^\nu. \tag{A4}$$

If $f(V)$ carries spinor indices and satisfies the equation

$$\Lambda_s f(V) = f(\Lambda V), \tag{A5}$$

then f will be called a covariant spinor function. The Pauli matrices σ_μ will be considered to have matrix elements $\sigma_{\mu\alpha\dot\beta}$. Then the function

$$g(v) \equiv \sigma \cdot v \tag{A6}$$

is, by virtue of the conventions adopted, a covariant spinor function.

A generalization of the Pauli $\sigma_{\mu\alpha\dot\beta}$ to higher dimension is defined by the recursion formula

$$\begin{aligned} \sigma^a_{(\mu)\ \alpha\dot\alpha'} = C_{bc}(a,\alpha;\beta,\gamma)C_{bc}(a,\dot\alpha';\dot\beta',\dot\gamma') \\ \times \sigma^b_{(\mu')\ \beta\dot\beta'}\ \sigma^c_{(\mu'')\ \gamma\dot\gamma'}, \end{aligned} \tag{A7}$$

where $a = b + c$. Here $(\mu) = \{\mu_1, \ldots, \mu_{2a}\}$ is a set of $2a$ vector indices and $(\mu) = (\mu') + (\mu'')$. The $\sigma^a_{(\mu)}$ is symmetric and traceless in each pair of tensor indices.[16]

---

* This work done under the auspices of the U.S. Atomic Energy Commission.

1 H. P. Stapp, in *Seminar on High-Energy Physics and Elementary Particles (Trieste, 1965)* (International Atomic Energy Agency, Vienna 1965).

2 I use the notation of Ref. 3. See also Appendix A and A. O. Barut, I. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).

3 H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

4 D. Hall and A. S. Wightman, Kgl. Dansk. Vid. Sels. Mat.-Fys. Medd. **31**, No. 5 (1957).

5 A detailed discussion is given below for the case of $n(K_1) > r(K_1)$.

6 This has also been proved by E. Wigner, Ann. Math. **40**, 149 (1939).

7 R. Jost, *Theoretical Physics in the Twentieth Century*, edited by M. Fierz and V. F. Weisskopf (Interscience, New York, 1960).

8 F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1964), Vol. II, Theorem 1.

9 P. Minkowski, D. Williams, and R. Seiler, *Proceedings of the Sym-*

posium on the Lorentz Group, Seventh Annual Summer Institute for Theoretical Physics, Boulder, Colorado, June 1964 (U. of Colorado Press, Boulder, 1965).
[10] H. J. Bremmermann, "Complex Analysis in Several Variables," Mathematics Department, University of California, Berkeley, California (unpublished notes).
[11] S. Bochner and W. Martin, *Several Complex Variables* (Princeton U. P., Princeton, N.J., 1948), p. 39.
[12] L. Ahlfors and L. Sario, *Riemann Surfaces* (Princeton, U.P., Princeton, N.J., 1960), p. 14.
[13] For a proof that the continuation generated by a complex Lorentz transformation takes one to the point that corresponds to the CPT inverse physical process see Ref. 14. In Ref. 3 this fact was a direct consequence of the stronger assumptions made there.
[14] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968).
[15] K. Hepp, Helv. Phys. Acta **36**, 355 (1963).
[16] D. Williams, Ph.D. thesis (University of California, Berkeley, 1963), or University of California, Lawrence Radiation Laboratory Report UCRL 11113, 1963.
[17] R. Seiler, Helv. Phys. Acta **39**, 641 (1966). [See also *Lectures in Theoretical Physics* (Gordon and Breach, New York, 1968), Vol. XID, p. 195.]

The $\lambda\varphi_2^4$ Quantum Field Theory without Cutoffs. IV. Perturbations of the Hamiltonian

James Glimm*
Courant Institute of Mathematical Sciences, New York University, New York, New York

and

Arthur Jaffe†
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received 3 March 1972)

We introduce an inductive method to estimate the shift δE in the vacuum energy, caused by a perturbation δH of the $\mathcal{P}(\varphi)_2$ Hamiltonian H. We prove that if δH equals the field bilinear form $\varphi(x,t)$, then δE is finite. We show that the vacuum expectation values of products of fields (Wightman functions) exist and are tempered distributions. They determine, via the reconstruction theorem, essentially self-adjoint field operators $\varphi(f)$, for real test functions $f \in \mathcal{S}(R^2)$. We also bound the perturbation of the $\mathcal{P}(\varphi)_2$ Hamiltonian by a polynomial $(\mathcal{P}_1(\varphi))(h) = \delta H$, so long as $\mathcal{P} + \mathcal{P}_1$ is formally positive. In that case, and with $\|h\|_\infty \leq 1$, δE is bounded by const(1 + diam supp h).

1. THE MAIN RESULTS

We derive new estimates on the vacuum energy in the $\mathcal{P}(\varphi)_2$ quantum field model. These estimates permit us to establish one of the missing Wightman axioms for $\mathcal{P}(\varphi)_2$.

Theorem A: The Wightman functions or vacuum expectation values

$$\langle \Omega, \varphi(x_1,t_1)\varphi(x_2,t_2)\cdots\varphi(x_n,t_n)\Omega\rangle, \tag{1.1}$$

for the $\mathcal{P}(\varphi)_2$ model exist as tempered distributions in $\mathcal{S}'(R^{2n})$.

Our estimates bound the shift in the vacuum energy due to a local perturbation of the Hamiltonian H. The simplest such estimate is

Theorem B: The field $\varphi(x,t)$ is a bilinear form on the domain $\mathfrak{D}(H^{1/2}) \times \mathfrak{D}(H^{1/2})$, and $H + \varphi(x,t)$ is bounded from below.

Other new results for the $\mathcal{P}(\varphi)_2$ quantum field model are established in Theorem 1.1–Corollary 1.3. The methods and estimates of this paper appear to have other applications. For example, it should be possible to show that the Wightman functions (1.1) have exactly the singularities predicted by perturbation theory. Furthermore, the methods used here appear useful in the study of the more singular φ_3^4 model.

Solutions to the $\mathcal{P}(\varphi)_2$ model were constructed in earlier work of the authors and of Rosen. They are known to satisfy the Haag–Kastler axioms for quantum fields, and most of the Wightman axioms, as was shown by the authors, Cannon and Rosen. See Refs. 1 and 2 for references, notation, and proofs.

The Hamiltonian operator in the $\mathcal{P}(\varphi)_2$ theory is defined as the limit of approximate, or cutoff, Hamiltonians $H(g)$,

$$H(g) = H_0 + H_1(g) - E(g)I. \tag{1.2}$$

Here $H_0 = H(0)$ is the free Hamiltonian for bosons of mass $m > 0$, and

$$H_1(g) = \int :\mathcal{P}(\varphi(x)):g(x)dx. \tag{1.3}$$

The space cutoff g has compact support, and

$$0 \leq g(x) \leq 1. \tag{1.4}$$

The vacuum energy $E(g)$ is a finite constant (divergent as $g \to 1$) and chosen so that

$$0 = \text{inf spectrum } H(g). \tag{1.5}$$

Certain uniform estimates on $H(g)$ allow us to pass to a limit $g = 1$ and so obtain a Hamiltonian without cutoffs,

$$H = \lim H(g), \tag{1.6}$$

see Refs. 1 and 3.

We now describe our new uniform estimates for cutoff Hamiltonians. We use these estimates to establish properties of the theory without cutoffs ($g \equiv 1$), such as Theorems A and B.

Case 1. Perturbations of arbitrary degree: Let h be a real function $\|h\|_\infty \leq 1$, $D = 1 +$ diam. supp. $h < \infty$. Let $p = \max\{\deg\mathcal{P}, \deg\mathcal{P}_1\}$, and let

$$0 \leq \mathcal{P}(\zeta)g(x), \qquad 0 \leq \mathcal{P}(\zeta)g(x) + \mathcal{P}_1(\zeta)h(x) \tag{1.7}$$

for all real ζ, x. If $\deg\mathcal{P} = \deg\mathcal{P}_1$ and $\mathcal{P} \neq \mathcal{P}_1$, we assume that the coefficient of the degree p term in $\mathcal{P}g + \mathcal{P}_1 h$ is bounded away from zero on a neighborhood of supp h. Let

$$\delta H = \int :\mathcal{P}_1(\varphi(x)): h(x)dx = \delta H(h), \tag{1.8}$$

$$\delta E = \text{inf spectrum}\{H(g) + \delta H\} = \delta E(g,h). \tag{1.9}$$

Case 2. Linear perturbations: Let h be a real C^∞ function with

posium on the Lorentz Group, Seventh Annual Summer Institute for Theoretical Physics, Boulder, Colorado, June 1964 (U. of Colorado Press, Boulder, 1965).
[10] H. J. Bremmermann, "Complex Analysis in Several Variables," Mathematics Department, University of California, Berkeley, California (unpublished notes).
[11] S. Bochner and W. Martin, *Several Complex Variables* (Princeton U. P., Princeton, N.J., 1948), p. 39.
[12] L. Ahlfors and L. Sario, *Riemann Surfaces* (Princeton, U.P., Princeton, N.J., 1960), p. 14.
[13] For a proof that the continuation generated by a complex Lorentz transformation takes one to the point that corresponds to the CPT inverse physical process see Ref. 14. In Ref. 3 this fact was a direct consequence of the stronger assumptions made there.
[14] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968).
[15] K. Hepp, Helv. Phys. Acta **36**, 355 (1963).
[16] D. Williams, Ph.D. thesis (University of California, Berkeley, 1963), or University of California, Lawrence Radiation Laboratory Report UCRL 11113, 1963.
[17] R. Seiler, Helv. Phys. Acta **39**, 641 (1966). [See also *Lectures in Theoretical Physics* (Gordon and Breach, New York, 1968), Vol. XID, p. 195.]

The $\lambda\varphi_2^4$ Quantum Field Theory without Cutoffs. IV. Perturbations of the Hamiltonian

James Glimm*
Courant Institute of Mathematical Sciences, New York University, New York, New York

and

Arthur Jaffe†
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received 3 March 1972)

We introduce an inductive method to estimate the shift δE in the vacuum energy, caused by a perturbation δH of the $\mathcal{P}(\varphi)_2$ Hamiltonian H. We prove that if δH equals the field bilinear form $\varphi(x,t)$, then δE is finite. We show that the vacuum expectation values of products of fields (Wightman functions) exist and are tempered distributions. They determine, via the reconstruction theorem, essentially self-adjoint field operators $\varphi(f)$, for real test functions $f \in \mathcal{S}(R^2)$. We also bound the perturbation of the $\mathcal{P}(\varphi)_2$ Hamiltonian by a polynomial $(\mathcal{P}_1(\varphi))(h) = \delta H$, so long as $\mathcal{P} + \mathcal{P}_1$ is formally positive. In that case, and with $\|h\|_\infty \leq 1$, δE is bounded by const(1 + diam supp h).

1. THE MAIN RESULTS

We derive new estimates on the vacuum energy in the $\mathcal{P}(\varphi)_2$ quantum field model. These estimates permit us to establish one of the missing Wightman axioms for $\mathcal{P}(\varphi)_2$.

Theorem A: The Wightman functions or vacuum expectation values

$$\langle \Omega, \varphi(x_1,t_1)\varphi(x_2,t_2)\cdots\varphi(x_n,t_n)\Omega\rangle, \tag{1.1}$$

for the $\mathcal{P}(\varphi)_2$ model exist as tempered distributions in $\mathcal{S}'(R^{2n})$.

Our estimates bound the shift in the vacuum energy due to a local perturbation of the Hamiltonian H. The simplest such estimate is

Theorem B: The field $\varphi(x,t)$ is a bilinear form on the domain $\mathfrak{D}(H^{1/2}) \times \mathfrak{D}(H^{1/2})$, and $H + \varphi(x,t)$ is bounded from below.

Other new results for the $\mathcal{P}(\varphi)_2$ quantum field model are established in Theorem 1.1–Corollary 1.3. The methods and estimates of this paper appear to have other applications. For example, it should be possible to show that the Wightman functions (1.1) have exactly the singularities predicted by perturbation theory. Furthermore, the methods used here appear useful in the study of the more singular φ_3^4 model.

Solutions to the $\mathcal{P}(\varphi)_2$ model were constructed in earlier work of the authors and of Rosen. They are known to satisfy the Haag–Kastler axioms for quantum fields, and most of the Wightman axioms, as was shown by the authors, Cannon and Rosen. See Refs. 1 and 2 for references, notation, and proofs.

The Hamiltonian operator in the $\mathcal{P}(\varphi)_2$ theory is defined as the limit of approximate, or cutoff, Hamiltonians $H(g)$,

$$H(g) = H_0 + H_1(g) - E(g)I. \tag{1.2}$$

Here $H_0 = H(0)$ is the free Hamiltonian for bosons of mass $m > 0$, and

$$H_1(g) = \int :\mathcal{P}(\varphi(x)):g(x)dx. \tag{1.3}$$

The space cutoff g has compact support, and

$$0 \leq g(x) \leq 1. \tag{1.4}$$

The vacuum energy $E(g)$ is a finite constant (divergent as $g \to 1$) and chosen so that

$$0 = \text{inf spectrum } H(g). \tag{1.5}$$

Certain uniform estimates on $H(g)$ allow us to pass to a limit $g = 1$ and so obtain a Hamiltonian without cutoffs,

$$H = \lim H(g), \tag{1.6}$$

see Refs. 1 and 3.

We now describe our new uniform estimates for cutoff Hamiltonians. We use these estimates to establish properties of the theory without cutoffs ($g \equiv 1$), such as Theorems A and B.

Case 1. Perturbations of arbitrary degree: Let h be a real function $\|h\|_\infty \leq 1$, $D = 1 +$ diam. supp. $h < \infty$. Let $p = \max\{\deg\mathcal{P}, \deg\mathcal{P}_1\}$, and let

$$0 \leq \mathcal{P}(\zeta)g(x), \qquad 0 \leq \mathcal{P}(\zeta)g(x) + \mathcal{P}_1(\zeta)h(x) \tag{1.7}$$

for all real ζ, x. If $\deg\mathcal{P} = \deg\mathcal{P}_1$ and $\mathcal{P} \neq \mathcal{P}_1$, we assume that the coefficient of the degree p term in $\mathcal{P}g + \mathcal{P}_1 h$ is bounded away from zero on a neighborhood of supp h. Let

$$\delta H = \int :\mathcal{P}_1(\varphi(x)): h(x)dx = \delta H(h), \tag{1.8}$$

$$\delta E = \text{inf spectrum}\{H(g) + \delta H\} = \delta E(g,h). \tag{1.9}$$

Case 2. Linear perturbations: Let h be a real C^∞ function with

$$\| \mu^{-(1/2)-\epsilon} h \|_2 \leq 1 \tag{1.10}$$

for $\epsilon = (8p)^{-1}$, $p = \deg\mathcal{P}$, and let $D = 1 + \text{diam supp } h < \infty$. Let

$$\delta H = \varphi(h) = \int \varphi(x)h(x)dx, \tag{1.11}$$

and let

$$\delta E = \delta E(g,h) = \inf \text{spectrum } \{H + \varphi(h)\}. \tag{1.12}$$

Theorem 1.1: There exists a constant M, independent of g and h, such that in Cases 1 and 2

$$|\delta E| \leq MD. \tag{1.13}$$

Remarks: In Case 2, if $1 \leq \|\mu^{-1/2-\epsilon}h\|_2 < \infty$, we prove (1.13) with $M = M(\|\mu^{-1/2-\epsilon}h\|_2)$. Formal reasoning indicates that $\epsilon = \frac{1}{2}$ is possible; but a proof requires improvements in Sec. 3B.

Theorems A and B are corollaries to Theorem 1.1.

We consider space–time averaged fields

$$\varphi(f) = \int \varphi(x,t)f(x,t)dxdt$$

in the theory with the cutoff Hamiltonian $H(g)$ and vacuum Ω_g, $H(g)\Omega_g = 0$. The g-cutoff vacuum expectation values are tempered distributions.[4] The analog to Theorem A in the cutoff model is:

Theorem A_g: There exist translation invariant Schwartz space norms $|\cdot|_\nu$, independent of g, such that

$$|\langle\Omega_g, \varphi(f_1)\cdots\varphi(f_n)\Omega_g\rangle| \leq |f_1|_1\cdots|f_n|_n. \tag{1.14}$$

The new aspect of (1.14) is a bound that is independent of g. In order to ensure space translation invariance of a limiting vacuum, we average the vacuum expectation values $\langle\Omega_g, \varphi(f_1)\cdots\varphi(f_n)\Omega_g\rangle$ over space translations in an interval of length 0 (diam supp g). Since the norms $|\cdot|_\nu$ are translation invariant, the space averaged expectation values are also bounded by $|f_1|_1\cdots|f_n|_n$. By the methods of Refs. 3 and 5 we pass to a subsequence as $g \to 1$. As in Refs. 3 and 5, the limiting vacuum expectation values (1.1) satisfy the bound (1.14) with $g = 1$. Furthermore, the limiting vacuum expectation values (1.1) are obtained as limits as $g \to 1$ of a subsequence of the space averaged $\langle\Omega_g, \varphi(f_1)\cdots\varphi(f_n)\Omega_g\rangle$. Applying the Schwartz nuclear theorem to (1.14) proves that the vacuum expectation values are tempered distributions. Thus Theorem A follows from Theorem A_g, which we establish below.

We remark that the estimate of Theorem 1.1 is also valid with a space cutoff given by a periodic box of volume V. (See Refs. 1 and 5 for notation.) In Cases 1_V and 2_V we define δH and δE as in (1.8)–(1.12), with an additional dependence on V. We assume supp $h \subset [-\frac{1}{2}V, \frac{1}{2}V]$.

Theorem 1.1_V: There is a constant M independent of g, h, and V such that in Cases 1_V and 2_V

$$|\delta E| \leq MD.$$

Our proof in this paper of Theorem 1.1, combined with the methods of Ref. 5, Sec. 2 that deal with estimates in a periodic box, prove Theorem 1.1_V. Likewise, we can establish a relevant Theorem A_V, and we pass to the $V = \infty$ limit in the vacuum expectation values. With this method, it is not necessary to average the vacuum expectation values

$$\langle\Omega_V, \varphi_V(f_1)\cdots\varphi_V(f_n)\Omega_V\rangle$$

over space translations in order to ensure translation invariance of the limiting vacuum Ω.

We now assume Theorem 1.1 and turn to the proof of Theorems A_g and B. We use

Lemma 1.1: Let $0 \leq H = H^*$. Let A and $\dot{A} = [iH, A]$ be real bilinear forms with domain $\mathcal{C}^\infty(H) \times \mathcal{C}^\infty(H)$. Let $R = (H + I)^{-1}$, and let

$$\| R^{1/2}AR^{1/2}\| + \| R^{1/2}\dot{A}R^{1/2}\| \leq M. \tag{1.15}$$

Then A uniquely determines a symmetric operator (also called A) with domain $\mathfrak{D}(H)$ and for a universal constant α,

$$\|AR\| \leq \alpha M. \tag{1.16}$$

Proof: Let $R(y) = [H + (y+1)I]^{-1}$. Then

$$R^{1/2} = \frac{1}{\pi}\int_0^\infty \lambda^{-1/2}R(\lambda)d\lambda.$$

On $\mathcal{C}^\infty(H) \times \mathcal{C}^\infty(H)$,

$$\begin{aligned} AR^{1/2} &= R^{1/2}A + \frac{1}{\pi}\int_0^\infty \lambda^{-1/2}[A, R(\lambda)]d\lambda \\ &= R^{1/2}A - \frac{i}{\pi}\int_0^\infty \lambda^{-1/2}R(\lambda)\dot{A}R(\lambda)d\lambda. \end{aligned} \tag{1.17}$$

By (1.15), $R(\lambda)\dot{A}R(\lambda)$ is bounded, norm continuous in λ, with norm bounded by $(\lambda+1)^{-1}M$. Hence the integral over λ in (1.17) is norm convergent, and by (1.17)

$$\begin{aligned} |\langle\theta, AR\theta\rangle| &\leq \|R^{1/2}AR^{1/2}\|\|\theta\|^2 + O(M)\|\theta\|^2 \\ &\leq \alpha M\|\theta\|^2. \end{aligned}$$

The constant α depends only on the integral of $\lambda^{-1/2}(\lambda+1)^{-1}$. Lemma 1.1 then follows by the Riesz representation theorem.

Proof of Theorem A_g: Let g be fixed. We apply Theorem 1.1, Case 2, for an arbitrary real element $h/\|\mu^{-(1/2)-\epsilon}h\|_2$ of $\mathcal{C}_0^\infty(R^1)$. This yields

$$\begin{aligned} \pm\varphi(h) &\leq \text{const}\,\|\mu^{-(1/2)-\epsilon}h\|_2[H(g) + I] \\ &\leq \text{const}\,\|h\|_1[H(g) + I] \end{aligned} \tag{1.18}$$

with constants independent of g, but possibly depending on D. By taking a partition of unity, $1 = \sum\zeta_i$, where ζ_i is the translate of a fixed $\mathcal{C}_0^\infty$ function, we obtain

$$\pm\varphi(h\zeta_i) \leq \text{const}\,\|h\zeta_i\|_1[H(g) + I],$$

and summing

$$\pm\varphi(h) \leq \text{const}\,\|h\|_1[H(g) + I],$$

with a constant independent of g and h. For real $f \in \mathcal{S}(R^2)$, we integrate over time to obtain

$$\pm\varphi(f) \leq \text{const}\,\|f\|_1[H(g) + I] \tag{1.18'}$$

with a constant independent of f, g. Likewise

$$\pm \dot{\varphi}(f) = \mp \varphi(D_t f) \leq \text{const} \|D_t f\|_1 [H(g) + I],$$

so by Lemma 1.1

$$\|\varphi(f) R_g\| \leq \alpha(\|f\|_1 + \|D_t f\|_1), \tag{1.19}$$

where $R_g = [H(g) + I]^{-1}$. On the domain $\mathcal{C}^\infty(H(g)) \times \mathcal{C}^\infty(H(g))$,

$$\langle \chi, [H(g) + I]^r \varphi(f) \theta \rangle = \sum_{j=0}^{r} \binom{r}{j} i^j \langle \chi, \varphi(D_t^j f)[H(g) + I]^{r-j} \theta \rangle. \tag{1.20}$$

Thus by (1.19)

$$\|(H(g) + I)^r \varphi(f) \theta\| \leq |f|_{r+1} \|(H(g) + I)^{r+1} \theta\|, \tag{1.21}$$

where

$$|f|_{r+1} = \alpha 2^{r+1} \sum_{j=0}^{r+1} \|D_t^j f\|_1 \tag{1.22}$$

is a translation invariant Schwartz space norm. The bound (1.14), and hence Theorem A_g follows from (1.21) and (1.22).

The Hilbert space $\mathcal{H}_{\text{GNS}}$ associated with the infinite volume limit is defined by means of the GNS construction, see Refs. 3 and 5. The limiting vacuum expectation values (1.1) can also be used to construct an infinite volume Hilbert space $\mathcal{H}_W$, by the Wightman reconstruction theorem. The vacuum expectation values in these two theories coincide. Hence by the uniqueness portion of the reconstruction theorem, $\mathcal{H}_W$ can be identified as a subspace of $\mathcal{H}_{\text{GNS}}$.

Proposition 1.1: $\mathcal{H}_W = \mathcal{H}_{\text{GNS}}$: That is, the Wightman and GNS constructions yield the same infinite volume Hilbert space. In particular the cyclic subspace $\mathfrak{D}$ generated from Ω by polynomials in $\varphi(f)$, $f \in \mathcal{S}(R^2)$, is dense in $\mathcal{H}_{\text{GNS}}$.

Proof: Let $\mathcal{H}_{\alpha\beta}$ be the subset of the real elements $\mathcal{S}_R(R^2)$ of $\mathcal{S}(R^2)$ satisfying

$$\|D_t^n f\|_1 \leq \alpha \beta^n.$$

Clearly $\cup_{\alpha\beta} \mathcal{H}_{\alpha\beta}$ is dense in $\mathcal{S}_R(R^2)$. Let $\mathfrak{F}_{\alpha\beta}$ be the linear span of vectors

$$e^{i\varphi(f_1)} \cdots e^{i\varphi(f_n)} \Omega$$

for $f_j \in \mathcal{H}_{\alpha\beta}$, $n = 0, 1, \cdots$. By definition $\cup_{\alpha\beta} \mathfrak{F}_{\alpha\beta}$ is dense in $\mathcal{H}_{\text{GNS}}$. Using (1.19) and (1.20), we establish for $f_j \in \mathcal{H}_{\alpha\beta}$,

$$\|\varphi(f_1) \cdots \varphi(f_k) \Omega\| \leq B^k k! \tag{1.23}$$

with a constant $B = B(\alpha, \beta)$. (See Ref. 4, p. 389). We now prove that $\mathfrak{D}$ is dense. For any χ we define

$$f(\lambda) = \langle \chi, e^{i\lambda_1 \varphi(f_1)} \cdots e^{i\lambda_n \varphi(f_n)} \Omega \rangle$$

as an iterated power series, summing first λ_1, then $\lambda_2, \cdots$. These series converge on account of (1.23). Thus if $\chi \perp \mathfrak{D}$, $f(\lambda) = 0$. In that case $\chi \perp \cup_{\alpha\beta} \mathfrak{F}_{\alpha\beta}$, so $\chi = 0$ and $\mathfrak{D}$ is dense.

We give three short corollaries to Theorem 1.1 which establish further properties of the space-time averaged fields $\varphi(f)$.

Corollary 1.1: Let $f, D_t f \in \mathcal{L}_1$. Then $\varphi(f)$ is an operator on $\mathfrak{D}(H)$ and

$$\|\varphi(f)(H + I)^{-1}\| \leq \alpha(\|f\|_1 + \|D_t f\|_1). \tag{1.24}$$

Corollary 1.2: Let f be real, and let $f, D_t f, D_{tt} f \in \mathcal{L}_1$. Then $\varphi(f)$ is essentially self-adjoint on any core for H.

Corollary 1.3: Let $f \in \mathcal{S}(R^2)$ be real, and let $\varphi(f)$ be the operator obtained from the vacuum expectation values (1.1) by the reconstruction theorem. [The domain of $\varphi(f)$ is $\mathfrak{D}$.] Then $\varphi(f)$ is essentially self-adjoint.

Proofs: By (1.18′), for real f,

$$\pm \langle \theta_g, \varphi(f) \theta_g \rangle \leq O(1) \|f\|_1 \langle \theta_g, [H(g) + I] \theta_g \rangle \tag{1.25}$$

for $\theta_g = \varphi(f_1) \cdots \varphi(f_n) \Omega_g \in \mathfrak{F}$. Let

$$\theta = \varphi(f_1) \cdots \varphi(f_n) \Omega \in \mathfrak{D} \subset \mathcal{H}_{\text{GNS}} = \mathcal{H}_W = \mathfrak{F}_{\text{ren}}.$$

We recall that the vacuum expectation value $\langle \theta, \varphi(f) \theta \rangle$ is obtained as a limit of a subsequence from space translation averages of $\langle \theta_g, \varphi(f) \theta_g \rangle$, and similarly $\langle \theta, H\theta \rangle$ is obtained as a limit from the cutoff expectation values $\langle \theta_g, H(g) \theta_g \rangle$. Thus we obtain in the limit $g = 1$,

$$\pm \langle \theta, \varphi(f) \theta \rangle \leq O(1) \|f\|_1 \langle \theta, (H + I) \theta \rangle.$$

Furthermore, we have $\dot{\varphi}(f) = -\varphi(D_t f)$ on $\mathfrak{D} \times \mathfrak{D}$, as a consequence of the corresponding identity in the cutoff theory. Hence Corollary 1.1 follows by Lemma 1.1.

By Proposition 1.1, the domain $\mathfrak{D}$ is dense. Since $\mathfrak{D}$ is invariant under $\exp(itH)$, $\mathfrak{D}$ is a core for H. Hence Corollary 1.3 is a consequence of Corollary 1.2. We infer Corollary 1.2 from Corollary 1.1 and a general result:

Theorem 1.2: Let $0 \leq H = H^*$ and $R = (H + I)^{-1}$. Let A be a symmetric operator with domain $\mathfrak{D}(H)$. Suppose that

$$\|AR\| + \|\dot{A}R\| < \infty, \tag{1.26}$$

where $\dot{A} = [iH, A]$ is defined in $\mathfrak{D}(H^2) \times \mathfrak{D}(H^2)$. Then A is essentially self-adjoint on any core for H.

Proof: Let $H = \int \lambda dE_\lambda$ be the spectral resolution of H. Let $C_n = E_n A E_n$. Then C_n is a bounded, self-adjoint operator with resolvent $K_n = (C_n - iy)^{-1}$. We first prove that K_n converges to the resolvent of a self-adjoint operator C. We then show that A^- extends C, so A is essentially self-adjoint on $\mathfrak{D}(H)$. The estimate $\|AR\| \leq O(1)$ then yields essential self-adjointness on any core for H.

(a) Graph convergence: We note that given (1.26), it follows by (1.17) that $R^{1/2} A R^{1/2}$ is bounded. Since $\|(E_n - I) R^{1/2}\| \leq O(n^{-1/2})$ as $n \to \infty$,

$$\|R(C_n - A)R\| \leq O(n^{-1/2}). \tag{1.27}$$

Furthermore, using (1.27) and the identity $C_n R^2 = RC_n R - iRC_n R$, we have

$$\|(C_n - A)R^2\| \leq O(n^{-1/2}), \quad n \to \infty. \tag{1.28}$$

Hence the operators C_n have a strong graph limit, a closed operator C. (See Ref. 1, Sec. 6.2). By (1.28) C agrees with A on $\mathfrak{D}(H^2)$. Since AR is bounded, and C is closed, C agrees with A on $\mathfrak{D}(H)$.

(b) *Smoothness preserving:* For real y, $|y|$ sufficiently large, we prove

$$\|(H + I)K_n R\| \leq 1. \tag{1.29}$$

In fact, for $\theta \in \mathfrak{D}(H)$ and $|y| - \|\dot{A}R\| \geq 1$,

$$\begin{aligned}\|(H + I)(C_n - iy)\theta\| &\geq \|(C_n - iy)(H + I)\theta\| - \|\dot{C}_n\theta\| \\ &\geq (|y| - \|\dot{C}_n R\|)\|(H + I)\theta\| \\ &\geq \|(H + I)\theta\|,\end{aligned}$$

proving (1.29).

(c) *Resolvent convergence:* For $\theta \in \mathfrak{D}(H)$, $\|(K_n - K_m)\theta\|^2 = \langle K_n\theta, (K_n - K_m)\theta\rangle - \langle K_m\theta, (K_n - K_m)\theta\rangle$ By (1.29), for $|y|$ sufficiently large,

$$\begin{aligned}\|(K_n - K_m)\theta\|^2 &\leq 2\|R(C_n - C_m)R\|\,\|(H + I)\theta\|^2 \\ &\leq o(1)\|(H + I)\theta\|^2\end{aligned}$$

as $n, m \to \infty$ by (1.27). Since $C_n = C_n^*$, we have $\|K_n\| \leq |y|^{-1}$, for y real. Hence for $|y|$ sufficiently large,

$$\text{s-lim}\, K_n = K. \tag{1.30}$$

By (a) and (c), $C_n = C_n^*$ has a graph limit C and a strong resolvent limit. We infer from Ref. 1, Theorem 6.2.3 that $C = C^*$ and $\lim K_n = K$ is the resolvent of C. Furthermore, since H is self-adjoint, the uniform bound (1.29) ensures $K: \mathfrak{D}(H) \to \mathfrak{D}(H)$ and

$$\|(H + I)KR\| \leq 1.$$

(d) *Essential self-adjointness:* We prove that $C \subset A^-$. Hence $C = A^-$. Let $\theta \in \mathfrak{D}(C)$ or $\theta = K\chi$. Choose $\chi_n \in \mathfrak{D}(H)$, $\chi_n \to \chi$. Since K is bounded, $\theta_n = K\chi_n \to \theta$. By (c), $\theta_n \in K\mathfrak{D}(H) \subset \mathfrak{D}(H) = \mathfrak{D}(A)$. By (a), C and A agree on $\mathfrak{D}(H)$. Thus

$$(A - iy)\theta_n = (C - iy)\theta_n = \chi_n \to \chi.$$

Hence $\theta \in \mathfrak{D}(A^-)$ and $A^-\theta = C\theta$.

Proof of Theorem B: In the case of the space–time averaged field operator $\varphi(f)$, we have identified $\varphi(f)$, acting on $\mathfrak{D} \subset \mathcal{K}_{GNS} = \mathcal{K}_W$, with a restriction of the self-adjoint field operator $\varphi(f)$ introduced in Ref. 3. In the case of the bilinear form $\varphi(x, t)$, we do not attempt such an identification. Bilinear forms, in contrast with self-adjoint operators, do not have natural maximal domains of definition. The domains for $\varphi(x, t)$ introduced in Ref. 3 could very possibly meet $\mathfrak{D}$ in the zero vector alone. Thus we proceed as follows.

By Theorem A and the space–time translation invariance of the vacuum, there exists a bilinear form $\varphi(x, t)$ on $\mathfrak{D} \times \mathfrak{D}$ such that

$$\langle\theta, \varphi(x, t)\chi\rangle \in \mathcal{C}^\infty(R^2), \tag{1.31}$$

and

$$\int \langle\theta, \varphi(x, t)\chi\rangle f(x, t)\,dx\,dt = \langle\theta, \varphi(f)\chi\rangle, \tag{1.32}$$

for all $\theta, \chi \in \mathfrak{D}$ and all $f \in \mathcal{S}(R^2)$. It is with this definition of $\varphi(x, t)$ that we prove Theorem B.

Let χ be a real, positive $\mathcal{C}_0^\infty$ function of total integral one and $\chi_\kappa(x) = \kappa\chi(\kappa x)$. In one space dimension, $\|\mu^{-1/2-\epsilon}\delta_x\|_2 \leq$ const, where δ_x is the Dirac measure concentrated at x. Then

$$\|\mu^{-1/2-\epsilon}(\chi_\kappa * \delta_x)\|_2 \leq \|\mu^{-(1/2)-\epsilon}\delta_x\|_2 \leq M, \quad \|\chi_\kappa\|_1 \leq 1, \tag{1.33}$$

$$\|\mu^{-(1/2)-\epsilon}(\chi_\kappa * \delta_x - \delta_x)\| \leq O(\kappa^{-\epsilon/2}), \tag{1.34}$$

$$\|\mu^{-(1/2)-\epsilon}(\delta_x - \delta_{x+a})\| \leq O(|a|^{\epsilon/2}). \tag{1.35}$$

We let $f_\kappa(x, t) = h_\kappa(x)\alpha_\kappa(t)$. By the remark below Theorem 1.1, we have for vectors θ_g as in (1.25)

$$\pm\langle\theta_g, \varphi(f_\kappa)\theta_g\rangle \leq (2M)^{-1}\|\mu^{-(1/2)-\epsilon}h_\kappa\|_2\|\alpha_\kappa\|_1 \times \langle\theta_g, [H(g) + \text{const}]\theta_g\rangle. \tag{1.36}$$

As in the proof of Corollary 1.1, we space average and pass to a limit $g \to 1$ to obtain for $\theta \in \mathfrak{D}$

$$\pm\langle\theta, \varphi(f_\kappa)\theta\rangle \leq (2M)^{-1}\|\mu^{-(1/2)-\epsilon}h_\kappa\|_2\|\alpha_\kappa\|_1 \times \langle\theta, (H + \text{const})\theta\rangle. \tag{1.37}$$

We now choose $h_\kappa = \chi_\kappa * \delta_x$ and $\alpha_\kappa = \chi_\kappa * \delta_t$. Passing to the limit $\kappa \to \infty$, we have (1.31)–(1.37)

$$\pm\langle\theta, \varphi(f_\kappa)\theta\rangle \to \pm\langle\theta, \varphi(x, t)\theta\rangle \leq \tfrac{1}{2}\langle\theta, (H + \text{const})\theta\rangle. \tag{1.38}$$

By taking limits of θ, (1.38) extends to $\mathfrak{D}(H^{1/2}) \times \mathfrak{D}(H^{1/2})$, and defines $\varphi(x, t)$ on that domain. We note by (1.35) that $\varphi(x, t)$ so defined is continuous in (x, t), and hence satisfies (1.32). This completes the proof of Theorem B.

The remainder of this paper is devoted to proving Theorem 1.1. For convenience, we suppress the κ cutoff or g cutoff and write

$$H(h) = H(g) + \delta H(h), \quad \delta E = \text{inf spectrum } H(h).$$

Both H_0 and $H(h)$ are self-adjoint operators on Fock space, and each has a unique ground state vector:

$$H_0\Omega_0 = 0, \quad H(h)\Omega(h) = \delta E\Omega(h).$$

The vectors Ω_0 and $\Omega(h)$ are not orthogonal. See Refs. 1 and 4 for proofs. It follows that $\delta E(h)$ is given by the exact formula

$$\delta E(h) = -\lim_{T\to\infty} (1/T)\log\langle\Omega_0, e^{-TH(h)}\Omega_0\rangle.$$

Our proof of Theorem 1.1. is based on estimating

$$\langle\Omega_0, e^{-TH(h)}\Omega_0\rangle \tag{1.39}$$

and bounding it by exp (MDT). The best bound previously established was $|\delta E| \leq O(1 + \text{diam supp } g)$, which diverges as $g \to 1$. The stronger bound proved here estimates the finite correction δE to the divergent vacuum energy $E(g)$ of (1.2).

We base our estimate of (1.39) on function space integration and the Feynman–Kac formula. Our new estimates can also be proved on Fock space (Q space) without appealing to probability theory methods. With some simple modifications to the present proof, we can eliminate the path space methods and substitute methods of Refs. 6 and 2.

In our proof of a uniform bound on δE, a major idea is to take advantage of localization in configuration space, as well as in momentum space. We isolate the contributions to $\delta E(h)$ from parts of the interaction localized near the support of h. We bound these contributions by the inductive construction of Sec. 2. We bound (1.39) by a sum of terms; each term represented by a graph. Each step in the induction replaces one graph by a sum of graphs with new vertices and new lines. At the end of each inductive step we decide for each graph whether to terminate the induction or whether to continue the expansion for that graph. If the inductive construction continues, the new vertices and lines introduce convergence factors into the expansion. One possibility is that the new vertices have a large lower momentum cutoff κ in some new vertex and thus contribute the small tail $O(\kappa^{-\delta})$ of a convergent integral to our estimates. Such vertices are said to be localized in high momentum regions. The remaining new vertices are localized in space–time regions far from the vertices of the original graph, so that at least one line connects vertices with space–time localizations separated by a large Euclidean distance d. By construction, d grows rapidly for successive inductive steps. In Sec. 3, we show that each line so localized in space–time contributes a convergence factor $\exp(-md/2)$ to our estimates of the corresponding graph. Here m is the mass of the particles in H_0 (bare mass). The small factors $\kappa^{-\delta}$ or $\exp(-md/2)$ lead to the proof of convergence of our inductive estimate. In particular they dominate the large number of graphs that occur and yield a convergent sum of small terms to bound δE.

2. THE INDUCTIVE CONSTRUCTION

A. Introduction

The proof of Theorem 1.1 is based upon an inductively constructed upper bound to (1.39). The inductive construction consists of repeating three basic substeps:

(a) the graph expansion,
(b) the squaring operation,
(c) the path space construction.

The estimates which prove convergence of the inductive construction are established in Secs. 3 and 4.

It is convenient to define

$$H_{II} = H_I + \delta H. \tag{2.1}$$

We also introduce a momentum cutoff indexed by κ into the interaction Hamiltonian H_{II}. Let

$$H_{I,\kappa} = \int : \mathcal{P}(\varphi_\kappa(x)) : g(x)dx,$$

$$H_{II,\kappa} = H_{I,\kappa} + \int \mathcal{P}_1(\varphi_\kappa(x)) : h(x)dx.$$

We now describe the localization in space and time of functions on path space. We index the space–time localization by $i = \{i_0, i_1\} \in \mathcal{S}^\tau \times \mathcal{S}$. We define the space–time square $\Delta_i = [i_0, i_0 + 1) \times [i_1, i_1 + 1)$. Here the first factor $[i_0, i_0 + 1)$ is a time interval, and the second factor is a space interval. We define the space–time localized interaction Hamiltonian $H_{I,i}$ as a time-dependent operator on Fock space

$$H_{I,i}(t) = H_I[\xi_i(\cdot, t)g(\cdot)]$$

or as a function on path space,

$$H_{I,i}(t, q(t)) = H_I[\xi_i(\cdot, t)g(\cdot)](q(t)).$$

Here ξ_i is the characteristic function of Δ_i. Thus

$$H_I = \sum_i H_{I,i}. \tag{2.2}$$

Similarly we define $\delta H_i, H_{II,i}$. Also we may localize the interaction both in space–time and in momentum space. We define

$$H_{I,\kappa,i} = \int : \mathcal{P}(\varphi_\kappa(x)) : \xi_i g(x)dx,$$

and similarly we define $H_{II,\kappa,i}$.

Let X be a union of space intervals $[j, j + 1)$, $j \in \mathcal{S}$. We define

$$H_I(X) = \sum_{i_1 \in X} H_I(\xi_i g),\ H(X) = H_0 + H_I(X),$$

and

$$E(X) = \inf \text{spectrum}\{H_0 + H_I(X)\}.$$

(Note that we now omit the vacuum energy from H.) Furthermore, we introduce sets $Y \subset R^2$ which are unions of squares $\Delta_i \subset R^2$. Let $Y(t)$ be the time slice of Y at time t. Let $|X|$ be the length of X and let $|Y|$ be the area of Y.

With the hypothesis of Theorem 1.1, we establish the inequality

$$-MD \leq \delta E. \tag{2.3}$$

The constant M is independent of g and h. For Theorem 1.1, Case 2, the bound (2.3) ensures

$$0 \leq H - E(g) \pm \varphi(h) + MD.$$

Thus if Ω is the vacuum for H,

$$\delta E \leq \langle \Omega, \{H - E(g) + \varphi(h)\}\Omega \rangle$$
$$\leq \langle \Omega, \varphi(h)\Omega \rangle \leq \langle \Omega, \{H - E(g) + MD\}\Omega \rangle \leq MD,$$

proving $|\delta E| \leq MD$. For Case 1 of the theorem, the desired inequality $|\delta E| \leq MD$ follows by an interchange in the roles of H and $H + \delta H$ in the proof of (2.3). The bound (2.3) is a consequence of two inequalities.

Proposition 2.1: There is a constant a_0 such that for all intervals X with $|X| \geq a_0$,

$$E(g) \leq E(\sim X) + |X|. \tag{2.4}$$

There is a constant M such that for all intervals X_0 containing supp h,

$$\inf\{E(\sim X) + |X| : X_0 \subset X\} \leq E(g) + \delta E(g,h) + M|X_0|. \quad (2.5)$$

The constants M and a_0 are independent of g and h, subject to the conditions of Theorem 1.1.

We choose $\max\{D, a_0\} \leq |X_0| \leq \max\{D, a_0\} + 1$. By Proposition 2.1,

$$E(g) \leq E(g) + \delta E(g,h) + M|X_0|. \quad (2.6)$$

Thus with a new constant M, we have proved (2.3) and hence Theorem 1.1.

In Secs. 2 and 3 we prove (2.5). In Sec. 4 we modify this proof to establish (2.4). We now describe the inductive construction used to bound (1.39). At each step in the inductive construction, we dominate (1.39) by a sum of terms of the form

$$\int ds \int_Q R(s) \exp[-\int H_{II}(\sim Y(\sigma), \sigma)(q(\sigma))d\sigma]\, dq(\cdot), \quad (2.7)$$

where R is a polynomial on path space and where $Y(\sigma)$ defines a localization of the exponent, as above. The connection between the Fock space inner product (1.39) and the path space integral (2.7) is given by the Feynman–Kac formula. Each inductive step replaces (2.7) by a sum of similar terms, and our final bound dominates (1.39) by the sum of a convergent series.

The first part of each inductive step, the graph expansion, is an expansion of (2.7) using the pull through formula. Each term in the expansion is represented by a graph. For some terms, the polynomial R is a constant. The associated graphs have no external legs and are called vacuum graphs. For these graphs, the inductive construction terminates. In terms for which the polynomial R is not a constant, new vertices have been introduced during the graph expansion, as will be explained below.

The second part of each inductive step is the squaring operation, in which we replace R by a positive polynomial R', $|R| \leq R'$. The fact that R' is positive makes the next path space construction possible, and the fact that R' is a polynomial allows us to perform the next graph expansion step.

The third portion of each inductive step is the path space construction, which we use to bound (and remove) those parts of the exponent that are localized near the localization of R. We use the near positivity of $\int H_{I,i}(q(s))ds$ to give an upper bound for localized portions of the exponent. This upper bound has the form of a convergent sum of polynomials on path space. Hence we remove part of the exponent and replace R by a sum of new polynomials R'.

In the first inductive step, we omit the graph expansion and the squaring operation and begin with the path space construction, see Sec. 2D. We then repeat the three substeps and arrive at a bound for (2.7) as a sum of vacuum graphs. Each vacuum graph is multiplied by a factor

$$\int_Q \exp(-\int_0^T H_I(\sim Y(\sigma))(q(\sigma))d\sigma)$$

$$= \langle \Omega_0, e^{-H(\sim Y(0))} e^{-H(\sim Y(1))} \cdots e^{-H(\sim Y(T))} \Omega_0 \rangle$$

$$\leq \exp(-\int_0^T E(\sim Y(\sigma))d\sigma).$$

The vacuum graphs are small, for reasons given in Sec. 1.

B. The Graph Expansion

The graph expansion is generated by the pull through formula

$$a(k)e^{-tH} = e^{-t(H+\mu)}a(k) - \int_0^t ds\, e^{-s(H+\mu)}[a(k), H_I]\, e^{-(t-s)H}. \quad (2.8)$$

Here $H = H(X) = H_0 + H_I(X)$ and $\mu = (k^2 + m^2)^{1/2}$. To derive (2.8) we set

$$F(s) = \int e^{-s(H+\mu)} a(k) e^{-(t-s)H} f(k)\, dk,$$

where $f \in L_2$, and we evaluate

$$F(0) = F(t) - \int_0^t F'(s)\, ds.$$

By Rosen's higher-order estimates, $F(s)$ is continuously norm differentiable on $[0, t]$. From the more elementary first order estimates, one can show that $F(s)$ is strongly differentiable and norm bounded on $[0, t]$ and continuously norm differentiable on the open interval $(0, t)$.

We call our expansion the graph expansion because each term in it can be labeled by a graph; the graphs provide a convenient way to visualize the expansion. In order to accommodate our time-dependent Hamiltonians, we rewrite (2.8) as an anti-time-ordered integral

$$\overline{T}a(k,0) \exp\left(-\int_0^t H(\sigma)\, d\sigma\right) = \overline{T}a(k,t)e^{-t\mu} \exp\left(-\int_0^t H(\sigma)\, d\sigma\right) - \overline{T}\int_0^t ds[a(k,s)e^{-s\mu}, H_I(X,s)] \exp\left(-\int_0^t H(\sigma)\, d\sigma\right). \quad (2.9)$$

Here $\overline{T}$ denotes the anti-time ordering. Later times, denoted by larger values of σ, are written to the right and precede in the operator product. The time parameter s in $a(k,s)$ and $H_I(X,s)$ indicate that the operators $a(k)$ or $H_I(X)$ occur at time s. We replace X by the space–time region Y, so that $H(\sigma) = H(Y(\sigma))$, and then (2.9) remains valid.

We say that (2.9) is the sum of a term with $a(k)$ "pulled through the exponent" and a term in which $a(k)$ "contracts to the exponent." When we use (2.9) to generate the graph expansion, (2.9) occurs multiplied by a polynomial in the fields R. New terms occur, arising from commutators between a and R, and we call these terms "contractions to existing vertices." For instance, if $R(s)$ equals a product of monomials $R_1(s_1)\cdots R_n(s_n)$, then for $t_1 \leq t, s_1, \ldots, s_n \leq t_2$,

$$\overline{T}a(k,t)R(s)\exp\left(-\int_{t_1}^{t_2} H(\sigma)\, d\sigma\right) = \overline{T}a(k,t_2)e^{-(t_2-t)\mu} R(s) \exp\left(-\int_{t_1}^{t_2} H(\sigma)\, d\sigma\right) - \overline{T}R(s)\int_{t_1}^{t_2} ds'[a(k,s')e^{-(s'-t)\mu}, H_I(x,s')] \times \exp\left(-\int_{t_1}^{t_2} H(\sigma)\, d\sigma\right)$$

$$+ \overline{T}\sum_{j=1}^{n}[a(k,s_j)e^{-(s_j-t)\mu}, R_j(s_j)]$$
$$\times \prod_{i\neq j} R_i(s_i)\exp\left(-\int_{t_1}^{t_2}H(\sigma)\,d\sigma\right). \quad (2.10)$$

We apply (2.10) to the basic term in the inductive construction, which we can write either as an integral on path space or as an anti-time ordered integral

$$\int ds\int_Q R(s)\exp\left[-\int H_I(Y(\sigma))d\sigma\right]dq(\cdot)$$
$$= \int ds\left\langle\Omega_0, \overline{T}\,R(s)\exp\left(-\int_0^T H(Y(\sigma))\,d\sigma\right)\Omega_0\right\rangle. \quad (2.11)$$

The integration over $s_1 \le s_2 \le \cdots \le s_n$ in (2.11) arises from previous use of (2.10). In this section we consider the integrands of the ds integration in (2.11). We choose a linear factor φ in R and write it as the sum of an annihilation operator a and a creation operator a^*. Then we use (2.10) to move a to the right and the adjoint of (2.10) to move a^* to the left. The terms in which a or a^* are pulled through equal zero, since $a\Omega_0 = 0$. The remaining terms from (2.10) are functions of φ, but not π, and hence they can be written as integrals on path space.

We now explain the graph terminology. Vertices in our graphs arise from monomials φ^n which occur in the interaction Hamiltonian H_{II}. These vertices may be produced in R during the graph expansion, the squaring operation, or the path space construction. In applying our expansions, we always use the representation (2.2) to express H_I as a sum of parts $H_{I,i}$ localized in space-time cubes Δ_i. Hence each vertex in our graphs has a space-time localization index i. Each vertex also occurs at a definite time, corresponding to the time s' in the integrand of (2.10). Later vertices are placed to the right, in accordance with the anti-time ordering. A vertex arising from φ^n has n legs connected to it. Of these legs, j are contracted (connected) to legs of other vertices and $n-j$ are not contracted. The uncontracted legs have no specified directions and label linear fields φ in φ^n. The contracted legs necessarily point to the left or to the right, according to whether the vertex to which they contract lies at an earlier or at a later time. The legs pointing to the left are creation legs, and those pointing to the right are annihilation legs. These legs label creation operators $a(k)^*$ and annihilation operators $a(-k)$, respectively. Each use of the commutation relations $[a(-k), a(l)^*] = \delta(k+l)$, replaces R by a new monomial. The graph of the new monomial R' is formed by contracting the legs of $a(-k)$ and $a(l)^*$ in the graph of R. Two such contracted *legs* are called a *line*. Each use of the pull through formula (2.10) also replaces R by a sum of new monomials. The second term on the right of (2.10) has a new graph formed by introducing a new vertex at time s', and one leg of this vertex is contracted to the leg of $a(k,s')$. We also say that this new vertex results from the contraction of the leg of $a(k,s')$ to the exponent.

Associated with each vertex is a kernel. If the vertex comes from φ^n and if it has j contracted legs, then the kernel is

$$v(k_1,\ldots,k_j) = [a^\#(k_1),[a^\#(k_2),\cdots,[a^\#(k_j),\varphi^n]\cdots]],$$

where $a^\#(k) = a(-k)$ or $a^*(k)$. If $j=n$, then v is fully contracted.

We label the vertices of the graph of R by a variable ν, $\nu = 1,2,\ldots,N$, and the momenta $k_{r,\nu}$ at vertex ν. For a fully contracted φ^n vertex we introduce the kernel

$$v_\nu(k_{1,\nu},\ldots,k_{n,\nu}) = (\zeta_i g)^{\sim}\left(\sum_{r=1}^{n} k_{r,\nu}\right)$$
$$\times \prod_{r=1}^{n}[\mu(k_{r,\nu})^{-1/2}\,\tilde{\chi}(k_{r,\nu}\kappa^{-1})], \quad (2.12)$$

where ζ_i gives the space-time localization of the vertex and $\tilde{\chi}(\cdot\kappa^{-1})$ is the upper momentum cutoff function for the $H_{I,\kappa,i}$ vertex. In the path space construction of Sec. 2D, we will encounter lower cutoff vertices of the form $H_{I,i} - H_{I,\kappa,i}$. For such a vertex, we replace

$$\prod_{r=1}^{n}\tilde{\chi}(k_{r,\nu}\kappa^{-1}) \quad \text{by} \quad 1 - \prod_{r=1}^{n}\tilde{\chi}(k_{r,\nu}\kappa^{-1}) \quad (2.13)$$

in (2.12). For such vertices we define the lower momentum cutoff $\kappa(v_\nu) = \max\{1,\kappa\}$.

A φ^n vertex which is not fully contracted has a kernel obtained as follows: Multiply (2.12) by a Wick ordered product of fields $\tilde{\varphi}(k)$, one field for each uncontracted momentum variable. Then integrate this product over the uncontracted momenta.

We also introduce energy factors $\mu(k_{r,\nu})^{\epsilon_{r,\nu}} = \mu_{r,\nu}{}^{\epsilon_{r,\nu}}$ and define the kernel

$$\bar{v}_\nu = v_\nu \prod_r \mu_{r,\nu}{}^{\epsilon_{r,\nu}}. \quad (2.14)$$

We use these factors μ^ϵ in order to transfer powers of the energy from one kernel to another. Hence they satisfy

$$\prod_{\nu,r}\mu_{r,\nu}{}^{\epsilon_{r,\nu}} = 1, \quad (2.15)$$

and then $(\prod_\nu v_\nu) = (\prod_\nu \bar{v}_\nu)$.

We choose $-\frac{1}{2} \le \epsilon_{r,\nu}$ and

$$\epsilon_\nu = \sum_r \epsilon^+_{r,\nu} < \tfrac{1}{2},$$

where $\epsilon^+_{r,\nu} = \max\{0, \epsilon_{r,\nu}\}$. We note that for $\delta > 0$,

$$\|\bar{v}_\nu\|_2 \le O[\kappa(v_\nu)^{-(1/2)+\epsilon_\nu+\delta}].$$

In addition to the vertices, each line (two contracted legs) has associated a function of the momenta and an integration over the momenta of the two contracted legs. Suppose the annihilation leg with momentum $k_{r,\nu}$ of vertex ν (localized at time $s_\nu = s_{r,\nu}$) contracts with the creation leg with momentum $k_{r',\nu'}$ of vertex ν' (localized at time $s_{r',\nu'}$). Associated with this line is the factor

$$\delta(k_{r,\nu}+k_{r',\nu'})\exp[-(s_{r',\nu'}-s_{r,\nu})\mu_{r,\nu}]$$
$$= \delta(k_{r,\nu}+k_{r',\nu'})\exp(-\delta s_{r,\nu}\mu_{r,\nu}), \quad (2.16)$$

where $0 \le \delta s_{r,\nu}$ is the time difference of the contraction. The factor $\exp(-\delta s\mu)$ introduces a time localization. We insert this factor in the kernel by defining

$$\hat{v}_\nu = \bar{v}_\nu \exp(-\tfrac{1}{2}\sum_r \delta s_{r,\nu}\mu_{r,\nu}). \quad (2.17)$$

Hence we assign half the time localization factor for a given line to each of the two vertices which the line connects.

A polynomial $\hat{R}(s)$ is then constructed from its graph by the formula

$$\hat{R}(s) = \overline{T}\int[\prod_\nu \hat{v}_\nu(k_{r,\nu}; q(\cdot))][\prod_{\text{lines}} \delta(k_{r,\nu} + k_{r',\nu'})]dk. \tag{2.18}$$

We define the polynomial

$$R(s) = BL\hat{R}(s), \tag{2.19}$$

where B and L are constants, depending on the inductive construction. B is a product of constants e^j, e^{j_0}, and η from (2.32) or (2.34), while L is a product of distance localization factors l_ν, coming from (2.22) or (2.24).

The polynomial R and the kernels $\hat{v}_\nu$ contributing to it are unbounded operators in the anti-time ordered formula (2.11). When this formula is interpreted in path space, R and the kernels $\hat{v}_\nu$ become multiplication operators on path space. If $\hat{v}_\nu$ is fully contracted, then it is a constant as a function on path space. Similarly, if the graph of R has no uncontracted legs, then R is a constant as a function on path space. Such a graph is called a vacuum graph.

To complete the graph expansion, we must specify which legs of the polynomial are to be pulled through and which are not, since the infinite series expansion obtained by pulling through all legs appears to diverge. We treat on a different footing the vertices which were present at the start of the graph expansion part of the given inductive step. We pull through all these legs until they contract to yield lines. We do not pull through legs of vertices added in the course of the graph expansion of the given inductive step. Thus the graph expansion in a given inductive step yields a finite number of terms, and at the end of a given graph expansion step all vertices in R with uncontracted legs are new (pull through) vertices introduced during that graph expansion step.

C. The Squaring Operation

At the conclusion of the graph expansion for (2.11), we have replaced (2.11) by a sum of similar terms. We now replace R in each term of the form (2.11) by a positive polynomial on path space. We make this replacement for two reasons: During the PS construction we take absolute values, which could replace the polynomial R on path space by the nonpolynomial function $|R|$. In that case we could not perform the next inductive step, since the graph expansion is only defined for polynomial R. The second reason to modify R is to improve our estimates in Sec. 3 for the pull through vertices in R (vertices introduced by contractions to the exponent). For this reason we sometimes choose to replace R by a polynomial of high degree.

A simple estimate for R is obtained by repeated application of the Schwarz inequality to (2.18). We obtain

$$|R| \le \prod_\nu \|\hat{v}_\nu\|_2 \le \prod_\nu \|\bar{v}_\nu\|_2, \tag{2.20}$$

where the $\mathcal{L}_2$ norms $\|\hat{v}\|_2$ or $\|\bar{v}\|_2$ are $\mathcal{L}_2$ norms in the contracted momentum variables. Unless v_ν is fully contracted, $\|\bar{v}_\nu\|_2$ depends on the uncontracted legs as a function on path space. In Theorem 3.1 we improve (2.20) to take into account the space–time localization of the vertices. We prove

$$|R| \le \prod_\nu (\gamma e^{-md_\nu/3}\|\bar{v}_\nu\|_2), \tag{2.21}$$

where d_ν is the contraction distance for vertex v_ν, as defined below in Sec. 3B.

We use the bound (2.21) as the first step in our squaring operation. We give the graphical interpretation of $\|\bar{v}_\nu\|_2$. Suppose the set $\mathcal{C}$ of contracted legs $\bar{v}_\nu$ has $\bar{n}$ elements. We start with two identical $\bar{v}_\nu$ vertices and contract the $\bar{n}$ legs of the first set $\mathcal{C}$ to the $\bar{n}$ legs of the second set $\mathcal{C}$. We then take the square root, yielding the graph of $\|\bar{v}_\nu\|_2$. In case that $\bar{v}_\nu$ is fully contracted, $\|\bar{v}_\nu\|_2$ is the square root of a vacuum diagram. It is a constant on path space, so we do not need to square: Otherwise, $\|\bar{v}_\nu\|_2$ is the square root of a polynomial function on path space.

In the latter case, we use the bound $ab \le a(1 + b^n)$ with $a = \exp(-md_\nu/6) \le 1$ and $b = \gamma \exp(-md_\nu/6)\|\bar{v}\|_2$ to obtain

$$\gamma e^{-md_\nu/3}\|\bar{v}_\nu\|_2 \le l_\nu + (l_\nu\|\bar{v}_\nu\|_2)^{n_\nu}. \tag{2.22}$$

Here $l_\nu = \gamma e^{-md_\nu/6}$, and the squaring exponent n_ν is a positive even integer, chosen below. Since n_ν is even, $\|\bar{v}_\nu\|_2^{n_\nu}$ is a polynomial on path space.

We define the graph of the constant term l_ν in (2.22) to be a single vertex with no legs. We define the kernel corresponding to this vertex to equal one. The constant l_ν becomes a factor in L of (2.19), reflecting the history of previous contractions to $\bar{v}_\nu$. It assigns a space–time localization factor for each such contraction.

The second term in (2.22) yields a graph in which n_ν pull through vertices have uncontracted legs. The contracted legs of this graph are all localized in a single square Δ, the localization of the original vertex $\hat{v}_\nu$. Therefore these lines contribute no small localization factors. Each vertex, however, is accompanied by l_ν, which becomes a factor in L for the polynomial R.

Let $\hat{v}_\nu$ be a pull through vertex with uncontracted legs. We define $d_c(v_\nu)$ as the (integer part of the) Euclidean distance of the contraction giving rise to the vertex $\hat{v}_\nu$. We choose n_ν as follows: Either

$$n_\nu = 2 \quad \text{or} \quad n_\nu = [4d_c(v_\nu)]^6. \tag{2.23}$$

To choose between the two possibilities, consider a fixed cube Δ and apply (2.22) successively, in order of decreasing value of $d_c(v_\nu)$, to the kernels $\hat{v}_\nu$ localized in Δ. Each application of (2.22) yields two terms. Thus the j-fold application to j vertices in Δ yields 2^j terms. In these terms n_ν enters only when the second term in (2.22) is selected at least once. For the $(j+1)$th application of (2.22) we choose $n_\nu = 2$ in the above case, that is when the second term in (2.22) has been selected at least once in the first j applications of (2.22). Otherwise, we choose $n_\nu = [4d_c(v_\nu)]^6$. With N vertices in Δ we obtain 2^N terms. In each of these terms we define $d_c(\Delta) = d_c(v_\nu)$, where $\hat{v}_\nu$ is the vertex for which $n_\nu = [4d_c(v_\nu)]^6$.

For uniformity of notation, we use the bound

$$\gamma e^{-md_\nu/3}\|\bar{v}_\nu\|_2 \le l_\nu\|\bar{v}_\nu\|_2, \tag{2.24}$$

for fully contracted vertices in (2.21).

D. The Path Space Construction

The path space integrals

$$\int R(s)\,\exp\Big[-\int H_{\mathrm{II}}(q(\sigma))\,d\sigma\Big]dq(\cdot) \tag{2.25}$$

cannot be evaluated or estimated directly. On the other hand, integrals of polynomials on path space can be estimated easily. The purpose of the path space construction is to bound (2.25) by a sum of integrals of polynomials on path space. The term by term integration of the exponential series in (2.25) converges only if $p \le 2$. In the case of linear perturbations we make a power series expansion of $\exp[-\int \delta H(q(\sigma))\,d\sigma]$. Otherwise, we use an asymptotic expansion of $\exp[-\int H_{\mathrm{II}}(q(\sigma))d\sigma]$ to obtain our bound. We use the formal positivity of H_{II} and undo the Wick ordering in low momentum regions. This bounds the low momentum part of the exponent from above by a constant. The high momentum tail is bounded by a convergent sum of polynomials on path space.

More precisely, for Theorem 1.1, Case 1, we write

$$H_{\mathrm{II},i} = H_{\mathrm{II},i,\kappa} + \delta H_{\mathrm{II},i,\kappa}. \tag{2.26}$$

By undoing the Wick ordering in the low momentum part of $H_{\mathrm{II},i,\kappa}$, we obtain

$$-M_1(\log\kappa)^{p/2} + 2 \le H_{\mathrm{II},i,\kappa}. \tag{2.27}$$

The constant M_1 is independent of g, h, i and κ (see, e.g., Refs. 1 and 2.) It is convenient to choose a sequence of cutoffs $\kappa_j = \exp[(j/M_1)^{2/p}]$, so that by (2.27) and the localization of $H_{\mathrm{II},i}$ in a time interval of unit length,

$$-j + 2 \le \int H_{\mathrm{II},i,\kappa_j}(q(\sigma))\,d\sigma. \tag{2.28}$$

We bound $\exp(-\int H_{\mathrm{II},i}d\sigma)$ by writing path space Q as a countable disjoint union $Q = \cup_{j\ge j_0} Q_j$, where for $j > j_0$,

$$Q_j = \{Q : -j \le \int H_{\mathrm{II},i}d\sigma \le -j+1\}. \tag{2.29}$$

Then

$$Q_{j_0} = \{Q : -j_0 \le \int H_{\mathrm{II},i}d\sigma\}. \tag{2.30}$$

By (2.28), paths in Q_j, for $j > j_0$, yield the bound

$$1 \le \Big|\int \delta H_{\mathrm{II},i,\kappa_j}\,d\sigma\Big| = R_{i,j}. \tag{2.31}$$

Hence for $n \ge 1$, we have $1 \le (R_{i,j})^n$ on Q_j. For $j > j_0$, we choose $n = n(j)$ to be the largest even integer less than $\kappa_j^{1/8p}$. We define $n(j_0) = 0$. Then

$$\exp(-\int H_{\mathrm{II},i}d\sigma) \le \sum_{j\ge j_0} e^j R_{i,j}^{n(j)}. \tag{2.32}$$

For Theorem 1.1, Case 2, we write

$$\int \delta H_i\,d\sigma = \int \varphi(h)_i\,d\sigma = \int \varphi(h)_i(q(\sigma))\,d\sigma$$

and expand

$$\exp(-\int \delta H_i\,d\sigma) \le e^{J_0} + \eta \sum_{k\ge J_0} \frac{1}{(2k)!}\Big[\int \varphi(h)_i\,d\sigma\Big]^{2k}. \tag{2.33}$$

Here η is a universal constant, independent of J_0. We apply (2.32) to $H_{\mathrm{I},i}$ rather than $H_{\mathrm{II},i}$. Since $H_{\mathrm{II},i} = H_{\mathrm{I},i} + \delta H_i$, we have

$$\exp\Big(-\int H_{\mathrm{II},i}\,d\sigma\Big) \le \Big(e^{J_0} + \eta \sum_{k\ge J_0} \frac{1}{(2k)!}\Big(\int \varphi(h)_i\,d\sigma\Big)^{2k}\Big) \times \sum_{j\ge J_0} e^j R_{i,j}^{n(j)}. \tag{2.34}$$

The "path space construction" is the use of (2.32) or (2.34) in squares Δ_i. We also say that we "remove portions of the exponent" from (2.25). Our choice of $j_0 = j_0(\Delta_i)$ will be made in this and later sections.

We begin the inductive construction (the first inductive step) by applying (2.32) or (2.34) in each square $\Delta_i \in \mathfrak{D}_0$, where $\mathfrak{D}_0$ is a given union of squares Δ. For our proof of (2.5), we choose

$$\mathfrak{D}_0 = X_0 \times [0, T],$$

so the perturbation δH is localized in $\mathfrak{D}_0$. In these squares we choose $j_0(\Delta_i) = J_0$. After this initial use of the PS construction, only H_{I} remains in the exponent, δH has been completely removed. Further application of the PS construction introduces new H_{I} vertices into graphs, but no new δH vertices.

For each term in the PS construction we decide when to terminate the PS portion of an inductive step. An individual term has the form (2.25) and is labeled by a single graph. For each term we make some definitions: We let $\mathfrak{D}_e$ be the set of squares Δ_i with a nonzero exponent. For a square Δ, we let

$$d_0(\Delta) = \mathrm{dist}(\Delta, \mathfrak{D}_0), \qquad d_e(\Delta) = \mathrm{dist}(\Delta, \mathfrak{D}_e),$$

where $\mathrm{dist}(X, Y)$ is the Euclidean distance between X and Y. We let $n(\Delta)$ equal the number of vertices with uncontracted legs that are localized in Δ. If $n(\Delta) \ne 0$, let $\Gamma(\Delta)$ be the square of side length $8n(\Delta)^{1/3}$, centered at Δ, and with sides parallel to the space and time axes. Note that $n(\Delta)$, $\mathfrak{D}_e$, and $d_e(\Delta)$ change during the inductive construction. We let $n(\Delta)_r$ denote the value of $n(\Delta)$ at the end of the rth inductive step, etc.

Our condition for terminating the PS construction of a given inductive step (and hence for terminating the inductive step) is the following:

$$\text{If } n(\Delta) \ne 0, \quad \text{then } \Gamma(\Delta)\cap\mathfrak{D}_e = \emptyset, \tag{2.35}$$

for every square Δ in our space-time cover. If (2.35) is not met in some square Δ', we apply (2.32) to remove the exponent in each $\Delta \subset \Gamma(\Delta')\cap\mathfrak{D}_e$. We say that Δ' has forced the PS construction in Δ, and we choose $j_0(\Delta)$ in this case as

$$j_0(\Delta) = [J_1 n(\Delta')^{1/6}] + 1. \tag{2.36}$$

Here $[\cdot]$ denotes "integral part." Also we choose J_1 sufficiently large so that for $x \ge 1$ we have

$$(15)^3 x^2 \le \tfrac{1}{2}\exp(x^{1/3p}M_2),$$

where $M_2 = (8p)^{-1}(J_1/M_1)^{2/p}$. By (2.36), if $n(\Delta) \ne 0$,

$$(15)^3 n(\Delta')^2 \le \tfrac{1}{2}\exp[n(\Delta')^{1/3p}M_2] \le \kappa_{j_0(\Delta)}^{1/8p} \le n_{ps}(\Delta) \le n(\Delta). \tag{2.37}$$

Remark: Note $n(\Delta) = n_{pt}(\Delta) + n_{ps}(\Delta) + n_h(\Delta)$, where $n_{pt}(\Delta)$ is the number of PT vertices in Δ with uncontracted legs, $n_{ps}(\Delta)$ is the number of PS vertices in Δ

with uncontracted legs, and $n_h(\Delta)$ is the number of uncontracted $\varphi(h)$ vertices in Δ. Here we call the vertices from $R^n_{i,j}$ in (2.32) PS vertices in Δ_i, and we call the vertices from $\int \varphi(h)_i\, d\sigma$ in (2.34), the $\varphi(h)$ vertices in Δ_i.

Proposition 2.2: If $\mathfrak{D}_0 \cap \mathfrak{D}_e = \emptyset$ and $n(\Delta) \neq 0$ during the PS contruction, then

$$d_0(\Delta) \leq n(\Delta)^{1/3}. \tag{2.38}$$

Proof: Step 1: Assume that the proposition holds at some point during the PS construction of the rth inductive step. We show that the proposition remains valid after the application of PS construction in any square Δ and, hence, at the end of the inductive step. If $n_{pt}(\Delta) \neq 0$, the PT vertices in Δ were present throughout the rth PS construction. Hence, by assumption, $d_0(\Delta) \leq n_{pt}(\Delta)^{1/3} \leq n(\Delta)^{1/3}$ and (2.38) is valid. On the other hand, suppose $n_{ps}(\Delta) \neq 0$ and Δ' forced the PS construction in Δ. Since $\Gamma(\Delta')$ has diagonal length $\sqrt{2}\,8n(\Delta')^{1/3} \geq \mathrm{dist}(\Delta', \Delta)$ and since $d_0(\Delta') \leq n(\Delta')^{1/3}$, the triangle inequality yields

$$\begin{aligned} d_0(\Delta) &\leq n(\Delta') + \mathrm{dist}(\Delta, \Delta') \\ &\leq (1 + \sqrt{2}\,8)n(\Delta')^{1/3} \leq 15n(\Delta')^{1/3}. \end{aligned} \tag{2.39}$$

In this case (2.38) follows by (2.37). Since $\Delta \notin \mathfrak{D}_0$, $n_h(\Delta) = 0$ and step 1 is proved.

Step 2: The proposition holds for the first PS construction. In the configuration immediately after removing the exponent from $\mathfrak{D}_0$, if $n(\Delta) \neq 0$, then $\Delta \in \mathfrak{D}_0$ and $d_0(\Delta) = 0$. Hence the assertion follows by step 1.

Step 3: Assuming the proposition holds for the rth PS construction, we establish the proposition for the $(r+1)$th PS construction. By the definition of $\Gamma(\Delta)$, whenever $n(\Delta)_r \neq 0$,

$$d_0(\Delta) \leq n(\Delta)_r^{1/3} \leq \tfrac{1}{2} d_e(\Delta)_r. \tag{2.40}$$

We consider the start of the $(r+1)$th PS construction. Only PT vertices v formed in the graph expansion of the $(r+1)$th inductive step can have uncontracted legs. These vertices have been squared during the squaring operation of the $(r+1)$th step. Hence if such a vertex v is localized in Δ, we have $n(\Delta) \geq [4d_c(\Delta)]^6$ by (2.23). Suppose v was formed by the contraction to the exponent of v' localized in Δ'. By the definition of the graph expansion, the vertex v' must be present at the end of the rth PS construction. Thus $n(\Delta')_r \neq 0$, and by (2.40) we have $d_0(\Delta') \leq d_e(\Delta')_r$. Note also $d_c(v) = \mathrm{dist}(\Delta, \Delta') \leq d_c(\Delta)$ and $d_e(\Delta')_r \leq d_c(\Delta)$. Hence by the triangle inequality,

$$\begin{aligned} d_0(\Delta) &\leq d_0(\Delta') + \mathrm{dist}(\Delta, \Delta') \leq d_e(\Delta')_r + d_c(\Delta) \\ &\leq 2d_c(\Delta) \leq n(\Delta)^{1/6} \leq n(\Delta)^{1/3}, \end{aligned} \tag{2.41}$$

establishing (2.38) for Δ. By step 1 the proof of the proposition is complete. We have also proved

Corollary 2.1: At the end of the PS construction of each inductive step, (2.40) holds in every Δ with $n(\Delta) \neq 0$.

Let $n(\Delta) \neq 0$ during the PS portion of some inductive step. We assign a square

$$\Delta\# \in \mathfrak{D}_0, \quad n(\Delta\#)_1 \neq 0, \tag{2.42}$$

and an integer $k(\Delta) \in \mathfrak{Z}^+$ to Δ in the following manner: If $\Delta \in \mathfrak{D}_0$ let $\Delta\# = \Delta$, $k(\Delta) = 0$. If $\Delta \notin \mathfrak{D}_0$, suppose we are considering the rth PS construction. In case $n_{ps}(\Delta) \neq 0$, let $\Delta^{(1)}$ be the square that forced the PS construction in Δ, and we consider $\Delta^{(1)}$ at the time (during the rth PS construction) that the exponent is removed in Δ. In case $n_{ps}(\Delta) = 0$, then $n_{pt}(\Delta) \neq 0$, and we choose any PT vertex v in Δ with the maximum contraction distance, i.e., $d_c(v) = d_c(\Delta)$. We define $\Delta^{(1)}$ to be the square with the vertex $v^{(1)}$ that contracted to the exponent forming v, and we consider $\Delta^{(1)}$ at the end of the $(r-1)$th inductive step. Continuing in this fashion, we assign $\Delta^{(k)}$ to $\Delta^{(k-1)}$ until we arrive at a square $\Delta^{(k)} \in \mathfrak{D}_0$ (and the first inductive step). We let $\Delta\# = \Delta^{(k)}$, $k(\Delta) = k \geq r$.

Let

$$N_0 = \min\{n(\Delta)_1 : \Delta \in \mathfrak{D}_0, n(\Delta)_1 \neq 0\}. \tag{2.43}$$

We note that in the proof of (2.5), either $n_h(\Delta) \geq 2J_0$ or $n_{ps}(\Delta) \geq \frac{1}{2}\kappa_{J_0}^{1/8p}$ for $\Delta \in \mathfrak{D}_0$, $n(\Delta)_1 \neq 0$. Thus $N_0 \to \infty$ as $J_0 \to \infty$.

Corollary 2.2: If $n(\Delta) \neq 0$ during the PS construction of the rth inductive step,

$$n(\Delta) \geq N_0^{2^r}. \tag{2.44}$$

Proof: If $n_{ps}(\Delta) \neq 0$, we have by (2.37) that $0 \neq n(\Delta^{(1)})^2 \leq n(\Delta)$. If $n_{ps}(\Delta) = 0$, we have by Corollary 2.1 and Eq. (2.23),

$$0 \neq n(\Delta^{(1)})_{r-1}^2 \leq d_e(\Delta^{(1)})_{r-1}^6 \leq d_c(\Delta)^6 \leq n(\Delta).$$

By induction we obtain

$$n(\Delta) \geq n(\Delta\#)_1^{2^k} \geq N_0^{2^r}, \tag{2.45}$$

since $n(\Delta\#)_1 \neq 0$.

Corollary 2.3: For any PT vertex v,

$$d_c(v) \geq N_0. \tag{2.46}$$

Proof: Let v be a PT vertex formed by the contraction of a vertex v' in Δ' during the rth inductive step. By Corollary 2.1, $d_c(v) \geq d_e(\Delta')_{r-1} \geq n(\Delta')_{r-1}^{1/3}$ and, by Corollary 2.2, $d_c(v) \geq N_0$.

E. Special Procedures for $\varphi(h)$ Vertices

We modify the inductive construction slightly in order to improve our estimates on $\varphi(h)$ vertices. These vertices are introduced with the first path space construction by applying (2.34). They are contracted during the first graph expansion. All the $\varphi(h)$ vertices are localized in $\mathfrak{D}_0$; but they may contract to vertices localized in any square Δ.

The Schwarz inequality of Sec. 2C introduces into our estimates the factors $\|\mu^{-1/2}h\|_2$. We modify Sec. 2C to arrive at constants $\|\mu^{-1/2-\epsilon}h\|_2$ required for Theorem 1.1. The extra factor $\mu^{-\epsilon}$ is obtained from the convergence of the $\mathcal{L}_2$ norm of the kernels of H_I or from the integration over some times s. Formally the method yields $\epsilon = \frac{1}{2}$, while the bound should fail for $\epsilon > \frac{1}{2}$. We give the proof for $\epsilon = (8p)^{-1}$.

In the graph expansion, a $\varphi(h)$ vertex contracts either to an H_I vertex or to another $\varphi(h)$. If $\varphi(h)$ contracts to an H_I, we multiply the kernel $v_\nu(k)$ of the contracted $\varphi(h)$ vertex by $\mu(k)^{-\epsilon}$, and we multiply the kernel of H_I vertex by μ^ϵ. Hence our kernels, with the $\mu^{\pm\epsilon}$ factors, have the general form (2.17) that we estimate in Sec. 3. We choose $\epsilon = (8p)^{-1}$. [To deal with $\epsilon \geq (2p)^{-1}$, it would be necessary to integrate over times s in such graphs.]

The remaining $\varphi(h)$ vertices contract to each other. The polynomial of a corresponding connected component of the graph of $\hat{R}(s)$ is

$$r_{\nu\nu'} = \int v_\nu(k) v_{\nu'}(-k) e^{-\delta s\mu}\, dk ds_\nu ds_{\nu'}, \tag{2.47}$$

where the $s_\nu, s_{\nu'}$ integrals are restricted to the time localization of the vertices and also $0 \leqq \delta s$. We perform the s_ν integration. We obtain a denominator μ^{-1} and two terms coming from the two endpoints of integration. Thus we obtain two graphs, of the form (2.47), each with only one time integration and with the kernels $\mu^{-1/2} v_\nu$ replacing v_ν. The application of (2.21) to such graphs yields for the integrand of the $s_{\nu'}$ integration

$$|r_{\nu\nu'}(s_{\nu'})| \leqq 2 l_\nu^2 \|\mu^{-1/2} v_\nu\|_2 \, \|\mu^{-1/2} v_{\nu'}\|_2. \tag{2.48}$$

3. ESTIMATES

A. Estimates on Kernels

In this section we derive estimates on kernels in the inductive construction. These estimates deal quantitatively with localization in configuration and in momentum space.

Lemma 3.1: Let $0 \leq \alpha, \beta$, $\alpha^2 + \beta^2 < m^2$, $0 < t$, $p \in R^s$, and

$$f(x,t) = D_t^j D_x^k \int \mu^\nu e^{-\mu t + ipx}\, dp.$$

Then for $\lambda = [\nu + j + |k| + s]$,

$$\left| \left(e^{\beta|x|} - \sum_{n=0}^{\lambda} \frac{1}{n!} (\beta|x|)^n \right) f(x,t) \right| \leq O(e^{-\alpha t}). \tag{3.1}$$

Remarks: The function $f(x,t)$ is smooth except possibly at $x = 0, t = 0$. Our estimates are uniform for $t > 0$. For $|x| > \epsilon > 0$ we have

$$|f(x,t)| \leq O(e^{-\alpha t - \beta|x|}) \tag{3.2}$$

by the lemma. For $t > \epsilon > 0$, our proof of (3.1) proves (3.2). We take the limit $t \to 0$ and define $f(x,0)$ which satisfies (3.1) and (3.2) with $\alpha = 0$, $\beta < m$. Furthermore, if $\lambda < 0$, then $f(x,t)$ is continuous and (3.2) holds for all (x,t).

Proof: We establish the case $s = 1$, $\nu = j = k = 0$. Then $\lambda = 1$. The proof in general is similar. We use the Cauchy formula to estimate the derivatives of $\exp[-t\mu(p)]$

$$D_p^n e^{-t\mu(p)} = \frac{n!}{2\pi i} \oint \frac{e^{-t\mu(\zeta)}}{(\zeta - p)^{n+1}}\, d\zeta, \tag{3.3}$$

and we integrate over a circle centered at p. For $|p| \leq m$, we choose a circle of radius $\beta + \epsilon$, where $\alpha^2 + (\beta + \epsilon)^2 < m^2$. We note that if $\mathrm{Re}z \geq 0$, then

$$(\mathrm{Re}z)^{1/2} \leq \mathrm{Re}(z^{1/2}). \tag{3.4}$$

Thus on our circle of integration $(\mathrm{Re}\mu^2)^{1/2} \leq \mathrm{Re}\mu$ and

$$\alpha \leq [p^2 + m^2 - (\beta + \epsilon)^2]^{1/2} \leq \mathrm{Re}\mu. \tag{3.5}$$

Thus (3.3) yields for $|p| \leq m$,

$$|D_p^n e^{-t\mu(p)}| \leq n!(\beta + \epsilon)^{-n} e^{-\alpha t}.$$

For $|p| > m$, we use a circle of radius $|p|$. On this circle $m \leq (\mathrm{Re}\mu^2)^{1/2} \leq \mathrm{Re}\mu$, and

$$|D_p^n e^{-t\mu(p)}| \leq n! |p|^{-n} e^{-mt}.$$

Expanding $e^{\beta|x|}$ in (3.1) in a power series and using $\|x^r f\|_\infty \leq \|D^r \tilde{f}\|_1$,

$$\begin{aligned}
&|\{e^{\beta|x|} - (1 + \beta|x|)\} f(x,t)| \\
&\quad \leq \sum_{n=2}^{\infty} \frac{1}{n!} (\beta|x|)^n |f(x,t)| \\
&\quad \leq \sum_{n=2}^{\infty} \frac{\beta^n}{n!} \|D_p^n e^{-t\mu(p)}\|_1 \\
&\quad \leq \sum_{n=2}^{\infty} 2m \left(\frac{\beta}{\beta + \epsilon}\right)^n e^{-\alpha t} + 2\int_m^\infty dp \left(\frac{\beta}{p}\right)^2 \\
&\qquad \times e^{-mt} \sum_{n=0}^{\infty} \left(\frac{\beta}{|p|}\right)^n \\
&\quad \leq O(e^{-\alpha t}) + O(e^{-mt}) \int_m^\infty p^{-2} dp \leq O(e^{-\alpha t}).
\end{aligned}$$

Lemma 3.2: Let e_i be the characteristic function of $[i, i+1]$. Then for $0 \leq t$, $\alpha^2 + \beta^2 < m^2$, $\mu_x = \mathfrak{F}\mu(p)\mathfrak{F}^{-1}$, we have

$$\|e_i \exp(-t\mu_x) e_j\| \leq O(e^{-\alpha t - \beta|i-j|}).$$

Proof: If $|i - j| \leq 2$, the bound follows from $|e_i| \leq 1$, $\|e^{-t\mu_x}\| \leq e^{-tm}$. If $|i - j| > 2$, the kernel $k_t(x,y)$ of $e_i \exp(-t\mu_x) e_j$ is $k_t(x,y) = e_i(x) f_t(x - y) e_j(y)$ where f_t is the Fourier transform of $\exp[-t\mu(p)]$. Note that k_t has support in the region $|x - y| > 1$. We apply Lemma 3.1 with $\alpha^2 + (\beta + \epsilon)^2 < m^2$, $j = k = \nu = 0$,

$$\begin{aligned}
|k_t(x,y)| &\leq O(1) e^{-\alpha t - \beta|x-y| - \epsilon|x-y|} e_i(x) e_j(y) \\
&\leq O(e^{-\alpha t - \beta|i-j|}) e^{-\epsilon|x-y|}.
\end{aligned}$$

Since $\exp(-\epsilon|x - y|)$ is the kernel of a bounded operator, Lemma 3.2 follows.

Lemma 3.3: Let $\zeta \in \mathcal{C}_0^\infty$, $\beta < m$, $-1 \leq \tau \leq 1$. Then

$$\|e_i \mu_x^{-\tau} \zeta \mu^\tau\| \leq O(e^{-\beta|i|}).$$

Proof: We let $e = \sum_{|j| \leq N} e_j$, where $[-N, N]$ is sufficiently large to include the support of ζ. We write $w_i = e_i \mu^{-\tau} \zeta \mu^\tau$ as

$$w_i = w_i e + w_i(I - e)$$

and bound each term. Let $r(x,y)$ be the kernel of $w_i(I - e)$. By Lemma 3.1, for $|i|$ sufficiently large,

$$|r(x,y)| \leq O(e^{-\beta|i| - \epsilon|x-y|}).$$

Thus for $|i|$ large,

$$\|w_i(I-e)\| \le O(e^{-\beta|i|}).$$

We now assume $\tau \ge 0$, and note that $\|w_i\| \le$ const, since

$$w_i = e_i\zeta + e_i\mu^{-\tau}[\zeta, \mu^\tau],$$

and $[\zeta, \mu^\tau]$ is a bounded operator. In the momentum representation, the kernel $h(p,p')$ of $[\zeta, \mu^\tau]$ satisfies

$$|h(p,p')| = |\tilde{\zeta}(p-p')\{\mu(p)^\tau - \mu(p')^\tau\}| \le O(1)\mu(p-p')^\tau|\tilde{\zeta}(p-p')|,$$

which is a rapidly decreasing function of $p-p'$ and, hence, the kernel of a bounded operator. We use this to bound $\|w_i\|$ and $\|w_i e\|$ for $|i| \le O(1)$. If $\tau < 0$, we use the representation

$$w_i = e_i\zeta + [\mu^{-\tau}, \zeta]\mu^\tau.$$

For $|i|$ large, we bound the kernel $r_i(x,y)$ of $w_i e$ as follows:

$$r_i(x,y) = e_i(x)\int f_-(x-z)\zeta(z)f_+(z-y)dz e(y),$$

where $f_\pm$ is the Fourier transform of $\mu(p)^{\pm\tau}$. For $|x|$ sufficiently large, $f_-(x-z)\zeta(z) = h_x(z)$ is a $\mathcal{C}_0^\infty$ function of z, and, by Lemma 3.1, $D_z^n h_x(z)$ is bounded by $O(e^{-\beta|x|})$, $|x| > R$. Furthermore,

$$\int f_-(x-z)\zeta(z)f_+(z-y)dz = \langle f_{+,y}, h_x\rangle$$

is the evaluation of a tempered distribution f_+, translated by y, with the $\mathcal{C}_0^\infty$ function $h_x(\cdot)$. Since translation is continuous in $\mathcal{S}'$, $f_{+,y}e(y)$ varies over a bounded set in $\mathcal{S}'$ as y varies over the support of e. Hence for some Schwartz space norm $\|\cdot\|_s$ independent of x, y,

$$|e(y)\langle f_{+,y}, h_x\rangle| \le \|h_x\|_s \le O(e^{-\beta|x|}),$$

for $|x|$ sufficiently large. Thus for $|i|$ sufficiently large

$$|r_i(x,y)| \le O(e^{-\beta|i|})e_i(x)e(y),$$

and, since $e_i(x)e(y)$ is the kernel of a bounded operator,

$$\|w_i e\| \le O(e^{-\beta|i|}),$$

to complete the proof.

We now apply these bounds to a φ^n vertex $\hat{v}_\nu$ of the form (2.17). Suppose $\hat{v}_\nu$ has $\bar{n}$ contracted legs and is localized in Δ_i. (Thus the time $s_{r,\nu}$ of the vertex ν lies in $[i_0, i_0+1)$.) Let E_j be the localization projection

$$E_j = \prod_{r=1}^{\bar{n}} e_{j_r}(x_r),$$

and $\sum E_j = I$.

Proposition 3.1: Let $\alpha_r^2 + \beta_r^2 < m^2$. Then

$$\|E_j\hat{v}_\nu\|_2 \le \|\bar{v}_\nu\|_2\gamma \prod_{r=1}^{\bar{n}} (e^{-(1/2)\alpha_r\delta s_{r,\nu} - (1/2)\beta_r|j_r - i_1|}). \quad (3.6)$$

Proof: Let $\delta S = \frac{1}{2}\sum_r \mu_{r,\nu}\delta s_{r,\nu}$. For $0 < \delta$, δ sufficiently small, we have by Lemma 3.2,

$$\|E_j\hat{v}_\nu\|_2 \le \sum_{j'} \|E_j e^{-\delta S}E_{j'}\bar{v}_\nu\|_2 \le O(1)\sum_{j'} \prod_{r=1}^{\bar{n}} (e^{-(1/2)\alpha_r\delta s_{r,\nu} - (1/2)(\beta_r+\delta)|j_r - j_r'|})\|E_{j'}\bar{v}_\nu\|_2. \quad (3.7)$$

In order to display the localization of $\bar{v}_\nu$, we write

$$\bar{v}_\nu = \left(\prod_{r=1}^{\bar{n}} K_{r,i_1}\right)\bar{v}_\nu = \bar{K}\bar{v}_\nu. \quad (3.8)$$

Here

$$K_{r,i_1} = \mu_{x_r}^{-(1/2)+\epsilon_{r,\nu}}\zeta_{i_1}(x_r)\mu_{x_r}^{(1/2)-\epsilon_{r,\nu}},$$

$\zeta_{i_1} \in \mathcal{C}_0^\infty$, and ζ_{i_1} equals one on a neighborhood of $[i_1, i_1+1]$, sufficiently large so (3.8) holds. In fact, we take ζ_{i_1} to be the translate of a fixed $\mathcal{C}_0^\infty$ function, so the $\mathcal{L}_p$ norms of its derivatives are independent of i_1. By Lemma 3.3,

$$\|E_{j'}\bar{K}\| \le O(1)\prod_{r=1}^{\bar{n}} e^{-(1/2)(\beta_r+\delta)|j_r' - i_1|}.$$

Thus performing the sum in (3.7) yields (3.6).

We remark that for an optimal choice of (α,β) satisfying $\alpha^2 + \beta^2 = (m-\epsilon)^2$,

$$e^{-\alpha t - \beta|x|} \le e^{-(m-\epsilon)d},$$

where $d = (t^2+x^2)^{1/2}$ is the Euclidean distance.

B. Estimates on Graphs

We use the estimates of Sec. 3A to establish estimates on a single polynomial $\hat{R}_\alpha$. We consider kernels $\hat{v}_\nu$ of the form (2.17), where $\nu = 1, 2, \ldots, N$ labels the vertices of the graph of $\hat{R}_\alpha$. Each kernel $\hat{v}_\nu(k_{1,\nu}, \ldots, k_{s,\nu}, q(\cdot))$ is a function of contracted lines with momenta $k_{r,\nu}$ and of the path space variables $q(\cdot)$ for uncontracted legs.

Each vertex has a space–time localization and a lower momentum cutoff $\kappa(v_\nu)$. The norm $\|\hat{v}_\nu\|_2 = \|\hat{v}(\cdot; q(\cdot))\|_2$ is the $\mathcal{L}_2$ norm of $\hat{v}$ in the contracted momentum variables, and it is a function on path space. (It is a constant on path space for fully contracted graphs.)

Let the line (i.e., the pair of contracted legs) with momentum $k_{r,\nu}$ connect vertices v_ν and $v_{\nu'}$, with space–time localizations Δ and Δ', respectively. We define the contraction distance $d_{r,\nu}$ of this line as the Euclidean distance between Δ and Δ'. We define the total contraction distance $d(v_\nu) = d_\nu$ of the vertex v_ν as

$$d_\nu = d(v_\nu) = \sum_r d_{r,\nu}, \quad (3.9)$$

where the sum extends over the contracted legs at vertex ν. Given $\epsilon > 0$, we introduce the localization factor l_ν of the vertex v_ν:

$$\gamma e^{-md_\nu/3}. \quad (3.10)$$

Here γ is a sufficiently large constant, chosen below.

Let $\hat{R}_\alpha(s)$ be a polynomial of the form (2.18).

Theorem 3.1: Given $0 < \epsilon$, there exists a constant γ such that

$$|\hat{R}_\alpha(s)| \le \left(\prod_{\text{vertices}} \gamma\|\bar{v}_\nu\|_2\right)\left(\prod_{\text{lines}} e^{-(m-\epsilon)d_{r,\nu}}\right) \le \prod_\nu(\gamma e^{-md_\nu/3}\|\bar{v}_\nu\|_2). \quad (3.11)$$

Remarks: For a fully contracted graph, $\|\bar{v}_\nu\|_2 \leq O(\kappa(v_\nu)^{-(1/2)+\epsilon_\nu+\delta})$, $\delta > 0$. Thus for a new constant γ,

$$|\hat{R}_\alpha(s)| \leq \prod_\nu [\gamma\kappa(v_\nu)^{-(1/2)+\epsilon_\nu+\delta} e^{-md_\nu/3}]. \tag{3.12}$$

With our choice $\epsilon_\nu \leq \frac{1}{8}$ in Sec. 2E, fully contracted graphs arising in the inductive construction satisfy

$$|\hat{R}_\alpha(s)| \leq \prod_\nu [\gamma\kappa(v_\nu)^{-1/4} e^{-md_\nu/3}]. \tag{3.13}$$

The proof of Theorem 3.1 follows from Proposition 3.1 in an elementary fashion. We localize each leg in space, and apply the Schwarz inequality to obtain

$$|\hat{R}| \leq \sum_{i_\nu} \prod_\nu \|E_{i_\nu} \hat{v}_\nu\|_2.$$

Using the fact that each pair of contracted legs has the same space–time localization, we bound the summation over i_ν with Proposition 3.1. We choose $(\alpha_{r,\nu}, \beta_{r,\nu})$ to optimize our bound and so obtain (3.11).

C. Combinatoric Estimates

At the conclusion of the inductive construction, we obtain an upper bound for (1.39) as a sum of constants on path space, with each constant represented by a vacuum graph. We now estimate this sum, which involves among other things, counting the number of terms in the expansion. We assign a positive constant $c(v_\nu)$ to each vertex v_ν in the inductive construction; $c(v_\nu)$ is called a combinatoric factor.

One use of combinatoric factors is to estimate sums by supremums. For instance,

$$|\sum_\alpha a_\alpha| \leq \sup_\alpha |c_\alpha a_\alpha|,$$

provided $\sum c_\alpha^{-1} \leq 1$. We also use the combinatoric factors to absorb the constants e^j or η in (2.34), except those coming from squares $\Delta \in \mathfrak{D}_0$. Since each vertex is localized in some square Δ, the time integration in (2.11) extends over a region of volume at most one. Thus we replace the integral over s by the supremum norm of the integrand. We note that for each $\varphi(h)$–$\varphi(h)$ contraction, the integral over one $\varphi(h)$ time has already been estimated and is not a part of the ds integration in (2.11), see Sec. 2E. We have

$$\begin{aligned}\langle \Omega_0, e^{-T(H_0+H_I+\delta H)}\Omega_0\rangle &= \int_Q \exp\Big(-\int_0^T H_{II}(q(\sigma))d\sigma\Big)dq(\cdot) \\ &\leq \sum_\alpha \int ds R_\alpha(s) \exp\Big(-\int_0^T H_I(\sim Y_\alpha(\sigma))(q(\sigma))d\sigma\Big) \\ &\leq \sum_\alpha \int ds |R_\alpha(s)| \exp\Big(-\int_0^T E(\sim Y_\alpha(\sigma))d\sigma\Big).\end{aligned} \tag{3.14}$$

Each polynomial $R_\alpha(s)$ is defined in (2.19) and is estimated in Sec. B:

$$|R_\alpha(s)| \leq B_\alpha L_\alpha \prod_\nu (l_\nu \|\bar{v}_\nu\|_2).$$

The constants in B_α, with the exception of factors e^{J_0} from applying (2.32) or (2.34) in squares $\Delta \in \mathfrak{D}_0$, are parts of combinatoric factors $c(v_\nu)$. With $c(v_\nu)$ as chosen below we will obtain

$$\begin{aligned}\langle \Omega_0, e^{-T(H_0+H_I+\delta H)}\Omega_0\rangle \leq e^{2J_0|X_0|T} \sup_\alpha \prod_\nu (c(v_\nu) l_\nu \|\bar{v}_\nu\|_2) \\ \times \exp\Big(-\int_0^T E(\sim Y_\alpha(\sigma))d\sigma\Big).\end{aligned} \tag{3.15}$$

The factor $\exp(2J_0|X_0|T)$ bounds the product of the constants e^{J_0} from squares $\Delta \in \mathfrak{D}_0$. We also prove

Proposition 3.2: For N_0 of (2.43) sufficiently large, we have

$$\begin{aligned}\prod_\nu (c(v_\nu) l_\nu \|\bar{v}_\nu\|_2) &\leq e^{-N(\alpha)} \\ &\leq \exp\Big(-\int_0^T |Y_\alpha(\sigma)|d\sigma - |X_0|T\Big),\end{aligned} \tag{3.16}$$

for every graph R_α in (3.14), where $N(\alpha)$ is the number of vertices in R_α.

The inequality (2.5) follows immediately. By (3.15) and (3.16), we have

$$\begin{aligned}&\langle \Omega_0, e^{-T(H_0+H_I+\delta H)}\Omega_0\rangle \\ &\leq e^{(2J_0+1)|X_0|T} \sup_\alpha \exp\Big(-\int_0^T \{E(\sim Y_\alpha(\sigma)) + |Y_\alpha(\sigma)|\}d\sigma\Big) \\ &\leq e^{(2J_0+1)|X_0|T} \sup_{X \supset X_0} e^{-T\{E(\sim X)+|X|\}}.\end{aligned} \tag{3.17}$$

Thus, by the formula above (1.39),

$$-E(g) - \delta E(g,h) \leq M|X_0| - \inf_{X \supset X_0} \{E(\sim X) + |X|\},$$

where $M = 2J_0 + 1$, which proves (2.5).

We now assign the combinatoric factors $c(v_\nu)$ to vertices v_ν. We denote these factors c_{pt}, c_{ps}, or c_h for PT, PS, or $\varphi(h)$ vertices, respectively.

Proposition 3.3: Let $\beta_0 > 0$ be given, and let N_0 of (2.43) be sufficiently large. There are constants β, N, independent of β_0 and N_0 such that

$$c_{pt}(v) \leq \beta d(v)^N, \tag{3.18}$$

$$c_{ps}(v) \leq \beta d(v)^N \kappa(v)^{1/8}, \tag{3.19}$$

$$c_h(v) \leq \beta_0 d(v)^N \tag{3.20}$$

and such that (3.5) is valid.

Proof: We assign a finite product of combinatoric factors to each vertex. Each factor in the product has the form $O(1)d(v)^n\kappa(v)^\epsilon$ and, hence, so does the product. Therefore, we only need to keep track of which powers $\kappa(v)^\epsilon$ occur, and we need to show that some factor for $\varphi(h)$ vertices can be made small.

1. The Graph Expansion

We start by assigning combinatoric factors to a given graph expansion step. This is our major assignment of combinatoric factors, since we count the number of graphs R_α that are formed during a graph expansion step. At the start of a graph expansion of a particular term, $n(\Delta)$ vertices in Δ have uncontracted legs. These "old" vertices contract to each other during the graph expansion, and they also contract to the exponent to form "new" PT vertices. The old vertices may contract further with the new vertices; but the new vertices do not contract to the exponent. A given PS or $\varphi(h)$ vertex can contract during at most two inductive steps. We assign combinatoric factors to vertices from the time they are introduced until the graph expansion when they fully contract.

If the rth leg of the νth vertex contracts to a vertex localized in Δ_j, we set $j = j(r,\nu)$. Consider the set of

all graphs in which the contraction localizations $j = j(r, \nu)$ of the old vertices take on fixed values. We count the number of such graphs $N(\{j(r,\nu)\})$, and then we sum over the allowed $\{j(r,\nu)\}$. Suppose that vertices v and v', localized in Δ and Δ', respectively, are connected by a line with contraction distance dist $(\Delta, \Delta') \leq \min\{d(v), d(v')\}$. Since

$$\sum_{\Delta'} \text{dist}(\Delta, \Delta')^{-3} = \beta_1 < \infty,$$

a combinatoric factor $\beta_1 d(v)^3$ assigned to each leg is sufficient to deal with each sum over contraction localizations.

Thus we may consider a fixed set of $\{j(r,\nu)\}$. Suppose that in a given graph expansion step l_{ij} old legs localized in Δ_i contract to legs localized in Δ_j, and that l'_{ij} new legs localized in Δ_i contract to legs localized in Δ_j. (Either l_{ij} or l'_{ij} is zero.) The contraction localizations $\{j(r,\nu)\}$ determine the $\{l_{ij}\}$ and the $\{l'_{ij}\}$. However the $\{j(r,\nu)\}$ do not determine the number of new vertices, their degrees, nor the number of contracted legs at each new vertex. Since new vertices need not be fully contracted, the number of new vertices in Δ_i with legs contracting to Δ_j lies between l'_{ij}/p and l'_{ij}. Each new vertex is a monomial φ^n, $1 \leq n \leq p$, and l'_{ij} values of n may be chosen in $p^{l'_{ij}}$ ways for a total of at most $l'_{ij}\, p^{l'_{ij}}$ possible choices of the number and degree of new vertices in Δ_i. We assign a combinatoric factor $\beta_2 = p^{2p}$ per new PT vertex, or at least $\beta_2^{l'_{ij}/p} = p^{2l'_{ij}}$ to count new vertices in Δ_i, and we assign a factor 2^p per vertex to count whether each of the (at most) p legs at a new vertex contracts. For new vertices, the localization factor $\beta_1 d(v)^3$ per leg deals with the sum over localizations, as above. Thus we now may fix the contraction localizations $\{j(r,\nu)'\}$ of the new vertices as well as $\{j(r,\nu)\}$. Let $l_{ij}^{\#}$ denote l_{ij} or l'_{ij}.

For new or old vertices, the $l_{ij}^{\#}$ legs from Δ_i to Δ_j can be connected in $l_{ij}^{\#}!$ ways if $i \neq j$ or in $l_{ii}!! \leq (l_{ii}!)^{1/2}$ ways if $i = j$. We need to count the contractions from Δ_i to Δ_j or from Δ_j to Δ_i but not both, so

$$N\{j,j'\} \leq \prod_{i,j} (l_{ij}^{\#}!)^{1/2}. \tag{3.21}$$

If $l'_{ij} \neq 0$, we divide $(l'_{ij}!)^{1/2} \leq (l'_{ij})^{l'_{ij}/2}$ equally among the l'_{ij} legs and, thus, assign a factor $(l'_{ij})^{p/2}$ per new PT vertex. In the $(k+1)$th inductive step, by Corollary 2.1, any PT vertex v in Δ_i contributing to l'_{ij} satisfies

$$l'_{ij} \leq p n(\Delta_j)_k \leq p d_e(\Delta_j)_k^3 \leq p d(v)^3.$$

Hence a factor $[p d(v)^3]^{p/2}$ per new PT vertex in Δ_i dominates the contribution of Δ_i to (3.21).

We now consider the case of old vertices $l_{ij} \neq 0$. Let m_{il}, $l = 1, 2, 3$ denote the number of uncontracted legs in Δ_i at the end of the kth inductive step and belonging to PS vertices ($l = 1$), $\varphi(h)$ vertices ($l = 2$), or PT vertices ($l = 3$). Then $m_{i1} \leq p n_{ps}(\Delta_i)_k$, $m_{i2} \leq n_h(\Delta_i)_k$, and $m_{i3} \leq (p-1) n_{pt}(\Delta_i)_k$. Also

$$l_i = \sum_j l_{ij} = m_{i1} + m_{i2} + m_{i3},$$
$$\prod_j (l_{ij}!)^{1/2} \leq (m_i!)^{1/2} \leq \prod_{l=1}^{3} (9 m_{il})^{m_{il}/2}. \tag{3.22}$$

We assign the factor $(9m_{il})^{m_{il}/2}$ among the vertices of type l localized in Δ_i. For $l = 1$, we have $m_{i1} \leq p n_{ps}(\Delta_i) \leq \kappa(v)^{1/8p}$. We divide the factor $m_{i1}^{1/2}$ equally among the PS vertices in Δ_i, yielding a factor

$$m_{i1}^{p/2} \leq O(1)\kappa(v)^{1/16} \tag{3.23}$$

per vertex.

For $\varphi(h)$ vertices, we assign the factor

$$3 m_{i2}^{1/2} \leq 3 n_h(\Delta_i)^{1/2} \tag{3.24}$$

per vertex, as well as a factor 2 per vertex for the two terms formed by the time integration for each φ–φ contraction (see Sec. 2E).

Finally, we consider PT vertices m_{i3}. We assert that there exists a constant β_3, depending only on p, such that

$$m_{i3} \leq \beta_3 d_c(\Delta_i)^6/3. \tag{3.25}$$

Our combinatoric factor (3.22) for $l = 3$ is bounded by

$$[O(1) d_c(\Delta_i)]^{\beta_3 d_c(\Delta_i)^6}. \tag{3.26}$$

The squaring exponent (2.23) of Sec. 2C ensures that if $m_{i3} \neq 0$, then at least $d_c(\Delta_i)^6$ old PT vertices v, each with $d_c(\Delta_i) = d_c(v)$, are localized in Δ_i. We assign a combinatoric factor $O(1) d_c(v)^{\beta_3}$ to each old PT vertex, which dominates (3.26).

We now prove (3.25). Let N_k be the number of new PT vertices produced in the graph expansion of the kth inductive step, and which retain uncontracted legs after squaring. Hence by (2.23),

$$m_{i3} \leq (p-1) n_{pt}(\Delta_i)_k \leq (p-1)\{2N_k + [4 d_c(\Delta_i)]^6\}.$$

But the vertices N_k are produced by contractions to the exponent, so

$$N_k \leq p \sum_{\Delta_j} n(\Delta_j)_{k-1}, \tag{3.27}$$

where the sum extends over squares Δ_j satisfying $n(\Delta_j)_{k-1} \leq d_e(\Delta_j)_{k-1}^3$, by Corollary 2.1. Since the vertices remain after squaring, $d_e(\Delta_j)_{k-1}^3 \leq d_c(\Delta_i)^3$, and at most $4 d_c(\Delta_i)^2$ squares Δ_j enter the sum (3.27). Hence $N_k \leq 4p d_c(\Delta_i)^5$, proving (3.25).

2. *The Squaring Operation*

The squaring operation (2.22) introduces two terms; so a combinatoric factor 2 is sufficient for each application of (2.22). We assign this factor to the vertex v' which contracts to the exponent forming the new PT vertex v requiring the use of (2.22). Each vertex v' can form at most p new PT vertices during the graph expansion, and we assign a combinatoric factor 2^p per vertex to deal with the squaring operation.

3. *The Path Space Construction*

Let us first consider (2.32). If $j > j_0$, we assign a combinatoric factor $\exp(\beta_4)$ per PS vertex. This factor $\exp[\beta_4 n_{ps}(\Delta)]$ dominates both e^j and the factor $O(1) j^2$ for summation over j. In fact, $j = O(1)$

$(\log\kappa_j)^{p/2} \le O(1)\kappa_j^{1/8p}$, and $n_{ps}(\Delta) \ge \kappa_j^{1/8p}$ if the lower cutoff in Δ is κ_j. Thus for β_4 sufficiently large,

$$O(1)j^2 e^j \le \exp(\beta_4 \kappa_j^{1/8p}) \le \exp[\beta_4 n_{ps}(\Delta)].$$

In case $j = j_0$, there are no PS vertices in Δ_i. For $\Delta_i \subset \mathfrak{D}_0$, we include the factor $\exp(J_0)$ explicitly in our bound (3.15). For $\Delta_i \notin \mathfrak{D}_0$, we assign $\exp[j_0(\Delta_i)]$ as a combinatoric factor for the vertices localized in Δ_j, where Δ_j forced the PS construction in Δ_i. By (2.35), the $n(\Delta_j)$ vertices in Δ_j with uncontracted legs force the PS construction in at most $64n(\Delta_j)^{2/3}$ squares Δ_i. [This is the area of $\Gamma(\Delta_j)$.] By (2.36), we assign at most $\exp(64n(\Delta_j)^{2/3}\{[J_1 n(\Delta_j)^{1/6}] + 1\})$ among the vertices in Δ_j. Hence a factor $O(1)$ per vertex is sufficient.

Finally we consider (2.33), the part of (2.34) not analyzed above. The factor $\exp(J_0)$ is kept explicitly in (3.15). For $2k$ $\varphi(h)$ vertices we have a factor $\eta/(2k)! \le \eta k^{-2k}$ or a factor

$$O(1)n_h(\Delta_i)^{-1} \tag{3.28}$$

per vertex. A combinatoric factor $O(1)$ per vertex dominates the sum over k. We choose $N_0 \le n_h(\Delta_i)$ sufficiently large so that (3.28) dominates (3.24), and their product is sufficiently small to ensure (3.20). This completes the proof of the proposition.

Proof of Proposition 3.2: We use Theorem 3.1 and Proposition 3.3. For a PT, PS, or $\varphi(h)$ vertex v_ν, we have

$$c(v_\nu)l_\nu\|\bar{v}_\nu\|_2 \le \begin{cases} \beta\gamma d(v_\nu)^N e^{-md(v_\nu)/6}, & \text{PT} \\ \beta\gamma d(v_\nu)^N e^{-md(v_\nu)/6}\kappa(v_\nu)^{-1/8}, & \text{PS} \\ \beta_0\gamma d(v_\nu)^N e^{-md(v_\nu)/6}\|\mu^{-(1/2)-1/8p}h\|_2, & \varphi(h) \end{cases}$$

We use Theorem 3.1 to give the estimate $O[\kappa(v_\nu)^{-(1/2)+1/8+\delta}]$ for PS vertices, the $\kappa^{1/8}$ coming from Sec. 2E, and this dominates the $O[\kappa(v_\nu)^{1/8}]$ growth of (3.19). By Corollary 2.2, $\kappa(v_\nu)$ for PS vertices tends to infinity with N_0 of (2.43). Similarly, for PT vertices, $d(v_\nu) \to \infty$ as $N_0 \to \infty$ by Corollary 2.3. We choose β_0 sufficiently small and N_0 sufficiently large so

$$c(v_\nu)l_\nu\|\bar{v}_\nu\|_2 \le e^{-1}.$$

To complete the proof of (3.16), we show that each polynomial R_α has a graph with $N(\alpha)$ vertices, where

$$N(\alpha) \ge \int_0^T |Y_\alpha(\sigma)|d\sigma - |X_0|T.$$

The right side is the area in $\sim \mathfrak{D}_0$ in which the exponent has been removed. Let $\Delta \notin \mathfrak{D}_0$ and let Δ' force the PS construction in Δ. We count the vertices in Δ'. The square $\Gamma(\Delta')$ has area $64n(\Delta')^{2/3}$ and if $N_0 > (64)^3$, then by Corollary 2.2, $64n(\Delta')^{2/3} < n(\Delta')$. Hence R_α has more vertices than squares in $\sim \mathfrak{D}_0$ with the exponent removed.

4. COMPLETION OF THE BOUND ON δE

Theorem 4.1: There exists a_0 such that for all intervals X with $|X| \ge a_0$,

$$E(g) \le E(\sim X) + |X|. \tag{4.1}$$

We use the inductive construction to bound $E(\sim X)$ and prove (4.1). In this section we describe the modifications to Secs. 2 and 3 required to establish (4.1), i.e., (2.4) and, hence, to complete the proof of Proposition 2.1 and Theorem 1.1.

A. The Inductive Construction

Let $X_a = X_a(g)$ minimize $E(\sim X) + |X|$ under the restriction that X is an interval of length $|X| \ge a$. Then, for $|X| \ge a$,

$$-E(\sim X) \le -E(\sim X_a) + |X| - |X_a|. \tag{4.2}$$

We assume that (4.1) is false, namely for $a \ge a_0$,

$$-E(g) \le -E(\sim X_a) - |X_a|. \tag{4.3}$$

We use the inductive construction and (4.2) and (4.3) to bound $E(\sim X_a)$; we thereby obtain a contradiction, establishing (4.1). We bound

$$\int \exp\left(-\int_0^T H_I(\sim X_a)(q(\sigma))d\sigma\right)dq(\cdot). \tag{4.4}$$

The only change in the inductive construction occurs at the start of the first inductive step, i.e., the start of the first PS construction. Roughly, the idea is to begin the inductive construction by inserting and removing the factor $\exp(-H_I(X_a))$ in the integrand of (4.4). For this purpose we use the inequality

$$1 \le e^{-r+s} + (r/s)^{2n}, \tag{4.5}$$

valid for real r, for $s > 0$ and for $n \in \mathfrak{I}^+$. We choose $n = [|X_a|^{1/8p}]$, $s = \frac{1}{2}|X_a|$, and

$$r = \int_{i_0}^{i_0+1} H_I(X_a)(q(\sigma))d\sigma. \tag{4.6}$$

We begin the first PS construction by inserting (4.5) and (4.6) in the integrand of (4.4) at each time i_0 such that $0 \le i_0 \le T$, $i_0 \in \tau\mathfrak{I}^+$, where τ is a positive integer to be determined later. Thus we apply (4.5) and (4.6) in the union of squares

$$\mathfrak{D}_1 = \{\Delta_i : i_1 \in X_a, i_0 \in \tau\mathfrak{I}^+, i_0 \le T\},$$

producing $2^{[T/\tau]}$ terms. For each such term, let $\mathfrak{D}_0$ be the subset of $\mathfrak{D}_1$ in which the exponent is removed, i.e., the squares in which the second term in (4.5) is selected. Let $\mathfrak{D}_0'$ be the complementary subset of $\mathfrak{D}_1$ in which the exponent remains. Then $\mathfrak{D}_1 = \mathfrak{D}_0 \cup \mathfrak{D}_0'$, and each square $\Delta_i \in \mathfrak{D}_0$ contains $2[|X_a|^{1/8p}]$ vertices, each of the form $2|X_a|^{-1}H_{I,i}$. We call these vertices PS vertices, we include the factor $2|X_a|^{-1}$ with the kernel of these PS vertices, and we assign the function $\kappa(v) = |X_a|$ to such PS vertices. Then $\kappa(v)$ plays the role of a lower cutoff $n = [\kappa(v)^{1/8p}]$ and $\|v\|_2 \le O(\kappa(v)^{-1/2})$.

After starting the PS construction in this manner and obtaining PS vertices in $\mathfrak{D}_0$, we continue the inductive construction as in Sec. 2. With our definition of the PS vertices in $\mathfrak{D}_0$, the estimates of Sec. 2 and 3 concerning the inductive construction apply. In the case of (2.43), $N_0 = 2n \ge 2a_0^{1/8p} \to \infty$ as $a_0 \to \infty$.

At the end of the inductive construction, we will obtain $\exp[-E(\sim X)]$, which we bound by (4.2) or (4.3). Here X is the space region (at some time) in which the exponent is removed. The inequalities (4.2) or (4.3)

apply if X is an interval and either $|X| \geq a$ or $X = \emptyset$, respectively. We now verify these conditions for any X appearing at the end of the inductive construction. If $\Delta_i \not\subset X_a \times [0, T]$, then $d_0(\Delta_i)$ exceeds the horizontal, or equal time distance from Δ_i to $X_a \times [0, T]$. By Corollary 2.1, if $n(\Delta_i)_r \neq 0$, then at the end of the rth PS construction $d_e(\Delta_i)_r > d_0(\Delta_i)$. Thus we ensure that X has the desired form if whenever the exponent is removed from $\Delta \in \mathfrak{D}_0'$, it is also removed from all other squares in $\mathfrak{D}_0'$ with the same time localization as Δ.

Suppose that Δ' forces the PS construction in $\Delta \in \mathfrak{D}_0'$. If $d_0(\Delta') \geq 8^{-1}\,\mathrm{dist}(\Delta, \Delta') + 8^{-1}|X_a|$, then by (2.38)

$$n(\Delta')^{1/3} \geq d_0(\Delta') \geq 8^{-1}\,\mathrm{dist}(\Delta, \Delta') + 8^{-1}|X_a|.$$

Thus Δ' forces the PS construction in a square $\Gamma(\Delta')$ with side length $8n(\Delta')^{1/3} \geq \mathrm{dist}(\Delta, \Delta') + |X_a|$ and centered at Δ'. This $\Gamma(\Delta')$ includes all squares in $\mathfrak{D}_0'$ that are localized at the same time as Δ.

On the other hand, if $d_0(\Delta') \leq 8^{-1}\,\mathrm{dist}(\Delta, \Delta') + 8^{-1}|X_a|$, let $d_0(\Delta') = \mathrm{dist}(\Delta', \Delta'')$, $\Delta'' \in \mathfrak{D}_0$. Then

$$\begin{aligned}\mathrm{dist}(\Delta, \Delta') &\geq \mathrm{dist}(\Delta, \Delta'') - \mathrm{dist}(\Delta'', \Delta') \\ &\geq \tau - d_0 \geq \tau - 8^{-1}\,\mathrm{dist}(\Delta, \Delta') - 8^{-1}|X_a|.\end{aligned}$$

Suppose $\tau = 10|X_a|$. Thus

$$|X_a| \leq 8^{-1}\,\mathrm{dist}(\Delta, \Delta'). \tag{4.7}$$

Let $d_1(\Delta')$ be the distance between the space localization of Δ' and the center of X_a. {If $\Delta' = \Delta_i$ and $X_a = [\alpha, \beta]$, then $d_1(\Delta') = |i_1 - \frac{1}{2}(\alpha + \beta)|$.} Half the side length of $\Gamma(\Delta')$ is greater than $2^{-1/2}\,\mathrm{dist}(\Delta, \Delta')$. We show that

$$d_1(\Delta') + \tfrac{1}{2}|X_a| \leq 2^{-1/2}\,\mathrm{dist}(\Delta, \Delta'), \tag{4.8}$$

so Δ' forces the PS construction in all squares in $\mathfrak{D}_0'$ localized at the same time as Δ. By the triangle inequality, $d_1(\Delta') \leq d_0(\Delta') + \frac{1}{2}|X_a|$; so by (4.7),

$$\begin{aligned}d_1(\Delta') + \tfrac{1}{2}|X_a| &\leq d_0(\Delta') + |X_a| \\ &\leq \tfrac{1}{8}\,\mathrm{dist}(\Delta, \Delta') + \tfrac{9}{8}|X_a| \\ &\leq \tfrac{3}{8}\,\mathrm{dist}(\Delta, \Delta') \leq 2^{-1/2}\,\mathrm{dist}(\Delta, \Delta').\end{aligned}$$

Thus we choose $\tau = \tau(a) = 10|X_a|$ to ensure the desired geometry.

B. The Combinatorial Estimates

The combinatoric factors for the graph expansion and the squaring operation are assigned as in chapter 3. In the path space construction, we now consider the combinatoric factors from the application of (4.5) and (4.6) in $\mathfrak{D}_1$. Let $\mathcal{I}_1 = \{i_0 : 0 \leq i_0 \leq T, i_0 \in \tau\mathbb{Z}_+\}$. For a vacuum graph R_α at the end of the inductive construction, let $\mathcal{I}_0 = \mathcal{I}_0(\alpha) \subset \mathcal{I}_1$ be the subset of times for which the second term of (4.5) was chosen. Let $\mathcal{I}_0'(\alpha) \subset \mathcal{I}_1$ be the complementary subset of times for which the first term of (4.5) was chosen. Let $\mathcal{I}_0'' \subset \mathcal{I}_0'$ be the subset of times in $\mathcal{I}_0'$ when the PS construction was not used in X_a. We note that $Y_\alpha(\sigma)$ in (3.14) has the form

$$Y_\alpha(\sigma) = \begin{cases} \emptyset, & [\sigma] \in \mathcal{I}_0''(\alpha) \\ X, X \supset X_a, & [\sigma] \notin \mathcal{I}_0''(\alpha). \end{cases} \tag{4.9}$$

Thus $H_I(\sim Y_\alpha(\sigma)) = H_I(g), H_I(\sim X)$ in these respective cases.

Each application of (4.5) yields two terms, so we assign a combinatoric factor 2 to each time $i_0 \in \mathcal{I}_1$. If $i_0 \in \mathcal{I}_0$, there are vertices in X_a at time i_0 and we assign the factor 2 to any such vertex. If $i_0 \in \mathcal{I}_0' \sim \mathcal{I}_0''$, we assign the factor 2 to a vertex in the square forcing the PS construction, as with the factor e^s below. The remaining factor is $\exp(|\mathcal{I}_0''|\log 2)$, where $|\mathcal{I}_0''|$ is the number of elements of $\mathcal{I}_0''$.

The other relevant factor is the constant $\exp(\frac{1}{2}|X_a|) = e^s$ from the first term of (4.5). If $i_0 \in \mathcal{I}_0' \sim \mathcal{I}_0''$, then $\exp(\frac{1}{2}|X_a|)$ is a combinatoric factor $\exp(\frac{1}{2})$ per square $\Delta = \Delta_{i_0, i_1}$ for $i_1 \in X_a$. We distribute this factor among the vertices in Δ' that forced the PS construction in Δ. Since $n(\Delta')$ vertices in Δ' force the PS construction in at most $64n(\Delta')^{2/3}$ squares, we assign at most a factor $O(1)$ per vertex in Δ'. For times $i_0 \in \mathcal{I}_0''$, we obtain a factor $\exp(\frac{1}{2}|X_a||\mathcal{I}_0''|)$ in our expansion.

Thus the inductive construction yields

$$\begin{aligned}\langle \Omega_0, e^{-T[H_0 + H_1(\sim X_a)]}\Omega_0\rangle &= \int e^{-H_I[\sim(X_a)(q(\sigma))]d\sigma}\,dq(\cdot) \\ &\leq \sum_\alpha \int ds\,|R_\alpha(s)|\exp\Big(-\int_0^T H_I(\sim Y_\alpha(\sigma))(q(\sigma))d\sigma\Big) \\ &\leq \sup_\alpha e^{|\mathcal{I}_0''(\alpha)|\{\log 2 + \frac{1}{2}|X_a|\}} \\ &\quad \times \prod_\nu (c(v_\nu) l_\nu \gamma \|\bar{v}_\nu\|_2)\exp\Big(-\int_0^T E(\sim Y_\alpha(\sigma))d\sigma\Big).\end{aligned} \tag{4.10}$$

For $|\mathcal{I}_0''(\alpha)|$ time intervals $\sigma \in [i_0, i_0 + 1)$, we bound $-E(\sim Y_\alpha(\sigma)) = -E(g)$ by (4.3). For the remaining time intervals we use (4.2). Thus

$$\begin{aligned}&\exp\Big(-\int E(\sim Y_\alpha(\sigma))d\sigma\Big) \\ &\quad \leq e^{-|\mathcal{I}_0''(\alpha)||X_a|} e^{-TE(\sim X_a)} \exp\Big(\int_0^T |Y_\alpha(\sigma) \sim X_a|d\sigma\Big).\end{aligned} \tag{4.11}$$

By Proposition 3.2, for N_0 sufficiently large,

$$\prod_\nu (c(v_\nu) l_\nu \gamma \|\bar{v}_\nu\|_2) \leq e^{-N(\alpha)}.$$

In our case

$$N(\alpha) \geq \int_0^T \{|Y_\alpha(\sigma) \sim X_a|\}d\sigma + \{T/\tau - |\mathcal{I}_0''(\alpha)|\}. \tag{4.12}$$

The bound $N(\alpha) \geq \int_0^T \{|Y_\alpha(\sigma) \sim X_a|\}d\sigma$ follows as in the proof of Proposition 3.2, neglecting squares in $\mathfrak{D}_1$. In a strip $X_a \times [i_0, i_0 + 1]$, $i_0 \in \mathcal{I}_0$, there is at least one vertex. For a square $\Delta \in X_a \times \mathcal{I}_0' \sim \mathcal{I}_0''$, we count one of the $n(\Delta')$ vertices in Δ', where Δ' forced the PS construction in Δ. Thus we count at least one vertex for each time interval in $\mathcal{I}_1 \sim \mathcal{I}_0''$. There are $T/\tau - |\mathcal{I}_0''|$ such time intervals, establishing (4.12).

From (4.10)–(4.12), we find

$$\begin{aligned}&\langle \Omega_0, e^{-T[H_0 + H_I(\sim X_a)]}\Omega_0\rangle \\ &\quad \leq \sup_\alpha e^{-TE(\sim X_a)} e^{-T/\tau} e^{-|\mathcal{I}_0''(\alpha)|\{-\log 2 - 1 + \frac{1}{2}|X_a|}} \\ &\quad \leq e^{-T[E(\sim X_a) + \tau^{-1}]},\end{aligned}$$

if a_0 (and hence $|X_a| \geq a \geq a_0$ and N_0) are sufficiently large.

Taking the logarithm, dividing by T and letting $T \to \infty$, the left side converges to $-E(\sim X_a)$. This gives the contradiction $0 \leq -\tau^{-1}$, and proves (4.1).

5. A CORRECTION

Our proof[5] of the spectrum condition $P^2 \leq H^2$ contains a gap, as was pointed out by Fröhlich and Faris. Namely, we required the Lorentz rotated Hamiltonian in a periodic box to have a simple ground state, which does not follow, as claimed, from standard methods. The remaining results in Ref. 5 are either independent of this gap or are proved (and improved) by the present paper, with the exception of the estimate

$$\pm \nabla\varphi(h) \leq \text{const}\|h\|_2(H + I).$$

This bound is Theorem 1.1, Case 2, $\epsilon = \frac{1}{2}$, and it should follow from the methods of the present paper.

* Supported in part by the National Science Foundation, Grant GP-24003.
† Supported in part by the Air Force Office of Scientific Research, Contract AF 44620-70-C0030 and the National Science Foundation, Grant GP-31239X.

1 J. Glimm and A. Jaffe, *Quantum Field Theory Models*, in the 1970 Les Houches lectures, edited by C. Dewitt and R. Stora (Gordon and Breach, New York, 1971).
2 J. Glimm and A. Jaffe, "Boson Quantum Field Models," in the 1971 London lectures, edited by R. Streater (to appear).
3 J. Glimm and A. Jaffe, Acta Math **125**, 203 (1970).
4 J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970).
5 J. Glimm and A. Jaffe, J. Math. Phys. **11**, 3335 (1970); Commun. Math. Phys. **22**, (1971).
6 J. Glimm and A. Jaffe, Commun. Math. Phys. **22**, 253 (1971).

Some Remarks on Three-Waves Interaction

G. Casati and E. Lazzaro
Laboratorio di Fisica del Plasma del C.N.R., Istituto di Fisica dell'Università di Milano, Milano, Italy
(Received 24 April 1972)

The problem of nonlinear interaction among three waves is analyzed in the case in which one of the waves is initially absent. Clarification is made of some mathematical aspects of the interaction mechanism.

The study of the nonlinear interaction among different waves in a plasma is of primary importance, especially in connection with the possibility of exciting proper low-frequency modes, e.g., ion acoustic waves of the plasma by external transverse waves.

The subject has been extensively explored by many authors,[1-7] but there are some aspects of the way in which solutions are obtained that it is worthwhile to reconsider and clarify. In particular we refer to the case in which two waves excite an initially absent third wave, when matching conditions $\omega_1 + \omega_2 = \omega_3$, $k_1 + k_2 = k_3$ are fulfilled.

The procedure used by several authors in this case seems to be not completely correct, and we reconsider the problem through a different calculation scheme which yields the correct solution.

It is well known that the problem of nonlinear three-wave interaction generally leads, in the usual scheme of approximation[6-8] to the following truncated system of equations for the evolution of amplitudes:

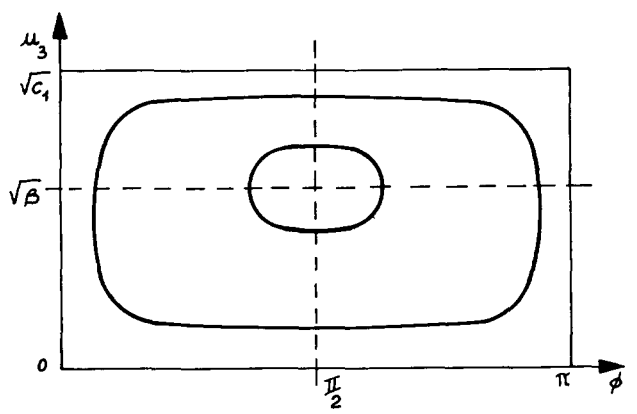


FIG. 1.

$$\frac{db_1}{d(\epsilon t)} = \lambda_1 b_2^* b_3, \quad \frac{db_2}{d(\epsilon t)} = \lambda_2 b_1^* b_3, \quad \frac{db_3}{d(\epsilon t)} = -\lambda_3 b_1 b_2, \tag{1}$$

where the b_j are the complex amplitudes of the three waves, ϵ is a small expansion parameter and the λ_j are the coupling parameters.[7]

This system is usually solved with the method employed in Ref. 9. Through the substitution

$$u_j = \sqrt{\lambda_k \lambda_l}\,|b_j|, \quad b_j = |b_j| e^{i\varphi_j} = (u_j/\sqrt{\lambda_k \lambda_l})\, e^{i\varphi_j}, \tag{2}$$

one obtains

$$\frac{du_1}{d(\epsilon t)} = u_2 u_3 \cos\phi, \quad \frac{du_2}{d(\epsilon t)} = u_1 u_3 \cos\phi,$$
$$\frac{du_3}{d(\epsilon t)} = -u_1 u_2 \cos\phi, \tag{3}$$

$$\frac{d\phi}{d(\epsilon t)} = \left(\frac{u_1 u_2}{u_3} - \frac{u_1 u_3}{u_2} - \frac{u_2 u_3}{u_1}\right) \sin\phi, \tag{4}$$

where

$$\phi = \varphi_3 - \varphi_2 - \varphi_1.$$

This system of equations is solved by a suitable set of elliptic functions.[6,7,9]

In the simplest case of physical interest, when the third wave is initially zero, it can be easily verified that the Lipschitz condition is violated in the equation for ϕ. On the other hand, it is evident that in this case the substitution (2) is meaningless as the phase is completely undetermined. It may be useful to show these considerations by looking at the curves in the plane u_3, ϕ.

From Eqs. (3) and (4), in fact, we easily obtain

$$\frac{du_3}{d\phi} = -\frac{(c_1 - u_3^2)(c_2 - u_3^2)\, u_3 \cos\phi}{3(\alpha - u_3^2)(\beta - u_3^2) \sin\phi}, \tag{5}$$

Taking the logarithm, dividing by T and letting $T \to \infty$, the left side converges to $-E(\sim X_a)$. This gives the contradiction $0 \leq -\tau^{-1}$, and proves (4.1).

5. A CORRECTION

Our proof[5] of the spectrum condition $P^2 \leq H^2$ contains a gap, as was pointed out by Fröhlich and Faris. Namely, we required the Lorentz rotated Hamiltonian in a periodic box to have a simple ground state, which does not follow, as claimed, from standard methods. The remaining results in Ref. 5 are either independent of this gap or are proved (and improved) by the present paper, with the exception of the estimate

$$\pm \nabla\varphi(h) \leq \text{const}\|h\|_2(H + I).$$

This bound is Theorem 1.1, Case 2, $\epsilon = \frac{1}{2}$, and it should follow from the methods of the present paper.

* Supported in part by the National Science Foundation, Grant GP-24003.
† Supported in part by the Air Force Office of Scientific Research, Contract AF 44620-70-C0030 and the National Science Foundation, Grant GP-31239X.

1 J. Glimm and A. Jaffe, *Quantum Field Theory Models*, in the 1970 Les Houches lectures, edited by C. Dewitt and R. Stora (Gordon and Breach, New York, 1971).
2 J. Glimm and A. Jaffe, "Boson Quantum Field Models," in the 1971 London lectures, edited by R. Streater (to appear).
3 J. Glimm and A. Jaffe, Acta Math **125**, 203 (1970).
4 J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970).
5 J. Glimm and A. Jaffe, J. Math. Phys. **11**, 3335 (1970); Commun. Math. Phys. **22**, (1971).
6 J. Glimm and A. Jaffe, Commun. Math. Phys. **22**, 253 (1971).

Some Remarks on Three-Waves Interaction

G. Casati and E. Lazzaro
Laboratorio di Fisica del Plasma del C.N.R., Istituto di Fisica dell'Università di Milano, Milano, Italy
(Received 24 April 1972)

The problem of nonlinear interaction among three waves is analyzed in the case in which one of the waves is initially absent. Clarification is made of some mathematical aspects of the interaction mechanism.

The study of the nonlinear interaction among different waves in a plasma is of primary importance, especially in connection with the possibility of exciting proper low-frequency modes, e.g., ion acoustic waves of the plasma by external transverse waves.

The subject has been extensively explored by many authors,[1-7] but there are some aspects of the way in which solutions are obtained that it is worthwhile to reconsider and clarify. In particular we refer to the case in which two waves excite an initially absent third wave, when matching conditions $\omega_1 + \omega_2 = \omega_3$, $k_1 + k_2 = k_3$ are fulfilled.

The procedure used by several authors in this case seems to be not completely correct, and we reconsider the problem through a different calculation scheme which yields the correct solution.

It is well known that the problem of nonlinear three-wave interaction generally leads, in the usual scheme of approximation[6-8] to the following truncated system of equations for the evolution of amplitudes:

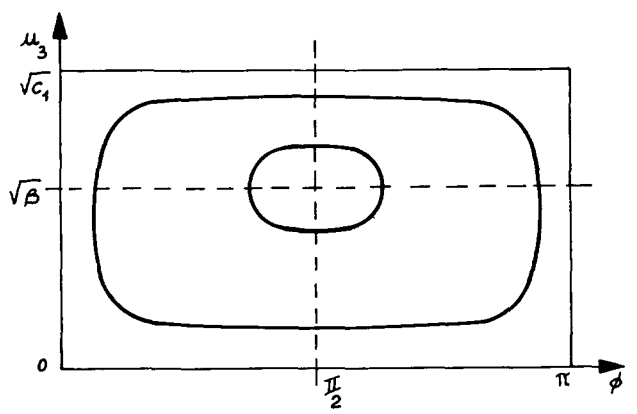


FIG. 1.

$$\frac{db_1}{d(\epsilon t)} = \lambda_1 b_2^* b_3, \quad \frac{db_2}{d(\epsilon t)} = \lambda_2 b_1^* b_3, \quad \frac{db_3}{d(\epsilon t)} = -\lambda_3 b_1 b_2, \tag{1}$$

where the b_j are the complex amplitudes of the three waves, ϵ is a small expansion parameter and the λ_j are the coupling parameters.[7]

This system is usually solved with the method employed in Ref. 9. Through the substitution

$$u_j = \sqrt{\lambda_k \lambda_l}\,|b_j|, \quad b_j = |b_j| e^{i\varphi_j} = (u_j/\sqrt{\lambda_k \lambda_l})\, e^{i\varphi_j}, \tag{2}$$

one obtains

$$\frac{du_1}{d(\epsilon t)} = u_2 u_3 \cos\phi, \quad \frac{du_2}{d(\epsilon t)} = u_1 u_3 \cos\phi,$$
$$\frac{du_3}{d(\epsilon t)} = -u_1 u_2 \cos\phi, \tag{3}$$

$$\frac{d\phi}{d(\epsilon t)} = \left(\frac{u_1 u_2}{u_3} - \frac{u_1 u_3}{u_2} - \frac{u_2 u_3}{u_1}\right) \sin\phi, \tag{4}$$

where

$$\phi = \varphi_3 - \varphi_2 - \varphi_1.$$

This system of equations is solved by a suitable set of elliptic functions.[6,7,9]

In the simplest case of physical interest, when the third wave is initially zero, it can be easily verified that the Lipschitz condition is violated in the equation for ϕ. On the other hand, it is evident that in this case the substitution (2) is meaningless as the phase is completely undetermined. It may be useful to show these considerations by looking at the curves in the plane u_3, ϕ.

From Eqs. (3) and (4), in fact, we easily obtain

$$\frac{du_3}{d\phi} = -\frac{(c_1 - u_3^2)(c_2 - u_3^2)\, u_3 \cos\phi}{3(\alpha - u_3^2)(\beta - u_3^2) \sin\phi}, \tag{5}$$

where c_1, c_2 are two constants of motion

$$c_1 = u_2^2 + u_3^2, \quad c_2 = u_1^2 + u_3^2, \tag{6}$$

and

$$\alpha = \tfrac{1}{3}[c_1 + c_2 + (c_1^2 + c_2^2 - c_1c_2)^{1/2}],$$
$$\beta = \tfrac{1}{3}[c_1 + c_2 - (c_1^2 + c_2^2 - c_1c_2)^{1/2}]. \tag{7}$$

The plot of u_3 versus ϕ is constituted by closed curves that degenerate to a rectangle when $u_3(0)$ is zero (see Fig. 1).

It is just on the border of the rectangle that the Lipschitz condition for Eq. (5) is violated.

In the current literature[6,7,9] these facts have been overlooked, and this may lead to consider a system of equations for the amplitudes that does not have solutions, as is the case of Eq. (27) in Ref. 7.

We show in the following that a simple way to overcome these difficulties is to avoid the use of polar representation of complex amplitudes. We put

$$b_j = (1/\sqrt{\lambda_k \lambda_l})\ (\kappa_j + iy_j), \tag{8}$$

thereby obtaining the system

$$\begin{aligned} \dot\kappa_1 &= \kappa_2\kappa_3 + y_2y_3, \quad & \dot y_1 &= \kappa_2 y_3 - \kappa_3 y_2, \\ \dot\kappa_2 &= \kappa_1\kappa_3 + y_1y_3, \quad & \dot y_2 &= \kappa_1 y_3 - \kappa_3 y_1, \\ \dot\kappa_3 &= -\kappa_1\kappa_2 + y_1y_2, \quad & \dot y_3 &= -\kappa_1 y_2 - \kappa_2 y_1, \end{aligned} \tag{9}$$

which we want to integrate with the condition $\kappa_3(0) = y_3(0) = 0$.

It is easily verified that a constant of the motion is

$$H = y_1y_2y_3 + y_1\kappa_2\kappa_3 + \kappa_1 y_2\kappa_3 - \kappa_1\kappa_2 y_3, \tag{10}$$

which is equal to zero with the considered condition.

From Eq. (10) we thus obtain

$$y_2 = \frac{\kappa_2(\kappa_1 y_3 - \kappa_3 y_1)}{y_1y_3 + \kappa_1\kappa_3}, \quad y_1 = \frac{\kappa_1(y_3\kappa_2 - \kappa_3 y_2)}{y_2y_3 + \kappa_2\kappa_3}. \tag{11}$$

By comparison of Eqs. (11) with Eqs. (9) it follows that

$$\dot\kappa_1/\kappa_1 = \dot y_1/y_1, \quad \dot\kappa_2/\kappa_2 = \dot y_2/y_2,$$

and therefore

$$y_1/\kappa_1 = s_1, \quad y_2/\kappa_2 = s_2; \tag{12}$$

by sutstituting Eqs. (12) into Eq. (10) it follows then that

$$y_3/\kappa_3 = (s_1 + s_2)/(1 - s_1s_2). \tag{13}$$

System (9) reads now

$$\begin{aligned} \dot\kappa_1 &= [1 + s_2(s_1 + s_2)/(1 - s_1s_2)]\kappa_2\kappa_3, \\ \dot\kappa_2 &= [1 + s_1(s_1 + s_2)/(1 - s_1s_2)]\kappa_1\kappa_3, \\ \dot\kappa_3 &= -(1 - s_1s_2)\kappa_1\kappa_2, \end{aligned} \tag{14}$$

and its solutions are the usual Jacobi functions. On the other hand, remembering that $y_i/\kappa_i = \tan\varphi_i$, from Eqs. (12) and (13) we obtain

$$\tan\varphi_3 = \tan(\varphi_1 + \varphi_2),$$

that is,

$$\phi = 0 \quad \text{or} \quad \phi = \pi.$$

It turns out that the phase ϕ is a discontinuous piecewise constant function which takes the value π during the time of growth of u_3 and the value zero during the decrease. Taking into account this behaviour of the phase ϕ, system (3) gives now the correct Eqs. for the evolution of the amplitudes.

1 R. E. Aamodt and W. E. Drummond, Phys. Fluids **7**, 1816 (1964).
2 A. A. Galeev, V. I. Karpman, and R. Z. Sagdeev, Nucl. Fusion **5**, 20 (1965).
3 R. E. Aamodt and W. E. Drummond, J. Nucl. Energy Pt. **C6**, 147 (1964).
4 K. Matsuura, J. Phys. Soc. Japan. **21**, 2011 (1966).
5 D. Montgomery, Physica **31**, 693 (1965).
6 R. Sugihara, Phys. Fluids **11**, 178 (1968).
7 K. P. Das, Phys. Fluids **14**, 124 (1971).
8 E. A. Frieman, J. Math. Phys. **4**, 410 (1963).
9 S. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).

Reduction of the Poincaré Group with Respect to the Lorentz Group*

W. W. MacDowell and Ralph Roskies
Physics Department, Yale University, New Haven, Conn. 06520

(Received 19 April 1972)

The representations of the Poincaré Group for spinless particles, reduced with respect to the Lorentz subgroup, are investigated. They involve the principal series of representations of $SL(2, C)$ and use is made of a basis, introduced by Gel'fand in which the states are labeled by a complex number z. The transformation matrices relating to the Wigner basis are derived. The matrix elements of the momentum operators are obtained. The general form of the S matrix in the new basis is discussed. This basis may be relevant for a field theoretical description of the Veneziano model.

1. INTRODUCTION

In the relativistic quantum theory of massive particles, physical states are generally described by vectors in Fock space, which is the Hilbert space of the direct product of single-particle states. The single-particle states form a basis for unitary irreducible representations of the Poincaré group. The basis most commonly used is the Wigner basis in which the Poincaré group is reduced with respect to the Abelian subgroup of translations. In addition, one operator commuting with the translation subgroup is given to specify the spin, such as in the helicity basis or in the

where c_1, c_2 are two constants of motion

$$c_1 = u_2^2 + u_3^2, \quad c_2 = u_1^2 + u_3^2, \tag{6}$$

and

$$\alpha = \tfrac{1}{3}[c_1 + c_2 + (c_1^2 + c_2^2 - c_1c_2)^{1/2}],$$
$$\beta = \tfrac{1}{3}[c_1 + c_2 - (c_1^2 + c_2^2 - c_1c_2)^{1/2}]. \tag{7}$$

The plot of u_3 versus ϕ is constituted by closed curves that degenerate to a rectangle when $u_3(0)$ is zero (see Fig. 1).

It is just on the border of the rectangle that the Lipschitz condition for Eq. (5) is violated.

In the current literature[6,7,9] these facts have been overlooked, and this may lead to consider a system of equations for the amplitudes that does not have solutions, as is the case of Eq. (27) in Ref. 7.

We show in the following that a simple way to overcome these difficulties is to avoid the use of polar representation of complex amplitudes. We put

$$b_j = (1/\sqrt{\lambda_k \lambda_l})\ (\kappa_j + iy_j), \tag{8}$$

thereby obtaining the system

$$\begin{aligned} \dot\kappa_1 &= \kappa_2\kappa_3 + y_2y_3, \quad & \dot y_1 &= \kappa_2 y_3 - \kappa_3 y_2, \\ \dot\kappa_2 &= \kappa_1\kappa_3 + y_1y_3, \quad & \dot y_2 &= \kappa_1 y_3 - \kappa_3 y_1, \\ \dot\kappa_3 &= -\kappa_1\kappa_2 + y_1y_2, \quad & \dot y_3 &= -\kappa_1 y_2 - \kappa_2 y_1, \end{aligned} \tag{9}$$

which we want to integrate with the condition $\kappa_3(0) = y_3(0) = 0$.

It is easily verified that a constant of the motion is

$$H = y_1y_2y_3 + y_1\kappa_2\kappa_3 + \kappa_1 y_2\kappa_3 - \kappa_1\kappa_2 y_3, \tag{10}$$

which is equal to zero with the considered condition.

From Eq. (10) we thus obtain

$$y_2 = \frac{\kappa_2(\kappa_1 y_3 - \kappa_3 y_1)}{y_1y_3 + \kappa_1\kappa_3}, \quad y_1 = \frac{\kappa_1(y_3\kappa_2 - \kappa_3 y_2)}{y_2y_3 + \kappa_2\kappa_3}. \tag{11}$$

By comparison of Eqs. (11) with Eqs. (9) it follows that

$$\dot\kappa_1/\kappa_1 = \dot y_1/y_1, \quad \dot\kappa_2/\kappa_2 = \dot y_2/y_2,$$

and therefore

$$y_1/\kappa_1 = s_1, \quad y_2/\kappa_2 = s_2; \tag{12}$$

by sutstituting Eqs. (12) into Eq. (10) it follows then that

$$y_3/\kappa_3 = (s_1 + s_2)/(1 - s_1s_2). \tag{13}$$

System (9) reads now

$$\begin{aligned} \dot\kappa_1 &= [1 + s_2(s_1 + s_2)/(1 - s_1s_2)]\kappa_2\kappa_3, \\ \dot\kappa_2 &= [1 + s_1(s_1 + s_2)/(1 - s_1s_2)]\kappa_1\kappa_3, \\ \dot\kappa_3 &= -(1 - s_1s_2)\kappa_1\kappa_2, \end{aligned} \tag{14}$$

and its solutions are the usual Jacobi functions. On the other hand, remembering that $y_i/\kappa_i = \tan\varphi_i$, from Eqs. (12) and (13) we obtain

$$\tan\varphi_3 = \tan(\varphi_1 + \varphi_2),$$

that is,

$$\phi = 0 \quad \text{or} \quad \phi = \pi.$$

It turns out that the phase ϕ is a discontinuous piecewise constant function which takes the value π during the time of growth of u_3 and the value zero during the decrease. Taking into account this behaviour of the phase ϕ, system (3) gives now the correct Eqs. for the evolution of the amplitudes.

1 R. E. Aamodt and W. E. Drummond, Phys. Fluids **7**, 1816 (1964).
2 A. A. Galeev, V. I. Karpman, and R. Z. Sagdeev, Nucl. Fusion **5**, 20 (1965).
3 R. E. Aamodt and W. E. Drummond, J. Nucl. Energy Pt. **C6**, 147 (1964).
4 K. Matsuura, J. Phys. Soc. Japan. **21**, 2011 (1966).
5 D. Montgomery, Physica **31**, 693 (1965).
6 R. Sugihara, Phys. Fluids **11**, 178 (1968).
7 K. P. Das, Phys. Fluids **14**, 124 (1971).
8 E. A. Frieman, J. Math. Phys. **4**, 410 (1963).
9 S. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).

Reduction of the Poincaré Group with Respect to the Lorentz Group*

W. W. MacDowell and Ralph Roskies
Physics Department, Yale University, New Haven, Conn. 06520

(Received 19 April 1972)

The representations of the Poincaré Group for spinless particles, reduced with respect to the Lorentz subgroup, are investigated. They involve the principal series of representations of $SL(2, C)$ and use is made of a basis, introduced by Gel'fand in which the states are labeled by a complex number z. The transformation matrices relating to the Wigner basis are derived. The matrix elements of the momentum operators are obtained. The general form of the S matrix in the new basis is discussed. This basis may be relevant for a field theoretical description of the Veneziano model.

1. INTRODUCTION

In the relativistic quantum theory of massive particles, physical states are generally described by vectors in Fock space, which is the Hilbert space of the direct product of single-particle states. The single-particle states form a basis for unitary irreducible representations of the Poincaré group. The basis most commonly used is the Wigner basis in which the Poincaré group is reduced with respect to the Abelian subgroup of translations. In addition, one operator commuting with the translation subgroup is given to specify the spin, such as in the helicity basis or in the

spinor basis. In principle, however, any basis is suitable for the description of one-particle states. The choice of a particular basis is dictated by its simplicity or convenience in exhibiting some important feature of physical systems. The Wigner basis is particularly useful because, by diagonalizing the generators of translations, one obtains a straightforward description of momentum conservation.

We propose here to investigate a basis which is obtained by reducing the Poincaré group with respect to the Lorentz subgroup. We shall call it the Lorentz basis. Our original motivation for doing this resulted from some work by Domokos *et al.*[1] in which they draw attention to the formal resemblance between the Koba–Nielsen representation[2] for the dual resonance model and irreducible representations of $SL(2C)$, in a basis introduced by Gelfand,[3] defined in terms of a complex number z. It suggests the possibility that the dual resonance model would correspond to a particularly simple form of the S matrix when expressed in this basis. Our efforts in establishing such a result were not successful. However, we have obtained the matrix elements of the momentum operators in this basis which turns out to be of mathematical interest in itself. This problem was, to our knowledge, first discussed by Chakrabarti *et al.*[4] using the (J^2, J_3) basis studied by Joos.[5] However, their results, although essentially in agreement with ours, are presented in a somewhat paradoxical form. They state that, for the case of spin-zero particles, $P_\mu|\lambda, z\rangle$ is a linear combination of states $|\lambda + i, z\rangle$ and $|\lambda - i, z\rangle$. However, Joos has shown that in the decomposition of a representation of the Poincaré group with respect to the Lorentz group only real positive values of λ occur. Thus the states $|\lambda + i, z\rangle$ and $|\lambda - i, z\rangle$ are not in the Hilbert space. The source of the paradox is that P_μ is an unbounded operator, and therefore, is defined only on a subspace of the Hilbert space. We feel that our analysis completely clarifies this question.

Similar problems arise in studying the Lie algebras of unitary representations of noncompact groups when diagonalizing operators which do not generate compact subgroups.[6] Some of our results are contained in Ruhl's work.[7] However, the two approaches are so much different and the subject so unfamiliar that we think the presentation of our methods to be warranted.

Our paper is organized as follows: In Sec. 2 we describe our choice of basis and show how it is related to the usual Wigner basis. For simplicity we confine the discussion to spinless particles. In Sec. 3 we obtain the matrix elements of the momentum operators, in the subspace of the Hilbert space where they are defined, for the principal series of unitary representations of $SL(2, C)$ with continuous eigenvalue λ and discrete eigenvalue $M = 0$. It is found that the momentum operators, considered as distributions, are defined on the set of functions $f_\lambda(z)$ which are analytic in the variable λ in the strip $|\mathrm{Im}\lambda| < 1$, and a characterization of these distributions is given. Another characterization is given in Appendix B. In Sec. 4 we indicate the implications of our results for the S matrix in the Lorentz basis, and in Sec. 5 we give the transformation of the T matrix from the Wigner basis to the Lorentz basis.

2. THE LORENTZ BASIS

Let $J_{\mu\nu}$ be the generators of the Lorentz group. The irreducible representations of $SL(2, C)$ are characterized by two numbers σ and M related to the Casimir operators of the group by

$$\tfrac{1}{2}J_{\mu\nu}J^{\mu\nu} = M^2 + \sigma(\sigma + 2), \tag{2.1}$$

$$(1/4!)g_{\mu\nu\rho\sigma}J^{\mu\nu}J^{\rho\sigma} = M(\sigma + 1). \tag{2.2}$$

Gel'fand[3] has shown that the basis vectors for a unitary irreducible representation of $SL(2, C)$ can be labeled by a complex variable z. If L is an element of $SL(2, C)$ given by

$$L = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha\delta - \beta\gamma = 1, \tag{2.3}$$

and if $D(L)$ is the operator associated with this element, on the space of vectors $|\sigma M; z\rangle$ of an irreducible representation of $SL(2, C)$, then these states transform as

$$D(L^{-1})|\sigma M; z\rangle = (\alpha - \beta z')^{-\sigma-M}(\alpha^* - \beta^* z'^*)^{-\sigma+M}|\sigma M; z'\rangle, \tag{2.4}$$

where

$$z' = (\alpha z + \gamma)/(\beta z + \delta). \tag{2.5}$$

The Poincaré group has generators $J_{\mu\nu}$ of Lorentz transformations and P_μ of translations. The irreducible representations are characterized by the Casimir operators $P^2 = P_\mu P^\mu$ and $W^2 = W_\mu W^\mu$, where $W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\nu\rho}P^\sigma$. For the unitary representations P^2 is a real number. If $P^2 = m^2 > 0$ then $W^2 = -m^2 s(s + 1)$ where s is an integer or half integer. The representation is then associated with a particle of mass m and spin s.

Joos[5] has shown that in the $SL(2, C)$ decomposition of a unitary irreducible representation of the Poincaré group corresponding to a massive particle with zero spin the only representations that occur are the representations in the principal series with $M = 0$ and $\sigma = -1 + i\lambda(\lambda > 0)$. For each λ the representation occurs once and only once. Therefore the basis states shall be labeled as $|\lambda z\rangle$ and are normalized in the following way:

$$\langle\lambda_2 z_2|\lambda_1 z_1\rangle = \delta(\lambda_2 - \lambda_1)\delta^2(z_2 - z_1), \tag{2.6}$$

where $\delta^2(z_2 - z_1) = \delta(\mathrm{Re}z_2 - \mathrm{Re}z_1)\delta(\mathrm{Im}z_2 - \mathrm{Im}z_1)$.

On the other hand the states in the Wigner basis for a representation of particles of mass m and zero spin are labeled by the eigenvalues $\mathbf{p}$ of the momentum operator $\mathbf{P}$, as $|\mathbf{p}\rangle$.

The two bases are connected by the transformation matrix $\langle\mathbf{p}|\lambda z\rangle$. In order to obtain this matrix let us consider a Lorentz transformation L and, taking (2.4) into account, write

$$\langle\mathbf{p}|\lambda z\rangle = \langle\mathbf{p}|D(L^{-1})^{-1}D(L^{-1})|\lambda z\rangle = |\alpha - \beta z'|^{-2\sigma}\langle\mathbf{p}'|\lambda z'\rangle, \tag{2.7}$$

where $\mathbf{p}', z'$ are obtained from $\mathbf{p}$ and z by the Lorentz transformation L^{-1}.

Now we notice that the transformation of z into z', as given by (2.5), is the same as that of the ratio $(-\psi_2/\psi_1)$ where ψ_1, ψ_2 are the components of a spinor ψ in the $\frac{1}{2}$ representation of $SL(2,C)$. Letting $\sigma_\mu = (1, \sigma)$ then $(\psi^\dagger, \sigma_\mu\psi)$ transforms under $SL(2,C)$ as a null 4-vector and if $\phi = \binom{1}{-z}$, then

$$n_\mu = (\phi^\dagger, \sigma_\mu\phi) = (1 + |z|^2, -2\,\mathrm{Re}z, -2\,\mathrm{Im}z, 1 - |z|^2) \tag{2.8}$$

transforms in the following way:

$$(\phi'^\dagger, \sigma_\mu\phi') = |\beta z + \delta|^{-2}\Lambda^{-1}_{\mu\nu}(\phi'^\dagger, \sigma^\nu\phi)$$

or

$$(\phi^\dagger, \sigma^\nu\phi) = |\alpha - \beta z'|^{-2}\Lambda_{\mu\nu}(\phi^\dagger, \sigma^\nu\phi'), \tag{2.9}$$

where $\Lambda_{\mu\nu}$ is the transformation matrix of vectors corresponding to the Lorentz transformation L. Therefore it follows from (2.7) and (2.9) that

$$\langle \mathbf{p}|\lambda z\rangle = ic(\lambda, m)[(p/m)\cdot n]^\sigma, \tag{2.10}$$

where $c(\lambda, m)$ depends only on λ and m and can be chosen to be real. In order to determine $c(\lambda, m)$ we use the normalization condition (2.6) together with the completeness condition

$$\int |\mathbf{p}\rangle\langle\mathbf{p}|\,\frac{d^3p}{2p_0} = 1; \tag{2.11}$$

then

$$c(\lambda_1 m)^2 \int \left(\frac{p}{m}\cdot n_2\right)^{\sigma_2 *}\left(\frac{p}{m}\cdot n_1\right)^{\sigma_1}\frac{d^3p}{2p_0} = \delta(\lambda_2 - \lambda_1)\delta^2(z_2 - z_1) = (2/\pi)\delta(\lambda_2 - \lambda_1)\delta(n_2\cdot n_1). \tag{2.12}$$

One obtains

$$c = \pi^{-3/2}(\lambda/m) \tag{2.13}$$

so that

$$\langle \mathbf{p}|\lambda z\rangle = \pi^{-3/2}\,(i\lambda/m)[(p/m)\cdot n]^\sigma. \tag{2.14}$$

In order to simplify the notation we shall from now on set all the masses equal to $m = 1$.

3. THE MOMENTUM OPERATORS

Since the generators of translations P_μ are unbounded, they are not defined in the whole Hilbert space but only in the subspace of vectors

$$|f\rangle = \int |\mathbf{p}\rangle f(\mathbf{p})\,\frac{d^3p}{2p_0} \tag{3.1}$$

for which $p_0 f(\mathbf{p})$ is square integrable. In the λ, z basis this corresponds to the subset $\mathfrak{F}$ of functions $\tilde{f}(\lambda, z)$ defined by

$$\tilde{f}(\lambda, z) = \int \frac{i\lambda}{\pi^{3/2}}(n\cdot p)^{-1-i\lambda} f(\mathbf{p})\,\frac{d^3p}{2p_0} \tag{3.2}$$

with $p_0 f(\mathbf{p}) \in L^2$ (with measure $d^3p/2p_0$).

One can show[7] that these functions are analytic in the strip $|\mathrm{Im}\lambda| < 1$. These functions also have the following property:

$$\tilde{f}(-\lambda, z) = (i\lambda/\pi)\int \tfrac{1}{2}(n\cdot n')^{-1+i\lambda}\tilde{f}(\lambda, z')d^2z', \tag{3.3}$$

which establishes the equivalence of the representations for values of λ of opposite signs.

Let us consider the matrix element

$$I = \langle \lambda_2 z_2|P_\mu|f\rangle = \int \langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle \tilde{f}(\lambda_1 z_1)d\lambda_1 d^2z_1 = \int \langle \lambda_2 z_2|P_\mu|\mathbf{p}\rangle f(\mathbf{p})\,\frac{d^3p}{2p_0} = \int \frac{-i\lambda_2}{\pi^{3/2}}\,(n_2\cdot p)^{-1-i\lambda_2}p_\mu f(\mathbf{p})\,\frac{d^3p}{2p_0}. \tag{3.4}$$

In Appendix A the following identity is proven:

$$-i\lambda_2(n_2\cdot p)^{-1-i\lambda_2}p_\mu = \tfrac{1}{2}(\lambda_2 - i)\,(n_2\cdot p)^{-1-i(\lambda_2 - i)}n_{2\mu} + (1/2\pi)\int \lambda_2(\tfrac{1}{2}n_1\cdot n_2)^{-1-i\lambda_2}\,n_{1\mu}[(-i-\lambda_2)/\pi^{3/2}] \times (n_1\cdot p)^{-1-i(-\lambda_2 - i)}d^2z_1. \tag{3.5}$$

Substituting in (3.4), one obtains

$$\int \langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle\tilde{f}(\lambda_1, z_1)d\lambda_1 d^2z_1 = \tfrac{1}{2}n_{2\mu}\tilde{f}(\lambda_2 - i, z_2) + (i\lambda_2/2\pi)\int (\tfrac{1}{2}n_1\cdot n_2)^{-1-i\lambda_2}n_{1\mu}\tilde{f}(-\lambda_2 - i, z_1)d^2z_1. \tag{3.6}$$

The λ_1 integration in (3.4) and (3.6) is over the interval $(0, \infty)$. Therefore one can write the following expression for $\langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle$ considered as a distribution on the set of analytic functions $\tilde{f}(\lambda_1, z_1)$:

$$\langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle = (i\lambda_2/2\pi)(\tfrac{1}{2}n_1\cdot n_2)^{-1-i\lambda_2}\delta(\lambda_1 + \lambda_2 + i)n_{1\mu} + \tfrac{1}{2}\delta(\lambda_1 - \lambda_2 + i)\delta(z_2 - z_1)n_{1\mu} \tag{3.7}$$

to be understood in the sense of (3.6).

Similarly, this matrix element considered as a distribution in the space of functions $\tilde{f}(\lambda_2, z_2)$ can be written as

$$\langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle = \tfrac{1}{2}\delta(\lambda_1 - \lambda_2 + i)\delta(z_2 - z_1)n_{2\mu} - (i\lambda_1/2\pi)(\tfrac{1}{2}n_1\cdot n_2)^{-1+i\lambda_1}\delta(\lambda_1 + \lambda_2 - i)n_{2\mu} \tag{3.8}$$

which is the Hermitian conjugate of (3.7).

Let us now consider the matrix element $\langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle$ as a distribution in the topological product of the space of functions $\tilde{f}(\lambda_1 z_1)$ and $\tilde{f}(\lambda_2 z_2)$. We have

$$\int \langle \lambda_2 z_2|P_\mu|\lambda_1 z_1\rangle\tilde{f}_1(\lambda_1 z_1)d\lambda_1 d^2z_1\tilde{f}_2(\lambda_2, z_2)^\dagger d\lambda_2 d^2z_2 = \tfrac{1}{2}\int \tilde{f}_1(\lambda_2 - i, z_2)n_{2\mu}\tilde{f}_2(\lambda_2, z_2)^\dagger d\lambda_2 dz_2 + \int (i\lambda_2/2\pi)(\tfrac{1}{2}n_1\cdot n_2)^{-1-i\lambda_2}n_{1\mu}\tilde{f}(-\lambda_2 - i, z_1) \times d^2z_1\tilde{f}_2(\lambda_2, z_2)^\dagger d\lambda_2 d^2z_2. \tag{3.9}$$

In the second term we have

$$\frac{i\lambda_2}{2\pi}\int\left(\frac{n_1\cdot n_2}{2}\right)^{-1-i\lambda_2}\tilde{f}_2(\lambda_1, z_2)^\dagger d^2z_2 = \frac{i\lambda_2}{2\pi}\int\left(\frac{n_1\cdot n_2}{2}\right)^{-1-i\lambda_2}\frac{-i\lambda_2}{\pi^{3/2}}\,(n_2\cdot p)^{-1+i\lambda_2}f_2(\mathbf{p})^\dagger \times \frac{d^3p}{2p_0}\,d^2z_2 = \tfrac{1}{2}\int \frac{-i\lambda}{\pi^{3/2}}\,(n_1\cdot p)^{-1-i\lambda_2}f_2(\mathbf{p})^\dagger\,\frac{d^3p}{2p_0} = \tfrac{1}{2}\tilde{f}_2(-\lambda_2, z_1)^\dagger. \tag{3.10}$$

The derivation is given in Appendix A. Therefore we can write

$$\langle \lambda_2 z_2 | P_\mu | \lambda_1 z_1 \rangle = \tfrac{1}{2} n_{1\mu} \delta(\lambda_1 - \lambda_2 + i)\delta(z_1 - z_2) \quad (3.11)$$

when considered as a distribution on the topological product of the space of functions $\tilde{f}(\lambda_1, z_1)$ and $\tilde{f}(\lambda_2, z_2)$. In Appendix B we give an alternate representation of the matrix element $\langle \lambda_1 z_1 | P_\mu | \lambda_2 z_2 \rangle$ defined as a limit of distributions with support entirely on the positive real λ axis. We consider the function

$$E(a,n) = \langle \lambda_2 z_2 | e^{-an\cdot P} | \lambda_1 z_1 \rangle, \quad (3.12)$$

where $a > 0$, and n is a null vector with $n_0 > 0$. Since P is timelike $n \cdot P > 0$ so that $e^{-an\cdot P}$ is a bounded operator, therefore, $E(a,n)$ is well defined. We expect to recover the matrix elements of $P\cdot n$ (and therefore P_μ) by formally writing

$$\langle \lambda_2 z_2 | P\cdot n | \lambda_1 z_1 \rangle = \lim_{a\to 0} -\frac{d}{da} E(a,n). \quad (3.13)$$

It turns out that this limit only exists on test functions analytic in the strip $|\mathrm{Im}\lambda| < 1$. An outline of the procedure is given in Appendix B.

4. THE T MATRIX IN THE (λz) BASIS

In this section we consider the matrix element of the T matrix for a scattering process $A_1 + A_2 \to A_3 + A_4$, with the states given in the (λz) basis. From Lorentz invariance one can readily see[1] that the matrix element of the T matrix is of the form

$$\langle \lambda_3 z_3; \lambda_4 z_4 | T | \lambda_1 z_1; \lambda_2 z_2 \rangle = \prod_{\substack{i=1,2\\ j=3,4}} (n_i \cdot n_j)^{(\sigma_i + \sigma_j^*)/2} \times (n_1\cdot n_2)^{-(\sigma_3^* + \sigma_4^*)/2} (n_3 \cdot n_4)^{-(\sigma_1 + \sigma_2)/2} M(\lambda_i, \xi), \quad (4.1)$$

where $\sigma_i = -1 + i\lambda_i$ and

$$\xi = \{[(z_1 - z_3)(z_2 - z_4)]/[(z_1 - z_4)(z_2 - z_2)]\} \quad (4.2)$$

so that

$$|1 - \xi|^2 = \frac{n_1\cdot n_2 n_3 \cdot n_4}{n_1\cdot n_4 n_2 \cdot n_3},$$
$$|1 - \xi^{-1}|^2 = \frac{n_1\cdot n_2 n_3\cdot n_4}{n_1\cdot n_3 n_2\cdot n_4}, \quad (4.3)$$

We want now to find the restrictions on $M(\lambda_i, \xi)$ that come from invariance of the T matrix under translations as given by the condition $[T, P_\mu] = 0$. We have

$$\int_0^\infty \prod_i d\lambda_i \int_0^\infty \prod_i d^2 z_i \, \langle \lambda_3 z_3; \lambda_4 z_4 | [T, P_\mu] | \lambda_1 z_1; \lambda_2 z_2 \rangle \times \tilde{f}_3(\lambda_3 z_3)^\dagger \tilde{f}_4(\lambda_4 z_4)^\dagger \tilde{f}_1(\lambda_1 z_1)\tilde{f}_2(\lambda_2 z_2) = 0. \quad (4.4)$$

From the unitarity of the S matrix we shall now show that the function

$$\begin{aligned}(p_3 + p_4)^2 \langle \mathbf{p}_3; \mathbf{p}_4 | T | f_1; f_2 \rangle &\equiv (p_3 + p_4)^2 \int \langle \mathbf{p}_3; \mathbf{p}_4 | T | \mathbf{p}_1; \mathbf{p}_2 \rangle \times f_1(\mathbf{p}_1) f_2(\mathbf{p}_2) \frac{d^3 p_1}{2p_{10}} \frac{d^3 p_2}{2p_{20}} \\ &\equiv \int \langle \mathbf{p}_3; \mathbf{p}_4 | T | \mathbf{p}_1; \mathbf{p}_2 \rangle (p_1 + p_2)^2 f_1(\mathbf{p}_1) f_2(\mathbf{p}_2) \times \frac{d^3 p_1}{2p_{10}} \frac{d^3 p_2}{2p_{20}} \end{aligned} \quad (4.5)$$

is square integrable in the pair of momenta $\mathbf{p}_3, \mathbf{p}_4$. Since $(p_1 + p_2)^2 = 2(1 + p_1\cdot p_2)$ and $p_{10} f_1(\mathbf{p}_1)$, $p_{20} f_2(\mathbf{p}_2)$ are by hypothesis square integrable, then it follows that

$$|f_{12}\rangle = \int |p_1, p_2\rangle (p_1 + p_2)^2 f_1(\mathbf{p}_1) f_2(\mathbf{p}_2) \frac{d^3 p_1}{2p_{10}} \frac{d^3 p_2}{2p_{20}} \quad (4.6)$$

is a vector in the Hilbert space where T is defined. Therefore, unitarity gives

$$\int |\langle p_3; p_4 | T | f_{12} \rangle|^2 \frac{d^3 p_3}{2p_{30}} \frac{d^3 p_4}{2p_{40}} \le \mathrm{Im}\langle f_{12} | T | f_{12} \rangle < \infty \quad (4.7)$$

which proves the assertion. It follows that $\langle \lambda_3 z_3; \lambda_4 z_4 | T | f_1 f_2 \rangle$ is analytic in the pair of variables λ_3, λ_4 inside the domain $\{|\mathrm{Im}\lambda_3| < 1, |\mathrm{Im}\lambda_4| < 1\}$. Moreover, as in (3.3), the following relation obtains:

$$\langle - \lambda_3 z_3; \lambda_4 z_4 | T | f_1 f_2 \rangle = (i\lambda_3/\pi) \int (\tfrac{1}{2} n_3 \cdot n_3')^{-1 + i\lambda_3} \times \langle \lambda_3 z_3'; \lambda_4 z_4 | T | f_1 f_2 \rangle \, d^2 z_3'. \quad (4.8)$$

Therefore one can use in (4.4) the expression (3.11) for the matrix element of P_μ. One obtains

$$\int \prod_i d\lambda_i d^2 z_i \langle \lambda_3 z_3; \lambda_4 z_4 | T | \lambda_1 z_1; \lambda_2 z_2 \rangle \Big\{ n_{1\mu} \frac{\tilde{f}_1(\lambda_1 - i, z_1)}{\tilde{f}_1(\lambda_1, z_1)} + n_{2\mu} \frac{\tilde{f}_2(\lambda_2 - i, z_2)}{\tilde{f}_2(\lambda_2, z_2)} - n_{3\mu} \frac{\tilde{f}_3(\lambda_3 - i, z_3)^\dagger}{\tilde{f}_3(\lambda_3, z_3)^\dagger} - n_{4\mu} \frac{\tilde{f}_4(\lambda_4 - i, z_4)^\dagger}{\tilde{f}_4(\lambda_4, z_4)^\dagger} \Big\} \tilde{f}_1(\lambda_1 z_1)\tilde{f}_2(\lambda_2 z_2)\tilde{f}_3(\lambda_3, z_3)^\dagger \times \tilde{f}_4(\lambda_4 z_4)^\dagger = 0. \quad (4.9)$$

In this integral each term that contains $n_{i\mu}$ as a factor is integrated in λ_1 from $-\infty$ to $+\infty$ and in the other λ's from 0 to $+\infty$.

Consistently with (4.8), one can define the matrix element of T with some of the λ's negative in terms of matrix elements with all the λ's positive, by expressions like

$$\langle - \lambda_3 z_3; \lambda_4 z_4 | T | \lambda_1 z_1; \lambda_2 z_2 \rangle = (i\lambda_3/\pi) \int (\tfrac{1}{2} n_3 \cdot n_3')^{-1 + i\lambda_3} \times \langle \lambda_3 z_3'; \lambda_4 z_4 | T | \lambda_1 z_1; \lambda_2 z_2 \rangle d^3 z. \quad (4.10)$$

With such a definition one can easily see that the integrations over all the λ's in (4.10) can be taken over the interval $(-\infty, +\infty)$. Replacing (4.1) and (3.2) into (4.9) one obtains

$$\begin{aligned} \int M(\lambda_i, \xi) \prod_{\substack{i=1,2\\ j=3,4}} (n_i\cdot n_j)^{-1 + (i/2)(\lambda_i - \lambda_j)} (n_1\cdot n_2)^{1 + (i/2)(\lambda_3 \lambda_4)} \\ \times (n_3\cdot n_4)^{1 - (i/2)(\lambda_1 + \lambda_2)} \frac{i\lambda_i}{\pi^{3/2}} (n_i\cdot p_i)^{-1 + i\lambda_i} \frac{-i\lambda_j}{\pi^{3/2}} \\ \times (n_j\cdot p_j)^{-1 + i\lambda_j} \tilde{f}_i(\mathbf{p}_i)\tilde{f}_j(\mathbf{p}_j)^\dagger \\ \times \Big(\frac{\lambda_1 - i}{\lambda_1} \frac{n_{1\mu}}{n_1\cdot p_1} + \frac{\lambda_2 - i}{\lambda_2} \frac{n_{2\mu}}{n_2\cdot p_2} - \frac{\lambda_3 + i}{\lambda_3} \frac{n_{3\mu}}{n_3\cdot p_3} \\ - \frac{\lambda_4 + i}{\lambda_4} \frac{n_{4\mu}}{n_4\cdot p_4} \Big) \prod_\kappa d\lambda_\kappa d^2 z_\kappa \frac{d^3 p_\kappa}{2p_{\kappa 0}} = 0. \end{aligned} \quad (4.11)$$

Since the $\tilde{f}(p)$'s form a dense set it follows that

$$\int M(\lambda_i, \xi) \prod_{\substack{i=1,2\\ j=3,4}} (n_i\cdot n_j)^{-1 + (i/2)(\lambda_i - \lambda_j)} (n_1\cdot n_2)^{1 + (i/2)(\lambda_3 + \lambda_4)} \times (n_3\cdot n_4)^{1 - (i/2)(\lambda_1 + \lambda_2)}$$

$$\times \frac{i\lambda_i}{\pi^{3/2}}(n_i\cdot p_i)^{-1-i\lambda_i}\frac{-i\lambda_j}{\pi^{3/2}}(n_j\cdot p_j)^{-1+i\lambda_j}$$
$$\times\left(\frac{\lambda_1-i}{\lambda_1}\frac{n_{1\mu}}{n_1\cdot p_1}+\frac{\lambda_2-i}{\lambda_2}\frac{n_{2\mu}}{n_2\cdot p_2}-\frac{\lambda_3+i}{\lambda_3}\frac{n_{3\mu}}{n_3\cdot p_3}\right.$$
$$\left.-\frac{\lambda_4+i}{\lambda_4}n_4\cdot p_4\right)\prod_\kappa d\lambda_\kappa d^2z_\kappa = \int M(\lambda_i,\xi)$$
$$\times\prod_{\substack{i=1,2\\j=3,4}}(n_i\cdot n_j)^{-1+(i/2)(\lambda_i-\lambda_j)}(n_1\cdot n_2)^{1+(i/2)(\lambda_3+\lambda_4)}$$
$$\times(n_3\cdot n_4)^{1-(i/2)(\lambda_1+\lambda_2)}\left(1+p_{i\rho}\cdot\frac{\partial}{\partial p_{i\rho}}\right)\left(1+p_{j\rho}\cdot\frac{\partial}{\partial p_{j\rho}}\right)$$
$$\times(n_i\cdot p_i)^{1-i\lambda_i}(n_j\cdot p_j)^{1+i\lambda_i}\left(\frac{n_{1\mu}}{n_1\cdot p_1}+\frac{n_{2\mu}}{n_2\cdot p_2}\right.$$
$$\left.-\frac{n_{3\mu}}{n_3\cdot p_3}-\frac{n_{4\mu}}{n_4\cdot p_4}\right)\prod_\kappa\frac{d\lambda_\kappa}{\pi^{3/2}}\frac{d^2z_\kappa}{(n_\kappa\cdot p_\kappa)^2}\Big] = 0. \quad (4.12)$$

(Here we have temporarily treated all four components of p_i as arbitrary.)

Now it is easy to verify that

$$\frac{d^2z}{(n\cdot p)^2} = \frac{d^3N}{4N_0}\,\delta(N\cdot p-1), \quad (4.13)$$

where N is a null vector such that $n\cdot N=0$ and $N_0>0$. We also have

$$\left(1+p_\rho\cdot\frac{\partial}{\partial p_\rho}\right)\delta(N\cdot p-1)=\delta'(N\cdot p-1). \quad (4.14)$$

Therefore (4.10) can be written in the following way:

$$\int M(\lambda_i,\xi)\prod_{\substack{i=1,2\\j=3,4}}(N_i\cdot N_j)^{-1+(i/2)(\lambda_i-\lambda_j)}(N_1\cdot N_2)^{1+(i/2)(\lambda_3+\lambda_4)}$$
$$\times(N_3\cdot N_4)^{1-(i/2)(\lambda_1+\lambda_2)}\,[N_{1\mu}+N_{2\mu}-N_{3\mu}-N_{4\mu}]$$
$$\times\prod_\kappa\frac{d\lambda_\kappa}{\pi^{3/2}}\frac{d^3N_\kappa}{4N_{\kappa0}}\,\delta'(N_\kappa\cdot p_\kappa-1)=0, \quad (4.15)$$

where ξ is defined in terms of the N's by homogeneous expressions like (4.3).

Let us introduce the variables u_i by

$$\begin{aligned} u_i &= \tfrac{1}{2}\ln(N_i\cdot N_3 N_i\cdot N_4)/N_3\cdot N_4 \quad i=1,2,\\ u_i &= \tfrac{1}{2}\ln(N_i\cdot N_1 N_i\cdot N_2)/N_1\cdot N_2 \quad i=3,4. \end{aligned} \quad (4.16)$$

Then performing the λ integrations in (4.15) one obtains

$$\int \mathfrak{M}(N_i)[N_{1\mu}+N_{2\mu}-N_{3\mu}-N_{4\mu}]\prod_\kappa\frac{d^3N_\kappa}{4N_{\kappa0}}\,\delta'(N_\kappa\cdot p_\kappa-1)=0, \quad (4.17)$$

where

$$\mathfrak{M}(N_i)=\frac{|1-\xi|^2|1-\xi^{-1}|^2}{N_1\cdot N_2 N_3\cdot N_4}\int_{-\infty}^{+\infty}M(\lambda_i,\xi)\prod_\kappa e^{iu_\kappa\lambda_\kappa}\times\frac{d\lambda_\kappa}{\pi^{3/2}}. \quad (4.18)$$

One obvious solution to (4.17) is

$$\mathfrak{M}(N_i)=\mathfrak{M}_0(N_i)\delta(N_1+N_2-N_3-N_4). \quad (4.19)$$

The condition

$$N_1+N_2-N_3-N_4=0 \quad (4.20)$$

implies

$$N_1\cdot N_2=N_3\cdot N_4,\quad N_1\cdot N_3=N_2\cdot N_4,\quad N_1\cdot N_4=N_2\cdot N_3, \quad (4.21)$$

and the coplanarity condition is equivalent to $\mathrm{Im}\xi=0$. The solutions of (4.20) with $N_{i0}>0$ restrict the values of ξ to the interval $(-\infty,0)$. The necessary and sufficient condition for $\mathfrak{M}(N_i)$ as given by (4.18) to be of the form (4.19) is that

$$M(\lambda_i,\xi)\prod_\kappa(\lambda_\kappa/\pi^{3/2})=M_0(\lambda,\xi)\delta(\mathrm{Im}\xi), \quad (4.22)$$

where $\lambda=\lambda_1+\lambda_2-\lambda_3-\lambda_4$.

But if $M(\lambda_i,\xi)$ has this form, the T matrix does not satisfy the constraint condition (4.10)! We have been unable to find solutions to (4.17) satisfying the constraint condition. However, the problem of determining the general form of the T matrix in the (λ,z) basis, under transformations of the Poincaré group, can be dealt with directly by transforming the T matrix in the Wigner basis to the (λ,z) basis. This is done in the next section.

5. TRANSFORMATION OF THE T MATRIX FROM THE WIGNER BASIS TO THE (λ,z) BASIS

The relation between the T matrix in the Wigner basis and in the (λ,z) basis for zero-spin particles (and unit mass) is

$$\langle\cdots\lambda_jz_j\cdots|T|\cdots\lambda_iz_i\cdots\rangle=\int\langle\cdots p_j\cdots|T|\cdots p_i\cdots\rangle\times\prod_i\frac{i\lambda_i}{\pi^{3/2}}(n_i\cdot p_i)^{-1+i\lambda_i}\prod_j\frac{-i\lambda_j}{\pi^{3/2}}(n_j\cdot p_j)^{-1-i\lambda_j}\prod_\kappa\frac{d^3p_\kappa}{2p_{\kappa0}}. \quad (5.1)$$

We shall take this relation to apply to both positive as well as negative values of λ, so that the conditions such as (4.10) hold.

In the case of scattering of two initial into two final particles we have

$$\langle p_3p_4|T|p_1p_2\rangle=F(s,t)\delta^4(p_1+p_2-p_3-p_4), \quad (5.2)$$

where s and t are the usual Mandelstam variables. After inserting (5.2) into (5.1) we have been unable to express the left-hand side as a two-dimensional transform of $F(s,t)$.

In order to obtain the transformation to the basis in terms of the vectors N_i introduced in Sec. 4 we multiply (5.1) by $\Pi_{i,j}\eta_i^{1-i\lambda_i}\eta_j^{1-i\lambda_j}$ and integrate in the λ's to obtain

$$\begin{aligned}\langle\cdots N_j\cdots|T|\cdots N_i\cdots\rangle &= \int\langle\cdots p_j\cdots|T|\cdots p_i\cdots\rangle\\ &\quad\times\prod_{i,j}\frac{i\lambda_i}{\pi^{3/2}}(N_i\cdot p_i)^{-1+i\lambda_i}\frac{-i\lambda_j}{\pi^{3/2}}(N_j\cdot p_j)\prod_\kappa\frac{d^3p_\kappa}{2p_{\kappa0}}\,d\lambda_\kappa\\ &= \left(\frac{4}{\pi}\right)^{\nu/2}\int\langle\cdots p_j\cdots|T|\cdots p_i\cdots\rangle\\ &\quad\times\prod_\kappa\delta'(N_\kappa\cdot p_\kappa-1)\frac{d^3p_\kappa}{2p_{\kappa0}}, \end{aligned} \quad (5.3)$$

where $N_\kappa=n_\kappa/\eta_\kappa$ and ν is the sum of the number of particles in the initial and final states. Since

$$\int\delta'(N\cdot p-1)\delta'(N\cdot p'-1)\frac{d^3N}{2N_0}=\pi^2\,2p_0\,\delta(\mathbf{p}-\mathbf{p}'), \quad (5.4)$$

then we also have

$$\langle\cdots p_j\cdots|T|\cdots p_i\cdots\rangle=(4\pi^3)^{-\nu/2}\int\langle\cdots N_j\cdots|T|\cdots N_i\cdots\rangle\times\prod_\kappa\delta'(N_\kappa\cdot p_\kappa-1)\frac{d^3N_\kappa}{2N_{\kappa0}}. \quad (5.5)$$

Upon the substitution of (5.2) into (4.18) and using the relation

$$\int \delta'(N\cdot p-1)\,\delta'(N\cdot p'-1)N_\mu \frac{d^3N}{2N_0} = \pi^2 2p_0\delta(\mathbf{p}-\mathbf{p}')p_\mu, \tag{5.6}$$

it results that (4.17) is identically satisfied.

APPENDIX A: EVALUATION OF CERTAIN INTEGRALS

Consider the expression

$$\int (\tfrac{1}{2}n_2\cdot n_1)^{-1-i\lambda_2}\, n_{1\mu}(n_1\cdot p)^{-2+i\lambda_2} d^2z_1. \tag{A1}$$

From its transformation properties under the group $SL(2,C)$ and taking into account that $n_2^2 = 0$ and $p^2 = 1$, one can write

$$\int (\tfrac{1}{2}n_2\cdot n_1)^{-1-i\lambda_2}\, n_{1\mu}(n_1\cdot p)^{-2+i\lambda_2} d^2z_1 = a(\lambda_2)(n_2\cdot p)^{-1-i\lambda_2}\, p_\mu + b(\lambda_2)(n_2\cdot p)^{-2-i\lambda_2}\, n_{2\mu}, \tag{A2}$$

where the coefficients a and b depend on λ_2 only. Taking the scalar product of both sides with n_2 and p, respectively, we obtain

$$a(n_2\cdot p)^{-i\lambda_2} = 2\int (\tfrac{1}{2}n_2\cdot n_1)^{-i\lambda_2}\,(n_1\cdot p)^{-2+i\lambda_2} d^2z_1 \tag{A3}$$

$$\times\,(a+b)(n_2\cdot p)^{-1-i\lambda_2} = \int \tfrac{1}{2}(n_2\cdot n_1)^{-1-i\lambda_2}\,(n_1\cdot p)^{-1+i\lambda_2} d^2z_1. \tag{A4}$$

Letting $n_2 = (1,0,0,-1)$ and $p = (1,0,0,0)$ one obtains

$$a = 2\pi\int_0^\infty (1+\rho^2)^{-2+i\lambda_2} d\rho^2 = -\frac{2\pi}{-1+i\lambda_2}, \tag{A5}$$

$$a+b = \pi\int_0^\infty (1+\rho^2)^{-1+i\lambda_2} d\rho^2 = -\frac{\pi}{i\lambda_2}. \tag{A6}$$

Then

$$b = \pi\{(1+i\lambda_2)/[i\lambda_2(-1+i\lambda_2)]\}. \tag{A7}$$

Taking (A5) and (A7) into (A2) one obtains the identity (3.5). The derivation of (3.10) can be done in a similar manner.

APPENDIX B

We wish to calculate

$$E(a,n) = \langle \lambda_2 z_2 | e^{-aP\cdot n} | \lambda_1 z_1\rangle, \tag{B1}$$

where $n^2 = 0$, $n_0 > 0$, $a > 0$. Via (2.14) (with $m = 1$) and (2.11) one obtains

$$E(a,n) = \int \frac{d^3p}{2p_0}\,\frac{\lambda_1\lambda_2}{\pi^3}\,(n_1\cdot p)^{-1+i\lambda_1}(n_2\cdot p)^{-1-i\lambda_2} e^{-ap\cdot n}. \tag{B2}$$

By using the integral representation

$$(p\cdot n)^\sigma = \frac{1}{\Gamma(-\sigma)}\int_0^\infty d\alpha e^{-\alpha p\cdot n}\alpha^{-1-\sigma}, \tag{B3}$$

one gets

$$E(a,n) = \frac{\lambda_1\lambda_2}{\pi^3}\,\frac{1}{\Gamma(1-i\lambda_1)\Gamma(1+i\lambda_2)}\int_0^\infty d\alpha_1\alpha_1^{-i\lambda_1} \times \int_0^\infty d\alpha_2\alpha_2^{-\lambda_2}\int \frac{d^3p}{2p_0}\, e^{-p\cdot(an+\alpha_1n_1+\alpha_2n)}. \tag{B4}$$

One can perform the last integral by noticing that

$$\beta = an + \alpha_1 n_1 + \alpha_2 n_2 \tag{B5}$$

is a timelike vector with positive time component. Transforming β to its "rest frame," we have[8] (recalling that $p^2 = 1$)

$$\int \frac{d^3p}{2p_0}\, e^{-p_0x} = 2\pi K_1\frac{(x)}{x}, \tag{B6}$$

where $x = \sqrt{\beta^2}$ and K_1 is the modified Bessel function. But[8]

$$\frac{K_1(x)}{x} = \tfrac{1}{2}\int_0^\infty dt\, e^{-(t^{-1}+x^2t)/2}, \tag{B7}$$

so that we have (recall that n, n_1, n_2 are lightlike)

$$E(a,n) = \frac{1}{\pi^2}\,\frac{\lambda_1}{\Gamma(1-i\lambda_1)}\,\frac{\lambda_2}{\Gamma(1+i\lambda_2)}\int_0^\infty d\alpha_1\alpha_1^{-i\lambda_1} \times\int_0^\infty d\alpha_2\alpha_2^{i\lambda_2}\int_0^\infty dt\, e^{-t-1/2}\, e^{-t[a\alpha_1 n\cdot n_1 + a\alpha_2 n\cdot n_2 + \alpha_1\alpha_2 n_1\cdot n_2]}. \tag{B8}$$

Let us write

$$n\cdot n_1\alpha_1 = \rho e^\kappa, \quad n\cdot n_2\alpha_2 = \rho e^{-\kappa}, \quad t = \tau/\rho. \tag{B9}$$

Then

$$E(a,n) = \frac{2}{\pi^2}\,\frac{\lambda_1(n\cdot n_1)^{-1+i\lambda_1}}{\Gamma(1-i\lambda_1)}\,\frac{\lambda_2(n\cdot n_2)^{-1-i\lambda_2}}{\Gamma(1+i\lambda_2)} \times\int_0^\infty d\tau\int_{-\infty}^\infty d\kappa\, e^{-i(\lambda_1+\lambda_2)\kappa}\, e^{-2\tau a\cosh\kappa} \times\int_0^\infty d\rho\,\rho^{i(\lambda_2-\lambda_1)} \times\exp\left[-\rho\left(\frac{1}{2\tau}+\tau\frac{n_1\cdot n_2}{n\cdot n_1 n\cdot n_2}\right)\right]. \tag{B10}$$

We now do the ρ and κ integrals to get[9]

$$E(a,n) = \frac{4}{\pi^2}\,\frac{\lambda_1}{\Gamma(1-i\lambda_1)}\,\frac{\lambda_2}{\Gamma(1+i\lambda_2)}(n\cdot n_1)^{-1+i\lambda_1} \times(n\cdot n_2)^{-1-i\lambda_2}\,\Gamma(1+i(\lambda_2-\lambda_1))\, I(\gamma), \tag{B11}$$

where

$$I(\gamma) = \int_0^\infty d\tau\left(\frac{1}{2\tau}+\tau\gamma\right)^{-[1+i(\lambda_2-\lambda_1)]}K_{i(\lambda_1+\lambda_2)}(2\tau a) \tag{B12}$$

and

$$\gamma = \{(n_1\cdot n_2)/[(n\cdot n_1)\,(n\cdot n_2)]\}. \tag{B13}$$

If $\gamma \neq 0$, then the integral can be performed to obtain[10]

$$\begin{aligned} I(\gamma) = {}& \gamma^{-1-i\lambda_2+i\lambda_1}\{(2\gamma)^{i(\lambda_2-\lambda_1)/2}/[4\Gamma(1+i\lambda_2-i\lambda_1)]\} \\ & \times[(8\gamma)^{-i(\lambda_1+\lambda_2)/2}\Gamma(-i(\lambda_1+\lambda_2)) \\ & \times\Gamma(1+i\lambda_2)\Gamma(-i\lambda_1)(2a)^{i(\lambda_1+\lambda_2)}\,{}_1F_2(1+i\lambda_2; \\ & 1+i\lambda_1, 1+i(\lambda_1+\lambda_2); -a^2/2\gamma) \\ & +(\lambda_1\longleftrightarrow-\lambda_2)\,(\lambda_2\longleftrightarrow-\lambda_1)] \\ & +2^{-2-i(\lambda_2-\lambda_1)}\Gamma(-i\lambda_2)\Gamma(i\lambda_1)(2a)^{i(\lambda_2-\lambda_1)} \\ & \times{}_1F_2(1+i(\lambda_2-\lambda_1); 1-i\lambda_1, 1+i\lambda_2, -a^2/2\gamma))\}, \end{aligned} \tag{B14}$$

where ${}_1F_2$ is a generalized hypergeometric function.

We are really interested only in $-(d/da)E(a,n)\,|_{a=0}$, which is the matrix element of $(P\cdot n)$. If we retain only those terms which contribute in the limit $a \to 0$ after differentiating, we obtain the first three terms of (B18)

There is still another term to consider, which arises when $\gamma = 0$. This term is singular, so to study it, we consider

$$I_1 = \int_0^\epsilon I(\gamma)d\gamma. \tag{B15}$$

Substituting (B12) for $I(\gamma)$, doing the γ integral first, and then doing the τ integral as we did before, we can evaluate I_1. Taking $-d/da$ of the resulting expression and keeping only the surviving terms in a, we then let ϵ tend to zero. We find a term which survives in the matrix element of $P\cdot n$. It is

$$(a^{-1-i(\lambda_2-\lambda_1)}/4\pi^2)\delta_+(\gamma)\tfrac{1}{2}(n_1\cdot n_2)^{-2+i\lambda_1-i\lambda_2}\,\Gamma(1+i\lambda_2-i\lambda_1), \tag{B16}$$

where $\delta_+(\gamma)$ is defined by

$$\int_0^\epsilon \delta_+(\gamma)f(\gamma)d\gamma = f(0).$$

But

$$\gamma = n_1\cdot n_2/n_1\cdot nn_2\cdot n = 2|z_1 - z_2|^2/n\cdot n_1 n\cdot n_2$$

so that

$$\delta_+(\gamma) = \tfrac{1}{2}n\cdot n_1 n\cdot n_2\delta_+(|z_1 - z_2|^2) = \tfrac{1}{2}\pi n\cdot n_1 n\cdot n_2\delta(z_1 - z_2), \tag{B17}$$

where the last step is obtained by using polar coordinates. Collecting all the terms we than obtain the following expression for the matrix element of $P\cdot n$:

$$\begin{aligned}\langle\lambda_2 z_2|P\cdot n|\lambda_1 z_1\rangle = \frac{i}{\pi^2}\lim_{a\to 0}\,[&2^{-i\lambda_1}\Gamma(1-i(\lambda_2+\lambda_1))\\ &\times\lambda_2(n\cdot n_1)^{i(\lambda_2+\lambda_1)}(n_1\cdot n_2)^{-1-i\lambda_2}a^{-1+i(\lambda_2+\lambda_1)}\\ &-2^{i\lambda_2}\,\Gamma(1+i(\lambda_2+\lambda_1))\lambda_1(n\cdot n_2)^{-i(\lambda_1+\lambda_2)}\\ &\times(n_1\cdot n_2)^{-1+i\lambda_1}a^{-1-i(\lambda_2+\lambda_1)}+(\lambda_1-\lambda_2)\\ &\times\Gamma(1-i(\lambda_2-\lambda_1))\frac{\Gamma(1+i\lambda_1)}{\Gamma(1-i\lambda_1)}\frac{\Gamma(1-i\lambda_2)}{\Gamma(1+i\lambda_2)}\\ &\times(n\cdot n_1)^{i\lambda_2}(n\cdot n_2)^{-i\lambda_1}(n_1\cdot n_2)^{-1-i(\lambda_2-\lambda_1)}a^{-1+i(\lambda_2-\lambda_1)}\\ &-i\pi 2^{-1+i(\lambda_2-\lambda_1)}\Gamma(1+i(\lambda_2-\lambda_1))(n\cdot n_2)^{-i(\lambda_2-\lambda_1)}\\ &\times\delta(z_2-z_1)a^{-1-i(\lambda_2-\lambda_1)}]\end{aligned} \tag{B18}$$

To make sense of the limit $a\to 0$ we should consider the right-hand side of (B18) as a distribution in either set of variables λ_1, z_1 or λ_2, z_2. We have then to find the set of test functions on which this distribution is defined. Let us consider as test functions the set of functions $\tilde{f}(\lambda_1 z_1)$ defined by (3.2). Then on performing the z_1 integration one finds that the integrand becomes a symmetric function of λ_1 analytic in the strip $|\mathrm{Im}\lambda_1| < 1$. It is then possible to extend the integration in λ_1 to the interval $(-\infty, \infty)$ and in each of the four terms move the contour to the line $\mathrm{Im}\lambda_1 = \pm 1$. Then the contribution to the λ_1 integral in the limit $a\to 0$ comes in each term, only from the region around the pole of the Γ function whose argument exactly coincides with the exponent of a^{-1}. One then obtains the result given by (3.7).

APPENDIX C: DETERMINATION OF $c(\lambda, m)$ IN (2.10).

We have to evaluate the integral in (2.12):

$$I = \int\left(\frac{p}{m}\cdot n_2\right)^{-1-i\lambda_2}\left(\frac{p}{m}\cdot n_1\right)^{-1+i\lambda_2}\frac{d^3p}{2p_0}. \tag{C1}$$

From transformation under $SL(2,C)$ and dilation of the null vectors n_i, it follows that I is of the form

$$I = c_1(\lambda_1)\delta(\lambda_1-\lambda_2)\delta(z_1-z_2) + c_2(\lambda_2)\delta(\lambda_1+\lambda_2)\times(n_1\cdot n_2)^{-1+i\lambda_1}, \tag{C2}$$

where the second term vanishes if the λ's are positive definite. In order to calculate the coefficient $c_1(\lambda_1)$ let us consider

$$J(p) = \int[(p/m)\cdot n_2]^{-1-i\lambda_2}\,(\tfrac{1}{2}a\cdot n_2)^{-1+i\lambda_2}d^2z_2, \tag{C3}$$

where a is a null vector and $a_0 > 0$. Because of invariance under $SL(2,C)$, $J(p)$ is of the form

$$J(p) = c_3(\lambda_2)[(p/m)\cdot a]^{-1+i\lambda_2}. \tag{C4}$$

Take

$$a = (1,0,0,-1), \quad p = (p_0,0,0,p),$$

then

$$J = \int_0^\infty[(1/m)(p_0-p) + (1/m)(p_0+p^2]^{-1-i\lambda_2} = \frac{\pi}{i\lambda_2}\,(1/m)(p_0+p))]^{-1+i\lambda_2}. \tag{C5}$$

Therefore $c_3(\lambda_2) = \pi/i\lambda_2$ and

$$J(p) = (\pi/i\lambda_2)[(p_0+p_z)/m]^{-1+i\lambda_2}. \tag{C6}$$

Then from (1), (2), and (3) one can write

$$\begin{aligned}\int\left(\frac{p}{m}\cdot n_1\right)^{-1+i\lambda_1}\frac{\pi}{i\lambda_2}\left(\frac{p_0+p_z}{m}\right)^{-1+i\lambda_2}\frac{d^3p}{zp_0}&\\ = c_1(\lambda_1)\delta(\lambda_1-\lambda_2) + c_2(\lambda_2)\delta(\lambda_1+\lambda_2)&\\ \times\int(n_1\cdot n_2)^{-1-i\lambda_2}\,(\tfrac{1}{2}a\cdot n_2)^{-1+i\lambda_2}d^2z.&\end{aligned} \tag{C7}$$

It is easy to see that the second term vanishes. In order to determine $c_1(\lambda_1)$, let us take $n_1 = (1,0,0,1)$; then

$$c_1(\lambda_1)\delta(\lambda_1-\lambda_2) = \int\left(\frac{p_0-p_z}{m}\right)^{-1-i\lambda_1}\frac{\pi}{i\lambda_2}\left(\frac{p_0+p_z}{m}\right)^{-1+i\lambda_2}\times\frac{d^3p}{2p_0}. \tag{C8}$$

Write

$$p_0 = m\rho c\mathrm{sh}\alpha, \quad p_z = m\rho\mathrm{sh}\alpha, \quad p_y = m\sqrt{\rho^2-1}\,\sin\psi,$$

$$p_x = m\sqrt{\rho^2-1}\,\cos\psi, \quad \frac{d^3p}{2p_0} = \tfrac{1}{2}dp_0dp_zd\psi = \tfrac{1}{2}m^2\rho d\rho d\alpha d\psi;$$

then

$$c_1(\lambda_1)\delta(\lambda_1-\lambda_2) = \frac{\pi^2m^2}{i\lambda_2}\int_1^\infty\rho^{-1+i(\lambda_1+\lambda_2)}d\rho \times\int_{-\infty}^{+\infty}e^{i\alpha(\lambda_2-\lambda_1)}d\alpha = \frac{\pi^3m^2}{\lambda_1^2}\,\delta(\lambda_2-\lambda_1). \tag{C9}$$

Therefore, $c(\lambda,m) = c_1(\lambda,m)^{-1/2} = (1/\pi^{3/2})(\lambda/m)$.

---

* Research (Report Yale 3075-5) supported by the U.S. Atomic Energy Commission under contract AT(11-1) 3075.

1 G. Domokos, S. Kovesi-Domokos, and E. Schonberg, Phys. Rev. D **2**, 1026 (1970).

2 Z. Koba and H. B. Nielsen, Nucl. Phys. **B10**, 633 (69); **B12** 517 (1969).

3 I. Gel'fand *et al*, *Generalized Functions* (Academic, New York, 1966), Vol. V, Chaps. III, IV.

4 A. Chakrabarti, M. Levy-Nahas, and R. Seneor, J. Math. Phys. **9**, 1274 (1968).

5 H. Joos, Fortschr. Physik **10**, 65 (1962).

6 C. Itzykson, J. Math. Phys. **10**, 1109 (1969). We wish to thank H. Abarbanel for bringing this paper to our attention.

7 W. Ruhl, Nuovo Cimento **63A**, 1131 (1969).

8 *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-

Hill, New York, 1953), Vol. II, p. 82.
[9] *Integral Transforms*, edited by A. Erdelyi (McGraw-Hill, New York, 1954), Vol. I, p. 35.
[10] Reference 9, Vol. II, p. 128.

Algebraic Solution for a Dirac Electron in a Plane-Wave Electromagnetic Field*

Brian Beers and H. H. Nickle
Department of Physics, Southern Illinois University at Carbondale, Carbondale, Illinois 62901
(Received 20 March 1972)

An algebraic classification is given for the solutions of the Dirac equation for an electron interacting with a classical plane-wave electromagnetic field. The solutions appear as the carrier space of the direct sum of the positive and negative energy, mass m, spin-$\frac{1}{2}$ representations of the restricted Poincaré group. An explicit construction is given for the generators of the representation. The explicit position space form of the solutions follows readily from the relatively simple form of these operators. Via these solutions, an expression for the propagator of the interacting electron is given.

1. INTRODUCTION

We have recently shown[1] that the solutions of certain quantum-mechanical problems involving electromagnetic fields can be deduced by an algebraic technique involving the generators of a realization of the restricted Poincaré group. In this note we show that the technique is not restricted to the Klein–Gordon equation, but can equally well be used to obtain solutions of the Dirac equation. The procedure is essentially identical to that given in I, the only complication being the introduction of spin, that is, the requisite algebra becomes somewhat more cumbersome due to the presence of the anticommuting gamma matrices. To make the connection with I transparent, we again confine our attention to classical plane-wave electromagnetic radiation. This has the additional merit of providing a simple algebraic interpretation of the solutions to this problem which were first given a long time ago by Volkov.[2]

Since we have given a detailed discussion of the necessary background in I, here we restrict ourselves to a few brief remarks concerning the technique. Simply stated, the method consists of finding a set of operators which commute with the "Hamiltonian" of the problem and which also obey the commutation relations of the generators of the Poincaré group. These operators are then used to classify the solutions according to their transformation properties under this "new" Poincaré group. That is, the solution space appears as a carrier space for a representation of the restricted Poincaré group $P_+^\uparrow$, the previously determined operators being the generators of this representation. Since the irreducible representations of $P_+^\uparrow$ have been completely classified,[3,4] and we have an explicit representation in hand, we are able to classify which irreducible representations occur. The space of solutions can then be constructed by using the "boost" technique.[3] The explicit position space form of the solutions is easily determined due to the relatively simple form of the operators which occur.

The solution presented in Sec. 2 for the interacting electron is identical in form to the following "solution" of the free Dirac equation. The "Hamiltonian" for the problem is given by $H_0 = \not{p}$.[5] The eigenvalue problem is to determine the solutions of the wave equation $H_0\psi = m\psi$. Each of the usual space–time translation operators $P^\mu = i(\partial/\partial x_\mu)$ commutes with H_0. The Lorentz generators for the "orbital" motion are given by $l^{\mu\nu} = i(x^\mu P^\nu - x^\nu P^\mu)$, while the Lorentz generators for the "spin" are given by $(i/2)\sigma^{\mu\nu} = -\frac{1}{4}[\gamma^\mu, \gamma^\nu]$. Although neither of the two sets of quantities $l^{\mu\nu}$, $(i/2)\sigma^{\mu\nu}$ commute with H_0, the Lorentz generators $M_0^{\mu\nu}$ for the "total" motion do, where $M_0^{\mu\nu} = l^{\mu\nu} + (i/2)\sigma^{\mu\nu}$.

Each of the sets of operators $l^{\mu\nu}$, $(i/2)\sigma^{\mu\nu}$, and $M_0^{\mu\nu}$ obey the commutation relations satisfied by the generators of the restricted Lorentz group $L_+^\uparrow$.[6] Since $(i/2)\sigma^{\mu\nu}$ commutes with both $l^{\mu\nu}$ and P^λ, it is easy to verify that the set of operators P^λ, $M_0^{\mu\nu}$ obey the commutation relations of the generators of $P_+^\uparrow$. The representation of $P_+^\uparrow$ which is generated by P^λ, $M_0^{\mu\nu}$ can be classified by calculating the two Casimir operators of the representation. The invariant mass operator is given by $P^\mu P_\mu = \not{P}\not{P} = m^2$, indicating that we must have a mass m representation. The other independent Casimir operator is provided by the square of the Pauli–Lubanski vector S_μ, where $S_\mu = (1/2)\epsilon_{\mu\nu\lambda\sigma} M_0^{\nu\lambda} P^\sigma$. Direct calculation shows that $S_\mu S^\mu = \frac{3}{4}m^2$. Since the eigenvalues of S^2 have the form $J(J+1)m^2$, where $J(J+1)$ is the eigenvalue of the Casimir operator $\mathbf{J}^2$ of the little group $SU(2)$, it follows that the solutions are labeled by the spin $\frac{1}{2}$ representation of $SU(2)$ (that is, $J = \frac{1}{2}$). The sign of the energy provides an additional label for irreducible representations. Since we wish to admit both positive and negative energy states in the Dirac theory, it follows that the solution space of $H_0\psi = m\psi$ is the carrier space for the direct sum of the positive and negative energy, spin $\frac{1}{2}$, mass m representations of $P_+^\uparrow$.

In the spin basis, the abstract vectors of the representation are labeled as $|p, \lambda, \pm\rangle$, where p^μ denotes the eigenvalue of the operator P^μ, $\lambda = \pm\frac{1}{2}$ refers to the component of the spin along the three-axis in the rest frame, and $\pm$ refers to the sign of the energy. The explicit form of the solutions is immediately evident due to the relative simplicity of the operators P^λ and $M_0^{\mu\nu}$. Since the construction of these states is, of course, well known,[5,6] we do not repeat the procedure here. In Sec. 3 we will mimic this technique in order to construct explicit solutions for the interacting Dirac electron problem.

The solution to the problem of a Dirac electron interacting with a plane-wave electromagnetic field will proceed in a fashion which is formally identical to the above discussion. In Sec. 2 we find the exact analog of each of the operators appearing in the free electron case. These provide the generators of a representation of $P_+^\uparrow$, which again turns out to be the direct sum of the forward and backward timelike, mass m, spin $\frac{1}{2}$ representations. In Sec. 3, this exact analogy is exploited in order to explicitly construct

Hill, New York, 1953), Vol. II, p. 82.
[9] *Integral Transforms*, edited by A. Erdelyi (McGraw-Hill, New York, 1954), Vol. I, p. 35.
[10] Reference 9, Vol. II, p. 128.

Algebraic Solution for a Dirac Electron in a Plane-Wave Electromagnetic Field*

Brian Beers and H. H. Nickle
Department of Physics, Southern Illinois University at Carbondale, Carbondale, Illinois 62901
(Received 20 March 1972)

An algebraic classification is given for the solutions of the Dirac equation for an electron interacting with a classical plane-wave electromagnetic field. The solutions appear as the carrier space of the direct sum of the positive and negative energy, mass m, spin-$\frac{1}{2}$ representations of the restricted Poincaré group. An explicit construction is given for the generators of the representation. The explicit position space form of the solutions follows readily from the relatively simple form of these operators. Via these solutions, an expression for the propagator of the interacting electron is given.

1. INTRODUCTION

We have recently shown[1] that the solutions of certain quantum-mechanical problems involving electromagnetic fields can be deduced by an algebraic technique involving the generators of a realization of the restricted Poincaré group. In this note we show that the technique is not restricted to the Klein–Gordon equation, but can equally well be used to obtain solutions of the Dirac equation. The procedure is essentially identical to that given in I, the only complication being the introduction of spin, that is, the requisite algebra becomes somewhat more cumbersome due to the presence of the anticommuting gamma matrices. To make the connection with I transparent, we again confine our attention to classical plane-wave electromagnetic radiation. This has the additional merit of providing a simple algebraic interpretation of the solutions to this problem which were first given a long time ago by Volkov.[2]

Since we have given a detailed discussion of the necessary background in I, here we restrict ourselves to a few brief remarks concerning the technique. Simply stated, the method consists of finding a set of operators which commute with the "Hamiltonian" of the problem and which also obey the commutation relations of the generators of the Poincaré group. These operators are then used to classify the solutions according to their transformation properties under this "new" Poincaré group. That is, the solution space appears as a carrier space for a representation of the restricted Poincaré group $P_+^\uparrow$, the previously determined operators being the generators of this representation. Since the irreducible representations of $P_+^\uparrow$ have been completely classified,[3,4] and we have an explicit representation in hand, we are able to classify which irreducible representations occur. The space of solutions can then be constructed by using the "boost" technique.[3] The explicit position space form of the solutions is easily determined due to the relatively simple form of the operators which occur.

The solution presented in Sec. 2 for the interacting electron is identical in form to the following "solution" of the free Dirac equation. The "Hamiltonian" for the problem is given by $H_0 = \not{p}$.[5] The eigenvalue problem is to determine the solutions of the wave equation $H_0\psi = m\psi$. Each of the usual space–time translation operators $P^\mu = i(\partial/\partial x_\mu)$ commutes with H_0. The Lorentz generators for the "orbital" motion are given by $l^{\mu\nu} = i(x^\mu P^\nu - x^\nu P^\mu)$, while the Lorentz generators for the "spin" are given by $(i/2)\sigma^{\mu\nu} = -\frac{1}{4}[\gamma^\mu, \gamma^\nu]$. Although neither of the two sets of quantities $l^{\mu\nu}$, $(i/2)\sigma^{\mu\nu}$ commute with H_0, the Lorentz generators $M_0^{\mu\nu}$ for the "total" motion do, where $M_0^{\mu\nu} = l^{\mu\nu} + (i/2)\sigma^{\mu\nu}$.

Each of the sets of operators $l^{\mu\nu}$, $(i/2)\sigma^{\mu\nu}$, and $M_0^{\mu\nu}$ obey the commutation relations satisfied by the generators of the restricted Lorentz group $L_+^\uparrow$.[6] Since $(i/2)\sigma^{\mu\nu}$ commutes with both $l^{\mu\nu}$ and P^λ, it is easy to verify that the set of operators P^λ, $M_0^{\mu\nu}$ obey the commutation relations of the generators of $P_+^\uparrow$. The representation of $P_+^\uparrow$ which is generated by P^λ, $M_0^{\mu\nu}$ can be classified by calculating the two Casimir operators of the representation. The invariant mass operator is given by $P^\mu P_\mu = \not{P}\not{P} = m^2$, indicating that we must have a mass m representation. The other independent Casimir operator is provided by the square of the Pauli–Lubanski vector S_μ, where $S_\mu = (1/2)\epsilon_{\mu\nu\lambda\sigma} M_0^{\nu\lambda} P^\sigma$. Direct calculation shows that $S_\mu S^\mu = \frac{3}{4}m^2$. Since the eigenvalues of S^2 have the form $J(J+1)m^2$, where $J(J+1)$ is the eigenvalue of the Casimir operator $\mathbf{J}^2$ of the little group $SU(2)$, it follows that the solutions are labeled by the spin $\frac{1}{2}$ representation of $SU(2)$ (that is, $J = \frac{1}{2}$). The sign of the energy provides an additional label for irreducible representations. Since we wish to admit both positive and negative energy states in the Dirac theory, it follows that the solution space of $H_0\psi = m\psi$ is the carrier space for the direct sum of the positive and negative energy, spin $\frac{1}{2}$, mass m representations of $P_+^\uparrow$.

In the spin basis, the abstract vectors of the representation are labeled as $|p, \lambda, \pm\rangle$, where p^μ denotes the eigenvalue of the operator P^μ, $\lambda = \pm\frac{1}{2}$ refers to the component of the spin along the three-axis in the rest frame, and $\pm$ refers to the sign of the energy. The explicit form of the solutions is immediately evident due to the relative simplicity of the operators P^λ and $M_0^{\mu\nu}$. Since the construction of these states is, of course, well known,[5,6] we do not repeat the procedure here. In Sec. 3 we will mimic this technique in order to construct explicit solutions for the interacting Dirac electron problem.

The solution to the problem of a Dirac electron interacting with a plane-wave electromagnetic field will proceed in a fashion which is formally identical to the above discussion. In Sec. 2 we find the exact analog of each of the operators appearing in the free electron case. These provide the generators of a representation of $P_+^\uparrow$, which again turns out to be the direct sum of the forward and backward timelike, mass m, spin $\frac{1}{2}$ representations. In Sec. 3, this exact analogy is exploited in order to explicitly construct

the solutions. This also permits an identification of the Green's function for a Dirac electron interacting with a plane-wave electromagnetic field.

2. ALGEBRAIC SOLUTION

According to the principle of minimal electromagnetic interaction, the Dirac equation for the interacting electron is obtained by making the substitution $P^\mu \to P^\mu - qA^\mu$, so that the new "Hamiltonian" is given by

$$H = \not{P} - q\not{A}, \tag{2.1}$$

where $P^\mu = i(\partial/\partial x_\mu)$. For simplicity[7] we choose the electromagnetic potential A^μ as

$$A^\mu = a^\mu \exp(ik\cdot x), \tag{2.2}$$

where a^μ is a constant vector and

$$k\cdot k = k\cdot A = 0. \tag{2.3}$$

The eigenvalue problem is to determine solutions of the equation $H\psi = m\psi$.

Our first task is to find a new set of translation operators π^μ which commute with the "Hamiltonian" (2.1). At this point we are guided by our previous solution of the Klein–Gordon equation. We anticipate that the translation generators will be similar in form to those for a spin-zero particle, with the exception that there must also be a term present which represents the interaction of the electron's magnetic moment with the electromagnetic field. Since the magnetic moment interaction is of the form $\sigma_{\mu\nu}F^{\mu\nu}$, where $F^{\mu\nu}$ is the antisymmetric field tensor, we are led to define the operator

$$\Omega = (i/2)\sigma^{\mu\nu}k_\mu A_\nu. \tag{2.4}$$

One can then easily verify that $\Omega = \frac{1}{2}\not{A}\not{k} = -\frac{1}{2}\not{k}\not{A}$, so that $k\Omega = \Omega k = 0$.

To obtain the generators π^μ, we consider the set of commutation relations between $H, P^\mu, A^\nu, A\cdot P, A\cdot A$, and Ω. In deriving this algebra it is important to note that $k\cdot k = k\cdot A = 0$, $A\cdot P = P\cdot A$ in the Lorentz Gauge, $\not{k}\Omega = \Omega\not{k} = 0$, and most important of all, that $k\cdot P$ commutes with all members of this algebra. We therefore replace $k\cdot P$ by a constant $\zeta = k\cdot P$. The relevant relations are

$$[P^\mu, H] = qk^\mu \not{A}, \tag{2.5a}$$

$$[A^\nu, H] = \not{k}A^\nu, \tag{2.5b}$$

$$[A\cdot P, H] = \not{k}(A\cdot P), \tag{2.5c}$$

$$[A\cdot A, H] = 2\not{k}(A\cdot A), \tag{2.5d}$$

$$[\Omega, H] = \zeta\not{A} - \not{k}(A\cdot P) + q\not{k}(A\cdot A), \tag{2.5e}$$

$$[P^\mu, A^\nu] = -k^\mu A^\nu, \tag{2.5f}$$

$$[P^\mu, (A\cdot P)] = -k^\mu (A\cdot P), \tag{2.5g}$$

$$[P^\mu, A\cdot A] = -2k^\mu (A\cdot A), \tag{2.5h}$$

$$[P^\mu, \Omega] = -k^\mu \Omega, \tag{2.5i}$$

with all remaining commutators vanishing. Inspection of Eqs. (2.5a)–(2.5e) shows that a new set of translation operators, commuting with H, is provided by the expressions

$$\pi^\mu = P^\mu - q\zeta^{-1}k^\mu(A\cdot P) + q^2(2\zeta)^{-1}k^\mu(A\cdot A) - q\zeta^{-1}k^\mu\Omega, \tag{2.6}$$

while inspection of the remaining equations shows that

$$[\pi^\mu, \pi^\nu] = 0.$$

It may further be verified that $\pi^\mu\pi_\mu = H^2$, where

$$H^2 = (\not{P} - q\not{A})^2 = \{P^2 - 2q(A\cdot P) + q^2(A\cdot A)\} - 2q\Omega. \tag{2.7}$$

The portion inside the curly brackets is the Klein–Gordon operator while the remaining term represents the magnetic moment interaction. Since $H^2\psi = m^2\psi$, we know we must have a mass m representation of $P_+^\uparrow$.

We must now find a new set of "coordinates" Q^ν which are canonical to the π^μ, that is, $[\pi^\mu, Q^\nu] = ig^{\mu\nu}$. These may readily be found by considering the commutator of x^ν with π^μ and by inspecting the commutation relations (2.5). It is necessary to note that $k\cdot P$ cannot be considered as a constant when commuted with arbitrary functions of x. However, the necessary commutator of x and $(k\cdot P)^{-1}$ can easily be determined by elementary techniques. Without much difficulty, the new "coordinates" are found to be

$$Q^\nu = x^\nu - iq\zeta^{-1}A^\nu + iq\zeta^{-2}k^\nu(A\cdot P) - iq^2(2\zeta)^{-2}k^\nu(A\cdot A) + iq\zeta^{-2}k^\nu\Omega. \tag{2.8}$$

One can easily verify that these "coordinates" have the desired property that $[Q^\mu, Q^\nu] = 0$.

In analogy to the case of the free Dirac electron, we define the "orbital" Lorentz operators $L^{\mu\nu}$ according to the rule

$$L^{\mu\nu} = i(Q^\mu\pi^\nu - Q^\nu\pi^\mu). \tag{2.9}$$

Since $[Q^\nu, H^2] = -2i\pi^\nu$, it immediately follows that $L^{\mu\nu}$ commutes with H^2. The set of operators $\pi^\mu, L^{\nu\lambda}$ obey the commutation relations of the Poincaré generators. Since they commute with H^2, it follows from the discussion given in I that the manifold of solutions of $H^2\psi = m^2\psi$ carries the spin zero, mass m representation of $P_+^\uparrow$. This does not, however, solve the original Dirac equation. We could, of course, use the expedient of Volkov and construct the solutions of $(H - m)\psi = 0$ by noting that if ψ' is a solution of $(H^2 - m^2)\psi' = 0$, then $\psi = (H + m)\psi'$ is a solution of $H\psi = m\psi$. Rather than do this, we shall continue the algebraic analysis.

As might be expected, $L^{\mu\nu}$ does not commute with H. The explicit expressions for these commutators are extremely cumbersome, and since it is never necessary to use these expressions, we do not give them here. In order to continue the analysis, we now need a new set of "spin" operators. If the free-particle analogy is to hold, these new "spin" operators must be constructed from objects which commute with π^μ and $L^{\nu\lambda}$. Noting that the γ matrices arise in the commutator $[x^\mu, H_0] = -i\gamma^\mu$, we define a new set of gamma matrices Γ^μ by the rule

$$[Q^\mu, H] = -i\Gamma^\mu. \tag{2.10}$$

Evaluation of the commutator yields

$$\Gamma^\mu = \gamma^\mu + q\zeta^{-1}(A^\mu \not{k} - k^\mu \not{A}) - (q^2/2)\zeta^{-2}k^\mu (A\cdot A)\not{k}. \tag{2.11}$$

A direct computation reveals the rather amazing fact that the Γ^μ also anticommute:

$$\{\Gamma^\mu, \Gamma^\nu\}_+ = 2g^{\mu\nu}. \tag{2.12}$$

Use of the Jacobi identity on $[\pi^\mu, [Q^\nu, H]]$ reveals that $[\pi^\mu, \Gamma^\nu] = 0$, while direct computation shows that $[Q^\mu, \Gamma^\nu] = 0$. Recalling that $[Q^\mu, H^2] = -2i\pi^\mu$ provides the important relation

$$\{\Gamma^\mu, H\}_+ = 2\pi^\mu. \tag{2.13}$$

Having verified that the new Γ matrices have all the appropriate properties, we now construct the new sigma matrices according to the rule

$$\Sigma^{\mu\nu} = (i/2)[\Gamma^\mu, \Gamma^\nu]. \tag{2.14}$$

The use of Eqs. (2.12) and (2.14) allows one to ascertain that the quantities $(i/2)\Sigma^{\mu\nu}$ obey the algebra of the generators of the Lorentz group.

In terms of the new Γ matrices, the "Hamiltonian" (2.1) takes the form

$$H = \Gamma^\mu \pi_\mu, \tag{2.15}$$

where we eschew the temptation to use the slash notation in hopes of avoiding confusion. In terms of the Σ matrices, we define the "total" Lorentz generators $M^{\mu\nu}$ by the rule

$$M^{\mu\nu} = L^{\mu\nu} + (i/2)\Sigma^{\mu\nu}. \tag{2.16}$$

With H in the form (2.15), the use of Eqs. (2.9), (2.10), and (2.12)–(2.14) permits us to verify that $[M^{\mu\nu}, H] = 0$ without direct evaluation of the rather cumbersome commutators which appear.

The set of objects $\pi^\mu, M^{\nu\lambda}$ commute with H and obey the algebra of the generators of the Poincaré group. In order to complete the algebraic description of the solutions, we need only to classify the representation which these operators generate. The mass-squared operator $\pi^\mu\pi_\mu = m^2$ tells us that we again have a mass m representation. The second Casimir operator is provided by $S_\mu S^\mu$,[4] where

$$S_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\lambda\delta}M^{\nu\lambda}\pi^\delta, \tag{2.17}$$

and $\epsilon_{\mu\nu\lambda\delta}$ is the completely antisymmetric Levi-Civita symbol. Direct computation in the rest frame shows that $S_\mu S^\mu = \frac{3}{4}m^2$. Referring to the discussion given in the introduction, we again have a mass m, spin $\frac{1}{2}$ representation of $P_+^\uparrow$.[3]

We of course wish to admit both forward and backward timelike solutions, so that we again have the direct sum of two irreducible representations. This completes the algebraic classification of the solutions of $H\psi = m\psi$.

3. EXPLICIT EIGENFUNCTIONS

Just as for the free Dirac electron, the abstract vectors in the representation for the interacting electron appear as $|p, \lambda, \pm\rangle$, where p^μ is the eigenvalue of π^μ, $\lambda = \pm\frac{1}{2}$ is (in the spin basis) the projection of the spin along the three-axis in the rest frame [the eigenvalue of $(i/2)\Sigma^{12}$], and $\pm$ refers to the sign of p^0. By using the operators π^μ, $M^{\nu\lambda}$ which were found in Sec. 2, it is now a relatively straightforward procedure to find the explicit eigenfunctions, since the form of the operators is simple.

Momentum eigenstates can be found by determining the eigenstate of π^0 in the "rest frame"[8] and then "boosting" to an arbitrary frame. Because of the simplicity of the operators and the known solutions of the free Dirac equation, however, it is easier to proceed directly. Inspection of the form of π^μ, given by Eq. (2.6), allows one to ascertain that the simultaneous eigenstates of the π^μ are given by

$$\psi \sim \exp\{-i[p\cdot x - iq\zeta_p^{-1}(A\cdot p) + iq^2(4\zeta_p)^{-1}(A\cdot A) - iq\zeta_p^{-1}\Omega]\}u, \tag{3.1}$$

where u is an arbitrary constant spinor, and ζ_p denotes the eigenvalue of $k\cdot P$ when acting on the function (3.1), that is, $\zeta_p = k\cdot p$ (in order to avoid an overly cumbersome notation, we henceforth drop the subscript p from ζ_p). Since the first three quantities in the argument of the exponential involve only the identity matrix, it is convenient to introduce

$$\phi_p = \exp\{-i[p\cdot x - iq\zeta^{-1}(A\cdot p) + iq^2(4\zeta)^{-1}(A\cdot A)]\} \tag{3.2}$$

(ϕ_p is, in fact, the solution of the Klein–Gordon equation when the particle has no magnetic moment). Since Ω is nilpotent ($\Omega^2 = 0$), the eigenfunctions of π^μ finally take the form

$$\psi \sim (1 - q\zeta^{-1}\Omega)\phi_p u. \tag{3.3}$$

The functions ψ given by (3.3) are solutions of the squared Dirac equation for arbitrary u. In order to satisfy the Dirac equation, we find [by direct substitution of expression (3.3) into the Dirac equation] that u must satisfy the equation

$$\Gamma^\mu p_\mu (1 - q\zeta^{-1}\Omega)u = m(1 - q\zeta^{-1}\Omega)u. \tag{3.4}$$

By using the fact that

$$(1 - q\zeta^{-1}\Omega)^{-1} = (1 + q\zeta^{-1}\Omega), \tag{3.5}$$

it is an easy matter to verify that $(1 - q\zeta^{-1}\Omega)$ is an intertwining operator for the two sets of gamma matrices, that is

$$\Gamma^\mu(1 - q\zeta^{-1}\Omega) = (1 - q\zeta^{-1}\Omega)\gamma^\mu. \tag{3.6}$$

From this, it is apparent that u must satisfy the free Dirac equation

$$(\not{p} - m)u = 0. \tag{3.7}$$

Combining all of this information together yields the final form of the solution[9]:

$$\psi_{p,\lambda} = (1 - q\zeta^{-1}\Omega)\phi_p u_{p,\lambda}. \tag{3.8}$$

The subscripts p, λ on ψ retain the same meaning as for the case of the free Dirac electron. This may be

seen by constructing in the usual fashion the positive and negative frequency projection operators, as well as the spin projection operators—the only difference being that the Γ matrices are used instead of the γ. The use of Eq. (3.6) then confirms this statement.

One can use solutions (3.8) to immediately write down the following expression for the exact Green's function:

$$G(x,x') \sim \int d^4p \frac{(\Gamma^\mu(x)p_\mu + m)}{(p^2 - m^2)} \times[1 - q\zeta^{-1}\Omega(x)]\phi_p(x)\phi_p^*(x')[1 + q\zeta^{-1}\Omega(x')]. \quad (3.9)$$

A detailed discussion of the properties of $G(x,x')$ will be given elsewhere.

As a final point, we wish to mention what it means for the free and interacting representations to be "identical." Mathematically, the representations are said to be unitarily equivalent (we are realizing the same representation with a different set of operators). This implies the existence of a unitary transformation U which carries the free solutions ψ and operators O into the interacting solutions ψ' and operators O' via the rule $\psi' = U\psi$, $O' = UOU^{-1}$.

Inspection of the preceding discussion shows that this transformation is accomplished by the operator[10]

$$U = \exp[-i(k\cdot P)^{-1}\{-iq(A\cdot P) + i(q^2/4)(A\cdot A)\}] \times [1 - q(k\cdot P)^{-1}\Omega].$$

Equation (3.6) provides a manifest example of this transformation.

Note added in proof: After submitting this manuscript, we noticed the recent article by Richard[11] in which some aspects of the present paper are discussed. Richard's article led us to the earlier work by Chakrabarti[12] in which the "dynamical representation" derived here was first noted. Chakrabarti noted the existence of this representation by constructing the operator U given above from the Volkov solutions. This should be contrasted to the first principles derivation given here which requires no *a priori* knowledge of the solutions.

---

* See B. L. Beers and H. H. Nickle, Bull. Amer. Phys. Soc. **17**, 565 (1972).

1 H. H. Nickle and B. L. Beers, Southern Illinois University Technical Report No. NB-1-72 (unpublished). This manuscript is henceforth referred to as I. See also B. L. Beers and H. H. Nickle, Lett. Nuovo Cimento **4**, 320 (1972).

2 D. M. Volkov, Z. Physik **94**, 250 (1935); Zh. Eksp. Teor. Fiz. **7**, 1286 (1937). In this regard, see also L. S. Brown and T. W. B. Kibble, Phys. Rev. **A133**, 705 (1964); I. M. Ternov, A. M. Khapaev, and Yu. I. Klimenko, Vestn. Mosk. Univ. Ser. Fiz., Moscow (1967), **22**, No. 1, 35 (1967) [Moscow Univ. Phys. Bull. **22**, No. 1, pp. 18–22].

3 E. P. Wigner, Ann. Math. **40**, 149 (1939).

4 Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **33**, 861, 1196, 1208 (1957); [Sov. Phys. JETP **6**, 664, 919, 929 (1958)].

5 For the Dirac equation and related quantities, we will follow the notation and conventions found in J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

6 For an elementary discussion of the Lorentz and Poincaré groups, see, e.g., S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), Chap. 2.

7 As shown in I, it is a trivial extension to consider vector potentials of the form $a_\mu f(\phi) + b_\mu g(\phi)$, where $\phi \equiv k\cdot x$, f and g are arbitrary functions, and $k\cdot a = k\cdot b = a\cdot b = 0$.

8 We use words like "rest frame" in the mathematical sense. We do not mean to imply that the interacting electron is at rest. With this warning, we will continue to use all the nomenclature for the free electron in the present case.

9 In comparing this result to other results given in the literature, please note that for the simple potential (2.2), the integrals (in terms of which the exact answer is usually given) can easily be integrated to yield the present form. For further clarification of this point, see I.

10 The fact that U is not manifestly unitary should not alarm the reader. The factors of i arise from our choice of the complex potential (2.2). This can easily be remedied and causes no difficulty.

11 J. L. Richard, Nuovo Cimento **8A**, 485 (1972).

12 A. Chakrabarti, Nuovo Cimento **56A**, 604 (1968).

The Scattering of Three Impenetrable Particles in One Dimension

J. B. McGuire*
Florida Atlantic University, Boca Raton, Florida
and
University of Adelaide, Adelaide, South Australia

and

C. A. Hurst
University of Adelaide, Adelaide, South Australia

(Received 5 May 1972)

A special one-dimensional quantum mechanical three-body problem with an interaction among the particles is solved exactly. The interaction is introduced in such a way that the particles retain their order along a line, and hence the particles are impenetrable. The assumption of impenetrability is consistant with the existence of bound states in all accessible channels, and the probabilities for direct and exchange scattering of one bound pair and one free particle are calculated both below and above the breakup threshold. Above the breakup threshold, the probabilities for breakup and free particle scattering are also found. The absolute squares of all elements of the scattering matrix are given explicitly, and the scattering matrix is shown to be unitary.

I. INTRODUCTION

In this paper we will discuss the problem of three particles in one dimension which satisfy

$$H\psi = \left(\frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{p_3^2}{2m_3}\right)\psi(x_1,x_2,x_3) = E\psi$$

$$= -\hbar^2\left(\frac{1}{2m_1}\frac{\partial^2}{\partial x_1^2} + \frac{1}{2m_2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{2m_3}\frac{\partial^2}{\partial x_3^2}\right)\psi = E\psi.$$

The particles, however, are not to be free, for we shall apply three constraints to the problem,

(1) $x_1 < x_2 < x_3$,

seen by constructing in the usual fashion the positive and negative frequency projection operators, as well as the spin projection operators—the only difference being that the Γ matrices are used instead of the γ. The use of Eq. (3.6) then confirms this statement.

One can use solutions (3.8) to immediately write down the following expression for the exact Green's function:

$$G(x,x') \sim \int d^4p \frac{(\Gamma^\mu(x)p_\mu + m)}{(p^2 - m^2)} \times[1 - q\zeta^{-1}\Omega(x)]\phi_p(x)\phi_p^*(x')[1 + q\zeta^{-1}\Omega(x')]. \quad (3.9)$$

A detailed discussion of the properties of $G(x,x')$ will be given elsewhere.

As a final point, we wish to mention what it means for the free and interacting representations to be "identical." Mathematically, the representations are said to be unitarily equivalent (we are realizing the same representation with a different set of operators). This implies the existence of a unitary transformation U which carries the free solutions ψ and operators O into the interacting solutions ψ' and operators O' via the rule $\psi' = U\psi$, $O' = UOU^{-1}$.

Inspection of the preceding discussion shows that this transformation is accomplished by the operator[10]

$$U = \exp[-i(k\cdot P)^{-1}\{-iq(A\cdot P) + i(q^2/4)(A\cdot A)\}] \times [1 - q(k\cdot P)^{-1}\Omega].$$

Equation (3.6) provides a manifest example of this transformation.

Note added in proof: After submitting this manuscript, we noticed the recent article by Richard[11] in which some aspects of the present paper are discussed. Richard's article led us to the earlier work by Chakrabarti[12] in which the "dynamical representation" derived here was first noted. Chakrabarti noted the existence of this representation by constructing the operator U given above from the Volkov solutions. This should be contrasted to the first principles derivation given here which requires no *a priori* knowledge of the solutions.

---

* See B. L. Beers and H. H. Nickle, Bull. Amer. Phys. Soc. **17**, 565 (1972).

1 H. H. Nickle and B. L. Beers, Southern Illinois University Technical Report No. NB-1-72 (unpublished). This manuscript is henceforth referred to as I. See also B. L. Beers and H. H. Nickle, Lett. Nuovo Cimento **4**, 320 (1972).

2 D. M. Volkov, Z. Physik **94**, 250 (1935); Zh. Eksp. Teor. Fiz. **7**, 1286 (1937). In this regard, see also L. S. Brown and T. W. B. Kibble, Phys. Rev. **A133**, 705 (1964); I. M. Ternov, A. M. Khapaev, and Yu. I. Klimenko, Vestn. Mosk. Univ. Ser. Fiz., Moscow (1967), **22**, No. 1, 35 (1967) [Moscow Univ. Phys. Bull. **22**, No. 1, pp. 18–22].

3 E. P. Wigner, Ann. Math. **40**, 149 (1939).

4 Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **33**, 861, 1196, 1208 (1957); [Sov. Phys. JETP **6**, 664, 919, 929 (1958)].

5 For the Dirac equation and related quantities, we will follow the notation and conventions found in J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

6 For an elementary discussion of the Lorentz and Poincaré groups, see, e.g., S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), Chap. 2.

7 As shown in I, it is a trivial extension to consider vector potentials of the form $a_\mu f(\phi) + b_\mu g(\phi)$, where $\phi \equiv k\cdot x$, f and g are arbitrary functions, and $k\cdot a = k\cdot b = a\cdot b = 0$.

8 We use words like "rest frame" in the mathematical sense. We do not mean to imply that the interacting electron is at rest. With this warning, we will continue to use all the nomenclature for the free electron in the present case.

9 In comparing this result to other results given in the literature, please note that for the simple potential (2.2), the integrals (in terms of which the exact answer is usually given) can easily be integrated to yield the present form. For further clarification of this point, see I.

10 The fact that U is not manifestly unitary should not alarm the reader. The factors of i arise from our choice of the complex potential (2.2). This can easily be remedied and causes no difficulty.

11 J. L. Richard, Nuovo Cimento **8A**, 485 (1972).

12 A. Chakrabarti, Nuovo Cimento **56A**, 604 (1968).

The Scattering of Three Impenetrable Particles in One Dimension

J. B. McGuire*
Florida Atlantic University, Boca Raton, Florida
and
University of Adelaide, Adelaide, South Australia

and

C. A. Hurst
University of Adelaide, Adelaide, South Australia

(Received 5 May 1972)

A special one-dimensional quantum mechanical three-body problem with an interaction among the particles is solved exactly. The interaction is introduced in such a way that the particles retain their order along a line, and hence the particles are impenetrable. The assumption of impenetrability is consistant with the existence of bound states in all accessible channels, and the probabilities for direct and exchange scattering of one bound pair and one free particle are calculated both below and above the breakup threshold. Above the breakup threshold, the probabilities for breakup and free particle scattering are also found. The absolute squares of all elements of the scattering matrix are given explicitly, and the scattering matrix is shown to be unitary.

I. INTRODUCTION

In this paper we will discuss the problem of three particles in one dimension which satisfy

$$H\psi = \left(\frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{p_3^2}{2m_3}\right)\psi(x_1,x_2,x_3) = E\psi$$

$$= -\hbar^2\left(\frac{1}{2m_1}\frac{\partial^2}{\partial x_1^2} + \frac{1}{2m_2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{2m_3}\frac{\partial^2}{\partial x_3^2}\right)\psi = E\psi.$$

The particles, however, are not to be free, for we shall apply three constraints to the problem,

(1) $x_1 < x_2 < x_3$,

$$(2)\quad -g'\psi(x_1,x_1,x_3)=\hbar\left(\frac{1}{m_1}\frac{\partial\psi}{\partial x_1}\frac{1}{m_2}\frac{\partial\psi}{\partial x_2}\right)\Big|_{x_1\to x_2},$$

$$(3)\quad -h'\psi(x_1,x_2,x_2)=\hbar\left(\frac{1}{m_2}\frac{\partial\psi}{\partial x_2}-\frac{1}{m_3}\frac{\partial\psi}{\partial x_3}\right)\Big|_{x_2\to x_3}.$$

These constraints have the effect of introducing an "interaction" between the particles. This interaction is such that the particles cannot penetrate each other. To see this point look at the component of the probability flux vector perpendicular to the surface $x_1 = x_2$. This component is proportional to

$$\lim_{x_1\to x_2}\left[\psi^*\left(\frac{1}{m_1}\frac{\partial\psi}{\partial x_1}-\frac{1}{m_2}\frac{\partial\psi}{\partial x_2}\right)-\psi\left(\frac{1}{m_1}\frac{\partial\psi^*}{\partial x_1}-\frac{1}{m_2}\frac{\partial\psi^*}{\partial x_2}\right)\right],$$

which vanishes because of (2). Thus we see that particle 1 is not allowed to penetrate particle 2. Similarly, of course, particle 2 is not allowed to penetrate particle 3.

We now re-express the partial differential equation in the center of mass system. We do this in two steps. First let

$$y_1=\sqrt{m_1}\,x_1/\hbar,\quad y_2=\sqrt{m_2}\,x_2/\hbar,\quad y_3=\sqrt{m_3}\,x_3/\hbar.$$

The Hamiltonian operator now has a simpler form,

$$H=-\frac{1}{2}\left[\frac{\partial^2}{\partial y_1{}^2}+\frac{\partial^2}{\partial y_2{}^2}+\frac{\partial^2}{\partial y_2{}^2}\right].$$

The surface $x_1 = x_2$ implies

$$y_1/\sqrt{m_1}=y_2/\sqrt{m_2},$$

and, similarly, $x_2 = x_3$ implies

$$y_2/\sqrt{m_2}=y_3/\sqrt{m_3}.$$

We now make a rotation in the y space. This, of course, leaves the form of the Hamiltonian invariant. We wish this rotation to separate the center of mass and also to make one of the new coordinates be proportional to $x_1 - x_2$. Such rotations are discussed in Ref. 1. The appropriate rotation is

$$z_1=\left(\frac{m_1}{m_1+m_2+m_3}\right)^{1/2}y_1+\left(\frac{m_2}{m_1+m_2+m_3}\right)^{1/2}\times y_2+\left(\frac{m_3}{m_1+m_2+m_3}\right)^{1/2}y_3,$$

$$z_2=\left(\frac{m_1m_3}{(m_1+m_2+m_3)(m_1+m_2)}\right)^{1/2}y_1+\left(\frac{m_2m_3}{(m_1+m_2)(m_1+m_2+m_3)}\right)^{1/2}y_2-\left(\frac{m_1+m_2}{m_1+m_2+m_3}\right)^{1/2}y_3,$$

$$z_3=\left(\frac{m_2}{m_1+m_2}\right)^{1/2}y_1-\left(\frac{m_1}{m_1+m_2}\right)^{1/2}y_2.$$

In this coordinate system

$$x_1-x_2=\hbar[(m_2+m_1/m_1m_2)]^{1/2}\,z_3,$$

$$x_2-x_3=\hbar[(m_2+m_3/m_2m_3)]^{1/2}(\sin\alpha z_2-\cos\alpha z_3),$$

$$\tan\alpha=\hbar[(m_1+m_2+m_3)m_2/m_1m_3]^{1/2}.$$

Finally, we express our Hamiltonian in cylindrical coordinates. Let

$$z_3=r\sin\theta,\quad z_2=r\cos\theta,\quad z_1=z_1.$$

In this coordinate system,

$$x_1-x_2=\hbar(\mu_{12})^{-1/2}r\sin\theta,$$

$$x_2-x_3=\hbar(\mu_{23})^{-1/2}r\sin(\theta-\alpha),$$

$$H=-\frac{1}{2}\left(\frac{\partial^2}{\partial z_1{}^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\right),$$

μ_{ij} = reduced mass of the ij subsystem.

Condition (1) now implies that we are to solve the partial differential equation $H\psi = E\psi$ within the wedge $0 < r < \infty, 0 < \theta < \alpha$. Conditions (2) and (3) read, respectively,

$$g'\psi(r,\alpha)=-(\mu_{12})^{-1/2}\frac{1}{r}\frac{\partial\psi}{\partial\theta}\Big|_{\theta=0},$$

$$h'\psi(r,0)=(\mu_{23})^{-1/2}\frac{1}{r}\frac{\partial\psi}{\partial\theta}\Big|_{\theta=\alpha}.$$

In this restated form the three particle scattering problem is formulated as the solution to the wave equation in two dimensions confined to the wedge $0 < \theta < \alpha$, and subject to the boundary conditions

$$g\psi(r,0)=-\frac{\partial\psi}{\partial n}\Big|_{\theta=0},\qquad(1a)$$

$$h\psi(r,\alpha)=\frac{\partial\psi}{\partial n}\Big|_{\theta=\alpha},\qquad(1b)$$

where

$$h=h'(\mu_{12})^{1/2},\quad g=g'(\mu_{23})^{1/2},$$

n = unit normal to the indicated surface.

In this form the problem has an electromagnetic analog, viz. the diffraction of an E-polarized electromagnetic wave by an imperfectly conducting wedge, where our strength constants g and h are related to the conductivity of the boundary material of the wedge. This analog problem has commanded some interest, notably that of Williams,[2] whose method we shall adopt. Nussensveig[3] recognized that this analogy existed and used the method of Williams' to describe a particular breakup problem.

The authors' interest in this problem was stimulated by its similarity to the problem of three penetrable particles in one dimension interacting via delta function potentials. In fact under certain conditions of high symmetry the two types of problems turn out to be identical.

It has been found, however, that the more limited class of impenetrable particle problems can stand in their own right as true three particle problems and as such provide an illustration of the type of mathematics which allows the effects present in more complicated problems.

II. THE METHOD OF SOLUTION

We adopt the method suggested by Williams, which is essentially Sommerfield's method for solving the perfectly conducting wedge. We assume that we may represent the solution as an integral over plane waves of a given energy $E = k^2/2$, which we take for the moment to be a positive number.

$$\psi(r,\theta) = \int_A^B f(\theta,w)e^{ikr\cos w}dw.$$

We must ensure that ψ satisfies the differential equation, and to this end we assume that the limits A and B are carried to infinity in some way so that we may differentiate under the integral sign. The limits A and B may be safely carried to infinity anywhere in the complex w plane where $\mathrm{Im}\cos w > 0$. This restriction implies that the end points of the integration must lie in a one of a series of strips in the w plane:

$$\text{for } \mathrm{Im}w > 0, \quad (2n-1)\pi < \mathrm{Re}w < 2n\pi,$$
$$\text{for } \mathrm{Im}w < 0, \quad 2n\pi < \mathrm{Re}w < (2n+1)\pi,$$

where n takes on all integer values from $-\infty$ to ∞.

We must evidently commence our integration in one of these strips and end it in another.

We note that at $r = 0$ the function ψ must be finite and independent of θ, for physically this is a single point and the solution must be single-valued. We anticipate that this independence must arise from our ability to deform the contour in the w plane for fixed θ. Since we further anticipate that the function f will have singularities in the w plane which give rise to bound states, etc., we recognize that we must choose our contour so that it surrounds none of these singularities. It turns out that all of the singularities introduced by the dynamics will lie within a finite distance of the real axis. We will therefore choose our contour so that it always has $\mathrm{Im}w > 0$ and is above any of the singularities of f. This basic contour is denoted C_1 in Fig. 1.

We now substitute the assumed form for ψ into the differential equation, integrate by parts, and obtain

$$(\nabla^2 + k^2)\psi = \frac{1}{r^2}\int_c\left(\frac{\partial^2 f}{\partial\theta^2} - \frac{\partial^2 f}{\partial w^2}\right)e^{ikr\cos w}dw = 0.$$

This is satisfied if

$$f(w,\theta) = G(w+\theta) + H(w-\theta);$$

thus

$$\psi = \int_c [G(w+\theta) + H(w-\theta)]e^{ikr\cos w}\,dw.$$

We take this form for ψ and calculate

$$\frac{1}{r}\frac{\partial\psi}{\partial\theta} = \int_c [G'(w+\theta) - H'(w-\theta)]e^{ikr\cos w}\,dw,$$

which, by partial integration, is

$$\frac{1}{r}\frac{\partial\psi}{\partial\theta} = ik\int_c \sin w[G(w+\theta) - H(w-\theta)]e^{ikr\cos w}\,dw.$$

We must, of course, ensure in what follows that our functions G and H are sufficiently regular to justify the partial integration.

The boundary condition (1b) along $\theta = 0$ will be satisfied if

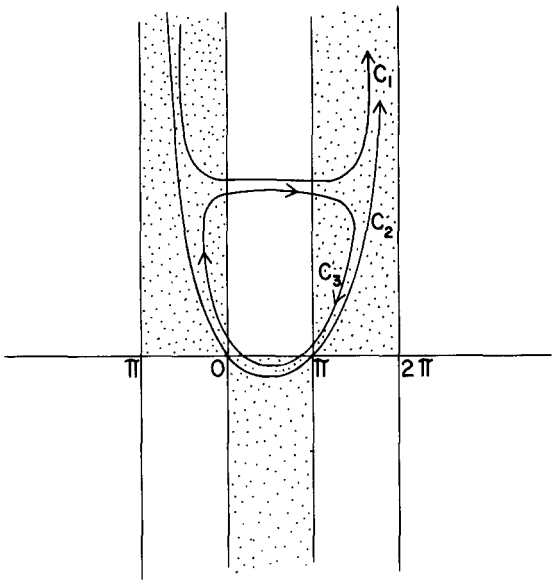


FIG. 1. The basic contour (C_1) and its deformation (C_2 plus C_3) for the integral form of the solution.

$$g[G(w) + H(w)] = ik\,\sin w[G(w) - H(w)];$$

likewise boundary condition (1a) along $\theta = \alpha$ requires that

$$-H[G(w+\alpha) + H(w-\alpha)] = ik\,\sin w[G(w+\alpha) - H(w-\alpha)].$$

Rewriting these conditions, we find

$$G(w+\alpha) = Y(w)H(w-\alpha),$$
$$H(w) = X(w)G(w),$$

where

$$X(w) = \frac{\sin w - ig/k}{\sin w + ig/k}, \qquad Y(w) = \frac{\sin w - ih/k}{\sin w - ih/k}.$$

Thus the boundary conditions on the sides of the wedge require that G and H satisfy a pair of coupled difference equations. The difference coefficients X and Y are unimodular functions for real w and have period 2π. We temporarily defer discussion of the difference equations and show first that a set of appropriate solutions to the difference equations completely determine the asymptotic properties of the wavefunction.

III. THE ASYMPTOTIC SOLUTION

In order that we have a proper solution, the wavefunction must satisfy:

(1) the boundary conditions at $\theta = 0$ and $\theta = \alpha$, which are expressed through the difference equations for G and H;

(2) regularity at $r = 0$, which is satisfied if G and H approach at most a constant as $w \to i\infty$;

(3) the condition that the wavefunction not increase exponentially anywhere within the wedge, which we shall see imposes conditions on the locations of poles in G and H.

We now wish to deform the contour C_1 so that it passes through the steepest descent points at $w = 0$ and $w = \pi$. This deformation will lead to sweeping the contour across singularities of G and H intro-

duced by the difference equations. The contribution from these singularities must be taken into account explicitly by use of the residue theorem. The contour deformation is made as shown in Fig. 1. The open contour C_2 passes through the steepest descent points, whereas the closed contour C_3 surrounds the singularities of G and H. The asymptotic solution will be the sum of a pair of steepest descent contributions plus some "plane wave" contributions from the poles of G and H.

What sort of singularities may G and H have? In the main this is governed by boundary condition (3), with the criterion being that no contributions are allowed which lead to exponential increases in the wavefunction anywhere in the wedge. Suppose that $G(w + \theta)$ has a pole at $w + \theta = u + iv$; this means that there will be a "plane wave" contribution to the integral at $w = u + iv - \theta$, with the corresponding wavefunction $\exp(ikr)\cos w$

$$ikr\cos w = ikr\cos(u + iv - \theta),$$
$$= ikr\cos(u - \theta)\cosh v + kr\sin(u - \theta)\sinh v.$$

If we take $v > 0$, corresponding to the upper half-plane, we see that $u - \theta$ must be negative. Since θ takes on values from 0 to α, we see that u must be less than 0 or greater than $\pi + \alpha$. At exactly the end points a pole of G would contribute a surface wave along the wall of the wedge. A pole at $u = 0$ would contribute an outgoing wave along the wall at $\theta = 0$ whereas a singularity at $u = \pi + \alpha$ would contribute an incoming surface wave along the wall at $\theta = \alpha$. We thus conclude that G may have no poles in the upper half-plane for $0 < \mathrm{Re}\, w < \pi + \alpha$. Poles may, however, occur at the boundaries of this region. In a similar manner we find that H must be free of poles in the upper half-plane for $-\alpha < \mathrm{Re}\, w < \pi$, with a pole at $u = \pi$ contributing an incoming wave along the surface $\theta = \alpha$.

An analysis of the two-body problem shows that the imaginary part of the location of a pole in G or H is dictated by the constants g and h. That is that the surface waves may only have an exponential decrease into the wedge region which is specific to the strength constant on that surface.

Thus we find, as indeed we would expect, that there are four surface wave contributions to the asymptotic form of the wavefunction. If we let

$$\sin i\beta = ig/k, \quad \sin i\gamma = ih/k,$$

then the appropriate amplitudes corresponding to these surface wave contributions are:

(1) For the incoming surface wave along $\theta = 0$: $\tilde{H}(\pi + i\beta) = \lim_{w\to\pi+i\beta}(w - \pi - i\beta)H(w)$;

(2) For the incoming surface wave along $\theta = \alpha$: $\tilde{G}(\pi + \alpha + i\gamma)$;

(3) For the outgoing surface wave along $\theta = 0$: $\tilde{G}(i\beta)$;

(4) For the outgoing surface along $\theta = \alpha$: $\tilde{H}(-\alpha + i\gamma)$.

The remainder of the contribution to the asymptotic wavefunction comes from the steepest descent points at $w = 0$ and $w = \pi$.

These contributions are:

from $w = 0$: $\sqrt{2\pi/kr}\, e^{i\pi/4} e^{ikr}[G(\theta) + H(-\theta)]$;

from $w = \pi$: $\sqrt{2\pi/kr}\, e^{-i\pi/4} e^{ikr}[G(\pi + \theta) + H(\pi - \theta)]$.

Thus, again as we would expect, we have both an incoming and an outgoing contribution for cylindrical "free waves" which correspond to all three of the particles being free.

Finally, we emphasize that we have now obtained an explicit representation of the asymptotic solution in terms of the solutions to the difference equations. If we know G and H, we can evaluate all of the above amplitudes. We thus proceed to a solution of the difference equations.

IV. SOLUTION OF THE DIFFERENCE EQUATIONS

The difference equations are

$$G(w + \alpha) = Y(w)H(w - \alpha), \quad H(w) = X(w)G(w).$$

A further factorization of these equations is possible. Let

$$G(w) = B(w)A(w + \alpha), \quad H(w) = B(w + 2\alpha)A(w + \alpha).$$

Substitution into the difference equations yields

$$B(w + 2\alpha) = X(w)B(w), \quad A(w + 2\alpha) = Y(w)A(w),$$

and thus the equations are separated.

We remark briefly on the theory of such equations. Birkhoff[4] discussed the equation

$$G(w + 2\alpha) = M(w)G(w),$$

and demonstrated that the most general solution is the product of a particular solution times a function with period 2α. Thus

$$G(w) = G_p(w)\phi(w),$$

where $G_p(w)$ is a particular solution and $\phi(w) = \phi(w + 2\alpha)$. A particular solution was found to be

$$G_p(w) = M^{-1}(w)M^{-1}(w + 2\alpha)\cdots M^{-1}(w + n\alpha),$$

provided, of course, that the infinite product converges.

In our case, where X is a unimodular periodic function for real w, we do not obtain convergence in this infinite product form. Jost[5] has made a study of such difference equations with periodic coefficients in connection with the penetrable delta function problem, which he approached by a different method. We shall find, however, that we will be able to modify Birkhoff's technique to our problem at hand.

At this stage we can make three simple observations about the properties of this equation and its solution:

(1) If $M(w)$ can be written as a product of two functions, say

$$M(w) = X_1(w)X_2(w),$$

then the solution is a product of solutions of simpler equations; i.e., let

$$G_1(w + 2\alpha) = X_1(w)G_1(w)$$

and

$$G_2(w + 2\alpha) = X_2(w)G_2(w);$$

then

$$G_1(w + 2\alpha)G_2)(w + 2\alpha) = X_1(w)X_2(w)G_1(w)G_2(w),$$

so that $G(w) = G_1(w)G_2(w)$ is a solution. This evidently holds for any number of terms in the product.

(2) If $M(w + 2\pi) = M(w)$, then $G(w + 2\pi)$ is a solution if $G(w)$ is a solution. Hence we can write

$$G(w + 2\pi) = G(w)\phi(w),$$

where

$$\phi(w + 2\alpha) = \phi(w).$$

(3) If $M(w)$ is unimodular for real w, so that

$$M^*(w) = M^{-1}(w),$$

and this relation is extended to complex w, then $G^*(w)G(w)$ is periodic with period 2α. For, from

$$G^*(w + 2\alpha) = M^*(w)G(w) = M^{-1}(w)G^*(w),$$

we have

$$G^*(w + 2\alpha)G(w + 2\alpha) = M^*(w)M(w)G^*(w)G(w) = G^*(w)G(w).$$

We concentrate on the separated difference equations one at a time. First we attempt to solve

$$B(w + 2\alpha) = X(w)B(w).$$

We note that X is an entire analytic function of w and as such has an infinite product representation in w plane. We shall couple this infinite product representation with the observation (1) above and attempt to write the solution for B as an infinite product.

To construct the infinite product representation of X, we write X in the form

$$X(w) = \sin\tfrac{1}{2}(w - i\beta)\cos\tfrac{1}{2}(w + i\beta)/\cos\tfrac{1}{2}(w - i\beta) \times \sin\tfrac{1}{2}(w + i\beta),$$

where, again,

$$\sin i\beta = ig/k.$$

Thus X is itself the product of two functions.

Using the infinite product forms of the sin and cosine, we write one of the factors of X as an infinite product.

$$\frac{\sin\frac{1}{2}(w - i\beta)}{\sin\frac{1}{2}(w + i\beta)} = \frac{(w - i\beta)}{(w + i\beta)}\prod_{n=1}^{\infty}\frac{1 - [(w - i\beta/2n\pi)]^2}{1 - [(w + i\beta/2n\pi)]^2} = \prod_{n=-\infty}^{\infty}\frac{w - i\beta + 2n\pi}{w + i\beta + 2n\pi}.$$

The infinite product form of the other factor of X is

$$\frac{\cos\frac{1}{2}(w + i\beta)}{\cos\frac{1}{2}(w - i\beta)} = \prod_{n=-\infty}^{\infty}\frac{w + i\beta + \pi + 2n\pi}{w - i\beta + \pi + 2n\pi}.$$

For each of the factors which appear in this infinite product, we can construct a solution. Consider the difference equation

$$G(w + 2\alpha) = [(w - w_0 - i\beta)/(w - w_0 + i\beta)]G(w).$$

This equation is satisfied by

$$G(w) = \Gamma((1/2\alpha)(w - w_0 - i\beta))/\Gamma((1/2\alpha)(w - w_0 + i\beta)),$$

because the gamma function satisfies the difference equation

$$\Gamma(z + 1) = z\Gamma(z).$$

The difference equation is equally well satisfied by

$$G(w) = \Gamma((1/2\alpha)(2\alpha - w + w_0 - i\beta))/\Gamma((1/2\alpha)(2\alpha - w + w_0 + i\beta)).$$

These two solutions differ by a multiple of period 2α as can be demonstrated using the property of the gamma function

$$\Gamma(z)\Gamma(1 - z) = \pi/\sin\pi z$$

Thus we may express the solution to the difference equation as an infinite product of gamma functions. Before completing this exercise, however, we should look ahead to the necessity of applying the pole condition which guarantees that there will be no exponentially increasing waves in the wedge.

Notice that the two solutions above represent two different extremes of the possible types of solutions to the difference equation. The first has a string of poles recurring every 2α in the upper half-plane beginning at $w_0 + i\beta$ and running to $-\infty$ parallel to the real axis, and a string of zeros recurring every 2α in the lower half-plane beginning at $w_0 - i\beta$ and running to $-\infty$ parallel to the real axis. The second solution puts zeros running to the right in the upper half-plane and poles running to the right in the lower half-plane, both commencing where $w = w_0 + 2\alpha$.

The solution to the whole difference equation will have an infinite number of these gamma function factors.

There is an entire strip of the complex w plane which must be free of poles. To satisfy this condition, we must choose solutions such that the elementary factors to the right of this strip lead to solutions with poles running to the right and the elementary factors to the left lead to poles running to the left; otherwise poles will fall in the forbidden strip.

We therefore split the elementary factors into two groups, those which occur to the left of the strip and those which occur to the right. Accordingly we write

$$\frac{\sin\frac{1}{2}(w - i\beta)}{\sin\frac{1}{2}(w + i\beta)}\prod_{n=0}^{\infty}\frac{w - i\beta + 2n\pi}{w + i\beta + 2n\pi}\prod_{n=1}^{\infty}\frac{w - i\beta - 2n\pi}{w + i\beta - 2n\pi},$$

where the first factor on the left contains all of the factors to the left of $\mathrm{Re}w = 0$ and the term on the right contains all of the factors to the right of $\mathrm{Re}w = 0$.

Thus a solution to

$$B_1(w + 2\alpha) = [\sin\tfrac{1}{2}(w - i\beta)/\sin\tfrac{1}{2}(w + i\beta)B_1(w),$$

which is free of poles in the strip $0 < \mathrm{Re}w < \pi + \alpha$ in the upper half-plane, is

$$B_1(w) = \prod_{n=0}^{\infty} \frac{\Gamma((1/2\alpha)(w - i\beta + 2nz))}{\Gamma((1/2\alpha)(w - i\beta + 2nz))} \times \prod_{n=1}^{\infty} \frac{\Gamma((1/2\alpha)(2\alpha - w - i\beta + 2n\pi))}{\Gamma((1/2\alpha)(2\alpha - w + i\beta + 2n\pi))}.$$

In fact this solution is free of poles in the upper half-plane from $\mathrm{Re}w = 0$ to $\mathrm{Re}w = 2\pi + 2\alpha$. This turns out to be necessary in the formation of the total solution since the total solution involves shifted products of functions such as the one above.

If we let

$$f(w, \beta) = \prod_{n=0}^{\infty} \frac{\Gamma((1/2\alpha)(w - i\beta + 2n\pi))}{\Gamma((1/2\alpha)(w - i\beta + 2n\pi))},$$

then

$$B_1(w) = f(w, \beta)f(2\pi + 2\alpha - w, \beta),$$

where this solution is free of poles in the upper half-plane for $0 < \mathrm{Re}w < 2\pi + 2\alpha$.

In a similar way we construct a solution to the remaining factors of $B(w)$. If

$$B_2(w + 2\alpha) = [\cos\tfrac{1}{2}(w + i\beta)/\cos\tfrac{1}{2}(w - i\beta)]\, B_2(w),$$

we know that the solution is

$$B_2(w) = f(w + \pi, -\beta)f(\pi + 2\alpha - w, -\beta).$$

We may now construct a solution to the whole difference equation which satisfies the pole criteria. Combining the terms to form $B(w)$, we find

$$\begin{aligned} B(w) &= B_1(w)B_2(w), \\ &= f(w, \beta)f(2\pi + 2\alpha - w, \beta)f(w + \pi, -\beta) \\ &\quad \times f(\pi + 2\alpha - w, -\beta). \end{aligned}$$

Similarly, for $A(w)$,

$$\begin{aligned} A(w) = f(w, \gamma)f(2\pi + 2\alpha - w, \gamma)f(w + \pi, \gamma) \\ \times f(\pi + 2\alpha - w, -\gamma). \end{aligned}$$

In turn we now combine A and B to get G and H.

$$\begin{aligned} G(w) &= B(w)A(w + \alpha) \\ &= f(w, \beta)f(2\pi + 2\alpha - w, \beta)f(w + \pi, -\beta) \\ &\quad \times f(\pi + 2\alpha - w, -\beta)f(w + \alpha, \gamma)f(2\pi + \alpha - w, \gamma) \\ &\quad \times f(w + \pi + \alpha, -\gamma)f(\pi + \alpha - w, -\gamma), \end{aligned}$$

$$\begin{aligned} H(w) &= B(w + 2\alpha)A(w + \alpha) \\ &= f(w + 2\alpha, \beta)f(2\pi - w, \beta)f(w + \pi + 2\alpha, -\beta) \\ &\quad \times f(\pi - w, -\beta)f(w + \alpha, \gamma)f(2\pi + \alpha - w, \gamma) \\ &\quad \times f(w + \pi + \alpha, -\gamma)f(\pi + \alpha - w, -\gamma). \end{aligned}$$

We have now constructed a particular solution to the difference equations. All other solutions are obtained by multiplying G and H by a function with period 2α. This particular solution is still not in the most convenient form for the manipulations which follow. We shall modify this solution in a way which will be motivated in the next section.

We seek to modify this particular solution so that it has the following properties:

(1) no incoming bound wave at $\theta = 0$, which implies that there is no pole of H at $w = \pi + i\beta$,

(2) no incoming bound wave at $\theta = \alpha$, which implies that there is no pole of G at $w = \pi + \alpha + i\gamma$,

(3) no incoming cylindrical wave, which implies no contribution from the steepest descent point at $w = \pi$, or $G(\pi + w) = -H(\pi - w)$.

This solution will not lead to a bounded wavefunction at $r = 0$, because there must be a source at that point since we have outgoing waves which do not arise from incoming waves.

We first seek a modification of our particular solution which satisfies condition (3). That is, we seek a function $\phi(w)$ which satisfies

$$\begin{aligned} &\phi(w) = \phi(w + 2\alpha), \\ &A(w + \alpha)B(w)\phi(w) \\ &\quad = -A(2\pi - w + \alpha)B(2\pi + 2\alpha - w)\phi(2\pi - w). \end{aligned}$$

Now we substitute for A and B in terms of the function f and find the equation for ϕ:

$$\begin{aligned} \frac{\phi(2 - w)}{\phi(w)} &= -\frac{f(w + \pi, -\beta)f(\pi + 2\alpha - w, -\beta)}{f(-\pi + w, -\beta)f(3\pi + 2\alpha - w, -\beta)} \\ &\quad \times \frac{f(w + \alpha + \pi, -\gamma)f(\pi + \alpha - w, -\gamma)}{f(3\pi - w + \alpha, -\gamma)f(-\pi + \alpha + w, -\gamma)}. \end{aligned}$$

Look at the first factor on the right:

$$\frac{f(w + \pi, -\beta)}{f(w - \pi, -\beta)} = \prod_{n=0} \frac{\Gamma((1/2\alpha)(w + \pi + i\beta + 2n\pi))}{\Gamma((1/2\alpha)(w - \pi + i\beta + 2n\pi))} \times \frac{\Gamma((1/2\alpha)(w - \pi - i\beta + 2n\pi))}{\Gamma((1/2\alpha)(w + \pi - i\beta + 2n\pi))};$$

almost all of the terms cancel and we are left with

$$\frac{f(w + \pi, -\beta)}{f(w - \pi, -\beta)} = \frac{\Gamma((1/2\alpha)(w - \; - i\beta))}{\Gamma((1/2\alpha)(w - \; + i\beta))}.$$

A similar cancellation takes place in all of the other factors as well, leaving us with

$$\begin{aligned} \frac{\phi(2\pi - w)}{\phi(w)} &= -\frac{\Gamma((1/2\alpha)(w - \pi - i\beta))\Gamma((1/2\alpha)(\pi + 2\alpha - w + i\beta))}{\Gamma((1/2\alpha)(w - \pi + i\beta))\Gamma((1/2\alpha)(\pi + 2\alpha - w - i\beta))} \\ &\quad \times \frac{\Gamma((1/2\alpha)(\pi + \alpha - w + i\gamma))}{\Gamma((1/2\alpha)(w + \alpha - \pi + i\gamma))} \\ &\quad \times \frac{\Gamma((1/2\alpha)(w + \alpha - \pi - i\gamma))}{\Gamma((1/2\alpha)(\pi + \alpha - w - i\gamma))}. \end{aligned}$$

Using the periodicity property of the gamma function,

$$\Gamma((1/2\alpha)(\pi + 2\alpha - w + i\beta))\Gamma((1/2\alpha)(w - \pi - i\beta)) = \pi[\sin(\pi/2\alpha)(w - \pi - i\beta)]^{-1},$$

we find

$$\frac{\phi(2\pi - w)}{\phi(w)} = -\frac{\sin(\pi/2\alpha)(w - \pi + i\beta)\sin(\pi/2\alpha)(w + \pi - \alpha + i\gamma}{\sin(\pi/2\alpha)(w - \pi - i\beta)\sin(\pi/2\alpha)(w - \pi - \alpha - i\gamma)}.$$

A solution to this equation is

$$\phi(w) = \sin(\pi/2\alpha)(w - \pi - i\beta)\sin(\pi/2\alpha)(w - \pi - \alpha - i\gamma).$$

Thus we have

$$G_0(w) = B(w)A(w + \alpha)\sin(\pi/2\alpha)(w - \pi - i\beta)\sin(\pi/2\alpha)(w - \pi - \alpha - i\gamma),$$

$$H_0(w) = B(w + 2\alpha)A(w + \alpha)\sin(\pi/2\alpha)(w - \pi - i\beta) \times \sin(\pi/2\alpha)(w - \pi - \alpha - i\gamma),$$

where the functions G_0 and H_0 satisfy conditions (1), (2), and (3), since the function $\phi(w)$ is zero at both $w = \pi + i\beta$ and $w = \pi + \alpha + i\beta$.

V. CALCULATION OF SCATTERING AMPLITUDES

In this section we will calculate the amplitudes for various events characteristic of this three-particle system. At this point we have succeeded in providing a solution to the difference equations which has the property that it corresponds to a point source at the origin. There are no incoming bound or free waves, yet outgoing waves exist in all channels.

This particular solution is useful because it is easily modified to fit any type of incoming conditions. The solution to the difference equations which gives rise to a particular physical situation in the incoming state (for example, an incoming state with particles 2 and 3 bound and particle 1 free) must arise from multiplying our basic solution by some function with period 2α. That is,

$$G(w) = G_0(w)\psi(w), \quad H(w) = H_0(w)\psi(w),$$
$$\psi(w) = \psi(w + 2\alpha).$$

If we wish to discuss all possible incoming states, we must have a complete set of functions ψ. By a complete set we mean a set of ψ's which span all the possible incoming physical states.

All of the possible incoming states must be labeled in some way. Implicitly we have already chosen this way to be all possible incoming states of a given three-particle energy. Even within this choice, however, there are infinitely many ways to choose the incoming states.

A similar situation arises in the two particle centro-symmetric scattering problem. We may characterize completely all of the scattering for a given incident energy in an infinity of ways. The two most popular are an angular momentum decomposition, where one specifies the scattering by giving an energy dependent phase-shift for each angular momentum channel, or a plane wave type of decomposition where one specifies a scattering amplitude $f(\theta, \psi)$, which gives the amplitude to scatter from a plane wave incoming at angle ψ to one outgoing at an angle θ. These two points of view are connected by a similarity transformation.

The case of two-particle centro-symmetric scattering has a simplification that has no obvious analog in our case. Since angular momentum is conserved, the scattering is diagonal in the angular momentum representation. We have no idea what peculiar set of states will give us a diagonal representation for the scattering, and so we choose a set of states (much like the plane wave basis) which is physically appealing.

We shall describe our scattering in a basis where the events are labeled by the momentum vector in the two-dimensional space of the problem. That is a basis where the quantum number tells how the energy is apportioned among the three particles. The magnitude of this momentum vector is fixed by the three-particle energy, and so we need only use the angle as a label. We shall denote the possible incoming states as:

$|0)$: an incoming state with particles 2 and 3 bound, particle 1 free (corresponding to $\theta = 0$),

$|\alpha)$: an incoming state with particles 1 and 2 bound, particle 3 free (corresponding to $\theta = \alpha$),

$|\psi)$: an incoming state with all particles free, where the parameter ψ tells how the three-particle energy is divided among the particles.

Similarly, the outgoing states are labeled:

$(0|$: an outgoing state where particles 2 and 3 are bound, particle 1 free,

$(\alpha|$: an outgoing state where particles 1 and 2 are bound, particle 3 free,

$(\theta|$: an outgoing state where all particles are free and the parameter θ tells how the three-particle energy is divided among the particles.

We need to make our difference equation formalism tell us the array of numbers

$$\begin{array}{lll} A(0|0) & A(\alpha|0) & A(\theta|0) \\ A(0|\alpha) & A(\alpha|0) & A(\theta|\alpha) \\ A(0|\psi) & A(\alpha|\psi) & A(\theta|\psi) \end{array}$$

where $A(\theta|0)$ is, for example, the coefficient of the asymptotic cylindrical wave at angle θ given unit amplitude for an incoming bound state of particles 2 and 3.

We now set up a method to calculate this array of numbers. The first row of this array is all conditional on unit amplitude for an incoming wave at $\theta = 0$, and no other incoming wave is present. We must ensure that the asymptotic form of the incoming solution provides an incoming wave along $\theta = 0$. It is therefore necessary to multiply our basic solution by a function $\psi_1(w)$ which has a pole at the point $w = \pi + i\beta$, for H must have a pole at this point to provide the proper incoming wave. The residue of this pole in

the function H will be the coefficient of the asymptotic wave incoming along $\theta = 0$.

The function ψ_1 must also be chosen so that there is no incoming cylindrical wave; thus we must preserve the property that the steepest descent point at $w = \pi$ does not contribute to the asymptotic form, and thus we must have

$$G_0(\pi + w)\psi_1(\pi + w) = -H_0(\pi - w)\psi_1(\pi - w).$$

Since G_0 and H_0 have been constructed with this property, we see that

$$\psi_1(\pi + w) = \psi_1(\pi - w),$$

and hence ψ is an even function about π.

Since there are no sources other than the bound wave at infinity the moduli of the functions G and H must be bounded as $w \to i\infty$. The functions G and H get exponentially large as $w \to i\infty$ like $\exp(\pi/2)\mathrm{Im}w$; thus the function ψ_1 must decrease at least as fast as $\exp(-\pi/2)\ \mathrm{Im}w$ in the same limit. Apart from a constant factor a function which satisfies these conditions is

$$\psi_1(w) = [\cos(\pi/\alpha)(w - \pi) - \cos(\pi i\beta/\alpha)]^{-1}.$$

In fact (apart from a constant factor) this function is unique. Due to the boundedness condition on G and H, we may only consider the possibility of introducing higher Fourier coefficients in the denominator, but these will inevitably introduce unwanted poles in G and H.

Now we see that the entire first row of our array of amplitudes is

$$A(0\,|\,0) = \tilde{G}_0(i\beta)\psi_1(i\beta)/H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta),$$

$$A(\alpha|0) = \tilde{H}_0(-\alpha + i\gamma)\psi_1(-\alpha + i\gamma)/H_0(\pi + i\beta) \times \tilde{\psi}_1(\pi + i\beta),$$

$$A(\theta\,|\,0) = G_0(\theta)\psi_1(\theta)\psi_1(\theta) + H_0(-\theta)\psi_1(-\theta)/H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta),$$

where we have chosen the arbitrary constant in ψ_1 to be such that the residue of the pole of $H(w)$ at $w = \pi + i\beta$ is unity.

The same line of reasoning allows us to construct the function $\psi_2(w)$ which conditions the amplitudes on an incident bound state of particles 1 and 2. We insist that ψ_2 have the following properties:

(1) $\psi_2(w) = \psi_2(w + 2\alpha)$,

(2) $\psi_2(w)$ has a pole where $w = \pi + \alpha + i\gamma$,

(3) $\psi_2(\pi + w) = \psi_2(\pi - w)$,

(4) $\psi_2(w)$ has no poles other than those required by (1), (2), and (3),

(5) $\psi_2(w)$ decreases at least as fast as $\exp(-\pi/\alpha)\ \mathrm{Im}w$ as $w \to i\infty$.

These conditions imply that apart from an arbitrary constant

$$\psi_2(w) = [\cos(\pi/\alpha)(w - \pi) - \cos(\pi/\alpha)(\alpha + i\gamma)]^{-1}.$$

The arbitrary constant is fixed by insisting that the residue of $G(w)$ at $\pi + \alpha + i\gamma$ is unity. This gives the second row of amplitudes in our array to be:

$$A(0\,|\,\alpha) = \tilde{G}_0(i\beta)\psi_2(i\beta)/G_0(\pi + \alpha + i\gamma)\tilde{\psi}_2(\pi + \alpha + i\gamma),$$

$$A(\alpha|\alpha) = \tilde{H}_0(-\alpha + i\gamma)\psi_2(-\alpha + i\gamma)/G_0(\pi + \alpha + i\gamma) \times \tilde{\psi}_2(\pi + \alpha + i\gamma),$$

$$A(\theta\,|\,\alpha) = [G_0(\theta)\psi_2(\theta) + H_0(-\theta)\psi_2(-\theta)]/G_0(\pi + \alpha + i\gamma)\tilde{\psi}_2(\pi + \alpha + i\gamma).$$

Finally we must produce a function which gives an incoming plane wave at $\theta = \psi$. This contribution must arise from a pole on the real w axis. Referring to the original contour integral, we see that the pole must fall in the region $0 < w < \pi$ to give a contribution. By the same sort of reasoning carried out in Sec. III, we conclude that the pole must fall in $H(w)$ at $w = \pi - \psi$, so that by the usual considerations

$$\psi_3(w) = [\cos(\pi/\alpha)(w - \pi) - \cos(\pi\psi/\alpha)]^{-1},$$

and

$$A(0\,|\,\psi) = \tilde{G}_0(i\beta)\psi_3(i\beta)/2\pi H_0(\pi - \psi)\tilde{\psi}_3(\pi - \psi),$$

$$A(\alpha|\psi) = \tilde{H}_0(-\alpha + i\gamma)\psi_3(-\alpha + i\gamma)/2\pi H_0(\pi - \psi)\tilde{\psi}_3(\pi - \psi),$$

$$A(\theta\,|\,\psi) = [G_0(\theta)\psi_3(\theta) + H_0(-\theta)\psi_3(-\theta)]/2\pi H_0(\pi - \psi) \times \tilde{\psi}_3(\pi - \psi).$$

The factor 2π in the denominators is required to preserve unitarity. There is a caveat in the interpretation of this last row of amplitudes. The point is best explained by examining a particular case.

VI. ANALYSIS OF A PARTICULAR CASE

Consider the case where we let both g and h tend to $-\infty$. This says that the wavefunction must vanish along the walls of the wedge, and thus we should obtain Sommerfield's solution for the perfectly conducting wedge. In our interpretation the solution will represent the scattering of three one-dimensional particles whose wavefunction must vanish when two particles are on top of one another.

The difference equations are

$$G(w + \alpha = -H(w - \alpha), \qquad H(w) = -G(w).$$

A particular solution is

$$G(w) = 1, \qquad H(w) = -1.$$

The considerations of the previous section lead us to multiply the particular solutions by a function which is period 2α, even about π, and has a pole of unit residue at $w = \pi - \psi$. Such a function is

$$\psi_3(w) = -\frac{\pi}{2\alpha}\sin\frac{\pi\psi}{\alpha}\left(\sin\frac{\pi}{2\alpha}(w - \pi - \psi) \times \sin\frac{\pi}{2\alpha}(w - \pi + \psi)\right)^{-1}.$$

The resulting scattering amplitude is

$$f(\theta,\psi) = G(\theta) + H(-\theta)$$
$$= -\frac{(1/4\alpha)\sin(\pi^2/\alpha)\sin(\pi\psi/\alpha)\sin(\pi\theta/\alpha)}{\sin(\pi/2\alpha)(\theta-\pi-\psi)\sin(\pi/2\alpha)(\theta-\pi-\psi)} \times [\sin(\pi/2\alpha)(\theta+\pi-\psi)\sin(\pi/2\alpha)(\theta+\pi+\psi)]^{-1}$$

which is Sommerfield's result.

Notice, however, that this amplitude has two singularities in the physical region of θ for fixed ψ. In the theory of electromagnetic diffraction these directions in θ are called the boundaries of geometric optics. They are the directions in θ which can arise from a set of specular reflections of the plane wave incident from the direction ψ. In the particle language they are the singularities of Rubin, Tiktopolis, and Sugar[6] which arise in the three-particle problem when an incoming and outgoing state are connected by a sequence of two-particle interactions each of which individually conserves energy and momentum and which is kinematically allowed by the order of the sequence.

When a probability or a flux of particles is computed it will depend quadratically upon the scattering amplitude. Unitarity will depend somehow upon the integral of the square of the amplitude, but in its present form the square of the amplitude is not an integrable function. We must see how to go around the poles to preserve unitarity.

In our formalism it is easy to see how to do this. Referring again to Fig. 1 the contour integral around the closed path (C_3) will yield plane waves which extend to infinity. We want to eliminate these plane waves from unitarity considerations. Since the closed contour covers the real axis from $0 < w < \pi$ we can do this by displacing all of the poles into the lower half-plane, which makes them exponentially decay at infinity. These plane waves will then not contribute to either the flux or the probability.

Thus

$$\psi_3(w) = -\frac{\sin(\pi)\psi/\alpha}{4\alpha} \times \frac{1}{\sin(\pi/2\alpha)(w-\pi-\psi+i\epsilon)\sin(\pi/2\alpha)(w-\pi+\psi+i\epsilon)}.$$

This displacement of the poles, however, has destroyed the symmetry which made the incoming cylindrical wave zero. We compute this incoming cylindrical wave in a consistent approximation for small ϵ and find

$$\psi_3(\pi+w) - \psi_3(\pi-w) = -\epsilon\pi\frac{\sin(\pi\psi/\alpha)}{\alpha^2} \times \left(\frac{\sin(\pi w/\alpha)}{[\cos(\pi w/\alpha)-\cos(\pi\psi/\alpha)]^2 + (\pi\epsilon/\alpha)^2\sin^2(\pi w/\alpha)}\right),$$

which approaches $\delta(w-\psi)$ (apart from an inessential phase factor) as $\epsilon \to 0$.

This procedure works also in general. When we invent a function like ψ_3 of Sec. V, we also displace the poles slightly into the exponential decay region. The result is to give us an incoming cylindrical wave whose θ dependence approaches a delta function as $\epsilon \to 0$. The outgoing amplitude squared is then interpreted as the outgoing particle flux given unit incident flux.

VII. FORMATION OF THE SCATTERING MATRIX

Unfortunately our array of amplitudes is not the end of the story. There remains a problem which is common to all three-particle or multichannel problems. This array is almost the scattering matrix, but in its present form it is not unitary. This absence of unitary is due to the ability of the system to change channels during the course of an interaction.

We will not reproduce here the arguments which tell one how to calculate properly the scattering matrix, but we will simply indicate how the arguments are made.

The problem is to connect the time independent results, which are the amplitudes in the array A, with a wavepacket outlook. One way to do this is to calculate the scattering, rearrangement, and breakup of wavepackets originating in the various channels, using wavepackets which are very sharp in momentum space and are very wide in configuration space. The time dependent theory of such wavepackets involves only time-independent amplitudes plus certain kinematical factors. One insists that probability be conserved in these wavepackets and that scattering be linear in the sense that amplitudes for distinct processes add.

A typical relation obtained by this prescription would be

$$1 = |A(0|0)|^2 + \left(\frac{k^2+h^2}{h^2}\right)^{1/2}\left(\frac{g^2}{k^2+g^2}\right)^{1/2}|A(\alpha|0)|^2 + \left(\frac{g^2}{k^2+g^2}\right)^{1/2}\frac{1}{\pi}\int_0^\alpha |A(0|0)|^2 d\theta.$$

This relation would arise when one guarantees unitarity of an amplitude which contains only incoming bound waves along $\theta = 0$. Guaranteeing that unitarity is maintained between a sum of wavepacket solutions which involve both an incoming wave along $\theta = 0$ and one along $\theta = \alpha$ produces a similar relation for the corresponding amplitudes conditional on $\theta = \alpha$. In addition a second type of result is obtained:

$$0 = A(0|0)A^*(0|\alpha)\left(\frac{k^2+g^2}{g^2}\right)^{1/2} + A(\alpha|0)A^*(\alpha|\alpha) \times \left(\frac{k^2+h^2}{h^2}\right)^{1/2} + \frac{1}{\pi}\int_0^\alpha A(\theta|0)A^*(\theta|\alpha)d\theta.$$

If we systematically cover all of the possibilities, we find that the array A is unitarized by similarity transforming with a diagonal matrix of the form

$$D(0|0) = [g^2(k^2+g^2)]^{1/4},$$
$$D(\alpha|\alpha) = [h^2(k^2+h^2)]^{1/4},$$
$$D(\theta|\theta) = \sqrt{\pi}.$$

The scattering matrix is

$$S = DAD^{-1} = \begin{bmatrix} A(0|0)\left(\frac{g^2}{k^2+g^2}\right)^{1/4}\left(\frac{k^2+h^2}{h^2}\right)^{1/4}A(\alpha|0) & \frac{1}{\sqrt{\pi}}\left(\frac{g^2}{k^2+g^2}\right)^{1/4} & A(\theta|0) \\ \left(\frac{h^2}{k^2+h^2}\right)^{1/4}\left(\frac{k^2+g^2}{g^2}\right)^{1/4}A(0|\alpha) & A(\alpha|\alpha)\frac{1}{\sqrt{\pi}}\left(\frac{h^2}{k^2+h^2}\right)^{1/4} & A(\theta|\alpha) \\ \sqrt{\pi}\left(\frac{k^2+g^2}{g^2}\right)^{1/4}A(\theta|\psi) & \sqrt{\pi}\left(\frac{k^2+h^2}{h^2}\right)^{1/4}A(\alpha|\psi) & A(\theta|\psi) \end{bmatrix}.$$

This matrix is not quite what one would ordinarily call the S matrix. One can see this by interpreting the diagonal elements of this matrix. A diagonal element of this matrix tells the amplitude for an incoming wave to go out along the same direction, that is, the amplitude that all three-particle momenta and the order of the particles along the line remain the same. One would ordinarily want the diagonal elements of the S matrix to give the amplitude that a particle which came in at θ went out at $\pi + \theta$, or the amplitude that all particle momenta remain the same, but that the order along the line is reversed. For impenetrable particles this type of S matrix would have zero diagonal elements; thus we find it more convenient to use the first type.

One knows from the usual general considerations of hermiticity of the Hamiltonian operator that unitarity is guaranteed, but it is not obvious that the amplitudes constructed from the difference equations have this property. We have put the proof of the difference equation unitarity in the Appendix.

The authors wish to emphasize that this unitary form for the scattering matrix may be manipulated in the usual way. Unitary changes of basis are quite acceptable and are equivalent re-expressions of this matrix even though such changes of basis mean a coherent mixing of channel states. Many of the standard references give the impression that this "major channel" basis is preferred, and that one must be particularly careful about transforming to some other basis which coherently mixes channel states. All of the care which need be taken has been taken by writing the matrix in unitary form.

VIII. PROBABILITIES AND PARTICLE FLUXES

We have given a complete prescription for the calculation of the elements of the scattering matrix. The absolute squares of these matrix elements are to be interpreted as a probability flux, that is, the number of events per second, given one event per second in the initial state.

In this section we will compute the absolute squares of the matrix elements of the previous section. We do this because the squares of the matrix elements are very much simpler functions than are the matrix elements themselves.

Let us calculate the probabilities associated with the first row of our scattering matrix. In this case there is no distinction between a probability of an event and the flux associated with that event since the incoming wave is bound in the channel. In terms of the solution to the difference equations the probabilities are

$$P(0|0) = |A(0|0)|^2$$

$$= |\tilde{G}_0(i\beta)\psi_1(i\beta)|^2/|H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta)|^2,$$

$$P(\alpha|0) = |A(\alpha|0)|^2 \tan(i\beta)/\tan(i\gamma)$$
$$= [\tan(i\beta)/\tan(i\gamma)]|\tilde{H}_0(-\alpha + i\gamma) \times \psi_1(-\alpha + i\gamma)|^2/|H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta)|^2,$$

$$P(\theta|0) = [\tan(i\beta)/\pi i]|A(\theta|0)^2$$
$$= [\tan(i\beta/\pi i]|G_0(\theta)\psi_1(\theta) + \tilde{H}_0(-\theta) \times \psi_1(-\theta)|^2/|H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta)|^2$$
$$= [\tan(i\beta)/\pi i]|G_0(\theta)|^2|\psi_1(\theta) - Q(\theta) \times \psi_1(-\theta)|^2/|H_0(\pi + i\beta)\tilde{\psi}_1(\pi + i\beta)|^2,$$

where

$$H_0(-\theta) = -G_0(2\pi + \theta) = -Q(\theta)G_0(\theta).$$

$Q(\theta)$ is a function with period 2α, by the results of Sec. III.

We appeal to our explicit representations for G_0 and H_0 to find the functions $G_0{}^*G_0, H_0{}^*H_0$, and Q. Note that

$$f^*(w,\beta)f(w,\beta) = f(w,-\beta)f(w,\beta) = 1,$$

and from this it follows that

$$G_0{}^*(w)G_0(w) = H_0{}^*(w)H_0(w)$$
$$= \sin(\pi/2\alpha)(w - \pi - i\beta)\sin(\pi/2\alpha)(w - \pi + i\beta) \times \sin(\pi/2\alpha)(w - \pi - \alpha - i\gamma) \times \sin(\pi/2\alpha)(w - \pi - \alpha + i\gamma).$$

Calculation of the function Q is straightforward and yields

$$Q = \frac{\sin(\pi/2\alpha)(w - i\beta)\sin(\pi/2\alpha)(w + \pi + i\beta)}{\sin(\pi/2\alpha)(w + i\beta)\sin(\pi/2\alpha)(w + \alpha + i\gamma)} \times \frac{\sin(\pi/2\alpha)(w + \alpha - i\gamma)\sin(\pi/2\alpha)(w + \pi + \alpha + i\gamma)}{\sin(\pi/2\alpha)(w - \pi - i\beta)\sin(\pi/2\alpha)(w - \pi + \alpha + i\gamma)}.$$

We may also calculate $|\tilde{G}_0(i\beta)|^2$ in terms of these functions:

$$|\tilde{G}_0(i\beta)|^2 = [\lim_{w\to i\beta}(w - i\beta)G_0(w)][\lim_{s^*\to -i\beta}(s + i\beta)G_0{}^*(s^*)].$$

Let $s^* = -w$, and then

$$|\tilde{G}_0(i\beta)|^2 = -\lim_{w\to i\beta}(w - i\beta)^2 G_0(w)G_0{}^*(-w).$$

Using the difference equations and the symmetry relation between G_0 and H_0 we obtain

$$|\tilde{G}_0(i\beta)|^2 = \lim_{w\to i\beta}(w - i\beta)^2 X^*(w)G_0(w)G_0{}^*(2\pi + w).$$

Evaluating the results of the function X^* explicitly,

we have

$$|\tilde{G}_0(i\beta)|^2 = 2\tan i\beta \lim_{w\to i\beta}(w - i\beta)Q^*(w)G_0(w)G_0{}^*(w).$$

Similarly

$$|\tilde{H}_0(-\alpha + i\gamma)|^2 = 2\tan i\gamma \lim_{w\to -\alpha+i\gamma}(w + \alpha - i\gamma)Q^*(w) \times H_0(w)H_0{}^*(w).$$

Thus all of the squares of the elements of this row of our scattering matrix may be found in terms of the known functions $\psi_1, G_0{}^*G_0$, and Q. The corresponding statement is true as well for the other rows. After some tedious calculation we find

$$P(0|0) = \frac{\sin^2(\pi i\beta/\alpha)\cos(\pi/2\alpha)[\pi + i(\beta-\gamma)]\cos(\pi/2\alpha)(\pi - i(\beta-\gamma)]}{\cos^2(\pi i/\alpha)(\beta-\gamma)\sin(\pi/2\alpha)(2i\beta-\pi)\sin(\pi/2\alpha)(2i\beta+\pi)},$$

$$P(\alpha|0) = -\frac{\sin(\pi^2/2\alpha)\sin(\pi i\gamma/\alpha)\sin(\pi i\beta/\alpha)}{\cos^2(\pi i/2\alpha)(\gamma-\beta)\cos(\pi/2\alpha)[\pi + i(\beta+\gamma)]\cos(\pi/2\alpha)[\pi - i(\beta+\gamma)]},$$

$$P(\theta|0) = -\frac{i}{4\alpha}\frac{\sin^2(\pi^2/2\alpha)\sin^2(\pi\theta/\alpha)\cos(\pi i/2\alpha)(\beta+\gamma))\cos(\pi/2\alpha)(\pi - i(\beta-\gamma))\sin(\pi i\beta/\alpha)}{\cos(\pi i/2\alpha)(\beta-\gamma)D(\theta, i\beta)B(\theta)},$$

$$P(0|\alpha) = P(\alpha|0),$$

$$P(\alpha|\alpha) = \frac{\sin^2(\pi i\gamma/\alpha)\cos(\pi/2\alpha)[\pi + i(\gamma-\beta)]\cos(\pi/2\alpha)[\pi - i(\gamma-\beta)]}{\cos^2(\pi i/2\alpha)(\beta-\gamma)\sin(\pi/2\alpha)(\pi + 2i\gamma)\sin(\pi/2\alpha(2i\gamma-\pi)},$$

$$P(\theta|\alpha) = -\frac{i}{4\alpha}\frac{\sin^2(\pi^2/2\alpha)\sin^2(\pi\theta/\alpha)\cos(\pi i/2\alpha)(\beta+\gamma)\cos(\pi/2\alpha)[\pi + i(\gamma-\beta)]\cos(\pi/2\alpha)[\pi - i(\gamma-\beta)]\sin(\pi i\gamma/\alpha)}{\cos(\pi i/2\alpha)(\gamma-\beta)D(\alpha-\theta, i\gamma)B(\theta)},$$

$$P(0|\psi) = P(\psi|0),$$

$$P(\alpha|\psi) = P(\psi|\alpha),$$

$$P(\theta|\psi) = \frac{1}{(16\alpha)^2}\sin^2\frac{\pi\theta}{\alpha}\sin^2\frac{\pi\psi}{\alpha}\left[\sin\frac{\pi^2}{\alpha}\cos\frac{\pi\theta}{\alpha}\cos\frac{\pi\psi}{\alpha} + \sin\frac{\pi^2}{2\alpha}\left(\cos\frac{\pi\theta}{\alpha} + \cos\frac{\pi\psi}{\alpha}\right)\right.$$
$$\times\left(\cos\frac{\pi}{2\alpha}(\pi + 2i\gamma) + \cos\frac{\pi}{2\alpha}(\pi + 2i\beta)\right) + \tfrac{1}{2}\sin\frac{\pi^2}{\alpha}\left(1 - \cos\frac{\pi^2}{\alpha} - 2\cos\frac{\pi}{2\alpha}(\pi + 2i\gamma)\cos\frac{\pi}{2\alpha}(\pi + 2i\beta)\right)\Big]$$
$$\times [D^2(\theta,\psi)B(\theta)B(\psi)]^{-1},$$

where

$$D(\theta,\psi) = \sin(\pi/2\alpha)(\theta + \pi + \psi)\sin(\pi/2\alpha)(\theta + \pi - \psi)\sin(\pi/2\alpha)(\theta - \pi + \psi)\sin(\pi/2\alpha)(\theta - \pi - \psi),$$

$$B(\theta) = \sin(\pi/2\alpha)(\theta + i\beta)\sin(\pi/2\alpha)(\theta - i\beta)\cos(\pi/2\alpha)(\theta + i\gamma)\cos(\pi/2\alpha)(\theta - i\gamma).$$

Many interesting properties of the solutions in this three-particle problem can be obtained by looking at the structure of these probabilities as analytic functions.

First we see that there is a particular value of three-particle energy for which the probability $P(\alpha|0)$ becomes infinite. That value of three-particle energy occurs when the factor $\cos(\pi/2\alpha)[\pi - (\beta+\gamma)]$ vanishes in the denominator. This signals the presence of a three-particle bound state when

$$i(\beta + \gamma) = \pi - \alpha$$

and hence

$$\sin i\beta\cos i\gamma + \cos i\beta\sin i\gamma = +\sin\alpha$$

or

$$\frac{ig}{k}\left(\frac{1+g^2}{k^2}\right)^{1/2} + \frac{ih}{k}\left(\frac{1+h^2}{k^2}\right)^{1/2} = \sin\alpha.$$

The solutions to this equation give the binding energy which can only be satisfied when k is pure imaginary.

A second class of interesting results may be seen by looking at probabilities such as $P(\theta|0)$. Here the function $D(\theta, i\beta)$, which accounts for the three-particle kinematic singularities (i.e., those of Rubin, Sugar, and Tiktopolis) appears as a factor in the denominator. As the energy gets large, β and γ become small, pushing these poles near to the real axis. This causes very large probabilities for certain preferred particle momenta in the outgoing state. These preferred momenta are just the ones which would be obtained if one calculated the result of a sequence of kinematically possible collisions between three particles where two of the three particles (i.e., the two in the initial bound state) are nearly at rest with respect to one another.

Of course, the function $P(\psi|\theta)$ has the same sort of kinematic singularity behaviour as the particular case of Sec. VI, and the problems are kinematically identical. The function $B(\theta)$ has singularities for complex values of θ also, but at high energy those singularities will not be felt because they approach the real ones at a point where the scattering probability is zero (i.e., the directions $\psi = 0, \psi = \alpha$).

IX. CALCULATION OF REARRANGEMENT PROBABILITIES BELOW THE THREE-PARTICLE BREAKUP THRESHOLD

If the energy is less than zero, no three-particle breakup can occur, that is, there must be at least one bound pair of particles in both the initial and final states. Below this threshold we must recalculate the rearrangement probabilities. One might think that these probabilities are analytic continuations of the corresponding probabilities above threshold, but we shall see that this is not the case.

When we calculated probabilities in the previous section we made use of the theorem

$$G^*(w)G(w) = G^*(w + 2\alpha)G(w + 2\alpha),$$

which assumed that the parameters β and γ were real. For k^2 less than zero we define λ to be a positive real number such that

$$k = i\lambda.$$

Since

$$\sin i\beta = ig/k = g/\lambda,$$

we see that β must be complex, which invalidates the theorem, since under these conditions G and G^* satisfy the same difference equation. We shall choose

$$i\beta = \tfrac{1}{2}\pi + iv,$$

which implies

$$\cosh v = g/\lambda,$$

and

$$v = \ln[g/\lambda + (g^2/\lambda^2 - 1)^{1/2}]$$

Thus v tends to zero at the two-particle g channel bound state threshold and tends to infinity at the three-particle breakup threshold.

Similarly, we let

$$i\gamma = \tfrac{1}{2}\pi + it, \quad \cosh t = h/\lambda,$$

$$t = \ln[h/\lambda + h^2/\lambda^2 - 1)^{1/2}].$$

In this energy domain we must have no contribution from the steepest descent point at $w = \pi$ because such a contribution would lead to an exponentially increasing wavefunction at infinity. Thus we must maintain the relation between G and H:

$$G(\pi + w) = H(\pi - w).$$

There will be two solutions in this domain and we shall label these solutions in the same way as in Sec. VI. Of course, we need not worry about amplitudes like $A(\theta|0)$ because these are amplitudes for exponentially decaying waves at infinity and hence do not contribute to the asymptotic probabilities.

Substituting the above expressions for β and γ into the appropriate amplitude expressions gives

$$A(0|0) = \tilde{G}_0(\tfrac{1}{2}\pi + iv)\psi_1(\tfrac{1}{2} + iv)/H_0(\tfrac{3}{2}\pi + iv)\tilde{\psi}_1(\tfrac{3}{2}\pi + iv),$$

$$A(\alpha|0) = \tilde{H}_0(-\alpha + \tfrac{1}{2}\pi + it)\psi_1(-\alpha + \tfrac{1}{2}\pi + it)/H_0(\tfrac{3}{2}\pi + iv)\tilde{\psi}_1(\tfrac{3}{2}\pi + iv),$$

$$A(0|\alpha) = \tilde{G}_0(\tfrac{1}{2}\pi + iv)\psi_2(\tfrac{1}{2}\pi + iv)/G_0(\tfrac{3}{2}\pi + \alpha + it)\tilde{\psi}_2 \times (\tfrac{3}{2}\pi + \alpha + it),$$

$$A(\alpha|\alpha) = \tilde{H}_0(\tfrac{1}{2}\pi - \alpha + it)\psi_{2\frac{1}{2}}\pi - \alpha + it)/G_0(\tfrac{3}{2}\pi + \alpha + it)\tilde{\psi}_2(\tfrac{3}{2}\pi + \alpha + it).$$

As before, the probabilities are the absolute squares of these amplitudes multiplied by a kinematical factor. We also substitute the expressions for β and γ into these kinematical factors and obtain

$$P(0|0) = |A(0|0)|^2,$$

$$P(\alpha|0) = (\cot iv/\cot it)|A(\alpha|0)|^2,$$

$$P(\alpha|\alpha) = |A(\alpha|\alpha)|^2,$$

$$P(0|\alpha) = (\cot it/\cot iv)|A(0|\alpha)|^2.$$

From the difference equations we now note that, under conditions of negative energy,

$$G(w)G^*(\pi + w) = G(w + 2\alpha)G^*(w + 2\alpha + \pi).$$

We shall express all of our probabilities for negative energy in terms of this periodic function. For example,

$$|\tilde{G}(\tfrac{1}{2}\pi + iv)|^2 = \lim_{w\to\pi/2+iv}(w - \tfrac{1}{2}\pi + iv)G(w) \lim_{w^*\to\pi/2-iv}(w^* - \tfrac{1}{2}\pi + iv)G^*(w^*),$$

and making an appropriate change of variable in each limit we obtain

$$|\tilde{G}(\tfrac{1}{2}\pi + iv)|^2 = -\lim_{w\to iv}(w - iv)^2G(\tfrac{1}{2}\pi + w)G^*(\tfrac{1}{2}\pi - w).$$

Using the connection between G and H, we have

$$|\tilde{G}(\tfrac{1}{2}\pi + iv)|^2 = \lim_{w\to iv}(w - iv)^2G(\tfrac{1}{2}\pi + w)H^*(\tfrac{3}{2}\pi + w).$$

Now we use the difference equations to eliminate H:

$$|\tilde{G}(\tfrac{1}{2}\pi + iv)|^2 = \lim_{w\to iv}(w - iv)^2X^*(\tfrac{3}{2}\pi + w)G(\tfrac{1}{2}\pi + w) \times G^*(\tfrac{3}{2}\pi + w).$$

Extracting the pole in X, we finally obtain

$$|\tilde{G}(\tfrac{1}{2}\pi + iv)|^2 = 2\cot iv \lim_{w\to iv}(w - iv)G(\tfrac{1}{2}\pi + w) \times G^*(\tfrac{3}{2}\pi + w),$$

which achieves the desired objective. By a similar process we can express all of the other probabilities in terms of this function.

In a straightforward but tedious way one calculates from the gamma function representation of $G_0(w)$

$$G_0(w)G_0{}^*(\pi + w) =$$
$$\sin(\pi/2\alpha)(w + \tfrac{1}{2}\pi - iv)\cos(\pi/2\alpha)(w + \tfrac{1}{2}\pi - it)$$
$$\times \sin(\pi/2\alpha)(w - \tfrac{3}{2}\pi - iv)\cos(\pi/2\alpha)(w - \tfrac{3}{2}\pi - it)$$
$$\times \sin(\pi/2\alpha)(w - \tfrac{1}{2}\pi + iv)\cos(\pi/2\alpha)(w - \tfrac{1}{2}\pi + it)$$
$$\times \sin(\pi/2\alpha)(w - \tfrac{1}{2}\pi - iv)\cos(\pi/2\alpha)(w - \tfrac{1}{2}\pi - it)]^{-1}.$$

The final results for the probabilities are

$$P(0|0) = P(\alpha|\alpha) = \frac{\cos(\pi/2\alpha)[\pi + i(v - t)]\cos(\pi/2\alpha)[\pi - i(v - t)]}{\cos(\pi/2\alpha)[\pi + i(v + t)]\cos(\pi/2\alpha)[\pi - i(v + t)]}$$

$$\times \frac{\cos^2(\pi i/2\alpha)(t+v)}{\cos^2(\pi i/2\alpha)(v+t)},$$

$$P(0|\alpha) = P(\alpha|0) = \frac{\sin^2(\pi^2/2\alpha)\sin(\pi it/2\alpha)\sin(\pi iv/2\alpha)}{\cos^2(\pi i/2\alpha)(t-v)\cos(\pi/2\alpha)[\pi + i(v+t)]} \times \{\cos(\pi/2\alpha)[\pi - i(v+t)]\}^{-1}.$$

These probability functions have the following properties:

(1) their sum is unity;

(2) if either v or t is zero, then $P(0|\alpha) = 0$ and $P(0|0) = 1$;

(3) these functions approach the same value at zero energy as do the corresponding functions above the three-particle breakup threshold.

Property (1) simply expresses the fact that only rearrangement or recoil is possible below the three-particle breakup threshold. Property (2) says that rearrangement from a channel of greater to a channel of lesser binding energy is impossible at the threshold of the lesser binding. Property (3) is a consequence of the analyticity of the amplitudes.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Australian Research Grants Committee.

APPENDIX: DEMONSTRATION OF UNITARITY FROM THE DIFFERENCE EQUATIONS

The evolution of a time-dependent wavepacket, as outlined in Sec. VII, asserts that the following relation among difference equation solutions must obtain in order that probability be conserved:

$$|\tilde{H}(\pi + i\beta)|^2\left(\frac{k^2+g^2}{g^2}\right)^{1/2} + |\tilde{G}(\alpha+\pi+i\gamma|^2 \times \left(\frac{k^2+h^2}{h^2}\right)^{1/2} + \frac{1}{\pi}\int_0^\alpha |G(\pi+\theta) + H(\pi-\theta)|^2 d\theta = |\tilde{G}(i\beta)|^2\left(\frac{k^2+g^2}{g^2}\right)^{1/2} + \frac{1}{\pi}\int_0^\alpha |G(\theta) + H(-\theta)|^2 d\theta,$$

where the term on the left represents the total incoming probability and the term on the right represents total outgoing probabilities for general G and H.

We demonstrate in this appendix that this relation is a consequence of the difference equation and the condition on the allowed poles of G and H.

First we rewrite the terms involving $\tilde{G}$ and $\tilde{H}$ so that the kinematical factor does not appear. This is done in the same way as in Sec. VIII. This yields

$$-2iG(\pi + i\beta)H^*(\pi - i\beta) - 2iG(\pi + \alpha + i\gamma)H^*(\pi - \alpha - i\gamma) + \frac{1}{\pi}\int_0^\alpha |G(\pi+\theta) + H(\pi-\theta)|^2 d\theta = -2iG(i\beta)H^*(-i\beta) - 2iG(-\alpha+i\gamma)H^* \times (-\alpha+i\gamma) + \frac{1}{\pi}\int_0^\alpha |G(\theta) + H(-\theta)|^2 d\theta.$$

A series of manipulations of the dummy variable allows us to rewrite the integrals on the right as

$$\int_0^\alpha |G(\theta) + H(-\theta)|^2 = \int_0^\alpha [G(\theta)G^*(\theta) + H(\theta-\alpha)H^*(\theta-\alpha)]d\theta + \int_0^\alpha [G(\theta)H^*(-\theta) + H(\theta-\alpha)G^*(\alpha-\theta)]d\theta.$$

The difference equations tell us that both of the functions appearing under the integral sign are period α. We may therefore rewrite the second of these integrals as

$$\int_0^\alpha [G(\ )H^*(-\theta) + H(\theta-\alpha)G^*(\alpha-\theta)]d\theta = \int_c [G(\theta)H^* \times (-\theta) + H(\theta-\alpha)G^*(\alpha-\theta)]d\theta.$$

If $\theta = x + iy$, the contour c runs:

Path 1: $x = 0$: y runs from infinity to 0.

Path 2: $y = 0$: x runs from 0 to α.

Path 3: $x = \alpha$: y runs from 0 to infinity.

Path 4: $y = \infty$: x runs from α to 0.

Path 4 does not contribute because of the boundedness condition on G and H. Path 1 exactly cancels Path 3 due to periodicity.

The pole condition on G and H states in Sec. III states that the only poles in this region are at the point $i\beta$ in the function $G(\theta)H^*(-\theta)$ and at the point $i\gamma$ in the function $H(\theta-\alpha)G^*(\alpha-\theta)$. These two pole terms exactly cancel the residue terms on the right-hand side. A similar cancelation occurs on the left-hand side. Unitarity is established provided that

$$\int_0^\alpha [G(\pi+\theta)G^*(\pi+\theta) + H(\pi+\theta-\alpha)H^*(\pi+\theta-\alpha)] \times d\theta = \int_0^\alpha [G(\theta)G^*(\theta+\alpha)H^*(\theta-\alpha)]d\theta,$$

which is manifestly true since the functions in the integral are period α.

[1] J. B. McGuire, J. Math. Phys. 5, 622 (1964).
[2] W. E. Williams, Proc. Roy. Soc. (London) **A252**, 376 (1959).
[3] H. M. Nussenzveig, Proc. Roy. Soc. (London) **A264**, 408 (1961).
[4] G. D. Birkhoff, Trans. Amer. Math. Soc. **12**, 243 (1911); see also N. E. Nörlund, *Differenzenrechnung* (Springer-Verlag, Berlin, 1924).
[5] R. Jost, Z. Angew. Math.-Phys. **6**, 316 (1955).
[6] M. Rubin, R. Sugar, and G. Tiktopoulos, Phys. Rev. **146**, 1130 (1966).

The Convex Structure of the Set of N-Representable Reduced 2-Matrices

R. M. Erdahl
Department of Mathematics, Queen's University, Kingston, Ontario, Canada
(Received 23 November 1971)

It is shown that in the finite-dimensional case the extreme points of the convex set of all N-representable 2-matrices D^2 are exposed points and that for the case $N = 3$ and $N = 4$ the preimage of an exposed point under the contraction map is unique. In addition, the convex structure of both D^2 and its associated polar is given in the case where the 1-rank is equal to $N + 2$. A long list of exposed points of D^2 is given which includes all previously published examples as well as some new ones. However, this list is shown not to be exhaustive.

1. INTRODUCTION

The idea of replacing the wavefunction by the second-order reduced density matrix in many-particle physics, and thereby greatly reducing the dimensionality of the problem has intrigued many-particle theorists for some time. The problem of giving a useful characterization of the set of all ensemble representable second-order reduced density matrices, which we denote by $D^2(H^N)$ using a notation we introduce in the second section, has been called the N-representability problem by Coleman.[1] $D^2(H^N)$ forms a subset of the set of all positive operators with trace equal to one on H^2, the Hilbert space of all two-fermion wave functions. It is a convex set and, in addition, compact in the trace topology. Basically there have been two approaches to the N-representability problem. Firstly, there has been much work on trying to obtain necessary and sufficient conditions for deciding whether a given two-particle operator is a member of $D^2(H^N)$. Such a problem has been solved, yielding a very elegant solution,[1,2] in the case of $D^1(H^N)$, the set of all ensemble representable first order density matrices. For the second-order density matrix many necessary conditions are known, but not much is known about sufficient conditions except that they will be difficult to obtain. However, further study of the N-representability conditions will undoubtedly lead to useful lower bound methods in many-particle physics.

This paper deals chiefly with a second approach, that of listing the extreme points of $D^2(H^N)$. As $D^2(H^N)$ is a compact convex set, by the Krein–Milman theorem[3] it is the convex closure of its extreme points. Thus an enumeration of the extreme points of $D^2(H^N)$ would serve to characterize it. In Theorem 7.3 we give a list of some extreme points of $D^2(H^N)$ which includes all previously known examples plus some new ones.

Interest in the extreme points of $D^2(H^N)$ is justified since they can always be used to described the ground state of a two-particle operator. In fact, we will see in Sec. 3 that the extreme points of $D^2(H^N)$ are also exposed points. If the ground state *wavefunction* of a two-particle operator is nondegenerate, then the corresponding density matrix is an exposed point. Moreover, if $d \in D^2(H^N)$ is an exposed point, then it will be the *unique* ground state *density matrix* for some two-particle operator. The theory is a bit awkward as it now stands since we have not excluded (except in the case where $N = 3$ or 4) the possibility that an exposed point of $D^2(H^N)$ corresponds to two or more orthogonal ground state wavefunctions. If such were the case, this degeneracy, occurring with the wavefunction, could not be removed by any two-particle operator. There is no such problem with $D^1(H^N)$. There, all of the exposed points are covered by unique ground state wavefunctions (up to an arbitrary phase factor). Moreover, we know that if the ground state of any one-particle operator is nondegenerate then it must correspond to a Slater determinant. Requiring that an element $d \in D^1(H^N)$ be an exposed point places severe restrictions on it. The exposed points of $D^1(H^N)$ are merely the one-densities corresponding to arbitrary Slater determinants. Presumably a similar situation holds in the case of $D^2(H^N)$, and the exposed points there form a very restricted subset of $D^2(H^N)$. More details about the connection between the ground states of two-particle operators and the exposed points of $D^2(H^N)$ are contained in the next section.

Many of our results are stated for $D^p(H^N)$, the set of all N-representable pth-order reduced density matrices, where p is arbitrary. In most situations we deal with the finite-dimensional case, that is, when $\dim H^1$ (the dimension of the space of all one-particle functions), $\dim H^2, \ldots, \dim H^N$ are all finite. For that reason topological considerations will not play an important role. When a result holds in the infinite-dimensional case we will draw that to the reader's attention. For a discussion of the N-representability problem and its associated polar problem in the infinite-dimensional case, the reader is referred to Kummer's paper[4] in which particular care is given to some of the topological problems that arise. There it is shown that the infinite-dimensional case can be approximated by the finite-dimensional one.

In Sec. 3 we establish the result that in the finite-dimensional case all of the extreme points of $D^p(H^N)$ are exposed.

In Sec. 4 we consider a particle–hole duality which will prove to be a useful tool in some of the later sections. A similar construction was given by Ruskai.[5] However, our treatment is simpler. In addition, our treatment can easily be applied to obtain information about the polar.

In Sec. 5 we give a complete description of the convex structure of both $D^p(H^N)$ and its polar $\tilde{D}^p(H^N)$, for the case where the dimension of the one-particle basis is equal to $N + p$. Previously, Ruskai has considered this problem.[5] Some of our results appear in her paper. Yoseloff and Kuhn[6] have considered a special case of this problem.

In Sec. 6 we obtain the result that in the case where the particle number N is 3 or 4, there is a unique wavefunction Ψ (up to an arbitrary phase factor) which covers an exposed point of $D^2(H^N)$.

In Sec. 7 we give in Theorem 7.3 an extensive list of exposed points of $D^2(H^N)$, many of which are new. All previously published examples of exposed points are contained in the statement of the theorem.

In Sec. 8 we give two examples of exposed points which are not of the type considered in Sec. 7, thus establishing that Theorem 7.3 does not exhaust the exposed points of $D^2(H^N)$.

2. SOME PRELIMINARY DEFINITIONS

If Ψ is a normalized antisymmetric N-particle wavefunction, then its *p-density* $D^p(\Psi)$ is a member of $L(H^p)$, the bounded linear operators on H^p. Here H^p denotes the physically relevant antisymmetric subspace of the p-fold tensor product of some Hilbert space H^1 with itself p times. The superscripts denote the particle number. $D^p(\Psi)$ may be represented as an integral operator, whose kernel is given by

$$D^p(\Psi)(12\cdots p\,|\,1'2'\cdots p') = \int_{p+1\cdots N}\bar{\Psi}(12\cdots N)\Psi(1'\cdots p'p+1\cdots N) = \sum\lambda_i^p\,\bar{\alpha}_i^p(12\cdots p)\alpha_i^p(1'2'\cdots p'),$$

where the λ_i^p are the eigenvalues and the α_i^p the eigenfunctions of $D^p(\Psi)$. The corresponding N-density $D^N(\Psi)\in L(H^N)$ is merely the projector onto Ψ.

With a few exceptions all of the results of this paper deal with the case where H^1 is finite-dimensional (much of Sec. 6 and 7 apply to the infinite-dimensional case; Proposition 7.2 deals exclusively with the infinite-dimensional case). We, therefore, make the assumption, that unless stated otherwise, H^1 is finite-dimensional. As was mentioned in the introduction, Kummer[4] has shown that the infinite-dimensional case can be approximated by the finite-dimensional one.

We denote by $D^p(H^N)$ the set of ensemble N-representable p densities. $D^p(H^N)$ is given by $\overline{\text{CONV}}\{D^p(\Psi)\,|\,\Psi\in H^N\}$, where $\overline{\text{CONV}}$ denotes the convex closure. The closure may be taken in the topology determined by the trace. If V is a subspace of H^N, then $D^p(V)$ denotes the convex set $\overline{\text{CONV}}\{D^p(\Psi)\,|\,\Psi\in V\}$. The elements of $D^p(H^N)$ are positive operators with trace equal to one. However, all positive operators of trace one are not members of $D^p(H^N)$. Given the trace topology, $D^p(H^N)$ is a compact convex set. The N-representability problem amounts to giving a useful characterization of this convex set. The compact convex set $D^N(H^N)$ consists of all positive operators in $L(H^N)$ with trace equal to 1.

The *pth-order transition density* corresponding to the states $\Psi_1,\Psi_2\in H^N$ is denoted by $D^p(\Psi_1,\Psi_2)$. It also may be represented as an integral operator, the kernel of which is

$$D^p(\Psi_1,\Psi_2)(12\cdots p\,|\,1'2'\cdots p') = \int_{p+1\cdots N}\bar{\Psi}_1(12\cdots N)\Psi_2(1'\cdots p'p+1\cdots N).$$

The linear map $L_N^p: L(H^N)\to L(H^p)$ defined by $L_N^p: D^N(\Psi)\to D^p(\Psi)$ and linear extension to all of $L(H^N)$ is Kummer's *contraction map*.[4] In both $L(H^N)$ and $L(H^p)$ one can introduce a scalar product by means of the trace, i.e., for $b,d\in L(H^N)$, $(b,d)\mapsto \text{Tr}(b^\dagger d)$. Thus we can define the adjoint of the contraction map Γ_p^N a map from $L(H^p)$ into $L(H^N)$, by the formula

$$\text{Tr}(\Gamma_p^N(b)^\dagger d) = \text{Tr}(b^\dagger L_N^p(d)),\qquad b\in L(H^p),\ d\in L(H^N).$$

We will refer to Γ_p^N as the *expansion map*. Kummer[4] has given the following formula for Γ_p^N:

$$\Gamma_p^N(b) = A_N(b\otimes I^{N-p})A_N,\qquad b\in L(H^p),$$

where A_N is the antisymmetrizer. Possibly a more familiar formula for the expansion operator is given by

$$\Gamma_p^N(a_{i_1}^\dagger\cdots a_{i_p}^\dagger a_{j_1}\cdots a_{j_p}) = \binom{N}{p}^{-1}a_{i_1}^\dagger\cdots a_{i_p}^\dagger a_{j_1}^\dagger\cdots a_{j_p},\tag{2.1}$$

where we have given the action of Γ_p^N on a single basis element of $L(H^p)$ and used second quantization notation. In order to keep the notation uncluttered we have used $a_{i_1}^\dagger\cdots a_{i_p}^\dagger a_{j_1}\cdots a_{j_p}$ to denote both an element of $L(H^p)$ (left-hand side of the above equality) and $L(H^N)$ (right-hand side). A similar formula can be given for the contraction operator for certain special elements of $L(H^N)$:

$$L_N^p(a_{i_1}\cdots a_{i_p}a_{j_1}^\dagger\cdots a_{j_p}^\dagger) = \binom{r-p}{N}\binom{r-p}{N}^{-1}a_{i_1}\cdots a_{i_p}a_{j_1}^\dagger\cdots a_{j_p}^\dagger,\tag{2.2}$$

where r denotes the dimension of H^1. This formula can be derived by using (2.1) and the fact that L_N^p and Γ_p^N are adjoint to one another. For the case $p=2$ we can establish this result by showing that if one chooses K equal to $\binom{r-2}{N}\binom{r-2}{2}^{-1}$ then $\text{Tr}_N(\Gamma_2^N(b)^\dagger a_i a_j a_k^\dagger a_l^\dagger) = K\,\text{Tr}_2(b^\dagger a_i a_j a_k^\dagger a_l^\dagger)$ holds for all $b\in L(H^2)$. Tr_2 denotes the trace in H^2 and Tr_N the trace in H^N. But this formula will hold for all b if it holds for the basis elements $a_\delta^\dagger a_\gamma^\dagger a_\beta a_\alpha$ of $L(H^2)$. We must show that

$$\binom{N}{2}^{-1}\text{Tr}_N(a_\alpha^\dagger a_\beta^\dagger a_\gamma a_\delta a_i a_j a_k^\dagger a_l^\dagger) = K\,\text{Tr}_2(a_\alpha^\dagger a_\beta^\dagger a_\gamma a_\delta a_i a_j a_k^\dagger a_l^\dagger)$$

for all $\alpha,\beta,\gamma,\delta$. We note that both sides are nonzero only if $\alpha\neq\beta$ and $k\neq l$. Also both sides will be nonzero only if the index set $\{\gamma,\delta,i,j\}$ is equal to $\{\alpha,\beta,k,l\}$. Thus the formula will hold if

$$\binom{N}{2}^{-1}\text{Tr}_N(a_\alpha^\dagger a_\alpha a_\beta^\dagger a_\beta a_k a_k^\dagger a_l a_l^\dagger) = K\,\text{Tr}_2(a_\alpha^\dagger a_\alpha a_\beta^\dagger a_\beta a_k a_k^\dagger a_l a_l^\dagger)$$

$\alpha\neq\beta$, $k\neq l$. But both sides will be nonzero only if $\alpha,\beta\neq k,l$. Thus for the nonzero case all indices must be different. But then

$$\text{Tr}_N(a_\alpha^\dagger a_\alpha a_\beta^\dagger a_\beta a_k a_k^\dagger a_l a_l^\dagger) = \binom{r-4}{N-2}$$

and $\text{Tr}_2(a_\alpha^\dagger a_\alpha a_\beta^\dagger a_\beta a_k a_k^\dagger a_l a_l^\dagger) = 1$. Thus $K = \binom{N}{2}^{-1}\binom{r-4}{N-2} = \binom{r-2}{N}\binom{r-2}{2}^{-1}$ and formula (2b) holds.

We define a *p-particle* operator on H^N to be any element of $L(H^N)$ which can be expressed as a linear combination of operators of the form $a_{i_1}^\dagger\cdots a_{i_p}^\dagger a_{j_1}\cdots a_{j_p}$ where we have p annihilators followed by p creators. Thus our p-particle operators contain all operators which are normally referred to as constants, one, two, ..., p-particle operators. For example, the constants can be written as p-particle operators since $\binom{N}{p} = \sum_{i_1<\cdots<i_p}a_{i_1}^\dagger\cdots a_{i_p}^\dagger a_{i_p}\cdots a_{i_1}$. Formula (2b) tells us how L_N^p acts on p-particle operators in $L(H^N)$. From formula (2.1) it is clear that Γ_p^N is an injection of $L(H^p)$ onto the set of all p-

particle operators in $L(H^N)$. We will show later (Proposition 4.1) that the kernel of the contraction operator L_N^p consists of all elements of $L(H^N)$ which are orthogonal to the set of all p-particle operators in $L(H^N)$, where by "orthogonal" we mean, orthogonal with respect to the trace scalar product. Thus formula (2b) tells us how the contraction map behaves on the orthogonal complement of its kernel.

A useful notion in the theory of convex sets is that of the *polar*. Our definition of the polar of $D^p(H^N)$ is

$$\tilde{D}^p(H^N) = \{b \mid \mathrm{Tr}(b^\dagger d) \geq 0,\ \forall d \in D^p(H^N),\ \mathrm{Tr}(\Gamma_p^N(b)^\dagger \Gamma_p^N(b))^{1/2} = 1\}.$$

Note the normalization condition which is included in this definition. In some circumstances we will find it more convenient to talk about the *polar cone*, which we denote by $\mathrm{CONE}\tilde{D}^p(H^N)$. It is the cone generated by $\tilde{D}^p(H^N)$. Both $\tilde{D}^p(H^N)$ and $\mathrm{CONE}\tilde{D}^p(H^N)$ are convex sets; $\tilde{D}^p(H^N)$ is compact. A useful characterization of $\tilde{D}^p(H^N)$ is that it consists of all elements $b \in L(H^p)$ such that (*i*) $\Gamma_p^N(b) \geq 0$ and (*ii*) $\mathrm{Tr}(\Gamma_p^N(b)^\dagger \Gamma_p^N(b))^{1/2} = 1$ or, equivalently, $\tilde{D}^p(H^N) = \{b \mid \Gamma_p^N(b) \in D^N(H^N)\}$. Interest in the polar arises from the fact that if we take the polar of $\tilde{D}^p(H^N)$ we get back $D^p(H^N)$:

$$D^p(H^N) = \tilde{\tilde{D}}^p(H^N) = \{d \mid \mathrm{Tr}(d^\dagger b) \geq 0,\ \forall b \in \tilde{D}^p(H^N),\ \mathrm{Tr}(d^\dagger d)^{1/2} = 1\}.$$

This is an application of the bipolar theorem.[3] The extreme points of $\tilde{D}^p(H^N)$ form a complete set of N-representability conditions as has been noted by several authors.[4,7]

Some special subsets of both $D^p(H^N)$ and $\tilde{D}^p(H^N)$ will play an important role. An extreme point d of $D^p(H^N)$ is one which is not interior to a line segment contained in $D^p(H^N)$, i.e., if $d = ta + (1-t)b$, $a, b \in D^p(H^N)$, $0 < t < 1$, then necessarily $a = b = d$. An *extreme subset* D of $D^p(H^N)$ is one such that if $d \in D$, $d = ta + (1-t)b$, $a, b \in D^p(H^N)$, $0 < t < 1$, then it follows that $a, b \in D$. Extreme subsets are convex sets. Each element $b \in \tilde{D}^p(H^N)$ determines a hyperplane $\{d \mid d \in L(H^p), \mathrm{Tr}(b^\dagger d) = 0\}$. If the intersection of this hyperplane with $D^p(H^N)$ is nonempty then it is an *exposed subset* of $D^p(H^N)$. If the intersection consists of a single point it is called an *exposed point*. The element b is called an *exposing operator*. Exposed subsets are convex.

It is the exposed subsets of $D^p(H^N)$ which are readily given a physical interpretation. Suppose that $b(G) \in \tilde{D}^p(H^N)$ and that the kernel of $\Gamma_p^N(b(G))$ is the subspace $G \subset H^N$. Then $D^p(G)$ is an exposed subset of $D^p(H^N)$ since $\mathrm{Tr}(b(G)^\dagger D^p(\Psi)) = \mathrm{Tr}(\Gamma_p^N(b(G))^\dagger D^N(\Psi)) = 0$ if and only if $\Psi \in G$. Moreover, all exposed subsets of $D^p(H^N)$ are related to subspaces that belong to the lowest eigenvalue of some p-particle operator in this way. The notions *extreme* and *exposed* are general and apply to any convex set. Exposed subsets are necessarily extreme subsets, and exposed points are necessarily extreme points. However it is not true, in general, that extreme subsets are exposed or, in particular, that extreme points are exposed points. We will denote the extreme points of $D^p(H^N)$ by $EXT\ D^p(H^N)$ and the exposed points by $EXP\ D^p(H^N)$.

If $\{a_i\}$ is the set of annihilators for some arbitrary choice of the one-particle basis set we can form the *pair annihilation operators* $\{\sigma_i\}$ where $\sigma_i = a_{2i-1}a_{2i}$. We denote by σ^N on the linear subspace of H^N all of whose elements are of the form $\Phi^\dagger \mid 0\rangle$, where Φ is a homogeneous polynomial of order $N/2$ in the σ_i. We call σ^N a *paired subspace* of H^N. There are, of course, many paired subspaces of H^N; one for each choice of pairing. All AGP (antisymmetrized geminal power) or projected BCS wavefunctions are elements of some paired subspace. Moreover, $D^2(\sigma^N) = \overline{\mathrm{CONV}}\{D^2(\phi) \mid \phi \in \sigma^N\}$ is an exposed subset of $D^2(H^N)$. One can easily verify this by observing that the operator $1 - (2/N)\sum \sigma_i^\dagger \sigma_i$ is (*i*) positive semidefinite on H^N and (*ii*) has a kernel equal to σ^N. The characterization of the exposed subset $D^2(\sigma^N)$ of $D^2(H^N)$ is an interesting subproblem of the N-representability problem which should be easier. There is some similarity with the problem of the characterization of $D^1(H^N)$. The occupation number of the pair σ_i may not exceed 1. However the σ_i commute with one another rather than anticommute.

We illustrate some of the ideas of this section with the simple example of $D^N(H^N)$. As was remarked above, $D^N(H^N)$ is the set of all positive operators with trace equal to 1. One may easily check that the polar $\tilde{D}^N(H^N)$ is equal to $D^N(H^N)$. The extreme points of $D^N(H^N)$ are merely all projectors onto one-dimensional subspaces of H^N. But these projectors are also exposed points of $D^N(H^N)$. If $P \in D^N(H^N)$ is a projector onto a one-dimensional subspace, then an exposing operator $b \in \tilde{D}^N(H^N)$ for P is given by $b = [\binom{r}{N} - 1]^{-1}(I - P)$, where I is the identity operator and r is the dimension of H^1. The factor $[\binom{r}{N} - 1]^{-1}$ is included so that $\mathrm{Tr}(b^\dagger b)^{1/2} = 1$. So in this special case $EXT D^N(H^N) = EXP D^N(H^N)$ and $\tilde{D}^N(H^N) = D^N(H^N)$. In the next section we will show that $EXT D^p(H^N) = EXP D^p(H^N)$ in the finite-dimensional case.

3. INFINITE DIMENSIONS $EXP D^p(H^N) = EXT D^p(H^N)$

In general, the extreme points of a compact convex set are not exposed points. However, in this section we show that if the dimension of H^1 is finite, then the extreme points of $D^p(H^N)$ are exposed points. This is important since, as we have seen above, it is the exposed subsets of $D^p(H^N)$ and, in particular, the exposed points of $D^p(H^N)$ which are physically interpretable. The exposed points of $D^p(H^N)$ are all of the form $D^p(G)$ where $D^p(G)$ consists of a single point and G is the linear space corresponding to the lowest eigenvalue of some p-particle operator. We are thus led to a new characterization of the extreme points of $D^p(H^N)$. The extreme points of $D^p(H^N)$ are precisely those points which are covered by the state ϕ, or in the case of degeneracy, the linear space G, corresponding to the lowest eigenvalue of some p-particle operator. In the degenerate case, all elements ϕ of G must yield the same p-density. The question of whether G is always one-dimensional is taken up in Sec. 6. In the infinite-dimensional case we have no such result. It is not known, in general, whether a similar result holds for $\tilde{D}^p(H^N)$ in either the finite- or infinite-dimensional case. However, in Theorem 5.1 we establish that the extreme points of $\tilde{D}^p(H^N)$ are exposed points for the special case where the dimension of H^1 is equal to $N + p$.

Let $A \subset B \subset C$ be compact, convex sets. From the definition of an extreme subset we can easily show

that if A is extreme in B and B is extreme in C, then A is extreme in C. If instead of extreme we say that A is an exposed subset of B and B is an exposed subset of C, then, in general, it does *not* follow that A is exposed in C. However, with $D^p(H^N)$ in the finite-dimensional case such a result holds.

Let $D^p(G) \subset D^p(H^N)$ be an exposed subset and $b(G) \in \text{CONE}\tilde{D}^p(H^N)$ be an operator which exposes $D^p(G)$. We have labeled both the exposed subset and the exposing operator by $G \subset H^N$, the kernel of the expanded operator $\Gamma_p^N(b(G))$. The relationship between G, $D^p(G)$, and $b(G)$ is given by the following set of equalities:

$$\begin{aligned} D^p(G) &= \{d_0 \mid d_0 \in D^p(H^N), \mathrm{Tr}(b(G)^\dagger d_0) = 0\} \\ &= \{L_N^p(d) \mid d \in D^N(H^N), \mathrm{Tr}(\Gamma_p^N(b(G))^\dagger d) = 0\} \\ &= \overline{\text{CONV}}\{D^p(\Psi) \mid \Psi \in G \equiv \text{kernel of } \Gamma_p^N(b(G))\}. \end{aligned}$$

Lemma 3.1: Let H^1 be finite dimensional. Then if $D^p(G_0)$ is an exposed subset of $D^p(H^N)$ and $D^p(G_1)$ is an exposed subset of $D^p(G_0)$, then $D^p(G_1)$ is an exposed subset of $D^p(H^N)$.

Proof: Let $b(G_0) \in \text{CONE}\,\tilde{D}^p(H^N)$ be an exposing operator for $D^p(G_0)$, and $b(G_1) \in \text{CONE}\,\tilde{D}^p(G_0)$ be an exposing operator for $D^p(G_1)$. Note that $G_1 \subset G_0 \subset H^N$ and that $\Gamma_p^N(b(G_1)) \mid G_0 \geq 0$ but that $\Gamma_p^N(b(G_1))$ need not be positive and thus $b(G_1)$ may not be contained in $\text{CONE}\,\tilde{D}^p(H^N)$. We now show that one can choose a $t \geq 0$ such that $b = tb(G_0) + b(G_1) \in \text{CONE}\,\tilde{D}^p(H^N)$, i.e., $\Gamma_p^N(b) \geq 0$ and $\mathrm{Tr}(b^\dagger d) = 0, d \in D^p(H^N)$ implies that $d \in D^p(G_1)$, thus establishing that $D^p(G_1)$ is an exposed subset of $D^p(H^N)$. This is equivalent to showing that a t can be chosen such that the kernel of $\Gamma_p^N(b)$ is precisely G_1 and $\Gamma_p^N(b)$ restricted to $G_1^\perp$ is strictly positive. Clearly G_1 lies in the kernel of $\Gamma_p^N(b)$, and we need only look at $\Gamma_p^N(b)$ on $G_1^\perp$.

$G_1^\perp$ can be written as $G_1^\perp = G_0^\perp \oplus (G_0 \cap G_1^\perp)$. Let $A = \{\phi \mid \phi \in G_0^\perp, \|\phi\| = 1\}$ and $B = \{\theta \mid \theta \in G_0 \cap G_1^\perp, \|\theta\| = 1\}$. Then define

$$\begin{aligned} \lambda &= \inf_{\phi \in A} \langle \phi \mid \Gamma_p^N(b(G_0))\phi\rangle, \\ \lambda_1 &= \inf_{\theta \in B} \langle \theta \mid \Gamma_p^N(b(G_1))\theta\rangle, \\ \alpha &= \inf_{\phi \in A} \langle \phi \mid \Gamma_p^N(b(G_1))\phi\rangle, \\ \beta &= \sup_{\phi \in A,\, \theta \in B} \langle \phi \mid \Gamma_p^N(b(G_1))\theta\rangle. \end{aligned}$$

It follows from finite dimensionality that λ and λ_1 are strictly greater than zero. Any $\Psi \in G_1^\perp$ can be written as $\Psi = v\phi + w\theta$, $\phi \in A$, $\theta \in B$. Then $\langle \Psi \mid \Gamma_p^N(b)\Psi\rangle = \langle \Psi \mid \Gamma_p^N(tb(G_0) + b(G_1))\Psi\rangle = |v|^2\langle\phi \mid \Gamma_p^N(tb(G_0) + b(G_1))\phi\rangle + |w|^2\langle\theta \mid \Gamma_p^N(b(G_1))\theta\rangle + \bar{v}w\langle\phi \mid \Gamma_p^N(b(G_1))\theta\rangle + \bar{w}v\langle\theta \mid \Gamma_p^N(b(G_1))\phi\rangle \geq |v|^2(\alpha + t\lambda) + |w|^2\lambda_1 - 2|v||w|\beta$. This last quantity will be strictly greater than zero if $(\alpha + t\lambda)\lambda_1 - \beta^2 > 0$. But clearly t can be chosen to satisfy this inequality since $\lambda, \lambda_1 > 0$.

We remark that finite dimensionality was used only to establish that zero is an isolated point in the spectrum of both $\Gamma_p^N(b(G_0))$ and $\Gamma_p^N(b(G_1)) \mid G_0$, thus allowing us to conclude that λ and λ_1 are strictly positive. We may substitute this weaker hypothesis in the statement of Lemma 3.1.

Theorem 3.1: Let H^1 be finite dimensional. The extreme points of $D^p(H^N)$ are exposed.

In order to prove this result we make use of the fact that all boundary points of $D^p(H^N)$ lir in some exposed subset of $D^p(H^N)$. This follows directly from the basic properties of compact, convex sets.[8]

Proof: Suppose that d is extreme in $D^p(H^N)$. Then by the above result on boundary points d is contained in some exposed subset $D^p(G_0)$ of $D^p(H^N)$. Also d is extreme in $D^p(G_0)$. By again applying the result on boundary points, d belongs to an exposed subset of $D^p(G_0)$ which we denote by $D^p(G_1)$. But by Lemma 3.1, $D^p(G_1)$ is exposed in $D^p(H^N)$. The dimension of $D^p(G_1)$ is strictly less than that of $D^p(G_0)$ which is strictly less than that of $D^p(H^N)$. We can continue this argument, constructing a sequence of exposed subsets $D^p(G_i), i = 0, 1, \ldots, m$, of decreasing dimensionality such that $D^p(H^N) \supset D^p(G_0) \supset D^p(G_1) \supset \cdots \supset D^p(G_m)$, all of which contain d. But by hypothesis H^1 and, therefore, $D^p(H^N)$ is finite dimensional, and thus the construction must stop with an exposed subset $D^p(G_m)$, containing d, which consists of a single point, i.e., the point d.

Theorem 3.2: In the finite-dimensional case all of the extreme points of $D^p(H^N)$ are of the form $D^p(G)$ where $D^p(G)$ consists of a single point, and G is the linear space corresponding to the lowest eigenvalue of some p-particle operator.

In passing we remark that in all known cases G consists of a single element (up to an arbitrary phase factor) of H^N. Thus these extreme points are covered by the unique ground state of a p-particle operator. It is a conjecture of Sec. 6 that this holds in general. If this were true then the extreme points of $D^p(H^N)$ would all be covered by elements $\Psi \in H^N$ which are nondegenerate ground states of some p-particle operator.

4. PARTICLE–HOLE DUALITY

In the case where r, the one-rank, is finite, one can convert the problem of the characterization of $D^p(H^N)$ and $\tilde{D}^p(H^N)$ into the equivalent problem of the characterization of $D^p(H^{r-N})$ and $\tilde{D}^p(H^{r-N})$. For example, the problem of the enumeration of the extreme points of $D^p(H^N)$ and $\tilde{D}^p(H^N)$ is equivalent to the enumeration of the extreme points of $D^p(H^{r-N})$ and $\tilde{D}^p(H^{r-N})$. Physically, we can think of this equivalent problem as a description of the original system in terms of holes rather than particles. As we shall see in the next section this can lead to considerable simplifications in the case where $r = N + p$. The equivalence is established by constructing two invertible linear maps Δ and Π from $L(H^p)$ onto $L(H^p)$ such that $\Delta(D^p(H^{r-N})) = D^p(H^N)$ and $\Pi(\tilde{D}^p(H^{r-N})) = \tilde{D}^p(H^N)$. But, as we are dealing with finite-dimensional spaces, both Π and Δ are homeomorphisms. Therefore, Δ and Π preserve both the topological and the linear properties of subsets of $L(H^p)$. Thus Δ maps the extreme (exposed) points of $D^p(H^{r-N})$ onto the extreme (exposed) points of $D^p(H^N)$. A similar statement holds for Π. Using the result of the preceding section that the extreme points of $D^p(H^N)$ are also exposed points, we have

$$\mathrm{EXT}D^p(H^N) = \mathrm{EXP}D^p(H^N) = \Delta(\mathrm{EXT}D^p(H^{r-N}))$$
$$= \Delta(\mathrm{EXP}D^p(H^{r-N}))$$

and

$$\mathrm{EXT}\tilde{D}^p(H^N) = \Pi(\mathrm{EXT}\tilde{D}^p(H^{r-N})),$$
$$\mathrm{EXP}\tilde{D}^p(H^N) = \Pi(\mathrm{EXP}\tilde{D}^p(H^{r-N})).$$

Without supplying all the arguments derived from the linearity of Δ and Π required to establish this important set of equalities, we will show how the linearity of Δ forces the fact that $\mathrm{EXT}D^p(H^N) = \Delta(\mathrm{EXT}D^p(H^N))$ given $\Delta(D^p(H^{r-N})) = D^p(H^N)$ and the invertibility of Δ.

Let d_0 be an extreme point of $D^p(H^{r-N})$. We will show that $d_1 = \Delta(d_0)$ is an extreme point of $D^p(H^N)$. Let $d_1 = ta + (1-t)b$, $a, b \in D^p(H^N)$, and $0 < t < 1$; then $d_0 = \Delta^{-1}(d_1) = t\Delta^{-1}(a) + (1-t)\Delta^{-1}(b)$ and $\Delta^{-1}(a) = \Delta^{-1}(b) = d_0$ since d_0 is extreme. Thus $a = b = d_1$. By precisely the same argument it can be established that extreme points of $D^p(H^N)$ are mapped by Δ^{-1} into extreme points of $D^p(H^{r-N})$.

Before constructing the maps Δ and Π we need to establish some properties of the contraction and expansion maps L_N^p and Γ_p^N. Recall from Sec. 2 that Γ_p^N is one-to-one onto the set of all p-particle operators in $L(H^N)$.

Proposition 4.1: Let K be the kernel of L_N^p.

(a) $K^\perp$ is the set of all p-particle operators.

(b) Γ_p^N is one-to-one from $L(H^p)$ onto $K^\perp$.

(c) $L_N^p \mid K^\perp$ is one-to-one onto $L(H^p)$.

Proof: (a) Let K_0 be the set of all p-particle operators. If $d \in K$, then $0 = \mathrm{Tr}(b^\dagger L_N^p(d)) = \mathrm{Tr}(\Gamma_p^N(b)^\dagger d)$ for all b and, therefore, $d \in K_0^\perp$. But if $d \in K_0^\perp$ then for all $b \in L(H^p)$, $0 = \mathrm{Tr}(\Gamma_p^N(b)^\dagger d) = \mathrm{Tr}(b^\dagger L_N^p(d))$ and $d \in K$. Thus $K = K_0^\perp$ and $K^\perp = K_0$.

(b) Γ_p^N is one-to-one onto the set of all p-particle operators K_0 by formula (2.1) of Sec. 2. But by (a) $K_0 = K^\perp$.

(c) L_N^p is one-to-one from the orthogonal complement of its kernel $K^\perp$ onto its range. Suppose its range K_1 is not all of $L(H^p)$. But then there would be an element $b \in L(H^p)$ such that $b \perp K_1$. The $0 = \mathrm{Tr}(b^\dagger L_N^p(d)) = \mathrm{Tr}(\Gamma_p^N(b)^\dagger d)$ for all d. But Γ_p^N is one-to-one, and thus $\Gamma_p^N(b) \neq 0$, and $\mathrm{Tr}(\Gamma_p^N(b)^\dagger d)$ cannot equal zero for all d.

We now construct a one-to-one linear map Δ_1 from $L(H^{r-N})$ onto $L(H^N)$ which converts particles into holes and holes into particles. Let R be the sequence $(1, 2, \ldots, r)$. Let I denote some ordered $(r-N)$-tuple of indices taken from R, and let $\mid I\rangle$ denote the corresponding Slater determinant. In second quantization notation, we have $\mid I\rangle = a_I^\dagger \mid 0\rangle$, where $a_I^\dagger = \Pi_{i\in I} a_i^\dagger$ and $\mid 0\rangle$ is the vacuum state. The order of the a_i in this product is important, and we assume that the indices increase from left to right. Denote by $\mid I^c\rangle$ the N-particle state $a_I \mid R\rangle$ where $a_I = \Pi_{i\in I} a_i$ and $\mid R\rangle = (\Pi_{i\in R} a_i^\dagger) \mid 0\rangle$. The same convention on ordering in these products is followed as before. The map Δ_0 defined on the $r-N$ particle states $\mid I\rangle$ by $\Delta_0 \mid I\rangle = \mid I^c\rangle$ can be extended linearly to all of H^{r-N}. The action of Δ_0 on an arbitrary element of H^{r-N} is given by

$$\Delta_0(\Sigma \alpha_I \mid I\rangle) = \Sigma \alpha_I \mid I^c\rangle.$$

Δ_0 is a bijection of H^{r-N} onto H^N. It also preserves the scalar product. The linear map $\Delta_1 : L(H^{r-N}) \to L(H^N)$ defined by $b \to \Delta_0 b \Delta_0^{-1}$ is a bijection of $L(H^{r-N})$ onto $L(H^N)$. More explicitly, if I and J are $(r-N)$-tuples of indices,

$$\Delta_1(a_I^\dagger a_J) = a_I a_J^\dagger. \tag{4.1}$$

The action of Δ_1 on an arbitrary element of $L(H^{r-N})$ can be computed using (4.1) and linearity. Δ_1 merely turns annihilators into creators and creators into annihilators.

Let K_0 be the kernel of L_{r-N}^p and K_1 the kernel of L_N^p. Then by Proposition 4.1 $K_0^\perp$ and $K_1^\perp$ are the set of all p-particle operators in $L(H^{r-N})$ and $L(H^N)$, respectively.

Proposition 4.2: The linear map $\Delta_1 : L(H^{r-N}) \to L(H^N)$

(a) is a bijection of $L(H^{r-N})$ onto $L(H^N)$,

(b) preserves multiplication, i.e., $\Delta_1(ab) = \Delta_1(a)\Delta_1(b)$,

(c) preserves taking adjoints, i.e., $\Delta_1(a^\dagger) = (\Delta_1(a))^\dagger$,

(d) maps the positive cone of $L(H^{r-N})$ onto the positive cone of $L(H^N)$,

(e) preserves the trace,

(f) maps $D^{r-N}(H^{r-N})$ onto $D^N(H^N)$, and

(g) maps K_0 onto K_1 and $K_0^\perp$ onto $K_1^\perp$.

Proof: (a) This follows from the discussion preceding the proposition.

(b) $\Delta_1(ab) = \Delta_0(ab)\Delta_0^{-1} = \Delta_0 a \Delta_0^{-1} \Delta_0 b \Delta_0^{-1} = \Delta_1(a)\Delta_1(b)$.

(c) Let $a \in L(H^{r-N})$, $\phi, \theta \in H^N$. Then $\langle \phi \mid \Delta_1(a)\theta\rangle = \langle \phi \mid \Delta_0 a \Delta_0^{-1}\theta\rangle = \langle \Delta_0^{-1}\phi \mid a\Delta_0^{-1}\theta\rangle = \langle a^\dagger \Delta_0^{-1}\phi \mid \Delta_0^{-1}\theta\rangle = \langle \Delta_0 a^\dagger \Delta_0^{-1}\phi \mid \theta\rangle = \langle \Delta_1(a^\dagger)\phi \mid \theta\rangle$.

(d) Any positive operator a in $L(H^{r-N})$ can be factored as $a = b^\dagger b$. But by (b) and (c) $\Delta_1(b^\dagger b) = \Delta_1(b^\dagger)\Delta_1(b) = (\Delta_1(b))^\dagger \Delta_1(b)$.

(e) $\mathrm{Tr}(b) = \Sigma_I \langle I \mid bI\rangle = \Sigma \langle \Delta_0 I \mid \Delta_0 b \Delta_0^{-1}\Delta_0^{-1}\Delta_0 I\rangle = \Sigma \langle I^c \mid \Delta_1(b) I^c\rangle = \mathrm{Tr}(\Delta_1(b))$, since Δ_0 preserves the scalar product.

(f) This follows from (d) and (e).

(g) The fact that $\Delta_1(K_0^\perp) = K_1^\perp$ follows directly from formula (4.1). But since Δ_1 preserves the trace, it preserves the trace scalar product and, therefore, maps K_0 onto K_1.

We are now in a position to define the maps Δ and Π. As before we let K_0 be the kernel of L_N^p and K_1 be the kernel of L_{r-N}^p. The maps

$$\Delta = (L_N^p \mid K_0^\perp)\Delta_1(L_{r-N}^p \mid K_1^\perp)^{-1}$$

and

$$\Pi = (\Gamma_p^N)^{-1}\Delta_1(\Gamma_p^{r-N})$$

are both well defined since by Proposition 4.1 $L_{r-N}^p \mid K_1^\perp$ is a bijection of $K_1^\perp$ onto $L(H^p)$ and Γ_p^{r-N} is a bijection from $L(H^p)$ onto $K_1^\perp$; similarly for $L_N^p \mid K_0^\perp$ and Γ_p^N. Also Δ_1 maps $K_1^\perp$ onto $K_0^\perp$ by Proposition 4.2.

Theorem 4.1: (properties of Δ and Π). (a) Δ and Π are bijections of $L(H^p)$ onto $L(H^p)$.

(b) If $b, d \in L(H^p)$ then $\mathrm{Tr}(b^\dagger d) = \mathrm{Tr}(\Pi(b)^\dagger \Delta(d))$.

(c) $\Delta(D^p(H^{r-N})) = D^p(H^N)$, $\Pi(\tilde{D}^p(H^{r-N})) = \tilde{D}(H^N)$.

(d) $\mathrm{EXT}D^p(H^N) = \mathrm{EXP}D^p(H^N) = \Delta(\mathrm{EXP}D^p(H^{r-N})) = \Delta(\mathrm{EXT}D^p(H^{r-N}))$, $\mathrm{EXT}\tilde{D}^p(H^N) = \Pi(\mathrm{EXT}\tilde{D}^p(H^{r-N}))$, $\mathrm{EXP}\tilde{D}^p(H^N) = \Pi(\mathrm{EXP}\tilde{D}^p(H^{r-N}))$.

Proof: (a) This follows from the discussion immediately preceding the theorem.

(b) This follows from the trace preserving properties of Δ_1 and the properties of the contraction and expansion maps.

(c) We first show that $\Pi(\tilde{D}^p(H^{r-N})) = \tilde{D}^p(H^N)$. But $\Gamma_p^{r-N}(\tilde{D}^p(H^{r-N}))$ is the set of all positive p-particle operators on H^{r-N} of trace one, and both Δ_1 and Δ_1^{-1} preserve positivity and trace and map p-particle operators into p-particle operators (Proposition 4.2). Thus $\Delta_1 \Gamma_p^{r-N}(\tilde{D}^p(H^{r-N}))$ is the set of all positive p-particle operators of trace one on H^N. Thus $\Pi(\tilde{D}^p(H^{r-N})) = \tilde{D}^p(H^N)$. Let $d_1 = \Delta(d_0)$, $d_0 \in D^p(H^{r-N})$, and let $b_1 = \Pi(b_0) \in \tilde{D}^p(H^{r-N})$ be arbitrary. Then by (b) and the fact that $b_0 \in \tilde{D}^p(H^{r-N})$ we have $\mathrm{Tr}(b_1^\dagger d_1) = \mathrm{Tr}(\Pi(b_0)^\dagger \Delta(d_0)) = \mathrm{Tr}(b_0^\dagger d_0) \geq 0$. By the bipolar theorem,[9] and since the contraction map as well as Δ_1 preserve the trace we have $\Delta(D^p(H^{r-N})) = D^p(H^N)$.

(d) These equalities follow from the discussion at the beginning of this paragraph.

The *total preimage* of a point $d_0 \in D^p(H^N)$ under the contraction mapping is given by $\{d \in D^N(H^N) \mid L_N^p(d) = d_0\}$. We denote this set by $(L_N^p)^{-1}(d_0)$. The following result regarding preimages of elements of $D^p(H^N)$ and particle-hole duality will be required in Sec. 6.

Theorem 4.2: Let $d_0 \in D^p(H^{r-N})$. Then $(L_N^p)^{-1}(\Delta(d_0)) = \Delta_1((L_{r-N}^p)^{-1}(d_0))$.

Proof: Clearly $\Delta_1((L_{r-N}^p)^{-1}(d_0)) \subset D^N(H^N)$. We show that $(L_N^p)^{-1}(\Delta(d_0)) \supset \Delta_1((L_{r-N}^p)^{-1}(d_0))$ by showing that $\Delta(d_0) = L_N^p \Delta_1((L_{r-N}^p)^{-1}(d_0))$. As before we let K_0 be the kernel of L_{r-N}^p and K_1 be the kernel of L_N^p. If $P(K_0^\perp)$ and $P(K_1^\perp)$ are projectors onto $K_0^\perp$ and $K_1^\perp$, respectively, then (i) $L_N^p = L_N^p P(K_1^\perp)$, (ii) $P(K_1^\perp)\Delta_1 = \Delta_1 P(K_0^\perp)$ by Proposition 4.2 (g), and (iii) the set $P(K_0^\perp)(L_{r-N}^p)^{-1}(d_0)$ consists of the single point $(L_{r-N}^p \mid K_0^\perp)^{-1}(d_0)$. But then $L_N^p \Delta_1((L_{r-N}^p)^{-1}(d_0)) = L_N^p P(K_1^\perp)\Delta_1((L_{r-N}^p)^{-1}(d_0)) = L_N^p \Delta_1 P(K_0^\perp)((L_{r-N}^p)^{-1}(d_0)) = L_N^p \Delta_1 (L_{r-N}^p \mid K_0^\perp)^{-1}(d_0) = (L_N^p \mid K_1^\perp)\Delta_1 (L_{r-N}^p \mid K_0^\perp)^{-1}(d_0) = \Delta(d_0)$ by the definition of the Δ map. The reverse inclusion can be obtained by a similar argument.

Explicit formulas for Δ and Π in the case where $p = 2$ are given by

Theorem 4.3:

$$\Delta(a_i a_j a_k^\dagger a_l^\dagger) = \binom{N}{2}^{-1}\Big[\binom{r-2}{2}(\delta_{jk}\delta_{il} - \delta_{ik}\delta_{jl}) + \binom{r-N}{2}(r-3)(\delta_{ik}a_l a_j^\dagger + \delta_{jl}a_k a_i^\dagger - \delta_{jk}a_l a_i^\dagger - \delta_{il}a_k a_j^\dagger) + \binom{r-N}{2} a_l a_k a_j^\dagger a_i^\dagger\Big].$$

$$\Pi(a_i^\dagger a_j^\dagger a_k a_l) = \binom{r-N}{2}^{-1}\Big[\delta_{jk}\delta_{il} - \delta_{ik}\delta_{jl} + \binom{N}{2}(\delta_{ik}a_l^\dagger a_j + \delta_{jl}a_k^\dagger a_i - \delta_{jk}a_l^\dagger a_i - \delta_{il}a_k^\dagger a_j) + \binom{N}{2} a_l^\dagger a_k^\dagger a_j a_i\Big].$$

Proof: The formulas can be obtained by application of formula (4.1) of this section and formulas (2.1) and (2.2) for Γ_p^N and L_N^p given in Sec. 2.

Elements $d \in D^p(H^N)$ are positive semidefinite. But as $\Delta^{-1}(d) \in D^p(H^{r-N})$, $\Delta^{-1}(d)$ must also be positive semi-definite. This yields a strong necessary (but not sufficient) condition that $d \in D^p(H^N)$. In fact, in the next section we will show that these conditions are complete in the case $r = N + p$. This condition may be formulated in a slightly different way by noting that the condition $\Delta^{-1}(d) \geq 0$ is equivalent to requiring that $\mathrm{Tr}(P^\dagger \Delta^{-1}(d)) \geq 0$ for all one-dimensional projectors $P \in \tilde{D}^p(H^{r-N})$ or, equivalently, (by Theorem 4.1 b) by requiring that $\mathrm{Tr}(\Pi(P)^\dagger d) \geq 0$ for all P. There is yet a third way to express this condition as we shall see in the following theorem. In the literature these N-representability conditions are frequently referred to as *Q-matrix conditions*.

Theorem 4.4: (Q-matrix conditions). The following are equivalent necessary conditions that $d \in D^p(H^N)$:

(a) $\Delta^{-1}(d) \geq 0$,

(b) $\mathrm{Tr}(\Pi(P)^\dagger d) \geq 0$ for all one-dimensional projectors $P \in \tilde{D}^p(H^{r-N})$,

(c) $1 - pN\,\mathrm{Tr}(D^1(\phi)D^1(d)) + \binom{p}{2}\binom{N}{2}\mathrm{Tr}(D^2(\phi)D^2(d)) - \binom{p}{3}\binom{N}{3}\mathrm{Tr}(D^3(\phi)D^3(d)) + \cdots + (-1)^p\binom{p}{p}\binom{N}{p}\mathrm{Tr}(D^p(\phi)D^p(d)) \geq 0$ for all $\phi \in H^p$.

Proof: Parts (a) and (b) follow from the discussion preceding the statement of the theorem. Part (c) can be proved directly by noting that if $\phi \in H^p$ and $d = D^p(\Psi)$, $\Psi \in H^N$, then $\| A_{N+p}\phi\Psi \|^2 \geq 0$. But by Sasaki's formula[1]

$$\begin{aligned}\| A_{N+p}\phi\Psi \|^2 &= \langle \phi\Psi \mid A_{N+p}\phi\Psi\rangle \\ &= 1 - pN\,\mathrm{Tr}(D^1(\phi)D^1(\Psi)) + \binom{p}{2}\binom{N}{2}\mathrm{Tr}(D^2(\phi)D^2(\Psi)) - \binom{p}{3}\binom{N}{3}\mathrm{Tr}(D^3(\phi)D^3(\Psi)) + \cdots \\ &\quad + (-1)^p\binom{p}{p}\binom{N}{p}\mathrm{Tr}(D^p(\phi)D^p(\Psi)).\end{aligned}$$

If we let $\phi = \Phi^\dagger \mid 0\rangle$, where $\Phi = \sum \alpha_I a_I$, $I = (i_1, i_2 \cdots i_p)$, then $P(\phi)$, the projector onto the one-dimensional subspace generated by ϕ, can be written as $\Phi^\dagger\Phi$. That (c) is equivalent to (a) can be seen as follows:

$$\begin{aligned}\| A_{N+p}\phi\Psi \|^2 \geq 0 &\iff \mathrm{Tr}(\Phi\Phi^\dagger D^N(\Psi)) \geq 0 \\ &\iff \mathrm{Tr}(\Phi^\dagger\Phi D^{r-N}(\Delta_0^{-1}(\Psi))) \geq 0 \\ &\iff \mathrm{Tr}(\Phi^\dagger\Phi D^p(\Delta_0^{-1}(\Psi))) \geq 0 \\ &\iff \mathrm{Tr}(P(\phi)^\dagger \Delta^{-1}(d)) \geq 0\end{aligned}$$

and

$$\mathrm{Tr}(P(\phi)^\dagger \Delta^{-1}(d)) \geq 0, \quad \forall \phi \iff \Delta^{-1}(d) \geq 0.$$

5. THE STRUCTURE OF $D^p(H^N)$ AND $\tilde{D}^p(H^N)$ WHERE $r = N + p$

Only when the rank r of the one-density is equal to $N + 2$ is the convex structure of both $D^2(H^N)$ and $\tilde{D}^2(H^N)$ completely understood. Actually the arguments used for the case $r = N + 2$ can be used without modification to determine the convex structure of $D^p(H^N)$ and $\tilde{D}^p(H^N)$ for $r = N + p$. We, therefore, consider that case in this section. This problem is solved by a simple application of the particle–hole duality established in the last section.

Theorem 5.1: In the case $r = N + p$, all of the extreme points of $D^p(H^N)$ ($\tilde{D}^p(H^N)$) can be written in the form $\Delta(P)$ ($\Pi(P)$), where $P \in L(H^p)$ is a projector onto a one–dimensional subspace of H^p. The extreme points of both $D^p(H^N)$ and $\tilde{D}^p(H^N)$ are exposed. The maps Δ and Π are those introduced in the preceding section.

Proof: This follows directly from Theorem 4.1 and the fact that the extreme points of $D^p(H^p) \equiv \tilde{D}^p(H^p)$ are simply the projections onto one-dimensional subspaces of H^p as was mentioned in Sec. 2. All the extreme points of $D^p(H^p)$ and thus of $\tilde{D}^p(H^N)$ are exposed (Sec. 2).

This theorem gives an alternative route to the result of Sec. 3 that $\mathrm{EXT}D^p(H^N) = \mathrm{EXP}D^p(H^N)$ in the case $r = N + p$. In addition it establishes the new result that $\mathrm{EXT}\tilde{D}^p(H^N) = \mathrm{EXP}\tilde{D}^p(H^N)$ under the stated condition on the rank.

Corollary 5.1: When $r = N + p$, the Q-matrix conditions, given by Theorem 4.4, are complete.

Let $\phi = \Phi^\dagger|0\rangle$, where $\Phi = \Sigma\alpha_I a_I$ and $\Sigma|\alpha_I|^2 = 1$, be an element of H^p. The range of summation is over all distinct ordered p-tuples $(i_1, i_2, \ldots, i_p)$ and $a_I = a_{i_1}a_{i_2}\cdots a_{i_p}$ as in Sec. 4. Then $\Phi^\dagger\Phi \in L(H^p)$ is a projector onto Φ. But then the extreme points of $D^p(H^N)$ are given by $\Delta(\Phi^\dagger\Phi) = L_N^p\Delta_1(\Phi^\dagger\Phi) = L_N^p(\Phi\Phi^\dagger) = \binom{N}{p}^{-1}\Phi\Phi^\dagger$ as ϕ ranges over H^p. The last equality results from an application of formula (2.1) of Sec. 2 for the contraction operator.

Corollary 5.2: When $r = N + p$, the extreme points of $D^p(H^N)$ are given by $\binom{N}{p}^{-1}\Phi\Phi^\dagger$ as ϕ ranges over H^p.

Theorem 5.2: When $r = N + p$, the contraction map is a bijection and $\mathrm{EXT}\, D^p(H^N) = L_N^p(\mathrm{EXT}\, D^N(H^N))$.

Proof: The map Δ_1 of Proposition 4.2 maps K_0, the kernel of L_N^p, onto K_1, the kernel of L_N^p. But as L_p^p is the identity map, K_0 and hence K_1 contain only 0. That $\mathrm{EXT}\, D^p(H^N) = L_N^p[\mathrm{EXT}\, D^N(H^N)]$ follows from the fact that L_N^p is a bijection.

When $r = N + 2$ and N is even, one can show that all wavefunctions are of "paired type." That is, there exists a set of pair annihilation operators $\{\sigma_i\}$, where $\sigma_i = a_{2i-1}a_{2i}$, $1 \le i \le \frac{1}{2}(N + 2)$, such that each element $\delta \in H^N$ can be written as $\Delta^\dagger|0\rangle$ where Δ is a homogeneous polynomial of order $(N/2)$ in the σ_i. Thus $\delta \in \sigma^N$ and $D^2(\delta)$ is contained in $D^2(\sigma^N)$. It is also true that when these conditions hold for r and N, elements $\delta \in H^N$ can be approximated arbitrarily closely by an AGP function (antisymmetrized geminal power) or projected BCS function. Theorem 5.3, which establishes these results, is due to Coleman. This theorem is based on a lemma due independently to Zumino[10] and Bloch.[11]

Lemma 5.1: Let $\phi \in H^2$, then there exists a one-particle basis such that if $\{a_i\}$ are the corresponding annihilation operators then ϕ can be expressed using the pair annihilation operators $\{\sigma_i\}$, where $\sigma_i = a_{2i-1}a_{2i}$, $1 \le i \le r/2$, in the following way:

$$\phi = \Sigma\,\xi_i\sigma_i^\dagger|0\rangle.$$

$D^1(\phi)$ is diagonal in the basis $\{a_i^\dagger|0\rangle\}$.

By the particle–hole duality established in the last section (see the construction of Δ_0) and using the notation introduced there, all elements of H^N can be expressed in the form

$$\Psi = \Sigma\alpha_I a_I|R\rangle, \quad \Sigma|\alpha_I|^2 = 1, \quad I = (i_1, i_2),$$

where $|R\rangle = (\Pi a_i^\dagger)|0\rangle$. But by the preceding lemma one can choose the one-particle basis so that

$$\Psi = \Sigma\xi_i\sigma_i|R\rangle, \quad \Sigma|\xi_i|^2 = 1.$$

Thus every Slater determinant in the expansion of Ψ is characterized by the absence of one of the pairs σ_i. If $\phi = \sum\eta_i\sigma_i^\dagger|0\rangle = \Phi^\dagger|0\rangle$, Ψ is expressible in the the form $\phi^N = (\Phi^\dagger)^{N/2}|0\rangle$ if and only if the equations

$$\xi_i = \Pi_j\eta_j/\eta_i$$

have a solution for the η's given the ξ's. But they clearly have a solution if all the ξ_i's $\neq 0$. If $\xi_1 = 0$ but all the other $\xi_i \neq 0$, then it is easy to see that Ψ is of the form $\sigma_1^\dagger(\Phi^\dagger)^{(N-2)/2}|0\rangle$, where $\Phi^\dagger = \sum_{i\neq 1}\eta_i\sigma_i^\dagger$; similarly for other cases where more than one of the $\xi_i = 0$. But these special cases where some of the ξ_i equal zero are of low dimensionality, and therefore all wavefunctions can be approximated by functions of the AGP type.

Theorem 5.3: ($r = N + 2, N$ even). All elements $\Psi \in H^N$ are of the form

$$\Psi = \left(\prod_{i\in I}\sigma_i^\dagger\right)(\Phi^\dagger)^s|0\rangle,$$

where $\Phi^\dagger = \sum_{j\in J}\eta_j\sigma_j^\dagger$, $I \cap J = 0$, $I \cup J = \{1, 2, \ldots, \frac{1}{2}(N + 2)\}$ and thus exhibit pairing. Here $s = (N/2) - |I|$ and $|I|$ is the number of elements in I. Elements of H^N can be approximated arbitrarily closely by AGP functions, i.e., wavefunctions of the form

$$(\Phi^\dagger)^{N/2}|0\rangle.$$

For odd N the wavefunction cannot exhibit complete pairing–there must be an odd man out. An extension of the above argument to cover this case allows one to give a canonical form for wavefunctions describing an odd number of fermions.

Corollary 5.3: ($r = N + 2, N$ odd) All $\Psi \in H^N$ are of the form

$$\Psi = a_0^\dagger\left(\prod_{i\in I}\sigma^\dagger\right)(\Phi^\dagger)^s|0\rangle,$$

where $\Phi^\dagger = \sum_{j\in J} \eta_j \sigma_j^\dagger$, $I \cap J = 0, I \cup J = \{1, 2, \ldots, \frac{1}{2}(N+1)\}$ and $s = (N-1)/2 - |I|$. Here, a_0 is the annihilator for an orbital strongly orthogonal to all the pairs.

6. UNIQUE PREIMAGES

For the case $r > N + p$ the contraction map L_N^p is not one-to-one and the preimage of an element $d_0 \in D^p(H^N)$, given by $(L_N^p)^{-1}(d_0) = \{d \in D^N(H^N) \mid L_N^p(d) = d_0\}$ is, in general, a set consisting of more than a single point. However, the total preimage of an exposed point of $D^1(H^N)$ under the contraction mapping is unique. It consists of the N-density corresponding to a Slater determinant. In all known cases, the exposed points of $D^2(H^N)$ have unique preimages. The author believes this situation to hold in general as stated in the following:

Conjecture: The total preimage of an exposed point of $D^p(H^N)$ under the contraction mapping L_N^p consists of a single point in $D^N(H^N)$.

In this section we prove the validity of this conjecture in certain special cases. In particular, we show that the preimage of an exposed point of $D^2(H^N)$ is unique when $N = 3$ or 4. The results of this and the next section depend on the following:

Lemma 6.1: If $\Psi_1, \Psi_2 \in H^N$, $\Psi_1 \perp \Psi_2$, and $D^N(\Psi_1)$ and $D^N(\Psi_2)$ are preimages under the contraction mapping of the same exposed point of $D^p(H^N)$, then the transition density $D^p(\Psi_1, \Psi_2) = 0$ or, equivalently, $\text{RANGE}D^q(\Psi_1) \subset [\text{RANGE}D^q(\Psi_2)]^\perp; N = p + q$.

Proof: Because $D^p(\Psi_1) = D^p(\Psi_2)$ is exposed, there is an operator $b \in \tilde{D}^p(H^N)$ whose associated hyperplane intersects $D^p(H^N)$ at the single point $D^p(\Psi_1)$. Each N-particle state lying in the linear space corresponding to the zero eigenvalue of $\Gamma_p^N(b)$ has a pth-order density matrix identical to $D^p(\Psi_1)$. Therefore $D^p(\alpha\Psi_1 + \beta\Psi_2) = D^p(\Psi_1) = D^p(\Psi_2)$, $(|\alpha|^2 + |\beta|^2 = 1)$. But $D^p(\alpha\Psi_1 + \beta\Psi_2) = |\alpha|^2 D^p(\Psi_1) + |\beta|^2 D^p(\Psi_2) + \bar{\alpha}\beta D^p(\Psi_1, \Psi_2) + \bar{\beta}\alpha D^p(\Psi_2, \Psi_1) = D^p(\Psi_1) + [\bar{\alpha}\beta D^p(\Psi_1, \Psi_2)] + [\bar{\alpha}\beta D^p(\Psi_1, \Psi_2)]^\dagger$. Since this equation is valid for all α, β with arbitrary relative phase, $D^p(\Psi_1, \Psi_2) = 0$, and this proves the first part of the theorem.

By a result of Erhardt Schmidt[12] which was rediscovered by Carlson and Keller,[13] and arbitrary wavefunction Ψ may be expanded in terms of the eigenfunctions ψ_i^p of $D^p(\Psi)$ and ψ_i^q of $D^q(\Psi)$ in the following way:

$$\Psi(s,t) = \Sigma\, \xi_i \psi_i^p(s)\psi_i^q(t),$$

where $|\xi_i|^2$ is simultaneously the eigenvalue of $D^p(\Psi)$ corresponding to ψ_i^p and the eigenvalue of $D^q(\Psi)$ corresponding to ψ_i^q. Making use of this expression and the fact that $D^p(\Psi_1, \Psi_2) = 0$ we have

$$0 = D^p(\Psi_1, \Psi_2)(s \mid s') = \int dt \bar{\Psi}_1(s,t)\Psi_2(s',t) = \sum_{i,j} \bar{\xi}_i \xi_j \bar{\psi}_{1i}^p(s)\psi_{2j}^p(s') \int dt \bar{\psi}_{1i}^q(t)\psi_{2j}^q(t).$$

Since the $\psi_{1i}^p(s)\psi_{2i}^p(s')$ are linearly independent, the coefficients $\langle \psi_{1i}^q | \psi_{2j}^q \rangle$ must all vanish. Therefore RANGE $D^q(\Psi_1) \subset [\text{RANGE } D^q(\Psi_2)]^\perp$.

Corollary 6.1: Under the assumptions of the lemma, $D^l(\Psi_1, \Psi_2) = 0$ for $l \le p$.

Theorem 6.1: The preimage of an exposed point of $D^p(H^N)$ under the contraction mapping is unique if

(i) $p \ge N/2, N$ even;

(ii) $p \ge (N+1)/2, N$ odd.

Proof: Assume the contrary. Let $b \in \tilde{D}^2(H^N)$ be an exposing operator for $d \in \text{EXP}D^p(H^N)$. Then, by assumption, there are two states Ψ_1 and Ψ_2 belonging to the kernel of the expanded operator $\Gamma_p^N(b)$ such that $D^N(\Psi_1)$ and $D^N(\Psi_2)$ are distinct and are both mapped by L_N^p onto d. As the kernel of $\Gamma_p^N(b)$ is a linear space we may assume that $\Psi_1 \perp \Psi_2$. By Corollary 6.1, $D^l(\Psi_1, \Psi_2) = 0$ for $l \le p$. In particular $D^q(\Psi_1, \Psi_2) = 0$ for $q = N - p$. It follows that

$$\begin{aligned}[D^p(\Psi_1)D^p(\Psi_2)](s' \mid s'') &= \int ds \int dt \bar{\Psi}_1(s',t)\Psi_1(s,t) \int dt' \bar{\Psi}_2(s,t')\Psi_2(s'',t') \\ &= \int dt\, dt' \bar{\Psi}_1(s',t)\Psi_2(s'',t') \int ds \Psi_1(s,t)\bar{\Psi}_2(s,t') \\ &= \int dt\, dt' \Psi_1(s',t)\Psi_2(s'',t')D^q(\Psi_1,\Psi_2)(t \mid t') \\ &= 0,\end{aligned}$$

which contradicts the fact that $D^p(\Psi_1) = D^p(\Psi_2) = d$.

In the finite rank case this result may be extended to cases not covered by the theorem ($p < N/2$, even N and $p < (N+1)/2$, odd N) when the rank r is small enough by using the particle–hole duality elaborated in Sec. 4. Assume that $d \in D^p(H^N)$ is exposed and that the rank r is finite. Then if Δ is the map defined in Sec. 4, $\Delta^{-1}(d)$ is exposed in $D^p(H^{r-N})$. Then by Theorem 6.1 $\Delta^{-1}(d)$ has a unique preimage if $p \ge (r-N)/2$ ($r-N$ even) or if $p \ge (r-N+1)/2$ ($r-N$ odd). But by Theorem 4.2 the preimage of d is related to that of $\Delta^{-1}(d)$ by the formula $(L_N^p)^{-1}(d) = \Delta_1(L_{r-N}^p)^{-1}(\Delta^{-1}(d))$. Thus d has a unique preimage when these conditions on the rank are satisfied.

Corollary 6.2: The preimage of an exposed point of $D^p(H^N)$ is unique if

(i) $p \ge (r-N)/2$, $(r-N)$ even,

(ii) $p \ge (r-N+1)/2$, $(r-N)$ odd.

For the case where $p = 2$ the above results may be summarized:

Theorem 6.2: The preimage of an exposed point of $D^2(H^N)$ is unique if either $N \le 4$ or $(r-N) \le 4$.

7. SOME EXPOSED POINTS OF $D^2(H^N)$ WHERE $r > N + 2$

In this section we produce a list of new exposed points and thereby add to our catalog of information on the convex structure of $D^2(H^N)$. The basic new result is contained in Theorem 7.1 which gives a prescription for manufacturing exposed points of $D^2(H^N)$ knowing some exposed points of $D^2(H^p)$ where $p < N$. In the simplest case the theorem states that if $D^2(\theta)$, $\theta \in H^p$ and $D^2(\phi)$, $\phi \in H^q$ are both exposed, then $D^2(\theta \wedge \phi)$ is exposed in $D^2(H^N)$, $N = p + q$ *if* θ and ϕ are strongly orthogonal. Even though this is a result that everyone would "suspect" to be true, it requires a lengthy argument to prove. Moreover, since θ and

ϕ are strongly orthogonal the wavefunction $\theta \wedge \phi$ seems to describe a system of two noninteracting parts. On physical grounds then it is clear that our theorem will not be powerful enough to allow us to construct all of the elements of $\mathrm{EXP}D^2(H^N)$ given say our complete understanding of the $r = N + 2$ case (c f., Sec. 5) or known examples of elements of $\mathrm{EXP}D^2(H^N)$.

Using Theorem 7.1 and Coleman's new result,[14] that arbitrary geminal power wavefunctions cover exposed points of $D^2(H^N)$, we give a long list of exposed points of $D^2(H^N)$ in Theorem 7.3 which includes all previously published exposed points and, in addition, some new ones. All of these exposed points are covered either by geminal power wavefunctions or wavefunctions that are manufactured by taking the Grassman product of several geminal power wavefunctions where the geminals are strongly orthogonal. In the next section we show that the list of Theorem 7.3 does not exhaust $\mathrm{EXP}D^2(H^N)$, thus substantiating our above intuition.

Three technical results are needed in order to establish the major results of this section. Let $H^1 = F^1 \oplus G^1$, where F^1 and G^1 are two mutually orthogonal subspaces of H^1. Denote by $F^p \wedge G^q$ the subspace of H^N $(N = p + q)$ generated by the functions $\theta \wedge \phi = A_N \theta\phi$, where $\theta \in F^p$ and $\phi \in G^q$ and A_N is the antisymmetrizer. Interpret $F^0 \wedge G^N$ as equal to G^N. Then $H^N = \oplus_{j=0}^{N} F^i \wedge G^{N-i}$. The product $(\theta, \phi) \to \theta \wedge \phi$ is usually called either the Grassman or wedge product.

Proposition 7.1: Suppose F^1 is perpendicular to G^1 and $H^1 = F^1 \oplus G^1$. Then $D^2(F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1})$ is an exposed subset of $D^2(H^N)$ for $N - 1 \geq p \geq 0$; $p + q = N$.

Proof: Let $N_F \in L(H^N)$ be the one-particle operator which counts the number of particles in F states. That is, for $\Psi \in F^p \wedge G^q$ we have $N_F\Psi = p\Psi$. [If $\{f_i\}$ is a set of annihilation operators corresponding to a complete set of orthonormal basis functions for F^1, then $N_F = \Sigma f_i^\dagger f_i$.] Let $b(p, p+1)$ be the unique element of $L(H^2)$ whose image under the expansion operator Γ_2^N is given by

$$\Gamma_2^N(b(p,p+1)) = \tfrac{1}{2}(N_F - p)(N_F - p - 1) \quad \text{for} \quad N - 1 \geq p \geq 0.$$

Then $\Gamma_2^N(b(p,p+1))$ has the following properties: (i) It is constant on each of the subspaces $F^i \wedge G^{N-i}$, $i = 0, \ldots, N$; (ii) its kernel is $F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1}$, and (iii) its lowest eigenvalue on the orthogonal complement of $F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1}$ is 1. The first two statements are easy to verify. To prove the third, note that the minimum eigenvalue on $(F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1})^\perp$ occurs for $\Psi \in F^{p-1} \wedge G^{q+1}$ $(1 \leq p \leq N - 1)$ and/or $\Psi \in F^{p+2} \wedge G^{q-2}$ $(0 \leq p \leq N - 2)$ and for such a Ψ, $\Gamma_2^N(b(p,p+1))\,\Psi = \Psi$. It follows, therefore, that $b(p,p+1) \in \mathrm{CONE}\tilde{D}^2(H^N)$ and for $\Psi \in H^N$, $\mathrm{Tr}(b(p,p+1)^\dagger D^2(\Psi)) = 0$ if and only if $\Psi \in F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1}$. Therefore, $D^2(F^p \wedge G^q \oplus F^{p+1} \wedge G^{q-1})$ is an exposed subset of $D^2(H^N)$.

Corollary 7.1: $D^2(F^p \wedge G^q)$ is an exposed subset of $D^2(H^N)$ for $0 \leq p \leq N$.

Proof: Let $b(p) \in L(H^2)$ be such that

$$\Gamma_2^N(b(p)) = \begin{cases} \Gamma_2^N((b(p-1,p) + b(p,p+1)) & \text{for } 1 \leq p \leq N-1 \\ N_F & \text{for } p = 0 \\ N_G & \text{for } p = N \end{cases}$$

(N_G is defined similarly to N_F.) Then $\Gamma_2^N(b(p))$ (i) is constant on each of the subspaces $F^i \wedge G^{N-i}$, $i = 0, \ldots, N$, (ii) has kernel $F^p \wedge G^q$, and (iii) its lowest eigenvalue on $(F^p \wedge G^q)^\perp$ is equal to 1. These properties follow directly from the properties of $\Gamma_2^N(b(p,p+1))$ listed in the proof of proposition 7.1 and the definition of N_F and N_G. Therefore, $b(p) \in \mathrm{CONE}\tilde{D}^2(H^N)$ and $D^2(F^p \wedge G^q)$ is an exposed subset.

The techniques employed in the proof of Proposition 7.1 and Corollary 7.1 can be employed to obtain

Corollary 7.2: Let $H^1 = \oplus_{i=1}^{m} F_i$ and the F_i be mutually orthogonal. Then $D^2(F_1^p \wedge F_2^q \wedge \cdots \wedge F_m^s)$ is an exposed subset of $D^2(H^N)$ where $N = p + q + \cdots + s$ and $H^N = \oplus_{i+j+\cdots+k=N} F_1^i \wedge F_2^j \wedge \cdots \wedge F_m^k$.

Corollary 7.3: Extreme points of $D^2(F^p \wedge G^q)$ are extreme in $D^2(H^N)$; $N = p + q$.

Proof: By Corollary 7.1, $D^2(F^p \wedge G^q)$ is an exposed subset of $D^2(H^N)$ and, therefore, an extreme subset. But by a general theorem in the theory of convex sets, the extreme points of an extreme subset of a convex set C are extreme in C. Therefore $\mathrm{EXT}D^2(F^p \wedge G^q) \subset \mathrm{EXT}D^2(H^N)$.

Corollary 7.4: Let F^1 be a subspace of some Hilbert space H^1. Then the extreme points of the convex set $D^2(F^N)$ are extreme in $D^2(H^N)$.

Proof: Let $G^1 = (F^1)^\perp$. Then $F^N = F^N \wedge G^0$. Apply Corollary 7.3.

By Corollary 7.3 we know that the extreme points of $D^2(F^p \wedge G^q)$ are extreme in $D^2(H^N)$. We have seen in Sec. 3 that these extreme points are also exposed points of $D^2(H^N)$ if we assume that H^1 has finite dimension. The following proposition does not assume that $\dim H^1 < \infty$ and asserts that the exposed points of $D^2(F^p \wedge G^q)$ are contained in $\mathrm{EXP}D^2(H^N)$. The proof of this result makes use of some special properties of density matrices. We recall again that if E is an exposed subset of a convex set C it is not, in general, true that exposed subsets of E are exposed in C. In particular, elements of $\mathrm{EXP}E$ are not necessarily elements of $\mathrm{EXP}C$.

Proposition 7.2: The exposed points of $D^2(F^p \wedge G^q)$ are exposed in $D^2(H^N)$ where $H^N = \oplus_{i=0}^{n} F^i \wedge G^{N-i}$. $\dim H^1 = \dim(F^1 \oplus G^1)$ may be infinite.

Proof: Let $D^2(\Psi)$ be an exposed point of $D^2(F^p \wedge G^q)$ which is exposed by $b(\Psi) \in \mathrm{CONE}\tilde{D}^2(F^p \wedge G^q)$. Note that $b(\Psi)$ need not be a member of $\mathrm{CONE}\tilde{D}^2(H^N)$ which is a proper subset of $\mathrm{CONE}\tilde{D}^2(F^p \wedge G^q)$ since $D^2(F^p \wedge G^q) \subset D^2(H^N)$. We may assume that $\Gamma_2^N(b(\Psi))$ leaves the subspaces $F^i \wedge G^{N-i}$, $i = 0, \ldots, N$, invariant or, equivalently, that $b(\Psi)$ contains no terms which do not conserve both F and G particles, that is, terms of the form $f_1^\dagger f_2^\dagger g_1 f_3$, $g_1^\dagger f_2^\dagger g_1 g_2$, etc., where f_1, f_2, f_3 and g_1, g_2 are annihilators for F and G states, respec-

tively, for if $b(\Psi)$ does not conserve F particles, it can be written as $b(\Psi) = b_0 + b_1$, where b_0 leaves the subspaces $F^i \wedge G^{N-i}$ invariant and b_1 maps elements of $F^i \wedge G^{N-i}$ into $(F^i \wedge G^{N-i})^\perp$. In second quantization notation, b_1 contains only terms of the form $f_1^\dagger f_2^\dagger g_1 f_3, f_1^\dagger f_2^\dagger g_1 g_2, \cdots$. But then, for $\xi \in F^p \wedge G^q$, $\mathrm{Tr}(b_1^\dagger D^2(\xi)) = \mathrm{Tr}(\Gamma_2^N(b_1)^\dagger D^N(\xi)) = \langle \xi | \Gamma_2^N(b_1)^\dagger \xi \rangle = 0$. Therefore if $b(\Psi)$ exposes $D^2(\Psi)$ then so does b_0; its F and G particle conserving part.

Let

$$\lambda = \inf_{\xi \in (F^p \wedge G^q)^\perp} \mathrm{Tr}(b(\Psi)^\dagger D^2(\xi))$$

and choose $1 > t > 0$ such that $t\lambda + (1-t) > 0$. Then if $b(p)$ is the operator defined in Corollary 7.1, $b = tb(\Psi) + (1-t)b(p)$ exposes $D^2(\Psi)$ in $D^2(H^N)$. We note that $\Gamma_2^N(b)$ leaves the subspaces $F^i \wedge G^{N-i}$, $i = 0, \ldots, N$, invariant. Since $\Gamma_2^N(b(p)) \equiv 0$ on $F^p \wedge G^q$ and ≥ 1 on $(F^p \wedge G^q)^\perp$, we have for $\xi \in F^p \wedge G^q$,

$$\begin{aligned}\mathrm{Tr}(b^\dagger D^2(\xi)) &= t\,\mathrm{Tr}(b(\Psi)^\dagger D^2(\xi)) + (1-t)\,\mathrm{Tr}(b(p)^\dagger D^2(\xi)) \\ &> 0 \text{ if } D^2(\xi) \neq D^2(\Psi) \text{ since } t > 0 \\ &= 0 \text{ if } D^2(\xi) = D^2(\Psi)\end{aligned}$$

for $\xi \in (F^p \wedge G^q)^\perp$,

$$\begin{aligned}\mathrm{Tr}(b^\dagger D^2(\xi)) &= t\,\mathrm{Tr}(b(\Psi)^\dagger D^2(\xi)) + (1-t)\,\mathrm{Tr}(b(p)^\dagger \\ &\quad \times D^2(\xi)) \geq t\lambda + (1-t) > 0\end{aligned}$$

by the choice of t. Therefore, $b \in \mathrm{CONE}\tilde{D}^2(H^N)$ and $\mathrm{Tr}(b^\dagger D^2(\Psi)) = 0$ if and only if $D^2(\xi) = D^2(\Psi)$.

Proposition 7.3: Suppose $F^1 \perp G^1$, $\theta_1, \theta_2 \in F^p (p \geq 2)$, and $\phi_1, \phi_2\ G^q (q \geq 2)$, then

$$\begin{aligned}N(N-1)&D^2(\theta_1 \wedge \phi_1, \theta_2 \wedge \phi_2)(xy\,|\,x'y') \\ &= p(p-1)\,\langle \phi_1 |\, \phi_2 \rangle\, D^2(\theta_1, \theta_2)(xy\,|\,x'y') \\ &\quad + q(q-1)\langle \theta_1 |\, \theta_2 \rangle D^2(\phi_1, \phi_2)(xy\,|\,x'y') \\ &\quad + pq[D^1(\theta_1, \theta_2)(x\,|\,x')D^1(\phi_1, \phi_2)(y\,|\,y') \\ &\quad - D^1(\theta_1, \theta_2)(x\,|\,y')D^1(\phi_1, \phi_2)(y\,|\,x') \\ &\quad - D^1(\theta_1, \theta_2)(y\,|\,x')D^1(\phi_1, \phi_2)(x\,|\,y') \\ &\quad + D^1(\theta_1, \theta_2)(y\,|\,y')D^1(\phi_1, \phi_2)(x\,|\,x')].\end{aligned}$$

Proof: Recall that all of the density matrices in the above expression are normalized to 1. The result follows directly from the definition of the transition density and the fact that the θ's are strongly orthogonal to the ϕ's.

The following theorem allows one to construct a large number of exposed points of $D^2(H^N)$ given some exposed points of $D^2(H^p)$ where p is smaller than N.

Theorem 7.1: Let $\phi_i, i = 1, \ldots, m$ be mutally strongly orthogonal N_i-particle functions such that $D^2(\phi_i)$ is an exposed point of $D^2(H^{N_i})$. Then

$$D^2(\phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_m)$$

is exposed in $D^2(H^N)$ where $N = \sum_{i=1}^m N_i$.

Proof: We give a proof for the case when $m = 2$ and $N_1 \geq 2$, $N_2 \geq 2$; the techniques used in the general case being a simple extension of those used here. Suppose that θ and ϕ are functions of p (≥ 2) and q (≥ 2), particles respectively. We show that *if* (i) θ and ϕ are strongly orthogonal and (ii) $D^2(\theta) \in \mathrm{EXP}D^2(H^p)$ and $D^2(\phi) \in \mathrm{EXP}D^2(H^q)$, *then* $D^2(\theta \wedge \phi)$ is exposed in $D^2(H^N)$ where $N = p + q$.

Since θ and ϕ are strongly orthogonal one may pick mutually orthogonal subspaces F^1 and G^1 of H^1 such that (i) $H^1 = F^1 \oplus G^1$ and (ii) $\theta \in F^p$ and $\phi \in G^q$. But then $\theta \wedge \phi \in F^p \wedge G^q$. By Proposition 7.2 one need only show that $D^2(\theta \wedge \phi)$ is exposed in $D^2(F^p \wedge G^q)$ in order to establish that $D^2(\theta \wedge \phi)$ is exposed in $D^2(H^N)$.

Let $\hat{b}(\theta) \in \mathrm{CONE}\tilde{D}^2(F^p) \subset L(F^2)$ be an exposing operator for $D^2(\theta) \in \mathrm{EXP}D^2(F^p)$. Let $b(\theta)$ be its extension to H^2 defined by

$$b(\theta)\,|\,F^2 \equiv \hat{b}(\theta)$$

$$b(\theta)\,|\,F^1 \wedge G^1 \oplus G^2 \equiv 0.$$

By letting $\{f_i\}$ denote a set of annihilation operators corresponding to a complete orthonormal basis for F^1, $b(\theta)$ can be written in the form $b(\theta) = \sum b(\theta)_{ijkl} f_i^\dagger f_j^\dagger f_k f_l$. By formula (2.1) for the expansion operator, $\Gamma_2^N(b(\theta)) = \binom{N}{2}^{-1} \sum b(\theta)_{ijkl} f_i^\dagger f_j^\dagger f_k f_l$. We suppose that $\{g_i\}$ is a set of annihilation operators corresponding to a basis for G^1. Assume that $\delta \in F^p$ and $\omega \in G^q$, then $\delta \wedge \omega$ can be written $\Delta^\dagger \Omega^\dagger |\, 0\rangle$, where $\Delta^\dagger$ is a homogeneous polynomial of order p in the creation operators $f_i^\dagger$ and $\Omega^\dagger$ is a homogeneous polynomial of order q in the creation operators $g_i^\dagger$. Thus

$$\begin{aligned}\Gamma_2^N(b(\theta))\delta \wedge \omega &= [\Gamma_2^N(b(\theta))\Delta^\dagger]\Omega^\dagger |\, 0\rangle \\ &= \binom{N}{2}^{-1}\binom{P}{2}[\Gamma_2^P(b(\theta))\delta] \wedge \omega. \quad (7.1)\end{aligned}$$

Similarly if $\hat{b}(\phi)$ is an element of $\mathrm{CONE}\tilde{D}^2(G^q)$ which exposes $D^2(\phi)$ and $b(\phi)$ is its extension to H^2, we have

$$\Gamma_2^N(b(\phi))\delta \wedge \omega = \binom{N}{2}^{-1}\binom{q}{2}\,\delta \wedge (\Gamma_2^q(b(\phi))\omega). \quad (7.2)$$

Let $\{\theta_i\}$ be an orthonormal basis spanning the kernel of $\Gamma_2^p(b(\theta))$ and $\{\phi_i\}$ be an orthonormal basis spanning the kernel of $\Gamma_2^q(b(\phi))$. From formulas (7.1) and (7.2) we conclude that (i) $\theta \wedge \phi$ lies in the kernel of $\Gamma_2^N(b(\theta) + b(\phi))$ and (ii) the kernel of $\Gamma_2^N(b(\theta) + b(\phi))$ is spanned by the orthonormal basis $\{\theta_i \wedge \phi_j\}$. We now show that if Ψ lies in the kernel of $\{\Gamma_2^N(b(\theta) + b(\phi))$, then necessarily $D^2(\Psi) = D^2(\theta \wedge \phi)$ thus showing that $D^2(\theta \wedge \phi)$ is an exposed point of $D^2(F^p \wedge G^q)$ with the exposing operator $b(\theta) + b(\phi)$ and thereby establishing the theorem.

As $\hat{b}(\theta) \in D^2(F^p)$ is an exposing operator for $D^2(\theta)$ we conclude that $D^2(\theta_i) = D^2(\theta)$. By Lemma 6.1' we can conclude that $D^2(\theta_i, \theta_j) = 0$. Similarly we have $D^2(\phi_j) = D^2(\phi)$ and $D^2(\phi_i, \phi_j) = 0$. Let $\Psi \in \mathrm{KER}\Gamma_2^N(b(\theta) + b(\phi))$. Then

$$\Psi = \sum \alpha_{ij}\theta_i \wedge \phi_j, \quad \sum |\alpha_{ij}|^2 = 1.$$

Then, applying Proposition 7.3, we have

$$\begin{aligned}D^2(\Psi) &= \sum_{ijkl} \bar{\alpha}_{ij}\alpha_{kl} D^2(\theta_i \wedge \phi_j, \theta_k \wedge \phi_l) \\ &= \sum_{ijkl} \alpha_{ij}\alpha_{kl}[N(N-1)]^{-1}\{p(p-1)\langle \phi_j |\, \phi_l \rangle D^2(\theta_i, \theta_k)\end{aligned}$$

$$\times (xy|x'y') + q(q-1)\langle\theta_i|\theta_k\rangle D^2(\phi_j,\phi_l)(xy|x'y')$$
$$+ pq\{D^1(\theta_i,\theta_k)(x|x')D^1(\phi_j,\phi_l)(y|y')$$
$$- D^1(\theta_i,\theta_k)(x|y')D^1(\phi_j,\phi_l)(y|x')$$
$$- D^1(\theta_i,\theta_k)(y|x')D^1(\phi_j,\phi_l)(x|y')$$
$$+ D^1(\theta_i,\theta_k)(y|y')D^1(\phi_j,\phi_l)(x|x')]\}$$
$$= \sum|\alpha_{ij}|^2[N(N-1)]^{-1}\{p(p-1)D^2(\theta) + q(q-1)$$
$$\times D^2(\phi) + pq[D^1(\theta)(x|x')D^1(\phi)(y|y')$$
$$- D^1(\theta)(x|y')D^1(\phi)(y|x') - D^1(\theta)(y|x')$$
$$\times D^1(\phi)(x|y') + D^1(\theta)(y|y')D^1(\phi)(x|x')]\}.$$

Coleman[14] has proved the following theorem regarding AGP (antisymmetrized general power) wavefunctions:

Theorem 7.2: If ϕ is an arbitrary geminal with rank $r \geq N$, then the AGP function $\phi^N(=\phi\wedge\phi\wedge\cdots\wedge\phi$, $\frac{1}{2}N$ factors) covers an exposed point $D^2(\phi^N)$ of $D^2(H^N)$, and the preimage of $D^2(\phi^N)$ under the contraction mapping is unique.

This result along with Theorem 7.1 allows one to construct a very large number of exposed points of $D^2(H^N)$. We recall that the linear space σ^N is spanned by elements of the form $\Phi^\dagger|0\rangle$, where $\Phi^\dagger$ is a homogeneous polynomial of order $\frac{1}{2}N$ in the pair creation operators $\{\sigma_i^\dagger\}$, $\sigma_i = a_{2i-1}a_{2i}$ and that $D^2(\sigma^N)$ is an exposed subset of $D^2(H^N)$ (see Sec. 2).

Theorem 7.3: If θ_i, $i=1,\ldots,s$ are one-particle functions and ϕ_j, $j=1,\ldots,t$ are geminals of rank greater than N_i, and if the θ_i and the ϕ_j are mutually strongly orthogonal, then

(i) $\Psi = \theta_1\wedge\theta_2\wedge\cdots\wedge\theta_s\wedge\phi_1^{N_1}\wedge\phi_2^{N_2}\wedge\cdots\wedge\phi_t^{N_t}$

covers an exposed point of $D^2(H^N)$, where $N = s + \sum_{i=1}^t N_i$,

(ii) the preimage of $D^2(\Psi)$ under the contraction mapping L_N^2 is unique and equal to $D^N(\Psi)$, and

(iii) when N is even, $D^2(\Psi)$ is a member of the exposed subset $D^2(\sigma^N)$ for some choice of pairing.

Proof: The proof of the fact that $D^2(\phi)$ is exposed is a direct application of Theorem 7.1. The uniqueness of the preimage of $D^2(\Psi)$ is an easy result which follows from the uniqueness of the preimages of the $D^2(\phi_j^{N_j})$ under the contraction mappings $L_{N_j}^2$. When N is even, it follows from Lemma 5.1 and strong orthogonality that we can construct a set of pair annihilation operators $\{\sigma_i\}$, $\sigma_i = a_{2i-1}a_{2i}$, such that (i) $\theta_1\wedge\theta_2\wedge\cdots\wedge\theta_s = \sum_{j\in J_0}\sigma_j^\dagger|0\rangle$, (ii) $\phi_i = \Phi_i^\dagger|0\rangle$ $\Phi_i^\dagger = \sum_{j\in J_i}\phi_{ij}\sigma_j^\dagger$, and (iii) $J_i\wedge J_j = 0$, $i\neq j$. Thus Ψ can be written

$$\Psi = \Big(\prod_{j\in J_0}\sigma_1^\dagger\Big)(\Phi_1^\dagger)^{N_1}(\Phi_2^\dagger)^{N_2}\cdots(\Phi_j^\dagger)^{N_t}|0\rangle$$

and is, therefore, a member of the paired subspace σ^N.

8. THERE ARE MORE EXTREME POINTS

The preceding leaves open the question whether the list of Theorem 7.3 exhausts all the extreme points of $D^2(H^N)$. Corollaries 8.1 and 8.2 show that such is not the case. For even N, each exposed point listed in Theorem 7.3 is an element of $D^2(\sigma^N)$ for some choice of pairing, i.e., is covered by an element of σ^N, one of the paired subspaces of H^N (see Sec. 2). In Corollary 8.1 we give an example of an exposed point of $D^2(H^N)$ which is not contained in the list of Theorem 7.3 and which for even N is not contained in any of the exposed subsets $D^2(\sigma^N)$. But the question remains as to whether, for even N, Theorem 7.3 supplies a comprehensive enough list to contain all the exposed points of $D^2(\sigma^N)$. Again the answer is no as is shown by Corollary 8.2.

In Theorem 8.1, we give an example of an extreme point where $N=3$ and $r=7$, which for certain choices of the defining parameters, is not of the type given by Theorem 7.3. For this case, Theorem 7.3 only yields extreme points which are either covered by a Slater determinant or are of the form $D^2(\theta\wedge\phi)$, where θ is an orbital strongly orthogonal to the geminal ϕ. In both cases the two-density contracts to a one-density that has at least one eigenvalue equal to $\frac{1}{3}$. When the two-density is covered by a wavefunction of the form $\theta\wedge\phi$, the eigenvalues which are less than $\frac{1}{3}$ are all doubly degenerate. In Theorem 8.1 we establish the existence of an extreme point of $D^2(H^3)$ which contracts to a one-density, all of whose eigenvalues are nondegenerate and less than $\frac{1}{3}$.

Theorem 8.1: $D^2(\Psi_0)$ is exposed if

$$\Psi_0 = \alpha[123] + \beta[145] + \gamma[167] + \delta[246] + \epsilon[357],$$

and [123] is the Slater determinant made from orbital one, two, and three. If $P(1)$ is the projector onto orbital number one, the corresponding one-density is given by

$$D^1(\Psi_0) = \tfrac{1}{3}[(\alpha^2+\beta^2+\gamma^2)P(1) + (\alpha^2+\delta^2)P(2)$$
$$+ (\alpha^2+\epsilon^2)P(3) + (\beta^2+\delta^2)P(4) + (\beta^2+\epsilon^2)P(5)$$
$$+ (\gamma^2+\delta^2)P(6) + (\gamma^2+\epsilon^2)P(7)].$$

If we choose $\alpha^2=\frac{1}{2}$, $\beta^2=\frac{1}{4}$, $\gamma^2=\frac{1}{16}$, $\delta^2=\frac{1}{8}$, and $\epsilon^2=\frac{1}{16}$, then all of the eigenvalues of $D^1(\Psi_0)$ are less than $\frac{1}{3}$ and are nondegenerate. In this case $D^2(\Psi_0)$ is not contained in Theorem 7.3.

Proof: The wavefunction Ψ_0 is annihilated by the positive operators:

$$b_1 = P(\beta[23] - \alpha[45]), \qquad b_8 = P(\epsilon[16] - \gamma[35]),$$
$$b_2 = P(\gamma[45] - \beta[67]), \qquad b_9 = P(25),$$
$$b_3 = P(\delta[13] + \alpha[46]), \qquad b_{10} = P(27)$$
$$b_4 = P(\epsilon[12] - \alpha[57]), \qquad b_{11} = P(34),$$
$$b_5 = P(\gamma[15] - \beta[26]), \qquad b_{12} = P(36),$$
$$b_6 = P(\epsilon[14] + \beta[37]), \qquad b_{13} = P(45),$$
$$b_7 = P(\delta[17] + \gamma[24]), \qquad b_{14} = P(56),$$

where, for example, $P(\beta[23]-\alpha[45])$ is given by $A^\dagger A$, $A = \beta a_2a_3 - \alpha a_4a_5$ and a_i is the annihilator for orbital i. Moreover, Ψ_0 is the unique ground state of the operator $b = \sum_{i=1}^{14} b_i$. This may be checked by finding the eigenfunctions and eigenvalues of the operators b_i and noticing that Ψ_0 is the only vector which is

simultaneously a ground state for all of them. Thus $D^2(\Psi_0)$ is an exposed point of $D^2(H^3)$. The remarks regarding the properties of $D^1(\Psi_0)$ are easy to verify.

By using the Ψ_0 of Theorem 8.1 and Theorem 7.1 we can construct exposed points of $D^2(H^N)$ for $N > 3$ which are not of the type given in Theorem 7.3. If θ_i, $i = 1, \ldots, s$ are mutually orthogonal one-particle functions strongly orthogonal to Ψ_0, then

$$\Psi = \theta_1 \wedge \theta_2 \wedge \cdots \wedge \theta_s \wedge \Psi_0$$

covers an exposed point of $D^2(H^N)$; $N = s + 3$. If we choose the defining parameters of Ψ_0 as in the second half of Theorem 8.1, the eigenvalues of $D^1(\Psi)$ which are less than N^{-1} are not doubly degenerate. But all of the exposed points of Theorem 7.3 contract to one-densities whose eigenvalues are either equal to N^{-1} or are doubly degenerate. Thus $D^2(\Psi)$ is not of the type given by Theorem 7.3. For any fixed pairing the elements of $D^2(\sigma^N)$ contract to one-densities whose eigenvalues are all doubly degenerate. Thus, for even N, $D^2(\Psi)$ is not a member of any of the exposed subsets $D^2(\sigma^N)$.

Corollary 8.1: For certain values of the defining parameters the wavefunction

$$\Psi = \theta_1 \wedge \theta_2 \wedge \cdots \wedge \theta_s \wedge \Psi_0 \in H^N, \qquad N = s + 3$$

given above, covers an exposed point of $D^2(H^N)$ which is not contained in the list of Theorem 7.3. For even N, $D^2(\Psi)$ is not a member of any of the exposed subsets $D^2(\sigma^N)$.

For $N = 4$, Theorem 7.3 establishes the existence of two types of exposed points of $D^2(H^N)$. They are of the form $D^2(\phi \wedge \phi)$ and $D^2(\theta, \phi)$, where ϕ and θ are strongly orthogonal two-particle functions. Both are elements of the exposed subset $D^2(\sigma^N)$ for some fixed pairing. In Theorem 8.3 we establish the existence of an exposed point of the form $D^2(\Psi)$, where $\Psi = \alpha\phi \wedge \phi + \beta\theta \wedge \phi$ and $\alpha, \beta \neq 0$. Here ϕ is again strongly orthogonal to θ. A similar argument to that in the proof of Theorem 7.3 shows that this exposed point is also a member of one of the exposed subsets $D^2(\sigma^N)$. Recalling that the exposed points of $D^2(H^2)$ have unique preimages under the contraction map L_4^2 (Theorem 6.2), we can demonstrate that $D^2(\alpha\phi \wedge \phi + \beta\theta \wedge \phi)$ is a new type of exposed point if $\alpha\phi \wedge \phi + \beta\theta \wedge \phi$ is *not* proportional to either a function of the form $\xi \wedge \xi$ or a function of the form $\xi \wedge \omega$ where ξ is strongly orthogonal to ω. Arguing as in the proof of Theorem 7.3 there exists a set of pair annihilation operators $\{\sigma_i\}$, $\sigma_i = a_{2i-1}a_{2i}$ such that

$$\Psi = \alpha\phi \wedge \phi + \beta\theta \wedge \phi = (\alpha(\Phi^\dagger)^2 + \beta\Theta^\dagger\Phi^\dagger)\,|\,0\rangle,$$

where $\Phi^\dagger = \sum_{i \in I} \phi_i\sigma_i^\dagger$, $\Theta^\dagger = \sum_{j \in J} \theta_j\sigma_j^\dagger$, $I \cap J = 0$. If Ψ were proportional to $\xi \wedge \omega$ then we could write it, using the pair annihilation operators, as $\Xi^\dagger\Omega^\dagger\,|\,0\rangle$, where $\Xi^\dagger = \sum_{k \in K} \xi_k\sigma_k^\dagger$, $\Omega^\dagger = \sum_{l \in L} \omega_l\sigma_l^\dagger$, and $K \cap L = 0$. Each term in the expansion of $\Xi^\dagger\Omega^\dagger$ has one pair from K and one pair from L. But Ψ is clearly not of this form, and Ψ is not proportional to $\xi \wedge \omega$. If we require that the rank of $D^1(\theta)$ be greater than 4 we can show that Ψ is not proportional to a function of the form $\xi \wedge \xi$. For proportionality to hold, $\xi \wedge \xi$ must be of the form $(\Xi^\dagger)^2\,|\,0\rangle$ where $\Xi^\dagger = \sum_{k \in K} \xi_k\sigma_k^\dagger$. But if there are nonzero terms involving $\sigma_m^\dagger$ and $\sigma_n^\dagger$ in the expansion of $(\Xi^\dagger)^2$, there is also a nonzero term of the form $\gamma\sigma_m^\dagger\sigma_n^\dagger$. If $D^1(\theta)$ has rank greater than 4, there are at least two terms in the expansion of θ which are nonzero. We can say that θ_m and $\theta_n \neq 0$. Thus there are nonzero terms in the expansion of Ψ involving both $\sigma_m^\dagger$ and $\sigma_n^\dagger$. However, it is clear that there are no terms in the expansion of Ψ of the form $\gamma\sigma_m^\dagger\sigma_m^\dagger\,|\,0\rangle$.

We establish the existence of an exposed point of the form $D^2(\Psi)$, where $\Psi = \alpha\phi \wedge \phi + \beta\theta \wedge \phi$, by first showing that a particular linear space V, containing Ψ, covers an exposed subset of $D^2(H^N)$ and then showing that $D^2(\Psi)$ is an exposed point of $D^2(V)$. But by Lemma 3.1 $D^2(\Psi)$ is then exposed in $D^2(H^N)$. The first step is to demonstrate that certain linear subspaces of H^N cover exposed subsets of $D^2(H^N)$. We assume some fixed pairing and as before denote the pair annihilation operators by $\{\sigma_i\}$, $i \in R$. Let I be a subset of R, and let G denote the linear subspace of H^2 spanned by $\{\sigma_i^\dagger\,|\,0\rangle \mid i \in I\}$. If J is the complement of I in R, i.e., $J = R - I$, then let F be the linear subspace spanned by $\{\sigma_j^\dagger\,|\,0\rangle \mid j \in J\}$. The linear subspace σ^N can be written as the direct sum of the linear subspaces $F^{2i} \wedge G^{N-2i}$, $i = 0, \ldots, \frac{1}{2}N$ where $F^{2p} \wedge G^{N-2p}$ is the linear space generated by all basis vectors of the form $\sigma_{j_1}^\dagger \cdots \sigma_{j_p}^\dagger\,\sigma_{i_1}^\dagger \cdots \sigma_{i_q}^\dagger\,|\,0\rangle$, $j_1, \ldots, j_p \in J$, $i_1, \ldots, i_q \in I$, and $p + q = \frac{1}{2}N$. The element of F^2 are strongly orthogonal of those of G^2.

Proposition 8.1: $D^2(\sigma^N)$ is an exposed subset of $D^2(H^N)$.

Proof: In Sec. 2 we demonstrated that the operator $1 - (2/N)\sum_{i \in R} \sigma_i^\dagger\sigma_i$ was an exposing operator for $D^2(\sigma^N)$.

Proposition 8.2: $D^2(G^N \oplus F^2 \wedge G^{N-2})$ is an exposed subset of $D^2(H^N)$.

Proof: Recall that $\sigma_j = a_{2j-1}a_{2j}$. The operator $N_F = \sum_{j \in J} a_{2j}^\dagger a_{2j}$, when restricted to σ^N, counts the number of F pairs. The operator $N_F(N_F - 1)$ when restricted to σ^N is (i) positive semi-definite and (ii) has kernel $G^N \oplus F^2 \wedge G^{N-2}$. Thus $D^2(G^N \oplus F^2 \wedge G^{N-2})$ is an exposed subset of $D^2(\sigma^N)$ and, therefore, by Lemma 3.1 an exposed subset of $D^2(H^N)$.

Let $\theta = \Theta^\dagger\,|\,0\rangle \in F^2$, $\phi = \Phi^\dagger\,|\,0\rangle \in G^2$, and $\xi = \Xi^\dagger|0\rangle \in G^2$, where $\Theta^\dagger = \sum_{j \in J} \theta_j\sigma_j^\dagger$, $\Phi^\dagger = \sum_{i \in I} \phi_i\sigma_i^\dagger$, $\Xi^\dagger = \sum_{i \in I} \xi_i\sigma_i^\dagger$, and $1 = \sum_{j \in J} |\theta_j|^2 = \sum_{i \in I} |\phi_i|^2 = \sum_{i \in I} |\xi_i|^2$. Let b be the restriction of $\Phi^\dagger\Theta + \Theta^\dagger\Phi$ to $G^4 \oplus F^2 \wedge G^2$. The eigenfunctions of the restricted operator either (i) lie in its kernel or (ii) are of the form $\alpha\phi \wedge \xi + \beta\theta \wedge \xi$. We denote the cardinalities of I, J, R by $|I|$, $|J|$, and $|R|$. Let $M = \dim(G^4 \oplus F^2 \wedge G^2)$.

Theorem 8.2: Let b be the restriction of $\Phi^\dagger\Theta + \Theta^\dagger\Phi$ to $G^4 \oplus F^2 \wedge G^2$. Then

(i) there are $2|I|$ eigenvalues and corresponding eigenfunctions of b occurring in pairs and given by

$$+\sqrt{\lambda}, \quad \phi \wedge \xi + \sqrt{\lambda}\ \theta \wedge \xi,$$
$$-\sqrt{\lambda}, \quad \phi \wedge \xi - \sqrt{\lambda}\ \theta \wedge \xi,$$

where λ is an eigenvalue and $\{\xi_i\}$, $i \in I$, the corresponding eigenvector of the matrix

$$\{(1 - 2|\phi_i|^2)\delta_{ij} + \phi_i\bar{\phi}_j\}, \quad i,j \in I.$$

(ii) There are $M - 2|I|$ additional eigenfunctions lying in the kernel of b.

(iii) The degeneracy of the top eigenvalue is equal to the degeneracy of the bottom eigenvalue. In the case of nondegeneracy both of these eigenfunctions cover exposed points of $D^2(H^N)$ which are elements of $D^2(\sigma^N)$. When the rank of $D^1(\theta)$ is greater than 4 or, equivalently, when $|J| \geq 2$, these exposed points are not contained in Theorem 7.3.

Proof: (i) As was pointed out above, an eigenfunction of b belonging to a nonzero eigenvalue must be of the form $\Psi = \alpha\phi \wedge \xi + \beta\theta \wedge \xi$. The action of b on $\theta \wedge \xi$ yields $\phi \wedge \xi$; the action of b on $\phi \wedge \xi$ yields $\Theta^\dagger\Xi^\dagger|0\rangle + \langle\phi|\xi\rangle\Theta^\dagger\Phi^\dagger|0\rangle - \sum 2|\phi_i|^2\xi_i\Theta^\dagger\sigma_i^\dagger|0\rangle$. In order for Ψ to be an eigenfunction, this latter vector must be proportional to $\theta \wedge \xi$. Thus we are led to the eigenvalue problem

$$(1 - 2|\phi_i|^2)\xi_i + \sum_{j\in I}\phi_i\bar{\phi}_j\xi_j = \lambda\xi_i, \quad i \in I. \quad (8.1)$$

For each solution of this eigenvalue problem one obtains two eigenfunctions of b. They are given in the statement of the theorem and obtained by an easy calculation. Since the eigenvalues of b are all real, those of (8a) must all be positive.

(ii) Since $M = \dim(G^4 \oplus F^2 \wedge G^2)$ there are $M - 2|I|$ eigenfunctions unaccounted for. These lie in the kernel of b.

(iii) If the bottom eigenvalue of b is nondegenerate then the corresponding eigenfunction Ψ covers an exposed point $D^2(G^4 \oplus F^2 \wedge G^2)$. But by Proposition 8.2 and Lemma 3.1 $D^2(\Psi)$ is exposed in $D^2(H^N)$. If the top eigenvalue of b is nondegenerate, then the same holds for the bottom eigenvalue of $-b$. The corresponding eigenfunction covers an exposed point of $D^2(H^N)$. Clearly these exposed points will be members of $D^2(\sigma^4)$. A modification of the argument immediately following corollary 8.1 will establish that these exposed points are not contained in Theorem 7.3 when the RANK$D^1(\theta) \geq 4$ or equivalently $|J| \geq 2$.

In the special case where $\phi = \Phi^\dagger|0\rangle$, $\Phi^\dagger = (|I|)^{-1/2}\sum_{i\in I}\sigma_i^\dagger$, (8a) can be easily solved and explicit formulas for the nonzero eigenvalues and corresponding eigenfunctions of b obtained.

Theorem 8.3: With ϕ as above and θ arbitrary, let b be the restriction of $\Theta^\dagger\Phi + \Phi^\dagger\Theta$ to $G^4 \oplus F^2 \wedge G^2$. Then

(i) the top and bottom eigenvalues of b and the corresponding eigenvectors are given by

$$+ (2 - 2/|I|)^{1/2}, \quad \phi \wedge \phi + (2 - 2/|I|)^{1/2}\theta \wedge \phi,$$
$$- (2 - 2/|I|)^{1/2}, \quad \phi \wedge \phi - (2 - 2/|I|)^{1/2}\theta \wedge \phi.$$

Both eigenvalues are nondegenerate. The two eigenfunctions cover exposed points of $D^2(H^N)$ which are contained in $D^2(\sigma^N)$. When the rank of $D^1(\theta)$ is greater than 4 these exposed points are not contained in Theorem 7.3.

(ii) There are $2|I| - 2$ eigenvalues and corresponding eigenfunctions given by

$$(1 - 2/|I|)^{1/2}, \quad \phi \wedge \xi + (1 - 2/|I|)^{1/2}\theta \wedge \xi,$$
degeneracy $|I| - 1 - (1 - 2/|I|)^{1/2}$,
$$\phi \wedge \xi - (1 - 2/|I|)^{1/2}\theta \wedge \xi, \quad \text{degeneracy } |I| - 1,$$

where $\xi \in G^2$ and is perpendicular to ϕ.

(iii) There are $M - 2|I|$ additional eigenfunctions lying in the kernel of b.

Proof: In order to prove the validity of these statements one needs merely to solve the eigenvalue problem (8a) and then apply Theorem 8.2.

Let K, I, and J be nonempty, nonintersecting subsets of R such that $I \cup J \cup K = R$. Define: $\omega = \Pi_{k\in K}\sigma_k^\dagger|0\rangle \in H^m$, $m = 2|K|$; $\phi = \Phi^\dagger|0\rangle$, $\Phi^\dagger = (|I|)^{-1/2}\sum_{i\in I}\sigma_i^\dagger$; $\theta = \Theta^\dagger|0\rangle$, $\Theta^\dagger = \sum_{j\in J}\theta_j\sigma_j^\dagger$. Then by Theorem 7.1 $\Psi = \omega \wedge \Psi_0 \in H^{m+4}$, where $\Psi_0 = \phi \wedge \phi \pm (1 - 2/|I|)^{1/2}\theta \wedge \phi$ covers an exposed point of $D^2(H^N)$; $N = m + 4$. Also $D^2(\Psi) \in D^2(\sigma^N)$, but is not contained in the list of Theorem 7.3 when $D^1(\theta) \geq 4$ by an argument similar to the one preceding Theorem 8.2

Corollary 8.2: $D^2(\Psi) \in D^2(H^N)$, where $\Psi = \omega \wedge \Psi_0$, ω and Ψ_0 are as above and $N = m + 4$ is an exposed point. $D^2(\Psi) \in D^2(\sigma^N)$, but is not contained in 7.10 when RANK $D^1(\theta) \geq 4$ or equivalently $|J| \geq 2$.

9. DISCUSSION

The list of extreme points of $D^2(H^N)$ given in Theorem 7.3 is of high dimension, but, as is shown in Sec. 8, it is not exhaustive. As there has been no systematic treatment of any other class of elements of EXT$D^2(H^N)$ we have no feeling for the number of extreme points which we have missed. It would be desirable to obtain an estimate of the portion of total volume of $D^2(H^N)$ which is occupied by the convex span of the elements of Theorem 7.3.

The characterization of $D^2(\sigma^N)$, the exposed subset of $D^2(H^N)$ covered by σ^N, a paired subspace of H^N, seems to be a much more tractable problem than the general N-representability problem. Moreover, this exposed subset is of considerable physical interest as it contains the AGP or projected BCS functions which have had considerable success in superconductivity and nuclear theory. Although the extreme points of Theorem 7.3 are all contained in $D^2(\sigma^N)$ for some choice of pairing, we have shown in Corollary 8.2 that there are elements of EXT$D^2(\sigma^N)$ which are not of this type. It is conceivable to the author that the techniques used to establish this counterexample and the extreme points of Theorems 8.2 and 8.3 might be exploited to obtain a much more general class of extreme points, containing those listed in Theorem 7.3 and exhausting $D^2(\sigma^N)$. The extreme points listed in Corollary 8.2 are of the form $D^2(\alpha\phi_1 + \beta\theta \wedge \phi_2)$, where $\phi_1 \in \sigma^N$, $\phi_2 \in \sigma^{N-2}$ both cover extreme points, and θ is a geminal strongly orthogonal to both ϕ_1 and ϕ_2. Thus an extreme point is constructed using both an extreme point $D^2(\phi_1) \in D^2(\sigma^N)$ of lower rank and one of lower rank and lower particle number $D^2(\phi_2)$. It is, therefore, tempting to look for some inductive process for constructing extreme points of higher rank starting from known extreme points of lower

rank. The functional form $\alpha\phi_1 + \beta\theta \wedge \phi_2$ seems flexible enough so that such a procedure might yield all of $\mathrm{EXT}D^2(\sigma^N)$.

An important unsolved problem is to establish whether the elements of $\mathrm{EXT}D^2(H^N)$ have unique preimages under the contraction mapping. As has been mentioned in the introduction, if such a property were to hold, one would be supplied with a satisfying characterization of the elements of $\mathrm{EXT}\,D^2(H^N)$. They would be precisely the elements of $D^2(H^N)$ which are covered by unique ground state wavefunctions of some two-particle operator. Unfortunately the techniques employed in Sec. 6 establishing this result for the case where N equals 3 or 4 cannot be extended to the case $N > 4$.

In the present paper we have given little attention to the convex structure of $\tilde{D}^2(H^N)$. However, this problem is of importance and hopefully will lead to efficient lower bound procedures in many-particle theory. We are planning to discuss a simpler subproblem of this problem, the Slater-hull or diagonal problem in a later paper. This essentially amounts to giving a characterization of $\tilde{D}^2(S^N)$, the polar of $D^2(S^N)$, the convex span of the two-matrices corresponding to Slater determinants in some fixed one-particle basis set.

Considerations based on the particle–hole duality established in Sec. 4 and the solution of the N-representability problem for the case $r = N + 2$ by reduction to the $N = 2$ case given in Sec. 5 has led us to consider the problem of obtaining the best $N + 2$ orbitals with which to describe the ground state of an N-electron system. This number of orbitals is precisely the number required for a valence bond description of a molecule with a single chemical bond. We are led to a simple iterative procedure similar to the closed shell SCF method. In fact, at each stage of the process, $N + 2$ orbitals are selected by a closed shell SCF calculation, and thus the convergence properties of our method are similar to those of the latter named method. A detailed study of this procedure is now being carried out by Brian Weiner of Queen's University.

ACKNOWLEDGMENTS

I am grateful to Professor A. J. Coleman for many helpful discussions bearing on the contents of this paper and to the National Research Council of Canada for financial support.

[1] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[2] H. W. Kuhn, Proc. Sym. Appl. Math. **10**, 141 (1960).
[3] G. Choquet, *Lectures on Analysis* (Benjamin, New York, 1969), Vol. II, Proposition 25.12.
[4] H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[5] M. B. Ruskai, J. Math. Phys. **11**, 3218 (1969).
[6] M. L. Yoseloff and H. W. Kuhn, J. Math. Phys. **10**, 703 (1969).
[7] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1756 (1964).
[8] Reference 3, Proposition 21.17.
[9] Reference 3, Theorem 22.7.
[10] B. Zumino, J. Math. Phys. **3**, 1055 (1962).
[11] F. Bloch and H. E. Rorschach, Phys. Rev. **128**, 1697 (1962).
[12] Erhardt Schmidt, Math. Ann. **63**, 433 (1907).
[13] B. C. Carlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).
[14] A. J. Coleman, J. Math. Phys. **13**, 214 (1972).

Asymptotic Expansions of Fourier Integrals with Light Cone Singularities

Fritz Schwarz
Universität Kaiserslautern, Fachbereich Physik, 675 Kaiserslautern, Postfach 1049, Germany
(Received 5 June 1972)

Let the integral transforms $\tilde{\phi}(y)$ and $\tilde{\phi}(y_1, y_2)$ be defined through $\tilde{\phi}(y) = \int d^4x e^{ixy} s(x)\phi(x)$ and $\tilde{\phi}(y_1, y_2) = \int d^4x_1 e^{ix_1y_1} \int d^4x_2 e^{ix_2y_2} s(x_1, x_2)\phi(x_1, x_2)$, respectively. Here all variables x and y are 4-vectors in Minkowski space. The functions ϕ are elements of $\mathcal{S}$, and the factors s contain certain types of light cone singularities. The integral transforms $\tilde{\phi}$ are investigated with respect to their characteristic properties implied by these light cone singularities using the method of van der Corput's neutralizers. It turns out that the behavior of $\tilde{\phi}$ is determined by the light cone singularities if one goes to infinity in the space of the y variables along an arbitrary straight line. All characteristically different cases are classified and for each case a complete asymptotic expansion is derived.

1. INTRODUCTION

There has been a considerable interest in the last couple of years in what is called "physics of the light cone." The starting point for the activity in this direction were two papers by Gribov *et al.*[1] and Ioffe.[2] More recent review type articles in this field are Refs. 3 and 4. See also the important work by Jackiw *et al.*[5,6] and by Gell-Mann and Fritzsch.[7]

What is meant by "physics of the light cone" can roughly be explained as follows: The S-matrix elements of certain processes in high energy physics can be written as the Fourier transform of the matrix elements of products (or commutators) of local operators in coordinate space. The matrix elements of these operator products in coordinate space contain singularities. More precisely, it is known from perturbation theory that the only singularities they contain are concentrated on the light cone. So it arises the question as to what the implications of this structure of the coordinate space matrix elements are—if any—for the Fourier transform in momentum space. There is one process which has attracted special attention in this connection, the inelastic scattering of electrons on protons. We refer the reader to the extensive literature on this subject, see, for example, Refs. 8–12, besides the papers quoted above. The way one usually argues to prove "light cone dominance" for certain limits is as follows: One takes a certain limit in momentum space, for inelastic e^-p-scattering; this is the so-called Bjorken or scaling limit, in which due to rapid oscillations of the exponential in the Fourier integral the contributions of most parts of the integration region are assumed to be canceled. It is suggested that the only region in coordinate space which gives a con-

rank. The functional form $\alpha\phi_1 + \beta\theta \wedge \phi_2$ seems flexible enough so that such a procedure might yield all of $\mathrm{EXT}D^2(\sigma^N)$.

An important unsolved problem is to establish whether the elements of $\mathrm{EXT}D^2(H^N)$ have unique preimages under the contraction mapping. As has been mentioned in the introduction, if such a property were to hold, one would be supplied with a satisfying characterization of the elements of $\mathrm{EXT}\,D^2(H^N)$. They would be precisely the elements of $D^2(H^N)$ which are covered by unique ground state wavefunctions of some two-particle operator. Unfortunately the techniques employed in Sec. 6 establishing this result for the case where N equals 3 or 4 cannot be extended to the case $N > 4$.

In the present paper we have given little attention to the convex structure of $\tilde{D}^2(H^N)$. However, this problem is of importance and hopefully will lead to efficient lower bound procedures in many-particle theory. We are planning to discuss a simpler subproblem of this problem, the Slater-hull or diagonal problem in a later paper. This essentially amounts to giving a characterization of $\tilde{D}^2(S^N)$, the polar of $D^2(S^N)$, the convex span of the two-matrices corresponding to Slater determinants in some fixed one-particle basis set.

Considerations based on the particle–hole duality established in Sec. 4 and the solution of the N-representability problem for the case $r = N + 2$ by reduction to the $N = 2$ case given in Sec. 5 has led us to consider the problem of obtaining the best $N + 2$ orbitals with which to describe the ground state of an N-electron system. This number of orbitals is precisely the number required for a valence bond description of a molecule with a single chemical bond. We are led to a simple iterative procedure similar to the closed shell SCF method. In fact, at each stage of the process, $N + 2$ orbitals are selected by a closed shell SCF calculation, and thus the convergence properties of our method are similar to those of the latter named method. A detailed study of this procedure is now being carried out by Brian Weiner of Queen's University.

ACKNOWLEDGMENTS

I am grateful to Professor A. J. Coleman for many helpful discussions bearing on the contents of this paper and to the National Research Council of Canada for financial support.

[1] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[2] H. W. Kuhn, Proc. Sym. Appl. Math. **10**, 141 (1960).
[3] G. Choquet, *Lectures on Analysis* (Benjamin, New York, 1969), Vol. II, Proposition 25.12.
[4] H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[5] M. B. Ruskai, J. Math. Phys. **11**, 3218 (1969).
[6] M. L. Yoseloff and H. W. Kuhn, J. Math. Phys. **10**, 703 (1969).
[7] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1756 (1964).
[8] Reference 3, Proposition 21.17.
[9] Reference 3, Theorem 22.7.
[10] B. Zumino, J. Math. Phys. **3**, 1055 (1962).
[11] F. Bloch and H. E. Rorschach, Phys. Rev. **128**, 1697 (1962).
[12] Erhardt Schmidt, Math. Ann. **63**, 433 (1907).
[13] B. C. Carlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).
[14] A. J. Coleman, J. Math. Phys. **13**, 214 (1972).

Asymptotic Expansions of Fourier Integrals with Light Cone Singularities

Fritz Schwarz
Universität Kaiserslautern, Fachbereich Physik, 675 Kaiserslautern, Postfach 1049, Germany
(Received 5 June 1972)

Let the integral transforms $\tilde{\phi}(y)$ and $\tilde{\phi}(y_1, y_2)$ be defined through $\tilde{\phi}(y) = \int d^4x e^{ixy} s(x)\phi(x)$ and $\tilde{\phi}(y_1, y_2) = \int d^4x_1 e^{ix_1y_1} \int d^4x_2 e^{ix_2y_2} s(x_1, x_2)\phi(x_1, x_2)$, respectively. Here all variables x and y are 4-vectors in Minkowski space. The functions ϕ are elements of $\mathcal{S}$, and the factors s contain certain types of light cone singularities. The integral transforms $\tilde{\phi}$ are investigated with respect to their characteristic properties implied by these light cone singularities using the method of van der Corput's neutralizers. It turns out that the behavior of $\tilde{\phi}$ is determined by the light cone singularities if one goes to infinity in the space of the y variables along an arbitrary straight line. All characteristically different cases are classified and for each case a complete asymptotic expansion is derived.

1. INTRODUCTION

There has been a considerable interest in the last couple of years in what is called "physics of the light cone." The starting point for the activity in this direction were two papers by Gribov *et al.*[1] and Ioffe.[2] More recent review type articles in this field are Refs. 3 and 4. See also the important work by Jackiw *et al.*[5,6] and by Gell-Mann and Fritzsch.[7]

What is meant by "physics of the light cone" can roughly be explained as follows: The S-matrix elements of certain processes in high energy physics can be written as the Fourier transform of the matrix elements of products (or commutators) of local operators in coordinate space. The matrix elements of these operator products in coordinate space contain singularities. More precisely, it is known from perturbation theory that the only singularities they contain are concentrated on the light cone. So it arises the question as to what the implications of this structure of the coordinate space matrix elements are—if any—for the Fourier transform in momentum space. There is one process which has attracted special attention in this connection, the inelastic scattering of electrons on protons. We refer the reader to the extensive literature on this subject, see, for example, Refs. 8–12, besides the papers quoted above. The way one usually argues to prove "light cone dominance" for certain limits is as follows: One takes a certain limit in momentum space, for inelastic e^-p-scattering; this is the so-called Bjorken or scaling limit, in which due to rapid oscillations of the exponential in the Fourier integral the contributions of most parts of the integration region are assumed to be canceled. It is suggested that the only region in coordinate space which gives a con-

tribution is the vicinity of the light cone. However, these discussions are not satisfactory for several reasons. For example, it is not clear from this type of arguments whether or not there are some other limits in momentum space which are "dominated" by the light cone, too. Further, it does not seem to be appropriate to introduce invariant variables as it is usually done in discussing inelastic e^-p scattering, because this changes the Fourier integrals one originally starts with into more complicated integral transforms. Finally, it seems not to be possible to generalize these arguments to more general Fourier transforms which contain matrix elements of products of more than two operators.

To clarify these questions we proceed as follows: Let $\phi(x)$ be a function of the 4-vector x which is an element of the space $\mathcal{S}$, i.e., it is differentiable an arbitrary number of times and falls off at infinity together with all its derivatives faster than any power. Then consider the integral transform

$$\tilde{\phi}(y) = \int d^4x e^{ixy} s(x)\phi(x). \tag{1.1}$$

The factor $s(x)$ is supposed to be singular on the light cone. We will discuss the types of singularities which are suggested by operator products of free fields, i.e., $s(x)$ contains the factor $\delta(x^2)$ or $\theta(x^2)$. In addition there may occur $\theta(x_0)$ or $\epsilon(x_0)$, which are defined through $\theta(x_0) = +1$ for $0 < x_0$ and $\theta(x_0) = 0$ for $x_0 < 0$, and $\epsilon(x_0) = +1$ for $0 < x_0$ and $\epsilon(x_0) = -1$ for $x_0 < 0$, respectively. Then we pose the following question: Are there any regions in y space where the behavior of $\tilde{\phi}(y)$ is determined essentially by the singularity $s(x)$, and what can be said about the behavior in these regions?

The proper mathematical framework for the discussion of questions of this kind has been developed by several authors.[13-23] Using these results we solve the above stated problem completely, i.e., we determine all regions where the light cone singularity $s(x)$ determines the Fourier transform (1.1) and what the behavior of $\tilde{\phi}(y)$ in these regions is. In other words, we get all possible information on $\tilde{\phi}(y)$ if one does not know more about $\phi(x)$ than the quite general restriction that it belongs to the function space $\mathcal{S}$. It turns out that, if one goes to infinity in y space along an arbitrary straight line, the behavior of $\tilde{\phi}(y)$ is essentially determined by the singularity $s(x)$, and a complete asymptotic expansion of $\tilde{\phi}(y)$ can be derived for these limits. A similar analysis is done for the more complicated case which may occur if the matrix element of the product of three operators is Fourier transformed:

$$\tilde{\phi}(y_1, y_2) = \int d^4x_1 e^{ix_1y_1} \int d^4x_2 e^{ix_2y_2} s(x_1, x_2) \times \phi(x_1, x_2). \tag{1.2}$$

Here again the function $\phi(x_1, x_2)$ belongs to the space $\mathcal{S}$, and $s(x_1, x_2)$ is the singular part which has in this case the structure

$$s(x_1, x_2) = \theta(x_{10})\theta(x_{20})\delta(x_1^2)\delta(x_2^2)\delta[(x_1 - x_2)^2]. \tag{1.3}$$

We show in Sec. 5 that (1.1) and (1.2) can be reduced to certain two- or three-dimensional Fourier integrals. These are treated separately in Sec. 3 and Sec. 4, respectively. We will need repeatedly the asymptotic expansions of certain types of one-dimensional Fourier transforms. Although they are known, we derive them in Sec. 2 because it gives a good illustration of the methods to be used later on. All our results are collected in some tables at the end of this paper.

2. ASYMPTOTIC EXPANSIONS OF ONE-DIMENSIONAL FOURIER TRANSFORMS

In this Section we will derive asymptotic expansions of certain types of one-dimensional Fourier transforms, if the argument of the transformed function gets large. The reason why we consider these Fourier integrals is that the many-dimensional Fourier transforms which we want to consider in Secs. 3 and 4, are always reduced to these one-dimensional cases. The results of this section are not new. They can be found in various places, see, for example, Refs. 24-27. However, we will use the same method of van der Corput's neutralizers as it will be used in the following Sections, and so we think it is instructive to explain this method on the comparatively simple one-dimensional case.

The integrals to be considered in this section are of the following types:

$$\tilde{f}(y) = \int dx\, e^{ixy} s(x) f(x). \tag{2.1}$$

The function $f(x)$ is assumed to belong to the space $\mathcal{S}$ which has already been defined in Sec. 1. The factor $s(x)$ is the singular part which may be $\theta(x)$, $\epsilon(x)$, or $\log x$. A function $N_i(x)$ is called a van der Corput neutralizer for the point x_i, or simply a neutralizer, if it has the following properties:

$$N_i(x) = \begin{cases} 1 & \text{for} \quad |x_i - x| \leq \epsilon \\ 0 & \text{for} \quad \delta \leq |x_i - x| \\ 0 \leq N_i(x) \leq 1 & \text{for} \quad \epsilon \leq |x_i - x| \leq \delta \end{cases}. \tag{2.2}$$

The constants ϵ and δ are fixed real numbers which obey the condition $\epsilon < \delta$. Further, $N_i(x)$ is assumed to be differentiable an arbitrary number of times. Then we also have

$$\frac{d^n N_i(x)}{dx^n} = 0 \tag{2.3}$$

for $|x_i - x| \leq \epsilon$ and $\delta \leq |x_i - x|$. The existence of such a neutralizer can be proven by explicit construction (see, for example, Ref. 24, p. 50). We say a point x_i does not contribute asymptotically to the integral (2.1), if there exists a neutrlizer $N_i(x)$ such that the integral

$$I_i = \int_{-\infty}^{+\infty} dx\, e^{ixy} N_i(x) s(x) f(x) \tag{2.4}$$

falls off faster than any power of y^{-1} if $|y| \to \infty$. We will employ frequently the so-called Landau symbols $O(x)$ and $o(x)$, which are defined as follows: The equation $f(x) = O(g(x))$ for $x \to x_0$ means that there exists a constant C so that $\lim_{x \to x_0} f(x) \leq Cg(x)$. We write $f(x) = o(g(x))$ for $x \to x_0$ if $\lim_{x \to x_0} f(x) \leq \epsilon g(x)$ for any positive number ϵ. More simply, if $g(x) \neq 0$, $f = O(g)$ means that f/g is bounded, $f = o(g)$ that f/g tends to zero as $x \to x_0$. If in a statement containing these Landau symbols the limit, for which it is valid, is not speci-

TABLE I. Asymptotic expansions of some one-dimensional Fourier transforms for large values of the variable in the transformed function.

Integral Transform	Asymptotic Expansion
$\int_{-\infty}^{+\infty} dx\, e^{ixy}\theta(x)f(x)$	$\sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1} f^{(\nu)}(0) + o(y^{-N})$
$\int_{-\infty}^{+\infty} dx\, e^{ixy}\epsilon(x)f(x)$	$\sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1} f^{(\nu)}(0) + o(y^{-N})$
$\int_{0}^{+\infty} dx\, \cos xy f(x)$	$2\sum_{\nu=1}^{N-1} \frac{(-1)^{\nu} f^{(2\nu-1)}(0)}{y^{2\nu}} + o(y^{-2N+1})$
$\int_{0}^{+\infty} dx\, \sin xy f(x)$	$\sum_{\nu=0}^{N-1} \frac{(-1)^{\nu} f^{(2\nu)}(0)}{y^{2\nu+1}} + o(y^{-2N})$
$\int_{-\infty}^{+\infty} dx\, e^{ixy}\theta(x)\log x f(x)$	$\sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1}\left(\psi(\nu+1) - \log y + \frac{i\pi}{2}\right) f^{(\nu)}(0) + O(y^{-N})$
$\int_{0}^{+\infty} dx\, \cos xy \log x f(x)$	$-\frac{\pi}{2}\sum_{\nu=0}^{N-1} \frac{(-1)^{\nu} f^{(2\nu)}(0)}{y^{2\nu+1}} + \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1} f^{(2\nu+1)}(0)}{y^{2\nu+2}}[\psi(2\nu+2) - \log y] + O(y^{-2N})$
$\int_{0}^{+\infty} dx\, \sin xy \log x f(x)$	$\sum_{\nu=0}^{N-1} \frac{(-1)^{\nu} f^{(2\nu)}(0)}{y^{2\nu+1}}[\psi(2\nu+1) - \log y] + \frac{\pi}{2}\sum_{\nu=0}^{N-1} \frac{(-1)^{\nu} f^{(2\nu+1)}(0)}{y^{2\nu+2}} + O(y^{-2N})$
$\int_{-\infty}^{+\infty} dx\, E_1(-ixy)\theta(\pm x)f(x)$	$\pm \sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1} \frac{f^{(\nu)}(0)}{\nu+1} + O(y^{-N})$

fied explicitly, it is always understood to be infinity. The same true if we use the expression "asymptotic" without further specification. Using the Landau symbols we can say that a point x_i does not contribute asymptotically to the integral (2. 1) if $I_i = o(y^{-n})$ for all n. We denote this behavior by $o(y^{-\infty})$. Otherwise it is said to contribute asymptotically and is called a critical point for the integral (2. 1).

The method of van der Corput's neutralizers for deriving an asymptotic expansion for an integral of the type (2. 1) consists essentially of two steps: In the first step one determines which points contribute asymptotically and which do not.

We consider only the simple case where the totality of critical points is a finite set. In the second step, the contributions of the critical points are expanded into asymptotic series, the sum of which is an asymptotic expansion for the integral considered.

We begin with the case $s(x) = \theta(x)$. Let us assume at first that $0 < x_i$. Then it is always possible to choose a neutralizer $N_i(x)$, the support of which is contained in the interval $0 \le x \le \infty$. So the contribution I_i can be written

$$I_i = \int_{-\infty}^{+\infty} dx\, e^{ixy} N_i(x)\theta(x)f(x) = \int_{x_i-\delta}^{x_i+\delta} dx\, e^{ixy} N_i(x)f(x). \tag{2.5}$$

Integrating N times by parts, we get

$$I_i = \left(\frac{i}{y}\right)^N \int_{x_i-\delta}^{x_i+\delta} dx\, e^{ixy} \frac{d^N}{dx^N}[N_i(x)f(x)] \tag{2.6}$$

because the integrated terms are zero as a consequence of the properties of the neutralizer $N_i(x)$. The integral at the right-hand side of Eq. (2. 6) tends to zero for $y \to \infty$ by the Riemann–Lebesgue lemma (see Ref. 26, p. 312, Theorem 9. 1). N is an arbitrary integer and so we have $I_i = o(y^{-\infty})$ if the point x_i lies in the interval $0 < x_i \le \infty$. This result could have been arrived at easier because it follows immediately from Ref. 28, p. 249, Theorem XII. We derived it explicitly because it will become clearer in this way why some points do contribute asymptotically and others do not. The former case occurs if $x_i = 0$. The contribution I_0 of the origin can be written as

$$I_0 = \int_0^{\delta} dx\, e^{ixy} N_0(x)f(x). \tag{2.7}$$

Integrating again N times by parts, we get now

$$I_0 = \sum_{\nu=1}^{N-1}\left(\frac{i}{y}\right)^{\nu+1} \frac{d^{\nu}f(x)}{dx^{\nu}}\bigg|_{x=0} + R_N \tag{2.8}$$

with

$$R_N = \left(\frac{i}{y}\right)^N \int_0^{\delta} dx\, e^{ixy} \frac{d^N}{dx^N}[N_0(x)f(x)] = o(y^{-N}). \tag{2.9}$$

Equation (2. 8) is an asymptotic power series for the contribution of the origin to the integral (2. 1) if $s(x) = \theta(x)$ and $|y| \to \infty$. Because the origin is the only critical point, it is at the same time an asymptotic expansion for the integral (2. 1). The whole discussion is nearly literally the same if $s(x) = \epsilon(x)$, and, therefore, this case is included without further discussion in Table I. If one separates the real and imaginary part in Eq. (2. 8), one gets an asymptotic expansion for the Fourier cosine and Fourier sine transform, respectively, which are included in Table I, too. Now we come to the case $s(x) = \theta(x)\log x$. It is immediately clear from the preceding discussion that the only point which contributes asymptotically in this case is again the origin. We have

$$I_0 = \int_0^{\delta} dx\, e^{ixy} \log x\, N_0(x)f(x). \tag{2.10}$$

Integrating this equation N times by parts leads to

$$I_0 = \sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1} \left[\left(e^{ixy}\log x + \sum_{\mu=1}^{\nu+1} E_{\mu}(-ixy)\right) \times \frac{d^{\nu}}{dx^{\nu}}(N_0(x)f(x))\right]_{x=0}^{x=\delta} + R_N \tag{2.11}$$

with

$$R_N = -\left(\frac{i}{y}\right)^N \int_0^\delta dx \left(e^{ixy}\log x + \sum_{\mu=1}^{N} E_\mu(-ixy)\frac{d^N}{dx^N} \times (N_0(x)f(x))\right). \quad (2.12)$$

The functions $E_n(z)$ are the exponential integrals as defined in Ref. 29, p. 228, Eq. (5.1.4). The logarithmic singularity at the lower integration limit of the first term in the square bracket in Eq. (2.11) cancels with the corresponding one in the exponential integral $E_1(-ixy)$, which is most easily seen from Eq. (5.1.11) on p. 229 in Ref. 29. Using Eqs. (5.1.12) and (5.1.23) of the same reference and the definition of the functions $E_n(z)$, one gets after some simple manipulations

$$I_0 = \sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1}\left(\psi(\nu+1) - \log y + \frac{i\pi}{2}\right)\frac{d^\nu f(x)}{dx^\nu}\bigg|_{x=0} + O(y^{-N}). \quad (2.13)$$

For the same reason as Eq. (2.8) this is an asymptotic expansion of the integral (2.1) with $s(x) = \theta(x) \times \log x$.

Finally, we consider another integral transform an asymptotic expansion of which will be needed in Sec. 4:

$$I_\pm = \int_{-\infty}^{+\infty} dx\, E_1(-ixy)\theta(\pm x) f(x). \quad (2.14)$$

The function $E_1(-ixy)$ is again the exponential integral as defined in Ref. 29, and $f(x)$ belongs to $\mathcal{S}$. Similar as in the preceding discussions, an asymptotic expansion is derived by integrating N times by parts where the relation

$$\int dx\, E_n(ax) = -(1/a)E_{n+1}(ax) \quad (2.15)$$

is used. This leads to

$$I_\pm = \pm \sum_{\nu=0}^{N-1}\left(\frac{i}{y}\right)^{\nu+1}\frac{1}{\nu+1}\frac{d^\nu f(x)}{dx^\nu}\bigg|_{x=0} + O(y^{-N}). \quad (2.16)$$

This and all other results of this Section are collected in Table I.

3. ASYMPTOTIC EXPANSIONS OF A CERTAIN TYPE OF TWO-DIMENSIONAL FOURIER TRANSFORM

To determine an asymptotic expansion for a multidimensional Fourier transform is much more complicated than in the case of a single variable. There are several reasons for this complication. At first, there are much more possibilities for the variables of the transformed function to approach a certain point. If this point lies at infinity, for example, one can approach it on an arbitrary radius or even along an arbitrary straight line, which does not necessarily go through the origin. For all these possibilities the asymptotic behavior has to be determined separately. A second source of complication is the fact that the region of integration may have a complicated shape. In analogy to the one-dimensional case, it might be suggested that the main contribution to a Fourier transform in several dimensions at a point far away from the origin comes from those parts of the integration region which are close to the boundary. It will turn out, however, that this is true only with some restrictions.

As already mentioned in the Introduction, the Fourier transforms in Minkowski space, which will be considered in Sec. 5, can be reduced to certain types of two- or three-dimensional Fourier transforms. In this section we determine, therefore, asymptotic expansions for the two-dimensional transform

$$\tilde{f}(\mathbf{y}) = \int_0^\infty dx_1 \int_{-x_1}^{+x_1} dx_2\, e^{i(x_1y_1+x_2y_2)} f(\mathbf{x}) \quad (3.1)$$

and leave the three-dimensional case to Sec. 4. In Eq. (3.1) we introduced the notation $\mathbf{x}$ and $\mathbf{y}$ for a point in the $x_1 - x_2$ plane or $y_1 - y_2$ plane, respectively. In the mathematical literature the behavior of multidimensional integrals of the type

$$g(k) = \int_G \cdots \int dx_1 \cdots dx_n\, e^{ikF(x_1\cdots x_n)} f(x_1 \cdots x_n), \quad (3.2)$$

if the parameter k goes to infinity, have been treated by several authors (see Refs. 13–23). If we specify in Eq. (3.1) a straight line in the $y_1 - y_2$ plane, along which we go to infinity, it has the form (3.2) with $F(x_1 \cdots x_n) = F(x_1, x_2)$, a linear function of the variables x_1 and x_2. Therefore, we can use the methods of the quoted papers to determine asymptotic expansions for the integral (3.1). The most suitable results for our purpose are given in the paper by Focke, Ref. 23. Focke treats only the case $n = 2$. However, it is possible to generalize his results to the case $n = 3$ for Fourier integrals. This will be done in the next section when we consider the integral (4.1).

From now on, if we talk about the asymptotic behavior of the integral (3.1), we always mean the behavior for large distances of the point $\mathbf{y}$ from the origin. The procedure of determining asymptotic expansions for (3.1) is completely analogous to Sec. 2. At first we determine the set of critical points, which does not necessarily consist of isolated points now, but may be a critical line also. It will turn out that the set of critical points depends on the limit considered. Then the contributions of these different critical regions will be expanded into asymptotic series.

Let the integration region of the integral (3.1) be R, i.e., $R = \{\mathbf{x}: 0 \le x_1 \le \infty, -x_1 \le x_2 \le +x_1\}$. Let P_i be an arbitrary point from R. We construct a neutralizer $N_i(\mathbf{x})$ for the point P_i in the following way:

$$N_i(\mathbf{x}) = \prod_{k=1,2} N_i^k(x_k), \quad (3.3)$$

where the $N_i^k(x_k)$ are neutralizers of the type (2.2). $N_i(\mathbf{x})$ is assumed to be partially differentiable an arbitrary number of times with respect to the variables x_1 and x_2. Then we have

$$\frac{\partial^{m+n} N_i(\mathbf{x})}{\partial x_1^m \partial x_2^n} = 0 \quad (3.4)$$

if $\mathbf{x}$ does not lie in the difference set of the two squares $|x_k - x_k^i| \le \delta$ and $|x_k - x_k^i| \le \epsilon$. If one takes for $N_i(\mathbf{x})$ the example given in Ref. 23, p. 28, the support of the neutralizer (3.3) is ΔR_i, where $\Delta R_i = \{\mathbf{x}: |x_k - x_k^i| \le \delta,\ k = 1, 2\}$. The intersection of ΔR_i with R is called ΔS_i, i.e., $\Delta S_i = \Delta R_i \cap R$. In analogy to the one-dimensional case, the contribution I_i of a point P_i to the integral (3.1) is defined to be

$$I_i = \int_{\Delta S_i} \int dx_1 dx_2\, e^{i(x_1y_1+x_2y_2)} N_i(\mathbf{x}) f(\mathbf{x}). \quad (3.5)$$

We want to determine now the conditions for a point P_i not to contribute to the integral (3.1) if one goes to infinity in the $y_1 - y_2$ plane along a radius, i.e., if $y_1 = my_2$ with m a fixed real number and $y_2 \to \pm\infty$. In this limit the integral (3.5) reads

$$I_i = \int_{\Delta S_i} \int dx_1 dx_2 \, e^{iy_2(mx_1+x_2)} N_i(\mathbf{x}) f(\mathbf{x}). \tag{3.6}$$

We consider at first a point P_i which does not lie on the boundary of R. Then it is always possible to choose ϵ and δ in Eq. (3.3) so that ΔR_i is completely contained in R. Therefore, the contribution of an inner point P_i is

$$I_i = \int_{x_1^i-\delta}^{x_1^i+\delta} dx_1 \int_{x_2^i-\delta}^{x_2^i+\delta} dx_2 \, e^{iy_2(mx_1+x_2)} N_i(\mathbf{x}) f(\mathbf{x}). \tag{3.7}$$

By applying again Theorem XII from Ref. 28, p. 249, it follows immediately that $I_i = o(y_2^{-\infty})$ without any restrictions for m. Next consider the case where P_i lies on the boundary of R but $x_k^i \neq 0$ $(k = 1, 2)$. Assume at first that $x_1^i = x_2^i$. Then the contribution of P_i is

$$I_i = \int_{x_1^i-\delta}^{x_1^i+\delta} dx_1 \, e^{ix_1 my_2} \int_{x_2^i-\delta}^{x_1} dx_2 \, e^{ix_2 y_2} N_i(\mathbf{x}) f(\mathbf{x}) \tag{3.8}$$

Integrating N times by parts with respect to x_2 leads to

$$I_i = -\sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \int_{x_1^i-\delta}^{x_1^i+\delta} dx_1 \, e^{ix_1(1+m)y_2} \frac{\partial^\nu}{\partial x_2^\nu} [N_i(\mathbf{x}) \times f(\mathbf{x})]_{x_2=x_1} + o(y_2^{-N}). \tag{3.9}$$

If $m \neq -1$ it follows from Theorem XII of Ref. 28 that P_i does not contribute asymptotically to the integral (3.1). Similarly it can be shown that P_i does not contribute asymptotically if it lies on the line $x_2 = -x_1$ and does not coincide with the origin and $m \neq +1$.

Now we consider the origin $x_1 = x_2 = 0$. We introduce a neutralizer $N_0(x_1)$ through

$$N_0(x_1) = \begin{cases} 1 & \text{for} \quad x_1 \leq \epsilon \\ 0 & \text{for} \quad \delta \leq x_1 \\ 0 \leq N_0(x_1) \leq 1 & \text{for} \quad \epsilon \leq x_1 \leq \delta \end{cases} \qquad (\epsilon < \delta). \tag{3.10}$$

In addition it has the usual differentiability properties. By using this neutralizer the contribution I_0 of the origin can be written as

$$I_0 = \int_0^\delta dx_1 \, e^{ix_1 my_2} N_0(x_1) \int_{-x_1}^{+x_1} dx_2 e^{ix_2 y_2} f(\mathbf{x}). \tag{3.11}$$

At first we integrate N times by parts with respect to x_2 and get

$$I_0 = \int_0^\delta dx_1 \, e^{ix_1 my_2} N_0(x_1) \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \times \left[e^{-ix_1 y_2} \frac{\partial^\nu f(\mathbf{x})}{\partial x_2^\nu}\bigg|_{x_2=-x_1} - e^{ix_1 y_2} \frac{\partial^\nu f(\mathbf{x})}{\partial x_2^\nu}\bigg|_{x_2=x_1} \right] + o(y_2^{-N}). \tag{3.12}$$

Then we integrate this equation N times by parts with respect to x_1 with the result $(m \neq \pm 1)$

$$I_0 = \sum_{\mu,\nu=0,1\cdots}^{\mu+\nu\leq N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \left[\frac{1}{(m-1)^{\mu+1}} \frac{d^\mu}{dx_{\mathbf{x}}^\mu} \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_2^\nu}\bigg|_{x_2=-x_1} \right) - \frac{1}{(m+1)^{\mu+1}} \frac{d^\mu}{dx_1^\mu} \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_2^\nu}\bigg|_{x_2=x_1} \right) \right]_{x_1=0} + o(y_2^{-N}). \tag{3.13}$$

So we have the following result for the integral (3.1): If one goes to infinity in the y plane along a straight line $y_1 = my_2$ and $m \neq \pm 1$, only the origin is a critical point. If $m = \pm 1$, also the points of the boundary of R may contribute. More precisely, if $m = \oplus 1$, the line $x_2 = -x_1$ $(0 \leq x_1)$ is a critical line, and if $m = -1$ the same is true for the line $x_2 = x_1$ $(0 \leq x_1)$.

We want to determine now asymptotic expansions for the integral (3.1) for various limits. To perform this, we have to expand the contributions of the critical regions into asymptotic series. Let us assume at first $m \neq \pm 1$. Then only the origin contributes and Eq. (3.13) is an asymptotic expansion for the integral (3.1). By changing the summation indices, it can be written more economically as

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-2} \sum_{\mu=0}^{\nu} \sum_{\pm} \left(\frac{i}{y_2}\right)^{\nu+2} \frac{1}{(m \pm 1)^{\mu+1}} \frac{d^\mu}{dx_1^\mu} \times \left[\frac{\partial^{\nu-\mu} f(\mathbf{x})}{\partial x_2^{\nu-\mu}}\bigg|_{x_2=\pm x_1} \right]_{x_1=0} + o(y_2^{-N}). \tag{3.14}$$

Separating the real and imaginary part gives

$$\mathrm{Re}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \sum_{\mu=0}^{2\nu} \sum_{\pm} \frac{\pm(-1)^\nu}{y_2^{2\nu+2}} \frac{1}{(m \pm 1)^{\mu+1}} \frac{d^\mu}{dx_1^\mu} \times \left(\frac{\partial^{2\nu-\mu} f(\mathbf{x})}{\partial x_2^{2\nu-\mu}}\bigg|_{x_2=\pm x_1} \right)_{x_1=0} + o(y_2^{-2N-1}), \tag{3.15a}$$

$$\mathrm{Im}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \sum_{\mu=0}^{2\nu+1} \sum_{\pm} \frac{\pm(-1)^\nu}{y_2^{2\nu+3}} \frac{1}{(m \pm 1)^{\mu+1}} \frac{d^\mu}{dx_1^\mu} \times \left(\frac{\partial^{2\nu+1-\mu} f(\mathbf{x})}{\partial x_2^{2\nu+1-\mu}}\bigg|_{x_2=\pm x_1} \right)_{x_1=0} + o(y_2^{-2N-1}). \tag{3.15b}$$

Now let m be $+1$. To separate off the contribution of the line $x_2 = -x_1$, we construct a neutralizer $N_+(\mathbf{x})$ with the properties

$$N_+(\mathbf{x}) = \begin{cases} 1 & \text{in the strip} -x_1 \leq x_2 \leq -x_1 + \epsilon \\ 0 & \text{for} -x_1 + \delta \leq x_2 \\ 0 \leq N_+(\mathbf{x}) \leq 1 & \text{for} -x_1 + \epsilon \leq x_2 \leq -x_1 + \delta \end{cases} \tag{3.16}$$

The partial derivatives of $N_+(\mathbf{x})$ with respect to x_1 and x_2 are supposed to exist up to arbitrary order and to obey

$$\frac{\partial^{\mu+\nu} N_+(\mathbf{x})}{\partial x_1^\mu \, \partial x_2^\nu} = 0 \tag{3.17}$$

outside the strip $-x_1 + \epsilon < x_2 < -x_1 + \delta$. By using this neutralizer the contribution of the line $x_2 = -x_1$ to the integral (3.1) in the limit $y_1 = y_2$ can be written as

$$\tilde{f}(\mathbf{y}) = \int_0^\infty dx_1 \int_{-x_1}^{+x_1} dx_2 \, e^{iy_1(x_1+x_2)} N_+(\mathbf{x}) f(\mathbf{x}). \tag{3.18}$$

By introducing new variables u and v through

$$u = 2^{-1/2}(x_1 + x_2), \qquad v = 2^{-1/2}(x_1 - x_2), \tag{3.19}$$

the integral (3.18) reads

$$\tilde{f}(\mathbf{y}) = \int_0^\delta du \, e^{i\sqrt{2}uy_2} M_+(u) \int_{-\infty}^0 dv \, g(u, v). \tag{3.20}$$

Evidently, $g(u, v)$ and $M_+(u)$ are the pictures of $f(\mathbf{x})$ and $N_+(\mathbf{x})$, respectively, under the mapping (3.19), and

$\bar{\delta} = \sqrt{2}^{-1}\delta$. The fact that $M_+(u)$ depends only on u is the consequence of a special choice of $N_+(\mathbf{x})$ which is always possible. With the results of Sec. 2 we get from (3.20) the asymptotic expansion

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \left(\frac{1}{\sqrt{2}}\right)^{\nu+1} \int_{-\infty}^{0} dv \, \frac{\partial^\nu g(u,v)}{\partial u^\nu}\bigg|_{u=0} + o(y_2^{-N}). \tag{3.21}$$

Returning to the old variables x_1 and x_2 gives

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \frac{1}{2^\nu} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^\nu \times f(\mathbf{x})\Big|_{x_2=-x_1} + o(y_2^{-N}). \tag{3.22}$$

For the real and imaginary part it follows from Eq. (3.22) that

$$\mathrm{Re}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu+1} y_2^{2\nu+2}} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^{2\nu+1} \times f(\mathbf{x})\Big|_{x_2=x_1} + o(y_2^{-2N}) \tag{3.23a}$$

$$\mathrm{Im}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu} y_2^{2\nu+1}} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^{2\nu} \times f(\mathbf{x})\Big|_{x_2=-x_1} + o(y_2^{-2N+1}). \tag{3.23b}$$

In the limit $y_1 = -y_2$, the contribution of the line $x_2 = x_1$ is separated off with the help of the neutralizer

$$N_-(\mathbf{x}) = \begin{cases} 1 & \text{in the strip } x_1 - \epsilon \le x_2 \le x_1 \\ 0 & \text{for } x_2 \le x_1 - \delta \\ 0 \le N_-(\mathbf{x}) \le 1 & \\ & \text{in the strip } x_1 - \delta \le x_2 \le x_1 - \epsilon \end{cases}. \tag{3.24}$$

All partial derivatives of $N_-(\mathbf{x})$ vanish outside the strip $x_1 - \delta \le x_2 \le x_1 - \epsilon$. Instead of Eq. (3.18) we have now

$$\tilde{f}(\mathbf{y}) = \int_0^\infty dx_1 \int_{-x_1}^{+x_1} dx_2 \, e^{-iy_2(x_1-x_2)} N_-(\mathbf{x}) f(\mathbf{x}). \tag{3.25}$$

By changing again the variables with Eqs. (3.19) it follows that

$$\tilde{f}(\mathbf{y}) = \int_0^{\bar{\delta}} dv \, e^{-i\sqrt{2}vy_2} M_-(v) \int_0^\infty du \, g(u,v). \tag{3.26}$$

From this equation we get the asymptotic expansion

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(-\frac{i}{y_2}\right)^{\nu+1} \left(\frac{1}{\sqrt{2}}\right)^{\nu+1} \int_0^\infty du \, \frac{\partial^\nu g(u,v)}{\partial v^\nu}\bigg|_{v=0} + o(y_2^{-N}). \tag{3.27}$$

The equivalent of Eq. (3.22) is

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(-\frac{i}{y_2}\right)^{\nu+1} \frac{1}{2^\nu} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}\right)^\nu \times f(\mathbf{x})\Big|_{x_2=x_1} + o(y_2^{-N}), \tag{3.28}$$

and instead of Eq. (3.23) we have now

$$\mathrm{Re}f(\mathbf{y}) = \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu+1} y_2^{2\nu+2}} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}\right)^{2\nu+1} \times f(\mathbf{x})\Big|_{x_2=x_1} + o(y_2^{-2N}), \tag{3.29a}$$

$$\mathrm{Im}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu} y_2^{2\nu+1}} \int_0^\infty dx_1 \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}\right)^{2\nu} \times f(\mathbf{x})\Big|_{x_2=x_1} + o(y_2^{-2N+1}). \tag{3.29b}$$

The asymptotic expansions (3.14), (3.15), (3.22), (3.23), and (3.28), (3.29) give a complete solution to the problem of determining the asymptotic behavior of the Fourier transform (3.1) if one goes to infinity in y space along an arbitrary radius. A more general problem is to determine asymptotic expansions for (3.1) if one goes to infinity in the y plane along an arbitrary straight line which does not necessarily pass through the origin, i.e., if

$$y_1 = my_2 + k, \tag{3.30}$$

where $|y_2| \to \infty$ and m and k are fixed real numbers. Because this limit is interesting for physical applications, we will determine asymptotic expansions of (3.1) for this possibility too. In the limit (3.30) the transform (3.1) can be written

$$\tilde{f}(\mathbf{y}) = \int_0^\infty dx_1 \int_{-x_1}^{+x_1} dx_2 \, e^{iy_2(mx_1+x_2)} e^{ikx_1} f(\mathbf{x}). \tag{3.31}$$

This is an integral of the same type as it occured in the limit $y_1 = my_2$, except that the original function $f(\mathbf{x})$ is multiplied by the factor e^{ikx_1}. This factor does not change the properties of the function $f(\mathbf{x})$ concerning its differentiability and the fact that it falls off asymptotically together with all its derivatives faster than any power. So we have essentially the same problem, only the separation into the real and the imaginary part is a little more complicated now. By Defining

$$f_{\substack{c\\s}}(\mathbf{x}) = \begin{matrix}\cos kx_1 \\ \sin kx_1\end{matrix} f(\mathbf{x}), \tag{3.32}$$

it follows that

$$\mathrm{Re}\tilde{f}(\mathbf{y}) = \mathrm{Re}\tilde{f}_c(\mathbf{y}) - \mathrm{Im}\tilde{f}_s(\mathbf{y}), \tag{3.33a}$$

$$\mathrm{Im}\tilde{f}(\mathbf{y}) = \mathrm{Im}\tilde{f}_c(\mathbf{y}) + \mathrm{Re}\tilde{f}_s(\mathbf{y}). \tag{3.33b}$$

The asymptotic expansions for the transforms at the right-hand side have already been determined, and so this problem is also solved.

Finally, we have to consider a certain variation of the transform (3.1) which is defined through

$$\bar{\tilde{f}}(\mathbf{y}) = \int_{-\infty}^{0} dx_1 \int_{x_1}^{-x_1} dx_2 \, e^{i(x_1y_1+x_2y_2)} f(\mathbf{x}). \tag{3.34}$$

The whole discussion is very similar to the previous case, and we do not repeat it here, but give only the results. In the limit $y_1 = my_2$ with $m \neq \pm 1$, only the origin contributes and this contribution is given by the right-hand side of Eq. (3.11); only the integration regions are different now. For x_1 it is $-\delta \le x_1 \le 0$ and for x_2 we have now $x_1 \le x_2 \le -x_1$. The same partial integrations are performed, and it turns out that the asymptotic expansion is exactly the same, i.e.,

$$\bar{\tilde{f}}(\mathbf{y}) = \tilde{f}(\mathbf{y}) \tag{3.35}$$

if $y_1 = my_2$, $m \neq \pm 1$ and $|y_2| \to \infty$. If $m = +1$, one gets for $\bar{\tilde{f}}(\mathbf{y})$ an equation which differs from (3.20) by the integration regions for u and v. They are now

$-\delta \le u \le 0$ and $-\infty \le v \le 0$, respectively. Clearly, also the neutralizer $M_+(u)$ has to be redefined appropriately. Similarly for the limit $y_1 = -y_2$, the integration regions in Eq. (3.26) have to be changed to $-\bar{\delta} \le v \le 0$, respectively. Then one gets the asymptotic expansions

$$\tilde{f}(\mathbf{y}) = -\sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \frac{1}{2^\nu} \int_{-\infty}^{0} dx_1 \left(\frac{\partial}{\partial x_1} \pm \frac{\partial}{\partial x_2}\right)^\nu f(\mathbf{x})\Big|_{x_2 = \pm x_1} + o(y_2^{-N}). \quad (3.36)$$

For the real and imaginary part, it follows that

$$\mathrm{Re}\tilde{f}(\mathbf{y}) = -\sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu+1} y_2^{2\nu+2}} \int_{-\infty}^{0} dx_1 \left(\frac{\partial}{\partial x_1} \pm \frac{\partial}{\partial x_2}\right)^{2\nu+1} \times f(\mathbf{x})\Big|_{x_2 = \pm x_1} o(y_2^{-2N}) \quad (3.37a)$$

$$\mathrm{Im}\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \frac{(-1)^{\nu+1}}{2^{2\nu} y_2^{2\nu+1}} \int_{-\infty}^{0} dx_1 \left(\frac{\partial}{\partial x_1} \pm \frac{\partial}{\partial x_2}\right)^{2\nu} \times f(\mathbf{x})\Big|_{x_2 = \pm x_1} + o(y_2^{-2N+1}). \quad (3.37b)$$

The $\pm$ signs refer to the limits $y_1 = \pm y_2$. The limits $y_1 = my_2 + k$ are again determined through Eqs. (3.33), where at the right-hand side the respective asymptotic expansions for the transforms $\tilde{f}(\mathbf{y})$ instead of $\tilde{f}(\mathbf{y})$ have to be inserted.

So we solved the problem of determining the asymptotic behavior of the integral transforms $\tilde{\tilde{f}}(\mathbf{y})$ and $\tilde{f}(\mathbf{y})$, if one goes to infinity in the y plane along an arbitrary straight line, completely. In Sec. 5 it will turn out that with these results it is easy to determine the asymptotic behavior of Fourier transforms in Minkowski space which contain singularities of the form $\delta(x^2)$ or $\theta(x^2)$ on the light cone.

4. ASYMPTOTIC EXPANSIONS OF A CERTAIN TYPE OF THREE-DIMENSIONAL FOURIER TRANSFORM

In the next Section it will turn out that a certain type of Fourier transform in Minkowski space, which depends on two 4-vectors, can be reduced to a three-dimensional Fourier transform of the following type:

$$\tilde{f}(\mathbf{y}) = \int_0^\infty dx_1 \int_0^\infty \frac{dx_2}{x_2} \int_{-x_2}^{+x_2} dx_3\, e^{i(x_1 y_1 + x_2 y_2 + x_3 y_3)} f(\mathbf{x}). \quad (4.1)$$

Now $\mathbf{x}$ and $\mathbf{y}$ are 3-vectors (x_1, x_2, x_3) and (y_1, y_2, y_3), respectively. The function $f(\mathbf{x})$ is an element of the space $\mathcal{S}$. If we talk about the asymptotic behavior of this integral we mean again the behavior if the point $\mathbf{y}$ is far away from the origin. The method to be used is the same as in the previous case. At first we determine the critical regions for the integral (4.1). It will turn out that they depend on the respective limit and might consist of a single point, a half line or part of a plane. Then their contributions are expanded into asymptotic series.

In this section, R denotes the integration region for the integral (4.1), i.e., $R = \{\mathbf{x}: 0 \le x_1 \le \infty, 0 \le x_2 \le \infty, -x_2 \le x_3 \le x_2\}$. For each point P_i from R a neutralizer $N_i(\mathbf{x})$ is defined through

$$N_i(\mathbf{x}) = \prod_{k=1,2,3} N_i^k(x_k), \quad (4.2)$$

where the $N_i^k(x_k)$ are neutralizers of the type (2.2). All partial derivatives with respect to the variables x_1, x_2, and x_3 are assumed to exist, and we have again

$$\frac{\partial^{l+m+n} N_i(\mathbf{x})}{\partial x_1^l \partial x_2^m \partial x_3^n} = 0 \quad (4.3)$$

if $\mathbf{x}$ does not lie in the difference set of the two cubes $|x_k - x_k^i| \le \epsilon$ and $|x_k - x_k^i| \le \delta$. If one takes for the $N_i^k(x_k)$ the example given in Ref. 23, the support of the neutralizer $N_i(\mathbf{x})$ is the small cube $R_i = \{\mathbf{x}: |x_k - x_k^i| \le \delta,\ k = 1, 2, 3\}$. Analogously to Sec. 3 we define $\Delta S_i = \Delta R_i \cap R$. We want to go to infinity in y space along a radius. This radius is specified as the intersection of the two planes $y_2 = my_1$ and $y_3 = ny_1$, where m and n are fixed real numbers and $|y_1| \to \pm\infty$. In this limit the contribution I_i of a point $P_i = (x_1^i, x_2^i)$ to the integral (4.1) is

$$I_i = \int\int_{\Delta S_i} \int dx_1 dx_2 dx_3\, e^{iy_1(x_1 + mx_2 + nx_3)} N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1}. \quad (4.4)$$

If P_i does not lie on the boundary of R, the corresponding neutralizer $N_i(\mathbf{x})$ may be chosen so that its support is completely contained in R, i.e., $\Delta S_i = \Delta R_i$ in Eq. (4.4). It follows by the above quoted Theorem XII from Ref. 28 that $I_i = o(y_1^{-\infty})$ for arbitrary m and n.

Next assume that P_i lies in that part of the plane $x_1 = 0$ which belongs to the boundary of R, but $x_3^i \ne \pm x_2^i$. The contribution of such a point may be written as

$$I_i = \int_0^\infty dx_1\, e^{ix_1 y_1} \int_{x_2^i - \delta}^{x_2^i + \delta} dx_2\, e^{ix_2 m y_1} \int_{x_3^i - \delta}^{x_3^i + \delta} dx_3\, e^{ix_3 n y_1} \times N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1}. \quad (4.5)$$

If $n \ne 0$, Theorem XII of Ref. 28 is applied to the x_3 integration. It follows that $I_i = o(y_1^{-\infty})$ for $n \ne 0$. If $n = 0$, the last exponential equals one and the same reasoning is applied to the x_2 integration, i.e., $I_i = o(y_1^{-\infty})$ if $n = 0$ and $m \ne 0$. If $m = n = 0$, Eq. (4.5) is N times integrated by parts with respect to x_1 with the result

$$I_i = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_1}\right)^{\nu+1} \frac{\partial^\nu}{\partial x_1^\nu} \left(\int_{x_2^i - \delta}^{x_2^i + \delta} dx_2 \int_{x_3^i - \delta}^{x_3^i + \delta} dx_3\, N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1} \right)_{x_1 = 0} + o(y_1^{-N}). \quad (4.6)$$

We see from this equation that the leading term at the right-hand side is y_1^{-1}. To include the possibility that the radius along which we go to infinity lies in the plane $x_1 = 0$ or $x_2 = 0$, the parametrization of the intersecting planes has to be changed. It turns out that we have considered already all characteristically different possibilities, i.e., a point on the boundary plane $x_1 = 0$ is critical only if one goes to infinity along the y_1 axis. Now we consider the case that P_i lies in that part of the plane $x_3 = x_2$ which belongs to the boundary, but $x_1^i, x_2^i \ne 0$. It is more appropriate now to specify the radius as the intersection of the two planes $y_3 = ny_2$ and $y_1 = my_2$. With this parametrization the contribution of the point P_i can be written in the form

$$I_i = \int_{x_3^i - \delta}^{x_3^i + \delta} dx_3\, e^{ix_3 n y_2} \int_{x_3}^{x_2^i + \delta} dx_2\, e^{ix_2 y_2} \int_{x_1^i - \delta}^{x_1^i + \delta} dx_1\, e^{ix_1 m y_2} \times N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1}. \quad (4.7)$$

If $m \ne 0$, it follows from Theorem XII of Ref. 28 that

$I_i = o(y_2^{-\infty})$. If $m = 0$, we integrate N times by parts and get

$$I_i = \int_{x_3^i-\delta}^{x_3^i+\delta} dx_3\, e^{ix_3(n+1)y_2} \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \frac{\partial^\nu}{\partial x_2^\nu} \times \left(\int_{x_1^i-\delta}^{x_1^i+\delta} dx_1 N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1}\right). \quad (4.8)$$

If $n \neq -1$, Theorem XII from Ref. 28 gives again the result $I_i = o(y_2^{-\infty})$. If $n = -1$, the leading term at the right-hand side is y_2^{-1}. It follows that a point P_i on the boundary plane $x_3 = x_2$ is critical only if one goes to infinity in y space along the line $y_3 = -y_2$ in the plane $y_1 = 0$. Similarly it follows that a point on the boundary plane $x_3 = -x_2$ is critical only if one goes to infinity along the line $y_3 = y_2$ in the plane $y_1 = 0$.

Up to now we considered the possibilities that P_i lies inside R or in a plane which belongs to the boundary of R; but it was excluded that P_i lies on the intersection of two boundary planes. These cases will be discussed now. We begin with the possibility that P_i lies on one of the two edges which are specified as the intersection of the planes $x_1 = 0$ and $x_3 = \pm x_2$. Our discussion will be for the case $x_3 = +x_2$. The origin is again excluded. The contribution of a point on this edge can be written in the form

$$I_i = \int_0^\delta dx_1\, e^{ix_1 m y_2} \int_{x_2^i-\delta}^{x_2^i+\delta} dx_2\, e^{ix_2 y_2} \int_{x_3^i-\delta}^{x_2} dx_3 e^{ix_3 n y_2} \times N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1}. \quad (4.9)$$

Here and in the remaining cases, the radius in y space is always specified as the intersection of the two planes $y_1 = my_2$ and $y_3 = ny_2$. If $n \neq 0$, this equation is N times integrated by parts with respect to the variable x_3 with the result

$$I_i = \int_0^\delta dx_1\, e^{ix_1 m y_2} \int_{x_2^i-\delta}^{x_2^i+\delta} dx_2\, e^{i(n+1)y_2} \sum_{\nu=0}^{N-1} \left(\frac{i}{ny_2}\right)^{\nu+1} \frac{\partial^\nu}{\partial x_3^\nu} \times (N_i(\mathbf{x}) f(\mathbf{x}) x_2^{-1})_{x_3=x_2} + o(y_2^{-N}). \quad (4.10)$$

If $n \neq -1$, it follows again from Theorem XII of Ref. 28 that $I_i = o(y_2^{-\infty})$. If $n = -1$, it is immediately clear that the leading term at the right-hand side of Eq. (4.10) is y_2^{-1} because the lower limit of the x_1 integration is not determined by the neutralizer. If $n = 0$, the partial integrations with respect to x_3 are superfluous. So we have the result that, for a point P_i on the edge $x_1 = 0$, $x_3 = x_2$ to be critical, it is necessary and sufficient that one goes to infinity in y space along a radius which is contained in the plane $y_3 = -y_2$. It is easy to see that a point on the edge $x_1 = 0$, $x_3 = -x_2$ is critical if the radius in y space lies in the plane $y_3 = y_2$.

Now let P_i be on the positive x_1 axis, but $x_1^i \neq 0$. The contribution of such a point to the integral (4.1) is

$$I_i = \int_0^\delta \frac{dx_2}{x_2} e^{ix_2 y_2} \int_{-x_2}^{x_2} dx_3\, e^{ix_3 n y_2} \int_{x_1^i-\delta}^{x_1^i+\delta} dx_1\, e^{ix_1 m y_2} \times N_i(\mathbf{x}) f(\mathbf{x}). \quad (4.11)$$

If $m \neq 0$, Theorem XII from Ref. 28 gives again $I_i = o(y_2^{-\infty})$. If $m = 0$, it is immediately clear from Eq. (4.11) that the point P_i is critical. The integration over x_3 cannot give enough powers of y_2^{-1} because the integration region is not limited by the neutralizer.

The x_2 integration is of the form (2.1) with $s(x) = \theta(x)$, the leading term of which in the asymptotic expansion is y_2^{-1}. So it is necessary and sufficient for a point on the positive x_1 axis to be critical that $m = 0$.

Finally, we have to determine the contribution I_0 of the origin. It can be written in the form

$$I_0 = \int_0^\delta \frac{dx_2}{x_2} e^{ix_2 y_2} \int_0^\delta dx_1\, e^{ix_1 m y_2} \int_{x_2}^{+x_2} dx_3\, e^{ix_3 n y_2} \times N_0(\mathbf{x}) f(\mathbf{x}). \quad (4.12)$$

If $m, n \neq 0$, this equation is N times integrated by parts with respect to the variables x_1 and x_3. This leads to

$$I_0 = \sum_{\mu,\nu=0}^{\mu+\nu\leq N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \frac{1}{m^{\mu+1} n^{\nu+1}} \int_0^\delta \frac{dx_2}{x_2} \times \left(e^{ix_2(1-n)y_2} \frac{\partial^{\mu+\nu} N_0(\mathbf{x}) f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\substack{x_1=0 \\ x_3=-x_2}} - e^{ix_2(1+n)y_2} \times \frac{\partial^{\mu+\nu} N_0(\mathbf{x}) f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\substack{x_1=0 \\ x_3=x_2}}\right) + o(y_2^{-N}). \quad (4.13)$$

If m or n are zero the respective integrations by parts are omitted. There is no trouble concerning the existence of the x_2 integral in Eq. (4.13) because the square bracket has a first-order zero for $x_2 = 0$. By using the relation

$$\int_a^b \frac{dx}{x} e^{ixy} f(x) = -[E_1(-ixy) f(x)]_{x=a}^{x=b} + \int_a^b dx \times E_1(-ixy) f^{(1)}(x) \quad (4.14)$$

and the results from Sec. 2, it follows from Eq. (4.13) after some simple manipulations that

$$I_0 = \log\frac{1+n}{1-n} \sum_{\mu,\nu=0}^{\mu+\nu\leq N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \frac{1}{m^{\mu+1} n^{\nu+1}} \frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\mathbf{x}=0} + \sum_{\lambda,\mu,\nu=0}^{\lambda+\mu+\nu\leq N-3} \left(\frac{i}{y_2}\right)^{\lambda+\mu+\nu+3} \frac{1}{m^{\mu+1} n^{\nu+1}} \frac{1}{\lambda+1} \left[\frac{1}{(1-n)^{\lambda+1}} \times \frac{d^{\lambda+1}}{dx_2^{\lambda+1}} \left(\frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\substack{x_1=0 \\ x_3=-x_2}}\right)_{x_2=0} - \frac{1}{(1+n)^{\lambda+1}} \frac{d^{\lambda+1}}{dx_2^{\lambda+1}} \times \left(\frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\substack{x_1=0 \\ x_3=x_2}}\right)_{x_2=0}\right] + o(y_2^{-N}). \quad (4.15)$$

So we have determined all the critical regions for the integral **(4.1)**. We distinguish the following cases:

(a) If one goes to infinity along the y_1 axis, the set $C_a = \{\mathbf{x}: x_1 = 0, |x_3| \leq x_2\}$ is critical.

(b) If one goes to infinity along the line $y_1 = 0, y_3 = \pm y_2$, the set $C_b^\pm = \{\mathbf{x}: 0 \leq x_1, 0 \leq x_2, x_3 = \mp x_2\}$ is critical.

(c) If one goes to infinity along a radius which is contained in the plane $y_3 = \pm y_2$, the set $C_c^\pm = \{\mathbf{x}: x_1 = 0, 0 \leq x_2, x_3 = \mp x_2\}$ is critical.

(d) If one goes to infinity in the plane $y_1 = 0$, the set $C_d = \{\mathbf{x}: 0 \leq x_1, x_2 = x_3 = 0\}$ is critical.

If the radius in y space is such that it does not coincide with one of the possibilities (a)–(d), the origin is the only critical point for the integral (4.1).

Now we have to expand the contributions of the various critical regions into asymptotic series. If we have none of the possibilities (a)–(d), only the origin contributes and Eq. (4.11) is an asymptotic expansion for the integral (4.1), i.e., we have in this case

$$\tilde{f}(\mathbf{y}) = I_0.$$

Next we consider case (a). To separate off the contribution of the critical region C_a we define the neutralizer $N(x_1)$ through

$$N(x_1) = \begin{cases} 1 & \text{for } x_1 \le \epsilon \\ 0 & \text{for } \delta \le x_1 \\ 0 \le N(x_1) \le 1 & \text{for } \epsilon \le x_2 \le \delta \end{cases} \tag{4.16}$$

and get in this limit

$$\tilde{f}(\mathbf{y}) = \int_0^\delta dx_1\, e^{ix_1y_1} N(x_1) \int_0^\infty \frac{dx_2}{x_2} \int_{-x_2}^{+x_2} dx_3 f(\mathbf{x}). \tag{4.17}$$

With the results of Sec. 2 it follows immediately

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_1}\right)^{\nu+1} \frac{\partial^\nu}{\partial x_1^\nu} \left(\int_0^\infty \frac{dx_2}{x_2} \int_{-x_2}^{+x_2} dx_3 f(\mathbf{x}) \right)_{x_1=0} + o(y_1^{-N}). \tag{4.18}$$

In case (b) our discussion will be for the limit $y_1 = 0$, $y_3 = +y_2$. The discussion for the limit $y_1 = 0$, $y_3 = -y_2$ is nearly literally the same; there are only some changes. To separate off the contribution of the plane $x_3 = -x_2$ we define the following two neutralizers:

$$N_1(x_2) = \begin{cases} 1 & \text{for } x_2 \le \epsilon \\ 0 & \text{for } \delta \le x_2 \\ 0 \le N_1(x_2) \le 1 & \text{for } \epsilon \le x_2 \le \delta \end{cases}, \tag{4.19}$$

$$N_2(x_2,x_3) = \begin{cases} 1 & \text{for } x_3 \le -x_2 + \epsilon \\ 0 & \text{for } -x_2 + \delta \le x_3 \\ 0 \le N_2(x_2,x_3) \le 1 & \text{for } -x_2 + \epsilon \\ & \le x_3 \le -x_2 + \delta \end{cases} \tag{4.20}$$

These neutralizers are assumed to have the usual differentiability properties. Using them, the Fourier transform $\tilde{f}(\mathbf{y})$ can be written in the limit under consideration as

$$\tilde{f}(\mathbf{y}) = \int_0^\delta \frac{dx_2}{x_2} e^{ix_2y_2} N_1(x_2) \int_{-x_2}^{+x_2} dx_3\, e^{ix_3y_2} \int_0^\infty dx_1 f(\mathbf{x}) + \int_\epsilon^\infty \frac{dx_2}{x_2} e^{ix_2y_2}[1 - N_1(x_2)] \int_{-x_2}^{-x_2+\delta} dx_3\, e^{ix_3y_2} \times N_2(x_2,x_3) \int_0^\infty dx_1 f(\mathbf{x}). \tag{4.21}$$

At first we integrate this equation N times by parts with respect to the variable x_3 and get

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \int_0^\infty dx_1 \int_0^\delta \frac{dx_2}{x_2} N_1(x_2) \times \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=-x_2} - e^{2ix_2y_2} \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=x_2} \right) \times \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \int_0^\infty dx_1 \int_\epsilon^\infty \frac{dx_2}{x_2} [1 - N_1(x_2)] \times \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=-x_2} + o(y_2^{-N}). \tag{4.22}$$

The cancellation of the singularities in the x_2 integrations under the first integral is most easily achieved through one partial integration with respect to x_2. The logarithmic singularity cancels with the corresponding one of the exponential integral. Using then the results of Sec. 2 leads after a few steps to

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \left(\log(-2iCy_2) \int_0^\infty dx_1 \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_2=x_3=0} - \int_0^\infty dx_1 \int_0^\infty dx_2 \log x_2 \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=-x_2} \right) - \sum_{\mu,\nu=0}^{\mu+\nu\le N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \frac{1}{2^{\mu+1}} \frac{1}{\mu+1} \int_0^\infty dx_1 \frac{\partial^{\mu+1}}{\partial x_2^{\mu+1}} \times \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=x_2} \right)_{x_2=0} + o(y_2^{-N}). \tag{4.23}$$

Here $C = \log\gamma$, and γ is Eulers constant (see Ref. 29, p. 229). In case (c) we assume that the radius in y space, along which infinity is approached, lies in the plane $y_3 = +y_2$. It is specified as the intersection of this plane with $y_1 = my_2$ $(m \neq 0)$. Using the neutralizers (4.16), (4.19), and (4.20), the contribution of the corresponding edge to the integral (4.1) is

$$\tilde{f}(\mathbf{y}) = \int_0^\delta \frac{dx_2}{x_2} e^{ix_2y_2} N_1(x_2) \int_{-x_2}^{+x_2} dx_3\, e^{ix_3y_2} \int_0^\delta dx_1 \times e^{ix_1my_2} N(x_1) f(\mathbf{x}) + \int_\epsilon^\infty \frac{dx_2}{x_2} e^{ix_2y_2}[1 - N_1(x_2)] \times \int_{-x_2}^{-x_2+\delta} dx_3\, e^{ix_3y_2} N_2(x_2,x_3) \int_0^\delta dx_1\, e^{ix_1my_2} \times N(x_1) f(\mathbf{x}). \tag{4.24}$$

Integrating by parts N times with respect to the variable x_1 leads to

$$\tilde{f}(\mathbf{y}) = \sum_{\mu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\mu+1} \frac{1}{m^{\mu+1}} \left(\int_0^\delta \frac{dx_2}{x_2} e^{ix_2y_2} N_1(x_2) \int_{-x_2}^{+x_2} dx_3 \times e^{ix_3y_2} \frac{\partial^\mu f(\mathbf{x})}{\partial x_1^\mu}\bigg|_{x_1=0} + \int_\epsilon^\infty \frac{dx_2}{x_2} e^{ix_2y_2}[1 - N_1(x_2)] \times \int_{-x_2}^{-x_2+\delta} dx_3\, e^{ix_3y_2} N_2(x_2,x_3) \frac{\partial^\mu f(\mathbf{x})}{\partial x_1^\mu}\bigg|_{x_1=0} + o(y_2^{-N}). \tag{4.25}$$

The expression in the square bracket is of the same form as Eq. (4.21), and so we get without further calculation

$$\tilde{f}(\mathbf{y}) = \sum_{\mu,\nu=0}^{\mu+\nu\le N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \frac{1}{m^{\mu+1}} \left(\log(-2iCy_2) \frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\mathbf{x}=0} - \int_0^\infty dx_2 \log x_2 \frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{\substack{x_1=0 \\ x_3=-x_2}} \right) - \sum_{\lambda,\mu,\nu=0}^{\lambda+\mu+\nu\le N-3} \left(\frac{i}{y_2}\right)^{\lambda+\mu+\nu+3} \frac{1}{2^{\lambda+1} m^{\mu+1}(\lambda+1)} \frac{\partial^{\lambda+1}}{\partial x_2^{\lambda+1}} \times \left(\frac{\partial^{\mu+\nu} f(\mathbf{x})}{\partial x_1^\mu \partial x_3^\nu}\bigg|_{x_3=x_2} \right)_{x_1=x_2=0} + o(y_2^{-N}). \tag{4.26}$$

Finally, we come to case (d). Using the neutralizer (4.19) we have now

$$\tilde{f}(\mathbf{y}) = \int_0^\delta \frac{dx_2}{x_2} e^{ix_2y_2} N_1(x_2) \int_{-x_2}^{+x_2} dx_3\, e^{ix_3ny_2} \int_0^\infty dx_1 f(\mathbf{x}). \tag{4.27}$$

N partial integrations with respect to x_3 lead to ($n \neq 0$)

$$\tilde{f}(\mathbf{y}) = \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \frac{1}{n^{\nu+1}} \int_0^{\delta} \frac{dx_2}{x_2} N_1(x_2) \int_0^{\infty} dx_1 \times \left(e^{ix_2(1-n)y_2} \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=-x_2} - e^{ix_2(1+n)y_2} \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=x_2} \right) o(y_2^{-N}). \tag{4.28}$$

If $n = 0$, the partial integrations with respect to x_3 are superfluous. With Eq. (4.14) and the results of Sec. 2, we get finally

$$\tilde{f}(\mathbf{y}) = \log\frac{1+n}{1-n} \sum_{\nu=0}^{N-1} \left(\frac{i}{y_2}\right)^{\nu+1} \frac{1}{n^{\nu+1}} \int_0^{\infty} dx_1 \frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_2=x_3=0} + \sum_{\mu,\nu=0}^{\mu+\nu\leq N-2} \left(\frac{i}{y_2}\right)^{\mu+\nu+2} \frac{1}{\mu+1} \int_0^{\infty} dx_1 \left[\frac{1}{(1-n)^{\mu+1}} \times \frac{\partial^{\mu+1}}{\partial x_2^{\mu+1}} \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=-x_2} \right)_{x_2=0} - \frac{1}{(1+n)^{\mu+1}} \frac{\partial^{\mu+1}}{\partial x_2^{\mu+1}} \times \left(\frac{\partial^\nu f(\mathbf{x})}{\partial x_3^\nu}\bigg|_{x_3=x_2} \right)_{x_2=0} \right] + o(y_2^{-N}). \tag{4.29}$$

So we have finished our discussion of the asymptotic behavior of the integral (4.1) if one approaches infinity along a radius. Similarly as in Sec. 3 we consider now the possibility that one goes to infinity in y space along an arbitrary straight line which is not necessarily a radius. Such a straight line may be specified as the intersection of two planes, for example,

$$y_1 = my_2 + a, \tag{4.30a}$$

$$y_3 = ny_2 + b, \tag{4.30b}$$

where a and b are fixed real numbers and $|y_2| \to \infty$. In this limit the exponential in the integral (4.1) reads $\exp[iy_2(mx_1 + x_2 + nx_3)] \times \exp[iax_1 + ibx_3)$. This factor does not change the properties which we assumed for $f(\mathbf{x})$. So we have again essentially the same problem. The asymptotic behavior is qualitatively the same if one goes to infinity along a radius or on a straight line parallel to it. The asymptotic expansions in the limit (4.30) are derived from those for $a = b = 0$ by replacing the function $f(\mathbf{x})$ through $\exp[iax_1 + ibx_3)f(\mathbf{x})$. We do not give them explicitly here because it would mean repeating the complicated formulas for the asymptotic expansions with a few changes. However, the first few terms for these limits are also given in Table IV.

At the end of this section let us make some remarks on the results which we have found. It has become clear what the relevant points in discussing the asymptotic behavior of Fourier transforms in several dimensions are. The asymptotic behavior is determined by the shape of the integration region and the singularities of the integrand. The effect of the shape of the integration region can be described as follows: The asymptotic behavior which is due to the shape of the integration region is different from $o(r^{-\infty})$, where r is the distance of a point in the space of the transformed variables from the origin, in those directions in which the planes of constant phase are tangent to the boundary of the integration region. If the integration region is limited by planes only, there are the following possibilities (in three dimensions): A plane of constant phase coincides with a boundary plane, or an edge, which is the intersection of two boundary planes, lies in a plane of constant phase. In these two cases the Fourier transform falls off slowly for large r, and in the former case it is slower than in the latter. A corner contributes always asymptotically. However, the corresponding contribution has a higher power of r^{-1}. This behavior can be nicely seen from the results of this section and Sec. 3 (see also the Tables). The Fourier transform (4.1) is more complicated because of the singularity x_2^{-1}, which is integrable as a consequence of the shape of the integration region. It gives rise to the logarithmic terms in Eqs. (4.23) and (4.26).

5. ASYMPTOTIC EXPANSIONS OF SINGULAR FOURIER INTEGRALS IN MINKOWSKI SPACE

In this section we come to our main task—to determine the asymptotic behavior of Fourier integrals in Minkowski space, if the original functions contains certain types of singularities on the light cone. We begin with the following integral

$$\tilde{\phi}(y) = \int dx\, e^{ixy}\theta(x_0)\delta(x^2)\phi(x). \tag{5.1}$$

Here x and y are 4-vector variables in Minkowski space. The function $\phi(x)$ is assumed to be continuously differentiable an arbitrary number of times and to fall off at infinity together with all its derivatives faster than any power, i.e., $\phi(x)$ belongs again to the space $\mathcal{S}$. We choose a special Lorentz frame such that y has only two nonzero components y_0 and y_3. Then the integral (5.1) can be written as

$$\tilde{\phi}(y) = \int dx_0\, \theta(x_0)e^{ix_0y_0} \int dx_3\, e^{-ix_3y_3} \int d\vec{x} \times \delta(x_0^2 - |\vec{x}|^2 - x_3^2)\phi(x_0, x_3, \vec{x}) \tag{5.2}$$

We introduced the notation $\vec{x}$ for a vector in the two-dimensional subspace of the 1- and 2-components, i.e., we have $x = (x_0, \mathbf{x}) = (x_0, \vec{x}, x_3) = (x_0, x_1, x_2, x_3)$. Introducing polar coordinates $|\vec{x}|$ and ψ in this two-dimensional subspace, Eq. (5.2) takes the form

$$\tilde{\phi}(y) = \int dx_0\, \theta(x_0)e^{ix_0y_0} \int dx_3\, e^{-ix_3y_3} \int d|\vec{x}|\,|\vec{x}| \times \delta(x_0^2 - |\vec{x}|^2 - x_3^2)\phi(x_0, x_3, |\vec{x}|), \tag{5.3}$$

where we have defined

$$\phi(x_0, x_3, |\vec{x}|) = \int_0^{2\pi} d\psi\, \phi(x_0, x_3, \vec{x}). \tag{5.4}$$

Now we use the δ functions to perform the integration over $|\vec{x}|$ with the result

$$\tilde{\phi}(y) = \int_0^{\infty} dx_0\, e^{ix_0y_0} \int_{-x_0}^{+x_0} dx_3\, e^{ix_3y_3} f(x_0, x_3). \tag{5.5}$$

The function $f(x_0, x_3)$ is connected with $\phi(x)$ through

$$f(x_0, x_3) = \tfrac{1}{2}\int_0^{2\pi} d\psi\, \phi(x_0, -x_3, |\vec{x}| = (x_0^2 - x_3^2)^{1/2}, \psi) \tag{5.6}$$

The integral (5.5) is of the type (3.1), and so we can apply the results of Sec. 3 to determine asymptotic expansions for $\tilde{\phi}(y)$. It is easy to see that Eqns. (3.15), (3.22), and (3.28) are asymptotic expansions for (5.5) for the respective limits if y_2 is replaced by y_3. These expansions look somehow complicated.

TABLE II. The leading terms of the asymptotic expansions of $\tilde{\phi}(y) = \int dx\, e^{ixy}\theta(x_0)s(x^2)\phi(x)$. If the singular factor $s(x^2) = \delta(x^2)$, the function $f(x_0, x_3)$ is defined through Eq. (5.6). If $s(x^2) = \theta(x^2)$ it is defined through Eq. (5.8). The constants m and k are fixed real numbers and $|y_3| \to \infty$.

Limit	Leading Terms of the Asymptotic Expansions of $\tilde{\phi}(y) = \int dx\, e^{ixy}\theta(x_0)s(x^2)\phi(x)$
$y_0 = my_3, m \neq \pm 1$	$\mathrm{Re}\tilde{\phi}(y) = \frac{2f(0,0)}{1-m^2}\frac{1}{y_3^2} - \frac{2}{(1-m^2)^3}\frac{1}{y_3^4}\left((3m^2+1)\frac{\partial^2 f(x_0,x_3)}{\partial x_0^2} + 8m\frac{\partial^2 f(x_0,x_3)}{\partial x_0 \partial x_3} + (m^2+3)\frac{\partial^2 f(x_0,x_3)}{\partial x_3^2}\right)_{x_0=x_3=0} + o(y_3^{-4})$
	$\mathrm{Im}\tilde{\phi}(y) = -\frac{4}{(1-m^2)^2}\frac{1}{y_3^3}\left(m\frac{\partial f(x_0,x_3)}{\partial x_0} + \frac{\partial f(x_0,x_3)}{\partial x_3}\right)_{x_0=x_3=0} + \frac{8}{(1-m^2)^4}\frac{1}{y_3^5}\left((m^2+1)m\frac{\partial^3 f(x_0,x_3)}{\partial x_0^3} + (5m^2+1)\frac{\partial^3 f(x_0,x_3)}{\partial x_0^2 \partial x_3} + m(m^2+5)\frac{\partial^3 f(x_0,x_3)}{\partial x_0 \partial x_3^2} + (m^2+1)\frac{\partial^3 f(x_0,x_3)}{\partial x_3^3}\right)_{x_0=x_3=0} + o(y_3^{-5})$
$y_0 = \pm y_3$	$\mathrm{Re}\tilde{\phi}(y) = -\frac{1}{2y_3^2}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)f(x_0,x_3)\Big\|_{x_3=\mp x_0} + \frac{1}{8y_3^4}\int_0^\infty dx\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^3 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-4})$
	$\mathrm{Im}\tilde{\phi}(y) = \frac{1}{y_3}\int_0^\infty dx_0 f(x_0, \mp x_0) - \frac{1}{4y_3^3}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-3})$
$y_0 = my_3 + k, m \neq \pm 1$	$\mathrm{Re}\tilde{\phi}(y) = \frac{2f(0,0)}{1-m^2}\frac{1}{y_3^2} - \frac{4mkf(0,0)}{(1-m^2)^5}\frac{1}{y_3^3} + \frac{2}{(1-m^2)^3}\frac{1}{y_3^4}\left(k^2(3m^2+1)f(0,0) - (3m^2+1)\frac{\partial^2 f(x_0,x_3)}{\partial x_0^2} - 8m\frac{\partial^2 f(x_0,x_3)}{\partial x_0 \partial x_3} - (m^2+3)\frac{\partial^2 f(x_0,x_3)}{\partial x_3^2}\right)_{x_0=x_3=0} + o(y_3^{-4})$
	$\mathrm{Im}\tilde{\phi}(y) = \frac{4m}{(1-m^2)^2}\frac{1}{y_3^3}\left(\frac{\partial f(x_0,x_3)}{\partial x_0} + m\frac{\partial f(x_0,x_3)}{\partial x_3}\right)_{x_0=x_3=0} - \frac{4k}{(1-m^2)^3}\frac{1}{y_3^4}\left((3m^2+1)\frac{\partial f(x_0,x_3)}{\partial x_0} + m\frac{\partial f(x_0,x_3)}{\partial x_3}\right)_{x_0=x_3=0} + o(y_3^{-4})$
$y_0 = \pm y_3 + k$	$\mathrm{Re}\tilde{\phi}(y) = -\frac{1}{y_3}\int_0^\infty dx\, \sin kx_0 f(x_0, \mp x_0) - \frac{1}{2y_3^2}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)\cos kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + \frac{1}{4y_3^3}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 \sin kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-3})$
	$\mathrm{Im}\tilde{\phi}(y) = \frac{1}{y_3}\int_0^\infty dx_0\, \cos kx_0 f(x_0, \mp x_0) - \frac{1}{2y_3^2}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)\sin kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} - \frac{1}{4y_3^3}\int_0^\infty dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 \cos kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-3})$

It seems desirable to have more explicit expressions for the lower terms. They are given therefore in Table II.

Another type of singularity on the light cone is the unit step function $\theta(x^2)$. Let us defined $\tilde{\phi}(y)$ now through

$$\tilde{\phi}(y) = \int dx\, e^{ixy}\theta(x_0)\theta(x^2)\phi(x). \tag{5.7}$$

If we go to the same Lorentz system as in the former case it follows easily that Eq. (5.7) can be brought into the form (5.5). However, instead of (5.6) we have in this case

$$f(x_0,x_3) = \int_0^{(x_0^2-x_3^2)^{1/2}} d|\vec{x}||\vec{x}| \int_0^{2\pi} d\psi\, \phi(x_0,-x_3,\vec{x}). \tag{5.8}$$

So the asymptotic expansions of (5.7) are given by the same formulas as for the transform (5.1). The only difference is that $f(x_0, x_3)$ is defined through Eq. (5.8) instead of Eq. (5.6). That does not mean, however, that the leading terms must be the same for the transforms (5.1) and (5.7) if the function $\phi(x)$ is given. This is easily seen by putting the respective definitions for $f(x_0, x_3)$ into the expressions for the coefficients in Table II or III.

Next we consider the transforms

$$\tilde{\phi}(y) = \int dx\, e^{ixy}\epsilon(x_0)s(x^2)\phi(x), \tag{5.9}$$

where $s(x^2) = \delta(x^2)$ or $\theta(x^2)$. Because $\epsilon(x_0) = \theta(x_0) - \theta(-x_0)$, this problem is equivalent to knowing the asymptotic expansions of (5.1) and (5.7) with $\theta(x_0)$ changed to $\theta(-x_0)$. Then it is easily seen that these integrals can be brought into the form

$$\tilde{\phi}(y) = \int_{-\infty}^0 dx_0\, e^{ix_0y_0}\int_{x_0}^{-x_0} dx_3\, e^{ix_3y_3}f(x_0,x_3) \tag{5.10}$$

with $f(x_0, x_3)$ given again by Eqs. (5.6) or (5.8). This transform is of the type (3.34). So the solution of this problem is also included in the results of Sec. 3. The lower terms of the asymptotic expansions for the transform (5.9) are given in Table III.

To make the meaning of our results clearer, let us illustrate them by an example. For $\phi(x)$ we take the function

$$\phi(x) = e^{-(px)^2}, \tag{5.11}$$

where p is a fixed timelike 4-vector, and for the singular part $s(x) = \theta(x_0)\delta(x^2)$. This function does not belong to the space $\mathcal{S}$ in the variable x because $px = \mathrm{const}$ for x on a certain spacelike hyperplane. However, the integration regions of the Fourier transforms considered in this section are the inside or the surface of the light cone, and so this does not affect our discussion. Going to the rest frame of p we get from (5.6)

$$f(x_0,x_3) = e^{-p_0^2x_0^2}. \tag{5.12}$$

TABLE III. The leading terms of the asymptotic expansions of $\tilde{\phi}(y) = \int dx\, e^{ixy}\epsilon(x_0)s(x^2)\phi(x)$. The function $f(x_0, x_3)$ is defined in the same way as in Table II. The constants m and k are again fixed real numbers and $|y_3| \to \infty$.

Limit	Leading Terms of the Asymptotic Expansions of $\tilde{\phi}(y) = \int dx\, e^{ixy}\epsilon(x_0)s(x^2)\phi(x)$
$y_0 = my_3, m \neq \pm 1$	$\mathrm{Re}\tilde{\phi}(y) = o(y_3^{-\infty}), \quad \mathrm{Im}\tilde{\phi}(y) = o(y_3^{-\infty})$
$y_0 = \pm y_3$	$\mathrm{Re}\tilde{\phi}(y) = -\frac{1}{2y_3^2}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)f(x_0,x_3)\Big\|_{x_3=\mp x_0} + \frac{1}{8y_3^4}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^3 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-4})$
	$\mathrm{Im}\tilde{\phi}(y) = \frac{1}{y_3}\int_{-\infty}^{+\infty} dx_0 f(x_0, \mp x_0) - \frac{1}{4y_3^3}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-3})$
$y_0 = my_3 + k,$ $m \neq \pm 1$	$\mathrm{Re}\tilde{\phi}(y) = o(y_3^{-\infty}), \quad \mathrm{Im}\tilde{\phi} = o(y_3^{-\infty})$
$y_0 = \pm y_3 + k$	$\mathrm{Re}\tilde{\phi}(y) = -\frac{1}{y_3}\int_{-\infty}^{+\infty} dx_0 \sin kx_0 f(x_0, \mp x_0) - \frac{1}{2y_3^2}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)\cos kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0}$
	$+ \frac{1}{4y_3^2}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 \sin kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-2})$
	$\mathrm{Im}\tilde{\phi}(y) = \frac{1}{y_3}\int_{-\infty}^{+\infty} dx_0 \cos kx_0 f(x_0, \mp x_0) - \frac{1}{2y_3^2}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)\sin kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0}$
	$- \frac{1}{4y_3^2}\int_{-\infty}^{+\infty} dx_0\left(\frac{\partial}{\partial x_0} \pm \frac{\partial}{\partial x_3}\right)^2 \cos kx_0 f(x_0,x_3)\Big\|_{x_3=\mp x_0} + o(y_3^{-2})$

Via the relation

$$\left.\frac{d^k e^{-ax^2}}{dx^k}\right|_{x=0} = \begin{cases} 0 & \text{for } k \text{ odd} \\ \dfrac{a^{k/2}k!(-1)^{k/2}}{(\frac{1}{2}k)!} & \text{for } k \text{ even} \end{cases}, \tag{5.13}$$

one gets easily from Table II for the limits $y_0 = my_3$, $|y_3| \to \infty$,

$$\mathrm{Re}\tilde{\phi}(y) = \begin{cases} \dfrac{2\pi}{1-m^2}\dfrac{1}{y_3^2} + 4\pi\dfrac{1+3m^2}{(1-m^2)^3}\dfrac{p_0^2}{y_3^4} + o(y_3^{-4}) \\ \quad \text{for } m \neq \pm 1 \\ \dfrac{\pi}{2y_3^2} + \dfrac{p_0^2}{4y_3^4} + o(y_3^{-4}) \quad \text{for } m = \pm 1 \end{cases} \tag{5.14a}$$

and

$$\mathrm{Im}\tilde{\phi}(y) = \begin{cases} o(y_3^{-\infty}) & \text{for } m \neq \pm 1 \\ \dfrac{\pi^{3/2}}{2p_0 y_3} & \text{for } m = \pm 1 \end{cases}. \tag{5.14b}$$

Similarly it follows for the limit $y_0 = my_3 + k$, $|y_3| \to \infty$ that

$$\mathrm{Re}\tilde{\phi}(y) = \begin{cases} \dfrac{2\pi}{(1-m^2)y_3^2} + \dfrac{4\pi mk}{(1-m^2)^2 y_3^3} + \dfrac{4\pi(3m^2+1)}{(1-m^2)^3} \\ \quad \left(1 + \dfrac{k^2}{2p_0^2}\dfrac{p_0^2}{y_3^4}\right) + o(y_3^{-5}) \quad \text{for } m \neq \pm 1 \\ -\dfrac{\pi k}{2p_0^2}\,{}_1F_1\left(1, \dfrac{3}{2}; \dfrac{k^2}{4p_0^2}\right)\dfrac{1}{y_3} + \dfrac{\pi}{2y_3^2} - \dfrac{k}{4y_3^3} \\ \quad + o(y_3^{-4}) \quad \text{for } m = \pm 1 \end{cases} \tag{5.15a}$$

and

$$\mathrm{Im}\tilde{\phi}(y) = \begin{cases} o(y_3^{-\infty}) & \text{for } m \neq \pm 1 \\ \dfrac{\pi}{2p_0 y_3} e^{-k^2/2p_0^2} + o(y_3^{-\infty}) & \text{for } m = \pm 1 \end{cases}. \tag{5.15b}$$

Alternatively, for the simple example (5.11) the Fourier transform (5.1) can also be calculated exactly. The result is, for $y_0 = my_3$,

$$\mathrm{Re}\tilde{\phi}(y) = \frac{\pi}{2p_0^2}(m+1)\,{}_1F_1\left(1, \frac{3}{2}; -y_3^2\frac{(m+1)^2}{4p_0^2}\right) - (m-1)\,{}_1F_1\left(1, \frac{3}{2}; -y_3^2\frac{(m-1)^2}{4p_0^2}\right), \tag{5.16a}$$

$$\mathrm{Im}\tilde{\phi}(y) = \frac{\pi^{3/2}}{2py_3}\left[\exp\left(-y_3^2\frac{(m-1)^2}{4p_0^2}\right) - \exp\left(-y_3^2\frac{(m+1)^2}{4p_0^2}\right)\right]. \tag{5.16b}$$

Via the asymptotic expansion for the confluent hypergeometric function ${}_1F_1(1, \frac{3}{2}; x)$ for $|x| \to \infty$, Ref. 29, p. 508, Eq. (13.5.1)

$${}_1F_1\left(1, \frac{3}{2}; x\right) = \frac{\sqrt{\pi}}{2}\frac{e^x}{\sqrt{x}} - \frac{1}{2\sqrt{\pi}}\frac{1}{x} \times \sum_{n=0}^{N-1}\frac{(-1)^n\Gamma(n+\frac{1}{2})}{x^n} + O(|x|^{-N+1}), \tag{5.17}$$

one gets from Eq. (5.16)

$$\mathrm{Re}\tilde{\phi}(y) = \begin{cases} \dfrac{\sqrt{\pi}}{y_3^2}\displaystyle\sum_{n=0}^{N-1}\frac{\Gamma(n+\frac{1}{2})2^{2n}p_0^{2n}}{y_3^{2n}}\left[\frac{1}{(m+1)^{2n+1}} - \frac{1}{(m-1)^{2n+1}}\right] + O(y_3^{-2N-2}) & \text{for } m \neq \pm 1 \\ \dfrac{\sqrt{\pi}}{2y_3^2}\displaystyle\sum_{n=0}^{N-1}\frac{\Gamma(n+\frac{1}{2})p_0^{2n}}{y_3^{2n}} + O(y_3^{-2N-2}) & \text{for } m = \pm 1 \end{cases} \tag{5.18}$$

and Eq. (5.15b) for the imaginary part. It is easy to see that the leading terms of this equation agree with Eq. (5.15a). Similarly, one can derive asymptotic expansions for the limit $y_0 = my_3 + k$ with the result

$$\mathrm{Re}\tilde{\phi}(y) = -\pi\sum_{n=0}^{N-1}\frac{2^{2n}n!p^{2n}}{y_3^{2n+2}}L_n^{(-1/2)}\left(\frac{-k^2}{4p_0^2}\right) \times \left(\frac{1}{(m-1)^{2n+1}} + \frac{1}{(m+1)^{2n+1}}\right) + \frac{\pi k}{4p_0^2}$$

TABLE IV. The leading terms in the asymptotic expansions of the integral transform (5.20) for various limits. The function $f(x_1, x_2, x_3)$ is defined through Eqs. (5.22) and (5.25). The constants m, n, a, and b are fixed real numbers.

Limit	Leading Terms of the Asymptotic Expansions of the Fourier Transform $\phi(y_1, y_2)$
$y_{10} = my_{20}, y_{23} = ny_{20}$ $\lvert y_{10}\rvert \to \infty$ Excluded are those values of m and n which match any other case in this table.	$\tilde{\phi}(y_1,y_2) = -\frac{1}{y_{20}^2}\log\frac{1+n}{1-n}\frac{f(0,0,0)}{mn} - \frac{i}{y_{20}^3}\frac{1}{mn}\left[\log\frac{1+n}{1-n}\left(\frac{1}{m}\frac{\partial}{\partial x_1}+\frac{1}{n}\frac{\partial}{\partial x_3}\right)f(x_1,x_2,x_3)\Big\rvert_{x_1=x_2=x_3=0}\right.$ $\left. + \frac{d}{dx_2}\left(\frac{f(0,x_2,-x_2)}{1-n} - \frac{f(0,x_2,x_2)}{1+n}\right)_{x_2=0}\right] + o(y_{20}^{-3})$
$y_{20} = y_{23} = 0$ $\lvert y_{10}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{i}{y_{10}}\int_0^\infty \frac{dx_2}{x_2}\int_{-x_2}^{+x_2} dx_3 f(0,x_2,x_3) - \frac{1}{y_{10}^2}\int_0^\infty \frac{dx_2}{x_2}\int_{-x_2}^{+x_2} dx_3 \frac{\partial f(x_1,x_2,x_3)}{\partial x_1}\Big\rvert_{x_1=0} + o(y_{10}^{-2})$
$y_{10} = 0, y_{23} = y_{20}$ $\lvert y_{23}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{i}{y_{20}}\log(-2iCy_{20})\int_0^\infty dx_1 f(x_1,0,0) - \frac{i}{y_{20}}\int_0^\infty dx_1\int_0^\infty dx_2 \log x_2 f(x_1,x_2,-x_2) + o(y_{20}^{-1})$
$y_{10} = my_{20}, m \neq 0$ $y_{23} = y_{20}$ $\lvert y_{20}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = -\frac{1}{y_{20}^2}\frac{1}{m}\log(-2iCy_{20})f(0,0,0) + \frac{1}{y_{20}^2}\frac{1}{m}\int_0^\infty dx_2 \log x_2 f(0,x_2,-x_2) + o(y_{20}^{-2})$
$y_{10} = 0$ $y_{23} = ny_{20}, n \neq \pm 1$ $\lvert y_{20}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{1}{y_{20}}\frac{i}{n}\log\frac{1+n}{1-n}\int_0^\infty dx_1 f(x_1,0,0) - \frac{1}{y_{20}^2}\left(\frac{1}{n^2}\log\frac{1+n}{1-n}\int_0^\infty dx_1 \frac{\partial f(x_1,0,x_3)}{\partial x_3}\Big\rvert_{x_3=0}\right.$ $+ \frac{1}{1-n}\int_0^\infty dx_1 \frac{\partial f(x_1,x_2,-x_2)}{\partial x_2}\Big\rvert_{x_2=0} - \left.\frac{1}{1+n}\int_0^\infty dx_1 \frac{\partial f(x_1,x_2,x_2)}{\partial x_2}\Big\rvert_{x_2=0}\right) + o(y_{20}^{-2})$
$y_{10} = my_{20} + a$ $y_{23} = ny_{20} + b$ $\lvert y_{20}\rvert \to \infty, m$ and n so that they do not match any other case	$\tilde{\phi}(y_1,y_2) = -\frac{1}{y_{20}^2}\log\frac{1-n}{1+n}\frac{f(0,0,0)}{mn} - \frac{i}{y_{20}^3}\frac{1}{mn}\left[\log\frac{1-n}{1+n}\left(\frac{1}{m}\frac{\partial}{\partial x_1}+\frac{1}{n}\frac{\partial}{\partial x_3}\right)f(x_1,x_2,x_3)\Big\rvert_{x_1=x_2=x_3}\right.$ $+ i\log\frac{1+n}{1-n}\left(\frac{a}{m}+\frac{b}{n}\right)f(0,0,0) + \frac{d}{dx_2}\left(\frac{f(0,x_2,-x_2)}{1-n} - \frac{f(0,x_2,x_2)}{1+n}\right)_{x_2=0} - \left.\frac{2ibf(0,0,0)}{1-n^2}\right] + o(y_{20}^{-3})$
$y_{20} = a, y_{23} = b$ $\lvert y_{10}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{i}{y_1}\int_0^\infty \frac{dx_2}{x_2}e^{iax_2}\int_{-x_2}^{+x_2} dx_3 e^{ibx_3} f(0,x_2,x_3) - \frac{1}{y_{10}^2}\int_0^\infty \frac{dx_2}{x_2}e^{iax_2}\int_{-x_2}^{+x_2} dx_3 e^{ibx_3}\frac{\partial f(x_1,x_2,x_3)}{\partial x_1}\Big\rvert_{x_1=0}$ $+ o(y_{10}^{-2})$
$y_{10} = a, y_{23} = y_{20} + b$ $\lvert y_{20}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{i}{y_{20}}\log(-2iCy_{20})\int_0^\infty dx_1 e^{iax_1} f(x_1,0,0) - \frac{i}{y_{20}}\int_0^\infty dx_1 e^{iax_1}\int_0^\infty dx_2 e^{-ibx_2}\log x_2 f(x_1,x_2,-x_2)$ $+ o(y_{20}^{-1})$
$y_{10} = my_{20} + a$ $y_{23} = y_{20} + b$ $\lvert y_{20}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = -\frac{1}{y_{20}^2}\frac{1}{m}\log(-2iCy_{20})f(0,0,0) + \frac{1}{y_{20}^2}\frac{1}{m}\int_0^\infty dx_2 e^{-ibx_2}\log x_2 f(0,x_2,-x_2) + o(y_{20}^{-2})$
$y_{10} = a, y_{23} = ny_{20} + b$ $n \neq \pm 1$ $\lvert y_{20}\rvert \to \infty$	$\tilde{\phi}(y_1,y_2) = \frac{1}{y_{20}}\frac{i}{n}\log\frac{1+n}{1-n}\int_0^\infty dx_1 e^{iax_1} f(x_1,0,0) - \frac{1}{y_{20}^2}\left[\frac{1}{n^2}\log\frac{1+n}{1-n}\int_0^\infty dx_1 e^{iax_1}\left(ibf(x_1,0,0)\right.\right.$ $\left. + \frac{\partial f(x_1,0,x_3)}{\partial x_3}\Big\rvert_{x_3=0}\right) - \frac{2ib}{1-n^2}\int_0^\infty dx_1 a^{iax_1} f(x_1,0,0) + \int_0^\infty dx_1 e^{iax_1}\left(\frac{1}{1-n}\frac{\partial f(x_1,x_2,-x_2)}{\partial x_2}\Big\rvert_{x_2=0}\right.$ $\left.\left. - \frac{1}{1+n}\frac{\partial f(x_1,x_2,x_2)}{\partial x_2}\Big\rvert_{x_2=0}\right)\right] + o(y_{20}^{-2})$

$$\times \sum_{n=1}^{N-1} \frac{2^{2n}(n-1)!\,p_0^{2n}}{y_3^{2n+1}} L_{n-1}^{(1/2)}\left(\frac{-k^2}{4p_0^2}\right) \times \left(\frac{1}{(m-1)^{2n}} + \frac{1}{(m+1)^{2n}}\right) + O(y_3^{-2N-1}) \tag{5.19a}$$

for $m \neq \pm 1$ and

$$\mathrm{Re}\tilde{\phi}(y) = -\frac{k}{2p_0^2 y_3}\,{}_1F_1\left(1, \frac{3}{2}; \frac{k^2}{4p^2}\right) + \frac{1}{2}\sum_{n=0}^{N-1} \times \frac{p_0^{2n}\,n!}{y_3^{2n+2}} L_n^{(-1/2)}\left(-\frac{k^2}{4p_0^2}\right) - \frac{k}{4}\sum_{n=1}^{N-1} \times \frac{p^{2n-2}(n-1)!}{y_3^{2n+1}} L_{n-1}^{(1/2)}\left(-\frac{k^2}{4p_0^2}\right) + O(y_3^{-2N-1}) \tag{5.19b}$$

for $m = \pm 1$. For the imaginary part we get again (5.15b). The functions $L_n^{(\alpha)}(x)$ are the generalized Laguerre polynomials, Ref. 29, p. 774. Again, the leading terms of Eq. (5.19) are easily seen to agree with the corresponding expansions (5.11) which were derived with the help of Table II. So we have a nice check of our results from Sec. 3.

Finally, we come to the following integral transform, which depends on two 4-vectors x_1 and x_2:

$$\tilde{\phi}(y_1,y_2) = \int dx_1 \int dx_2\, e^{ix_1y_1} e^{ix_2y_2}\theta(x_{10})\theta(x_{20})\delta(x_1^2) \times \delta(x_2^2)\delta[(x_1-x_2)^2]\phi(x_1,x_2). \tag{5.20}$$

We go to a special Lorentz frame. It is defined through the conditions that the vectors y_1 and y_2 have the form $y_1 = (y_{10}, \mathbf{0})$ and $y_2 = (y_{20}, 0, 0, y_{23})$. This is always possible if y_1 is timelike. It is not difficult to see the changes which have to be made in the following discussion if both 4-vectors y_1 and y_2 are spacelike. We choose the direction of $\mathbf{y}_2$ as polar axis and call the polar and azimuthal angles of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, $\Omega_1 = (\theta_1, \psi_1)$ and $\Omega_2 = (\theta_2, \psi_2)$,

respectively. Then the integral (5.1) reads

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_{10}\, e^{ix_{10}y_{10}} \int_0^\infty dx_{20} e^{ix_{20}y_{20}} \times \int_0^\infty d|\mathbf{x}_1|\,|\mathbf{x}_1|^2 \int d\Omega_1 \int_0^\infty d|\mathbf{x}_2|\,|\mathbf{x}_2|^2 \times \int d\Omega_2\, e^{-i|\mathbf{x}_2||\mathbf{y}_2|\cos\theta_2} \delta(x_1^2)\delta(x_2^2) \times \delta[(x_1-x_2)^2]\phi(x_{10},x_{20},|\mathbf{x}_1|,|\mathbf{x}_2|,\Omega_1,\Omega_2). \quad (5.21)$$

Using the δ functions to perform some of the integrations and defining

$$\bar{\phi}(x_{10},x_{20},\cos\theta_2) = \frac{\pi}{2}\int_0^{2\pi} d\psi \phi(x_{i0},|\mathbf{x}_i| = x_{i0},\Omega_1,\Omega_2), \quad (5.22)$$

We get from Eq. (5.21)

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_{10}\, e^{ix_{10}y_{10}} \int_0^\infty dx_{20}\, e^{ix_{20}y_{20}} \int_{-1}^{+1} d\cos\theta_2 \times e^{-ix_{20}|\mathbf{y}_2|\cos\theta_2} \phi(x_{10},x_{20},\cos\theta_2). \quad (5.23)$$

With the substitutions $x_{10} = x_1$, $x_{20} = x_2$, and $x_{20} \times \cos\theta_2 = x_3$, it follows that

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_1\, e^{ix_1y_{10}} \int_0^\infty \frac{dx_2}{x_2} e^{ix_2y_{20}} \times \int_{-x_2}^{+x_2} dx_3 e^{-ix_3y_{23}} f(x_1,x_2,x_3), \quad (5.24)$$

where

$$f(x_1,x_2,x_3) = \bar{\phi}(x_{10},x_{20},\cos\theta_2). \quad (5.25)$$

The integral transform (5.24) is of the same type as Eq. (4.1). By using the results of Sec. 4 the lower terms of the asymptotic expansions of $\tilde{\phi}(y_1,y_2)$ are given in Table IV.

ACKNOWLEDGMENT

The author thanks Professor W. Rühl for the suggestion of the problem and for numerous helpful discussions.

1 V. N. Gribov, B. L. Ioffe, and I. Ya. Pomeranchuk, Yad. Fiz. **2**, 768 (1965); [Sov. J. Nucl. Phys. **2**, 549 (1966)].
2 B. L. Ioffe, JETP Lett. **9**, 97 (1969)
3 R. A. Brandt, Lecture given at the Hamburg Summer School (1971).
4 Y. Frishman, Acta Phys. Austriaca **34**, 351 (1971).
5 J. M. Cornwall and R. Jackiw, Phys. Rev. D **4**, 367 (1971).
6 R. Jackiw, D. Dicus, and V. Teplitz, Phys. Rev. D **4**, 1733 (1971).
7 M. Gell-Mann and H. Fritzsch, in *Broken Scale Invariance and the Light Cone*, edited by M. Gell-Mann, Lectures from the Coral Gables Conference on Fundamental Interactions at High Energy, January 20-22 (1971) (Gordon and Breach, New York, 1971).
8 R. Jackiw, R. Van Royen, and G. B. West, Phys. Rev. D **2**, 2473 (1970).
9 R. A. Brandt, Phys. Rev. D **1**, 2808 (1970); Phys. Rev. Letters **23**, 1260 (1969).
10 W. Rühl, in *Springer Tracts in Modern Physics* (Springer, Berlin, 1971), Vol. 57.
11 P. Stichel, contribution to "Ecole Internationale de la Physique des Particules Elementaires, Basko Polje-Makarska (Yougoslavie), 1971."
12 H. Leutwyler and J. Stern, Nucl. Phys. B **20**, 77 (1970).
13 P. W. M. Boin, Dissertation Delft, Netherlands, 1965.
14 J. Berghuis, Dissertations Delft, Netherlands, 1955.
15 J. Berghuis, Mathematical Centre Amsterdam, Publication MR 11.
16 D. S. Jones, Proc. Roy. Soc. Edinburgh **69**, 205 (1970).
17 D. S. Jones and M. Kline, J. Math. Phys. **37**, 1 (1958).
18 D. S. Jones, J. Inst. Math. Appl. **2**, 197 (1966).
19 N. Bleistein and R. A. Handelsman, J. Math. Anal. Appl. **27**, 434 (1969).
20 N. Chako, J. Inst. Math. Appl. **1**, 372 (1965).
21 G. Braun, Acta Phys. Austriaca **10**, 8 (1955).
22 M. V. Fedorjuk, USSR Comput. Math. and Math. Phys. **2**, 152 (1962).
23 J. Focke, Math. Nat.-Wiss. Klasse **101**, (1953).
24 A. Erdelyi, *Asymptotic Expansions* (Dover, New York, 1956).
25 E. T. Copson, *Asymptotic Expansions* (Cambridge U.P., Cambridge 1967).
26 D. S. Jones, *Generalized Functions* (McGraw Hill, London, 1966).
27 M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge U.P., Cambridge, 1958).
28 L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).
29 M. Abramowitz and I. Stegun, *Handbook of Mathematical Formulas* (Dover, New York, 1964).

The $U_{n,1}$ and IU_n Representation Matrix Elements

Kurt Bernardo Wolf
C.I.M.A.S.S. Universidad Nacional Autónoma de México, México 20, D.F., México
(Received 27 March 1972)

We construct a realization of the $U_{n,1}$ and IU_n groups as multiplier representations of the space of functions on the U_n group manifold. Making use of the orthogonality and completeness of the U_n unitary irreducible representation matrix elements (UIRME's), we are able to express the $U_{n,1}$ boost and IU_n translation matrix elements (the generalized Wigner d-functions) of the principal series of UIR's as an integral over a compact domain (unit disc) of two U_n d-functions, phases, and the multiplier. This is an extension to the unitary groups of a method previously used [J. Math. Phys. 12, 197 (1971)] to find the SO_n, $SO_{n,1}$, and ISO_n UIRME's in a recursive fashion. We establish a number of symmetry properties, the asymtotic (Regge-like) and contraction ($U_{n,1} \to IU_n$) behavior of these functions.

1. INTRODUCTION

The unitary and pseudo-unitary groups in nuclear and elementary-particle physics have been used mainly through the associated Lie algebra.[1] The states of a system are identified with the components of the bases for unitary irreducible representations (UIR's) classified in some mathematically convenient or physically relevant chain of subalgebras. Interactions are then represented by operators with either irreducible tensor properties under the group or constructible in some simple fashion out of the universal enveloping algebra. Thus, the Wigner coefficients and the matrix elements of the generators of the Lie algebra[2] have played the main role in the applications of unitary groups.

The orthogonal, pseudo-, and inhomogeneous-orthogonal groups, on the other hand, have been widely used in connection with their finite transformations, either as a geometry group or in harmonic analysis on the $SO_{2,1}$ and $SO_{3,1}$ groups, whose UIR matrix elements (ME's) constitute a "best set" of functions in which to expand high-energy scattering data.[3] Also, a number of field theories have made use of the Poincaré group ($ISO_{3,1}$) manifold.[4]

respectively. Then the integral (5.1) reads

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_{10}\, e^{ix_{10}y_{10}} \int_0^\infty dx_{20} e^{ix_{20}y_{20}} \times \int_0^\infty d|\mathbf{x}_1|\,|\mathbf{x}_1|^2 \int d\Omega_1 \int_0^\infty d|\mathbf{x}_2|\,|\mathbf{x}_2|^2 \times \int d\Omega_2\, e^{-i|\mathbf{x}_2||\mathbf{y}_2|\cos\theta_2} \delta(x_1^2)\delta(x_2^2) \times \delta[(x_1-x_2)^2]\phi(x_{10},x_{20},|\mathbf{x}_1|,|\mathbf{x}_2|,\Omega_1,\Omega_2). \quad (5.21)$$

Using the δ functions to perform some of the integrations and defining

$$\bar{\phi}(x_{10},x_{20},\cos\theta_2) = \frac{\pi}{2}\int_0^{2\pi} d\psi \phi(x_{i0},|\mathbf{x}_i| = x_{i0},\Omega_1,\Omega_2), \quad (5.22)$$

We get from Eq. (5.21)

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_{10}\, e^{ix_{10}y_{10}} \int_0^\infty dx_{20}\, e^{ix_{20}y_{20}} \int_{-1}^{+1} d\cos\theta_2 \times e^{-ix_{20}|\mathbf{y}_2|\cos\theta_2} \phi(x_{10},x_{20},\cos\theta_2). \quad (5.23)$$

With the substitutions $x_{10} = x_1$, $x_{20} = x_2$, and $x_{20} \times \cos\theta_2 = x_3$, it follows that

$$\tilde{\phi}(y_1,y_2) = \int_0^\infty dx_1\, e^{ix_1y_{10}} \int_0^\infty \frac{dx_2}{x_2} e^{ix_2y_{20}} \times \int_{-x_2}^{+x_2} dx_3 e^{-ix_3y_{23}} f(x_1,x_2,x_3), \quad (5.24)$$

where

$$f(x_1,x_2,x_3) = \bar{\phi}(x_{10},x_{20},\cos\theta_2). \quad (5.25)$$

The integral transform (5.24) is of the same type as Eq. (4.1). By using the results of Sec. 4 the lower terms of the asymptotic expansions of $\tilde{\phi}(y_1,y_2)$ are given in Table IV.

ACKNOWLEDGMENT

The author thanks Professor W. Rühl for the suggestion of the problem and for numerous helpful discussions.

1 V. N. Gribov, B. L. Ioffe, and I. Ya. Pomeranchuk, Yad. Fiz. **2**, 768 (1965); [Sov. J. Nucl. Phys. **2**, 549 (1966)].
2 B. L. Ioffe, JETP Lett. **9**, 97 (1969)
3 R. A. Brandt, Lecture given at the Hamburg Summer School (1971).
4 Y. Frishman, Acta Phys. Austriaca **34**, 351 (1971).
5 J. M. Cornwall and R. Jackiw, Phys. Rev. D **4**, 367 (1971).
6 R. Jackiw, D. Dicus, and V. Teplitz, Phys. Rev. D **4**, 1733 (1971).
7 M. Gell-Mann and H. Fritzsch, in *Broken Scale Invariance and the Light Cone*, edited by M. Gell-Mann, Lectures from the Coral Gables Conference on Fundamental Interactions at High Energy, January 20-22 (1971) (Gordon and Breach, New York, 1971).
8 R. Jackiw, R. Van Royen, and G. B. West, Phys. Rev. D **2**, 2473 (1970).
9 R. A. Brandt, Phys. Rev. D **1**, 2808 (1970); Phys. Rev. Letters **23**, 1260 (1969).
10 W. Rühl, in *Springer Tracts in Modern Physics* (Springer, Berlin, 1971), Vol. 57.
11 P. Stichel, contribution to "Ecole Internationale de la Physique des Particules Elementaires, Basko Polje-Makarska (Yougoslavie), 1971."
12 H. Leutwyler and J. Stern, Nucl. Phys. B **20**, 77 (1970).
13 P. W. M. Boin, Dissertation Delft, Netherlands, 1965.
14 J. Berghuis, Dissertations Delft, Netherlands, 1955.
15 J. Berghuis, Mathematical Centre Amsterdam, Publication MR 11.
16 D. S. Jones, Proc. Roy. Soc. Edinburgh **69**, 205 (1970).
17 D. S. Jones and M. Kline, J. Math. Phys. **37**, 1 (1958).
18 D. S. Jones, J. Inst. Math. Appl. **2**, 197 (1966).
19 N. Bleistein and R. A. Handelsman, J. Math. Anal. Appl. **27**, 434 (1969).
20 N. Chako, J. Inst. Math. Appl. **1**, 372 (1965).
21 G. Braun, Acta Phys. Austriaca **10**, 8 (1955).
22 M. V. Fedorjuk, USSR Comput. Math. and Math. Phys. **2**, 152 (1962).
23 J. Focke, Math. Nat.-Wiss. Klasse **101**, (1953).
24 A. Erdelyi, *Asymptotic Expansions* (Dover, New York, 1956).
25 E. T. Copson, *Asymptotic Expansions* (Cambridge U.P., Cambridge 1967).
26 D. S. Jones, *Generalized Functions* (McGraw Hill, London, 1966).
27 M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge U.P., Cambridge, 1958).
28 L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).
29 M. Abramowitz and I. Stegun, *Handbook of Mathematical Formulas* (Dover, New York, 1964).

The $U_{n,1}$ and IU_n Representation Matrix Elements

Kurt Bernardo Wolf
C.I.M.A.S.S. Universidad Nacional Autónoma de México, México 20, D.F., México
(Received 27 March 1972)

We construct a realization of the $U_{n,1}$ and IU_n groups as multiplier representations of the space of functions on the U_n group manifold. Making use of the orthogonality and completeness of the U_n unitary irreducible representation matrix elements (UIRME's), we are able to express the $U_{n,1}$ boost and IU_n translation matrix elements (the generalized Wigner d-functions) of the principal series of UIR's as an integral over a compact domain (unit disc) of two U_n d-functions, phases, and the multiplier. This is an extension to the unitary groups of a method previously used [J. Math. Phys. 12, 197 (1971)] to find the SO_n, $SO_{n,1}$, and ISO_n UIRME's in a recursive fashion. We establish a number of symmetry properties, the asymtotic (Regge-like) and contraction ($U_{n,1} \to IU_n$) behavior of these functions.

1. INTRODUCTION

The unitary and pseudo-unitary groups in nuclear and elementary-particle physics have been used mainly through the associated Lie algebra.[1] The states of a system are identified with the components of the bases for unitary irreducible representations (UIR's) classified in some mathematically convenient or physically relevant chain of subalgebras. Interactions are then represented by operators with either irreducible tensor properties under the group or constructible in some simple fashion out of the universal enveloping algebra. Thus, the Wigner coefficients and the matrix elements of the generators of the Lie algebra[2] have played the main role in the applications of unitary groups.

The orthogonal, pseudo-, and inhomogeneous-orthogonal groups, on the other hand, have been widely used in connection with their finite transformations, either as a geometry group or in harmonic analysis on the $SO_{2,1}$ and $SO_{3,1}$ groups, whose UIR matrix elements (ME's) constitute a "best set" of functions in which to expand high-energy scattering data.[3] Also, a number of field theories have made use of the Poincaré group ($ISO_{3,1}$) manifold.[4]

There has been a corresponding increase of interest in considering the UIRME's—the generalized D and d-functions—as "special functions,"[5] that is, as orthogonal and complete[6] sets of functions in terms of which one can expand any well-behaved function on the group manifold which, furthermore, due to the group properties, exhibit summation and recursion formulae, the emphasis being placed not so much in their explicit expressions which, like the series expansion of a Bessel function, provides at best a limited insight into the aforementioned properties, but in the relations between functions which these properties imply.

It was in this spirit that we treated in Ref. 7 the generalized Wigner d-functions for the SO_n, $SO_{n,1}$, and ISO_n groups.[8] In the present paper we apply the techniques developed in Ref. 7 to the unitary (U_n), pseudo-unitary ($U_{n,1}$), and inhomogeneous unitary (IU_n) groups. As the method is essentially parallel, we shall skip most of the introductory material on multipliers as well as the detailed description of the U_n manifold and representation theory. In these, we use the concepts introduced in Ref. 9, giving a summary of notation in Sec. 2.

The U_2 UIRME's are essentially the classical Wigner d-functions. Bég and Ruegg[10] and T. J. Nelson[11] studied the U_3 harmonic functions, the analog of the SO_3 spherical harmonics on the (five-dimensional) manifold of the complex 3-sphere $C_3 \cong U_2 \backslash U_3$, the eigenfunctions of the Laplace–Beltrami operators of the manifold. Using these techniques, Fischer and Rączka[12] gave explicit expressions for the U_n and $U_{p,q}$ harmonic functions. These can be used in order to find the UIRME's themselves, as was done by Holland[13] for SU_3 and by Delbourgo, Koller, and Williams[14] for SU_n. This technique, however, can only give the $[J_1 J \cdots J\, J_n]$ UIR's of the general U_n ($n > 3$) groups since[9] only these can be realized on the $C_n \cong U_{n-1} \backslash U_n$ homogeneous space.

A different line of approach was followed by Chacón and Moshinsky,[15] who expressed the general U_3 transformation as a product of several U_2 transformations and transpositions. This method was extended to U_n by Flores and Niederle.[16] Taking these d-functions as known, our approach hinges in defining the action of a $U_{n,1}$ group as a group of transformations of the U_n manifold such that, while the canonical U_n subgroup of $U_{n,1}$ produces "rigid" mappings (leaving the Haar measure invariant), the boosts of $U_{n,1}$ produce "deformations" of the manifold. This is detailed in Sec. 3. By considering those transformations which commute with the canonical U_{n-1} subgroup, it is sufficient to define the "deformation" on the $C_n \cong U_{n-1} \backslash U_n$ manifold. This leads to a multiplier representation and to the expression, in Sec. 4, of the $U_{n,1}$ d-functions of the principal series of UIR's[17,18] in terms of an integral over a compact domain of two U_n d-functions, phases, and a multiplier. Some properties of the d-functions are exhibited in Sec. 5. In Sec. 6, a similar procedure gives the IU_n d-functions. These are checked to correspond to contractions[19] of the $U_{n,1}$ d-functions.

The formalism works best when we use the unitary analog of the Euler angles,[15,20] the "last latitude" angle in $U_{n,1}$ being a boost and, in IU_n, a real translation. As for the orthogonal groups,[7] we want to emphasize that our procedure gives the $U_{n,1}$ principal series of UIRME's classified by the canonical chain of subgroups. Several properties are apparent from the integral form. This method seems to be extendable to other groups and manifolds in essentially the same form.

2. THE UNITARY GROUP MANIFOLDS AND REPRESENTATIONS

The Euler-angle parametrization[15,20] of U_n can be defined, enclosing collective variables in curly brackets:

$$u_n(\{\phi, \theta\}^{(n)}) = u_{n-1}(\{\phi, \theta\}^{(n-1)})\, c_n(\{\phi^{(n)}, \theta^{(n)}\}), \tag{2.1a}$$

$$c_n(\{\phi^{(n)}, \theta^{(n)}\}) = \Phi_n(\phi_n^{(n)}) r_{n-1,n}(\theta_{n-1}^{(n)})\, c_{n-1}(\{\phi^{(n)}, \theta^{(n)}\}), \tag{2.1b}$$

$$u_1(\{\phi, \cdot\}^{(1)}) = c_1(\{\phi^{(1)}, \cdot\}) = \Phi_1(\phi_1^{(1)}), \tag{2.1c}$$

where $r_{pq}(\theta)$ are rotations by θ in the p-q plane of an n-dimensional complex coordinate space $Z^n \ni \mathbf{z}$, and $\Phi_k(\phi)$ are phase rotations by ϕ in the kth coordinate. Defining

$$\mathbf{z}(\{\phi, \theta\}) = c_n(\{\phi, \theta\})^{-1} \mathbf{z}_0$$

for a fixed $\mathbf{z}_0 \in Z^n$, we introduce complex-spherical coordinates in Z^n as

$$Z_k(\{\phi, \theta\}) \equiv r_k(\{\phi\}) e^{-i\psi_k} = re^{-i(\phi_n + \cdots + \phi_k)} \times \sin\theta_{n-1} \cdots \sin\theta_k \cos\theta_{k-1}, \tag{2.2a}$$

for $k = 2, \ldots, n-1$. For $k = 1$ we can put formally $\theta_0 \equiv 0$, while for $k = n$

$$Z_n(\{\phi, \theta\}) \equiv r_n(\{\theta\}) e^{-i\psi_n} = re^{-i\phi_n} \cos\theta_{n-1}. \tag{2.2b}$$

Choosing the ranges[15] $\theta_i \in [0, \pi/2]$ ($i = 1, \ldots, n-1$), $\phi_j \in [0, 2\pi)$ ($j = 1, \ldots, n$), we give to $r_k(\{\theta\})$ the meaning of the modulus of $\mathbf{z}$ and ψ_k as its phase.

For fixed r we have the $(2n-1)$-dimensional manifold of the complex n-sphere $C_n \cong U_{n-1} \backslash U_n$ with

$$dc_n(\{\phi, \theta\}) = d\mu_n(\phi_n, \theta_{n-1})\, dc_{n-1}(\{\phi, \theta\}), \tag{2.3a}$$

$$d\mu_k(\phi_k, \theta_{k-1}) = \sin^{2k-3}\theta_{k-1} \cos\theta_{k-1} d\phi_k d\theta_{k-1}, \tag{2.3b}$$

$$dc_1(\{\phi, \cdot\}) = d\mu_1(\phi_1, \cdot) = d\phi_1, \tag{2.3c}$$

and through (1.1) we construct the Haar measure for U_n. Integrating (2.3) over C_n we find its area to be $|C_n| = 2\pi^n/\Gamma(n)$. The volume of U_n is, from (2.1), $\mathrm{vol}U_n = \mathrm{vol}U_{n-1} \cdot |C_n|$, $\mathrm{vol}U_1 = 2\pi$.

For the $U_{n-1,1}$ group, the rotation angle in the $(n-1)$-n plane in (1.1b) is replaced by a boost $b_{n-1,n}(\xi)$, $\xi \in [0, \infty)$ in that plane, while for the IU_{n-1} group, it is replaced by a real translation $t_{n-1}(\xi)$, $\xi \in [0, \infty)$ in the $(n-1)$th direction.

The U_n Gel'fand kets[2,21] will be abbreviated[7] $|J_n, \overline{J_{n-1}}\rangle$, where $J_n \equiv [J_{n,1}, J_{n,2}, \ldots, J_{n,n}]$ labels the U_n UIR and $\overline{J_{n-1}}$ its row index: $\overline{J_{n-1}} \equiv \{J_{n-1}, J_{n-2}, \ldots, J_1\}$, where J_k denotes the UIR of the canonical U_k subgroup of U_n. The individual labels J_{km} obey the known "zig-zag" inequalities

$$J_{k,m-1} \geq J_{k-1,m-1} \geq J_{k,m}, \quad n \geq k \geq m \geq 2. \tag{2.4}$$

The U_n representation D-matrices are thus labeled as

$$D^{J_n}_{\overline{J_{n-1}},\overline{J'_{n-1}}}[u_n(\{\phi,\theta\}^{(n)})] \equiv \langle J_n \overline{J_{n-1}} | u_n(\{\phi,\theta\}^{(n)}) | J_n \overline{J'_{n-1}} \rangle, \tag{2.5}$$

and can be decomposed through (1.1) into sums of products of the phase functions

$$p^{J_k}_{J_{k-1}}(\phi) \equiv \langle J_k \overline{J_{k-1}} | \Phi_k(\phi) | J_k \overline{J_{k-1}} \rangle, \tag{2.6}$$

which are diagonal and independent of the $U_m (m > k$ and $m < k-1)$ labels, and the generalized Wigner d-functions

$$d^{J_k}_{J_{k-1},J_{k-2},J'_{k-1}}(\theta) \equiv \langle J_k J_{k-1} \overline{J_{k-2}} | r_{k-1,k}(\theta) | J_k J'_{k-1} \overline{J_{k-2}} \rangle, \tag{2.7}$$

diagonal in the U_k and U_{k-2} UIR labels and independent of the $U_m (m > k$ and $m < k-2)$ labels. The U_n D-functions (2.5) are orthogonal and complete on the U_n manifold with the U_n Haar measure and the Plancherel weight $\dim J_n / \mathrm{vol} U_n$.

For the $U_{n-1,1}$ and IU_{n-1} groups, the Gel'fand patterns[17,18] are similar to the U_n ones, except for the labels J_{n1} and J_{nn} which are, in general, complex and do not abide (1.4). The representations are thus infinite-dimensional. The d-functions we want to calculate, which we shall denote by ${}^P d$ and ${}^I d$ for the pseudo- and inhomogeneous-unitary groups, are the matrix elements respectively, of the boost $b(\zeta)$ and the real translation $t(\xi)$ in the corresponding Euler-angle parametrization.

3. THE $U_{n,1}$ ALGEBRA AND MULTIPLIER REPRESENTATION

The set of operators on the complex n-space Z^n

$$\mathcal{C}_j{}^k \equiv z_j \frac{\partial}{\partial z_k} - z^k \frac{\partial}{\partial z^j} \tag{3.1}$$

with $\overline{z_k} = z^k$ (complex conjugation), have the well-known commutation relations of the generators of the u_n algebra.[21] They leave the n-sphere $C_n \cong U_{n-1} \backslash U_n$ invariant. If we add the z_k and z^k $(k = 1, \ldots, n)$ to the set (3.1), we have the generators of an iu_n algebra. Using the second-order Casimir operator

$$\Psi^{(n)} \equiv \mathcal{C}_j{}^k \mathcal{C}_k{}^j \tag{3.2}$$

(sum over repeated indices, unless otherwise indicated), we can construct, out of the universal enveloping algebra of iu_n, the operators

$${}^{(\sigma)}\mathcal{C}^{n+1}_k \equiv \tfrac{1}{2}[\Psi^{(n)}, z_k] + \sigma z_k = z_l \mathcal{C}^l{}_k + (\tfrac{1}{2}n + \sigma) z_k, \tag{3.3a}$$

$${}^{(\sigma)}\mathcal{C}^k_{n+1} \equiv \tfrac{1}{2}[\Psi^{(n)}, z^k] + \overline{\sigma}\, z^k = z^l \mathcal{C}_l{}^k + (-\tfrac{1}{2}n + \overline{\sigma}) z^k, \tag{3.3b}$$

$${}^{(\sigma)}\mathcal{C}^{n+1}_{n+1} \equiv [{}^{(\sigma)}\mathcal{C}^k_{n+1}, {}^{(\sigma)}\mathcal{C}^{n+1}_k] - \mathcal{C}^k_k = z_l z^j \mathcal{C}^l_j + (\sigma + \overline{\sigma}) \tag{3.3c}$$

(no sum over k), where σ is an (as yet) arbitrary complex number. As the notation suggests, (3.3) together with (3.1) generate a $u_{n,1}$ algebra which leaves C_n invariant with $r = 1$. Each value of σ gives a different set of generators which will produce a correspondingly distinct UIR, as can be seen from the $u_{n,1}$ Casimir operator

$$\Psi^{(n,1)}(\sigma) = \Psi^{(n,1)}(0) + \sigma^2 + \overline{\sigma}^2 \tag{3.4a}$$

and the unitary invariant

$$\Omega^{(n,1)} \equiv \sum_{k=1}^{n} \mathcal{C}^k_k - \mathcal{C}^{n+1}_{n+1} = -\sigma - \overline{\sigma}. \tag{3.4b}$$

We can build an $so_{n,1} \subset u_{n,1}$ subalgebra generated by

$$M_{jk} \equiv \mathcal{C}_j{}^k - \mathcal{C}_j{}^k \quad (1 \leq j < k \leq n) \tag{3.5a}$$

and

$${}^{(\sigma)}M_{k,n+1} \equiv {}^{(\sigma)}\mathcal{C}^{n+1}_k - {}^{(\sigma)}\mathcal{C}^k_{n+1} \quad (k = 1, \ldots, n), \tag{3.5b}$$

anti-Hermitean under the measure (2.3). The operators (3.5a) will generate boosts in the kth direction. Now, since ${}^{(\sigma)}M_{n,n+1}$ commutes with the generators of u_{n-1}, its action on U_n can be fully studied as the action on $C_n \cong U_{n-1} \backslash U_n$. It is sufficient, therefore, to construct ${}^{(\sigma)}M_{n,n+1}$ in terms of the complex-spherical coordinates (2.2). Direct calculation through (2.2), (3.1), (3.3), and (3.5b) yields

$$\begin{aligned} {}^{(\sigma)}M_{n,n+1} &= \sin\theta_{n-1} \cos\phi_n \frac{\partial}{\partial \theta_{n-1}} \\ &+ (\sec\theta_{n-1} + \cos\theta_{n-1}) \sin\phi_n \frac{\partial}{\partial \phi_n} \\ &- \sec\theta_{n-1} \sin\phi_n \frac{\partial}{\partial \phi_{n-1}} + \cos\theta_{n-1} \\ &[(n + 2i\,\mathrm{Im}\sigma)\cos\phi_n - 2i\,\mathrm{Re}\sigma \sin\phi_n]. \end{aligned} \tag{3.6}$$

The exponentiation of (3.6), for $\sigma = 0$ yields the action of $b_{n,n+1}(\zeta)$ on C_n and can be found from the action of $U_{n,1}$ on itself (in the Iwasawa decomposition $U_{n,1} = U_n \cdot A \cdot N$) modulo N, in the same fashion as was done in Ref. 7, generating the following transformation of $z \in C_n$: the unit disc $|z_n| \leq 1$,

$$z_n \to z'_n = \frac{z_n \cosh\zeta - \sinh\zeta}{\cosh\zeta - z_n \sinh\zeta}, \tag{3.7a}$$

which defines $\phi_n \to \phi'_n$ and $\theta_n \to \theta'_n$, and

$$\begin{aligned} \phi_{n-1} &\to \phi'_{n-1} \\ &= \phi_{n-1} + \arg(\cos\theta_{n-1} \cosh\zeta - \exp i\phi_n \sinh\zeta) \\ &\equiv \phi_{n-1} + \chi(\phi_n, \theta_{n-1}, \zeta), \end{aligned} \tag{3.7b}$$

all other coordinates of z remaining unaffected. This can be seen as the "complexification" of the more familiar transformation $\tan\frac{1}{2}\theta \to \tan\frac{1}{2}\theta' = e^{\zeta} \tan\frac{1}{2}\theta$ which appears in connection with the pseudo-orthogonal groups[3,7,8,22]. The Jacobian of the transformation (3.7) is

$$\frac{dc_n(\{\phi', \theta'\})}{dc_n(\{\phi, \theta\})} = \frac{d\mu_n(\phi'_n, \theta'_{n-1})}{d\mu_n(\phi_n, \theta_{n-1})} = \left(\frac{\sin\theta'_{n-1}}{\sin\theta_{n-1}}\right)^{2n}. \tag{3.8}$$

We have thus, for $\sigma = i\tau$ (τ real), a unitary multiplier representation[23] of $b_{n,n+1}(\zeta)$ on the space of functions f on C_n (and therefore on U_n) as

$$T^{(\sigma)}(b_{n,n+1}(\zeta))f(z) = \exp[\zeta^{(\sigma)}M_{n,n+1}]f(z)$$
$$= [\sin\theta'_{n-1}/\sin\theta_{n-1}]^{n+\sigma} f(z'). \quad (3.9)$$

4. THE U_{n-1} MATRIX ELEMENTS

The phase functions (2.6) are the matrix elements of transformations generated by $\mathcal{C}_k{}^k$ (no sum). As[21] $\mathcal{C}_k{}^k|J_{n+1}\overline{J_n}\rangle = \omega_k|J_{n+1}\overline{J_n}\rangle$ with $\omega_k = \sum_{l=1}^{k} J_{kl} - \sum_{l=1}^{k-1} J_{k-1,l}$

$$p^{J_k}_{J_{k-1}}(\phi) = \exp(i\omega_k\phi), \quad k = j, \ldots, n+1. \quad (4.1)$$

The eigenvalue of the unitary invariant (3.4b) is $\omega_1 + \omega_2 + \cdots + \omega_n - \omega_{n+1}$. For σ pure imaginary, (3.4b) is zero and hence $\omega_{n+1} = \sum_{k=1}^{n} J_{nk}$.

The calculation of the $U_{n,1}$ ${}^Pd$-functions, however, will require the multiplier representation (3.9). Given a set $\{\phi^{(\mu)}_k\}$ $k \in N$ (N an index set determined by μ) of orthogonal functions on a manifold M, a representation of a group of transformations $G \ni g$ of M can be constructed as[7]

$$D^{(\lambda,\mu)}_{k,k'}(g) = [\omega(k)\omega(k')]^{1/2}(\phi^{(\mu)}_k, T^{(\lambda)}(g)\phi^{(\mu)}_{k'})_M, \quad (4.2)$$

where ω is the Plancherel weight of N. Using for M the U_n manifold and $D^{J_n}_{J_{n-1},\overline{J'_{n-1}}}$, $\overline{J'_{n-1}}$ as the set of orthogonal functions, we proceed to prove that, in close analogy with the orthogonal groups[7], the ${}^Pd$-functions can be found as

$${}^Pd^{\{\alpha,J'_{n-1},\beta\}}_{J_n J_{n-1} J'_n}(\zeta) = [\dim J_n \dim J'_n)^{1/2}/\mathrm{vol}U_n]$$
$$\times \left(D^{J_n}_{J'_{n-1},\overline{J_{n-1}}}, T^{(\sigma)}(b_{n,n+1}(\zeta))D^{J'_n}_{J'_{n-1},\overline{J_{n-1}}}\right), \quad (4.3)$$

where the connection between α, J'_{n-1}, β, and the $U_{n,1}$ UIR labels J_{n+1} will be clarified below.

At $\zeta = 0$, the orthogonality of the D's insures that ${}^Pd^{J_{n+1}}_{J_n J_{n-1} J'_n}(0) = \delta_{J_n,J'_n}$ (the Kronecker δ in the collective indices J_n and J'_n stands for a product of δ's in the individual indices J_{nk} and J'_{nk}, $k = 1, \ldots, n$). The completeness of the D's gives the addition formula

$$\sum_{J''_n} {}^Pd^{J_{n+1}}_{J_n J_{n-1} J''_n}(\zeta_1)\, {}^Pd^{J_{n+1}}_{J''_n J_{n-1} J'_n}(\zeta_2) = d^{J_{n+1}}_{J_n J_{n-1} J'_n}(\zeta_1 + \zeta_2), \quad (4.4)$$

hence (4.3) together with (4.1) and (4.2) for $g \in U_n$ provide us with a representation $U_{n,1}$. There is no invariant subspace. This construction gives us the classification through the Gel'fand patterns of the $U_{n,1}$ UIR $J_{n+1} = \{\alpha, J'_{n-1}, \beta\}$ since the individual indices $J_{n+1,1} = \alpha$, $J_{n+1,k+1} = J'_{n-1,k}$ $(k = 1, \ldots, n)$, $J_{n+1,n+1} = \beta$ restricted through the zig-zag inequalities (2.4) for $U_n \supset U_{n-1}$, when taken as the $U_{n,1}$ UIR's restrict in turn the UIR labels of $U_n \subset U_{n,1}$. The "end point" labels α and β will now be related to σ when we identify them as the continuation of the values of $J_{n+1,1}$ and $J_{n+1,n+1}$ entering into the expressions for (i) the unitary invariant (3.4b) eigenvalue

$$0 = \sum J_{n-1,k} = \alpha + \beta + \sum J'_{n-1,k} \quad (4.5a)$$

(the sum extending over the allowed values of the free index) and (ii) the second-order Casimir operator (3.4a) eigenvalue

$$\sum J_{n+1,k}(J_{n+1,k} - 2k + n + 2) = \alpha(\alpha + n) + \beta(\beta - n)$$
$$+ \sum J'_{n-1,k}(J'_{n-1,k} - 2k + n), \quad (4.5b)$$

which, if the representation is to be unitary, (iii) has to be real. Lastly, (iv) the dependence of (4.5b) on $\sigma = i\tau$ must be that given by (3.4a).

All four conditions (i)–(iv) can be satisfied by the choice $\alpha = -\frac{1}{2}(n + \sum J'_{n-1,k}) + i\tau$ and $\beta = \frac{1}{2}(n - \sum J'_{n-1,k}) - i\tau$. The parameter τ can be identified with Chakrabarti's[17] parameter ϵ, and seen to label the continuum of principal series UIR's of $U_{n,1}$. Values of τ and $-\tau$ give equivalent UIR's.

The integral over U_n in (4.3) can be simplified when the D's are written in terms of p's, d's, and the U_{n-1} D's as in (2.1). Orthogonality relations can be used to yield Kronecker δ's in the corresponding labels, and the multiple integral reduces to an integral over the unit disc:

$${}^Pd^{\{\alpha,J'_{n-1},\beta\}}_{J_n J_{n-1} J'_n}(\zeta) = \frac{(\dim J_n \dim J'_n)^{1/2}}{\dim J_{n-1} \dim J'_{n-1}} \times \frac{(\mathrm{vol}U_{n-1})^2}{\mathrm{vol}U_n \mathrm{vol}U_{n-2}}$$
$$\times \sum_{J_{n-2}} \dim J_{n-2} \int d\mu_n(\phi,\theta)\, \overline{p^{J_n}_{J_{n-1}}(\phi)}\; \overline{d^{J_n}_{J'_{n-1} J_{n-2} J_{n-1}}(\theta)}$$
$$\times \left(\frac{\sin\theta'}{\sin\theta}\right)^{n+i\tau} \exp\left[i\left(\sum J_{n-1} - \sum J_{n-2}\right)\chi(\phi,\theta,\zeta)\right] p^{J'_n}_{J'_{n-1}}(\phi')$$
$$\times d^{J'_n}_{J'_{n-1} J_{n-2} J_{n-1}}(\theta'), \quad (4.6)$$

where the primed variables are related to the unprimed one through the transformation (3.7).

5. SOME PROPERTIES OF THE ${}^Pd$–FUNCTIONS

We will not attempt here the explicit evaluation of (4.6). Several properties are apparent, however, from the integral form (4.3)–(4.6):
(i) the group property yields the addition formula (4.4);
(ii) unitarity of the representation gives

$${}^Pd^{J_{n+1}}_{J_n J_{n-1} J'_n}(-\zeta) = \overline{{}^Pd^{J_{n+1}}_{J'_n J_{n-1} J_n}(\zeta)}; \quad (5.1)$$

(iii) invariance of the scalar product (4.3) under the involution $u_n \leftrightarrow u_n^{-1}$ and the unitary of the $U_n D$'s imply

$${}^Pd^{\{\alpha,J'_{n-1},\beta\}}_{J_n J_{n-1} J'_n}(\zeta) = \overline{{}^Pd^{\{\alpha,J_{n-1},\beta\}}_{J_n J'_{n-1} J'_n}(\zeta)}; \quad (5.2)$$

(iv) the asymptotic behavior ($\zeta \to \infty$) is similar to the Regge behavior of the $SO_{n,1}$ d-functiond[4,7,8]: It is exponentialy decreasing in ζ. As the disc (3.7a) streches towards the point $z_n = -1$, $\sin\theta'/\sin\theta \sim e^{-\zeta}$ and

$${}^Pd^{\{\alpha,J'_{n-1},\beta\}}_{J_n J_{n-1} J'_n}(\zeta) \xrightarrow[\zeta\to\infty]{} \delta_{J_{n-1},J'_{n-1}} \Delta^{\alpha J'_{n-1}\beta}_{J_n J'_n} e^{-[n+i\tau]\zeta}, \quad (5.3)$$

where $\Delta^{\alpha J'_{n-1}\beta}_{J_n J'_n}$ are constants obtainable from (4.6).

6. THE IU_n MATRIX ELEMENTS

We consider now the finite translations generated by z_k and z^k as a multiplier representation on the space

of functions f on C_n. The real translation $t_n(\xi) \in ISO_n \subset IU_n$ [taking the place of $r_{n-1,n}(\theta)$ in (2.1b)] is generated by $x_n = \frac{1}{2}(z_n + z^n) = r\cos\theta_{n-1}\cos\phi_n$ and has the action

$$T^{(r)}(t_n(\xi))f(z) = \exp(i\xi x_n)f(z), \tag{6.1}$$

which is unitary for real r, but produces no deformation of the C_n manifold. Again, as x_n commutes with the generators of u_{n-1}, the action (6.1) of x_n on C_n can be used to construct the IU_n UIR's through (4.2) and, analogously to (4.3) and (4.6), we find the IU_n ^Id-functions as

$$\begin{aligned}
{}^Id^{\{r,J'_{n-1},s\}}_{J_n J_{n-1} J'_n}(\xi) &= \frac{(\dim J_n \dim J'_n)^{1/2}}{\mathrm{vol}U_n} \\
&\left(D^{J_n}_{J'_{n-1}\,\overline{J_{n-1}}}, T^{(r)}(t_n(\xi))D^{J'_n}_{J'_{n-1},J_{n-1}}\right) \\
&= \frac{[\dim J_n \dim J'_n]^{1/2}}{\dim J_{n-1}\dim J'_{n-1}} \frac{(\mathrm{vol}U_{n-1})^2}{\mathrm{vol}U_n\,\mathrm{vol}U_{n-2}} \\
&\sum_{J_{n-2}} \dim J_{n-2} \int d\mu_n(\phi,\theta) \\
&\times \overline{p^{J_n}_{J_{n-1}}(\phi)}\ \overline{d^{J_n}_{J'_{n-1}J_{n-2}J_{n-1}}(\theta)} \\
&\exp[ir\xi\cos\theta\cos\phi]p^{J'_n}_{J'_{n-1}}(\phi)d^{J'_n}_{J'_{n-1}J_{n-2}J_{n-1}}(\theta).
\end{aligned} \tag{6.2}$$

The iu_n second-order Casimir operator $z_l z^l$ has eigenvalues r^2, and thus r (real) labels the IU_n UIR's corresponding to Chakrabarti's[17] parameter κ. The ^Id-functions (6.2) are independent of the label s. This label enters into the picture when we consider the phase of the translation $\Phi_{n+1}(\phi)$. Its matrix elements follow from (4.1) and will not be considered again. Properties analogous to those presented in the last section follow.

As was the case for the orthogonal groups[7], the $U_{n,1}$ group can be defomed in the Inönü–Wigner sense[19] into the IU_n group when we consider UIR's with $\tau \to \infty$ while keeping $\tau\zeta = r\xi$. The multiplier (3.9) becomes then

$$(\sin\theta'/\sin\theta)^{n+i\tau} \xrightarrow[\tau\to\infty]{} \exp[ir\xi\cos\theta\cos\phi]$$

while, as $\zeta \to 0$, there is no deformation of the group manifold. Comparing (4.6) and (6.2) we see that

$${}^Pd^{\{\alpha(\tau),J'_{n-1},\beta(\tau)\}}_{J_n J_{n-1} J_n}(\zeta) \xrightarrow[\substack{\tau\to\infty \\ \zeta\tau = \xi r}]{} {}^Id^{\{r,J'_{n-1},s\}}_{J_n J_{n-1} J'_n}(\xi),$$

thus, characterizing the value of the last IU_n label when we maintain the eigenvalue of the unitary invariant (3.4b) as zero.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. Elpidio Chacón and Dr. Thomas Seligman for several discussions.

1 E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959); Y. Ne'eman, *Algebraic Theory of Particle Physics* (Benjamin, New York, 1967); M. Moshinsky, *The Harmonic Oscilator in Modern Physics: from Atoms to Quarks* (Gordon and Breach, New York, 1969).

2 J. P. Louck, Amer. J. Phys. **38**, 3 (1970).

3 M. Toller, Nuovo Cimento **53**, 671 (1968); **54**, 295 (1968); A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967); G. Consenza, A. Sciarrino, and M. Toller, Nuovo Cimento **57A**, 253 (1968).

4 F. Lurçat, Physics **1**, 95 (1964); N. X. Hai, Commun. Math. Phys. **12**, 331 (1969); A. Kihlberg, Arkiv Fysik **28**, 121 (1964); **32**, 263 (1966); Nuovo Cimento **53**, 592 (1968); J. D. Hind, *ibid.* **4**, 71 (1971).

5 N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations*, Mathematical Monographs, Vol. 22 (Amer. Math. Soc., Providence, R. I., 1968), Vol. 22; J. D. Talman, *Special Functions, A Group-Theoretic Approach* (Benjamin, New York, 1968); W. Miller *Lie Groups and Special Functions* (Academic, New York, 1968); R. Hermann, *Fourier Analysis on Groups and Partial-Wave Analysis* (New York, 1968); K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (Polish Scientific Publ., Warsaw, 1968).

6 Completeness of the UIRME's for compact groups is guaranteed by the Peter–Weyl theorem (see J. D. Talman, Ref. 5, Chap. 7; K. Maurin, Ref. 5, p. 157). For noncompact, semisimple groups, see R. Rączka "Operator Distributions in Group Representation-Theory and Their Applications," Lecture Notes, Institute of Theoretical Physics, Göteborg, Sweden, (1969).

7 K. B. Wolf, J. Math. Phys. **12**, 197 (1971).

8 These have also been treated from a similar standpoint by C. P. Boyer, J. Math. Phys. **12**, 1599 (1971); C. P. Boyer and F. Ardalan, *ibid.*, 2070 (1971).

9 R. L. Anderson and K. B. Wolf, J. Math. Phys. **11**, 3176 (1970).

10 M. A. B. Bég and H. Ruegg, J. Math. Phys. **6**, 677 (1965).

11 T. J. Nelson, J. Math. Phys. 8, 857 (1967).

12 J. Fischer and R. Rączka, Commun. Math. Phys. **3**, 233 (1966); **4**, 8 (1967).

13 D. F. Holland, J. Math. Phys. **10**, 531, 1963 (1969).

14 R. Delbourgo, K. Koller, and R. M. Williams, J. Math. Phys. **10**, 957 (1969).

15 E. Chacón and M. Moshinsky, Phys. Letters **23**, 567 (1966).

16 J. Flores and J. Niederle, Czech. J. Phys. **B20**, 1241 (1970).

17 A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).

18 U. Ottoson, Commun. Math. Phys. **10**, 114 (1968); I. M. Graev, Izv. Akad. Nauk SSSR, Ser. Mat., **29**, 1329 (1965).

19 E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).

20 A. Maduemezia, J. Math. Phys. **12**, 1681 (1971).

21 I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); J. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).

22 V. Bargmann, Ann. Math. **48**, 568 (1947).

23 Reference 7, Sec. 2.

Classification of Semisimple Subalgebras of Simple Lie Algebras*

M. Lorente
Departmento de Física Teórica, Universidad de Madrid, Madrid-3, Spain†
and
Physics Department, Saint Louis University, St. Louis, Mo. 63103‡

and

B. Gruber
Physics Department, Southern Illinois University, Carbondale, Illinois 62901
(Received 16 February 1971; Revised Manuscript Received 29 March 1972)

An explicit classification of the semisimple complex Lie subalgebras of the simple complex Lie algebras is given for algebras up to rank 6. The notion of defining vector, introduced by Dynkin and valid for subalgebras of rank 1, has been extended to the notion of defining matrix, valid for any semisimple subalgebra. All defining matrices have been determined explicitly, which is equivalent to the determination of the embeddings of the generators of the Cartan subalgebra of a semisimple subalgebra in the Cartan subalgebra K of the simple algebra containing this subalgebra. Moreover, the embedding of the root system of the subalgebras in the dual space K^* of an algebra is given for all subalgebras. For the S-subalgebras of the simple algebras (up to rank 6), the embedding of the whole subalgebra in an algebra is given explicitly. In addition, the decomposition (branching) of the defining (fundamental) and adjoint representations of an algebra with respect to the restriction to its S-subalgebras has been determined. In the first part of this article a brief review of Dynkin's theory of the classification of the semisimple Lie subalgebras of the simple Lie algebras is given. No proofs are repeated, and at places where concepts have been extended and new results derived, merely an indication for their proof is given. This part of the article will serve as a prescription for a classification of semisimple subalgebras of the simple Lie algebras of rank exceeding 6. Later in the article, explicit expressions are given for the index of an embedding of a simple Lie subalgebra in a simple Lie algebra. These expressions are valid for the classical Lie algebras of arbitrary rank as well as for the exceptional Lie algebras.

1. INTRODUCTION

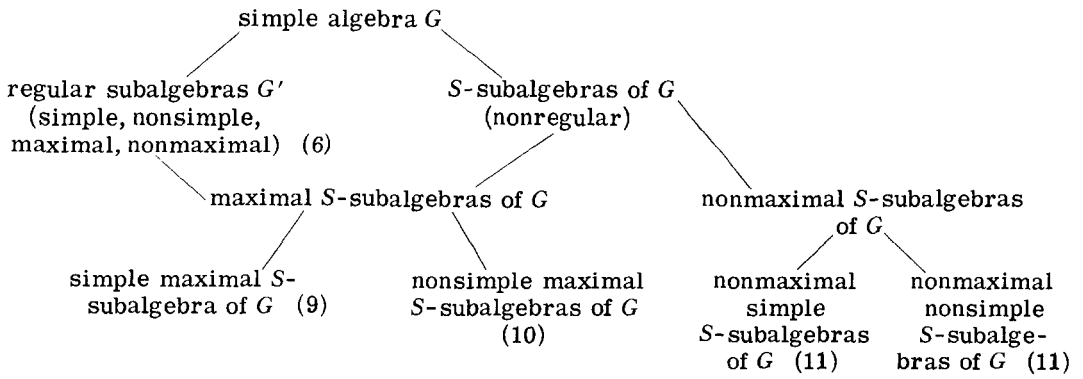
The semisimple Lie algebras and their linear representations play an important role in many branches of physics. The states of a physical system may be classifiable as states of some reducible or irreducible representation of a simple or semisimple Lie algebra. This happens in the case of the familiar angular momentum group $SO(3)$, as a consequence of the invariance of a physical system under rotations in ordinary 3-space, as well as for many other groups, which may leave invariant only part of the Hamiltonian of the system. Examples of groups of this type can be found, for instance, in nuclear physics,[1-3] atomic spectroscopy,[4] and elementary particle physics.[5,6]

If a particular Lie algebra is utilized in some model, two things may happen. On the one hand, it may become essential to know the Lie subalgebras of this Lie algebra (for example, in order to obtain physically meaningful labels for the states, as it happens in nuclear and atomic physics). On the other hand, it may become of interest to extend the symmetry to larger symmetries (as in the example of the extension of isotopic spin to unitary spin in particle physics), in which case the original Lie algebra becomes a subalgebra.

In recent years more and more Lie algebras and, for a given Lie algebra, various chains of subalgebras, have come into use in atomic and nuclear physics.[2-4] Moreover, the embedding of Lie algebras in the algebras of higher rank is of importance for both the search of larger symmetries[5] as well as for the (mathematical) state labeling problem.[7] It seems, therefore, to be justified to give an *explicit* classification of the semisimple Lie subalgebras of the simple Lie algebras, up to some reasonable rank (≤ 6), serving as a sort of catalog of all semisimple Lie subalgebras of a simple Lie algebra. Apart from a mere classification of the Lie subalgebras, their embedding in the Lie algebra is given explicitly, which should make this classification even more valuable. In view of Lie subalgebras of simple Lie algebras of rank >6, a brief resumé of Dynkin's theory is given, supplemented with some extensions of his concepts (the introduction of the defining matrix), in order to serve as a prescription for the classification of the semisimple Lie subalgebras of those simple Lie algebras not considered in this article.

The tables contained in this article are arranged in such a manner that the information they contain is given in "building blocks." That is, whenever the subalgebras of some simple algebra are to be classified, this classification can be achieved by putting together the various subalgebras obtained from the tables according to the rules given below. This has the advantage that the same tables can also be used (together with additional information) for the classification of the subalgebras of simple algebras with rank >6.

Below a scheme is given (with the definitions to



follow later) for the determination of all the semisimple subalgebras of a simple algebra. The numbers refer to the sections which treat the particular kind of subalgebra.

The scheme indicates that, in the first step, the subalgebras are separated into regular subalgebras and S-subalgebras *of* G. The part to the right then gives a classification of all S-subalgebras of G. This same classification has to be applied to each one of the regular subalgebras G' (in fact, G itself is a trivial case of a regular subalgebra). That is, for given regular subalgebra G', all subalgebras of G' have to be found which are S-subalgebras with respect to G', and this has to be done for every regular subalgebra G'. Proceeding in this manner all semisimple subalgebras of G are obtained. The S-subalgebras of a proper regular subalgebra G' of G are called R-subalgebras of G.

Some work previously done and related to ours can be found in Refs. 8–12.

2. SEMISIMPLE LIE ALGEBRAS

Let G be a semisimple complex Lie algebra. For every Cartan subalgebra K of the algebra G there is a canonical decomposition

$$G = K + \sum_{\alpha\in\Sigma} G_\alpha, \tag{2.1}$$

where Σ is the system of roots of the algebra G and G_α a one-dimensional root subspace. Every element X of the algebra G can be expressed in terms of the generators H_i of the Cartan subalgebra K and of the roots vectors E_α,

$$X = \sum_{i=1}^{n} b_i H_i + \sum_{\alpha\in\Sigma} c_\alpha E_\alpha, \qquad b_i, c_\alpha \in C. \tag{2.2}$$

In a semisimple algebra G, a *scalar product* can be introduced by

$$(X,Y) = \mathrm{Tr}[\mathrm{ad}(X)\mathrm{ad}(Y)], \qquad X,Y \in G, \tag{2.3}$$

where $\mathrm{ad}(X)$ is the adjoint representation of the element X. This scalar product is invariant with respect to all automorphisms of G. Similarly, for any representation ϕ of G a scalar product can be defined by

$$(X,Y)_1 = \mathrm{Tr}[\phi(X)\phi(Y)], \qquad X,Y \in G. \tag{2.4}$$

Both scalar products are related by a numerical factor

$$(X,Y)_1 = (X,Y) l_\phi, \tag{2.5}$$

where l_ϕ is *independent* of the elements X and Y and is called the *index of the linear representation* ϕ of G. Its value is given by[13]

$$l_\phi = [N(\phi)/N(G)](M, M + 2R_0), \tag{2.6}$$

where $N(\phi)$ is the dimension of the representation ϕ, $N(G)$ the dimension of the algebra G, M the highest weight of the representation ϕ, and $2R_0$ the sum of all positive roots of G. The indices of the fundamental representations of the simple algebras are given in Ref. 14.

Given a Cartan subalgebra K of G there exists its *dual space* K^*, consisting of all linear forms $\rho(H)$, $H \in K$. There is a one-to-one correspondence between the elements H_ρ and $\rho(H)$ of the two spaces K and K^* defined by

$$(H_\rho, H) = \rho(H), \qquad H, H_\rho \in K, \ \rho(H) \in K^*.$$

From this isomorphism it follows that a scalar product can be defined in K^* by setting

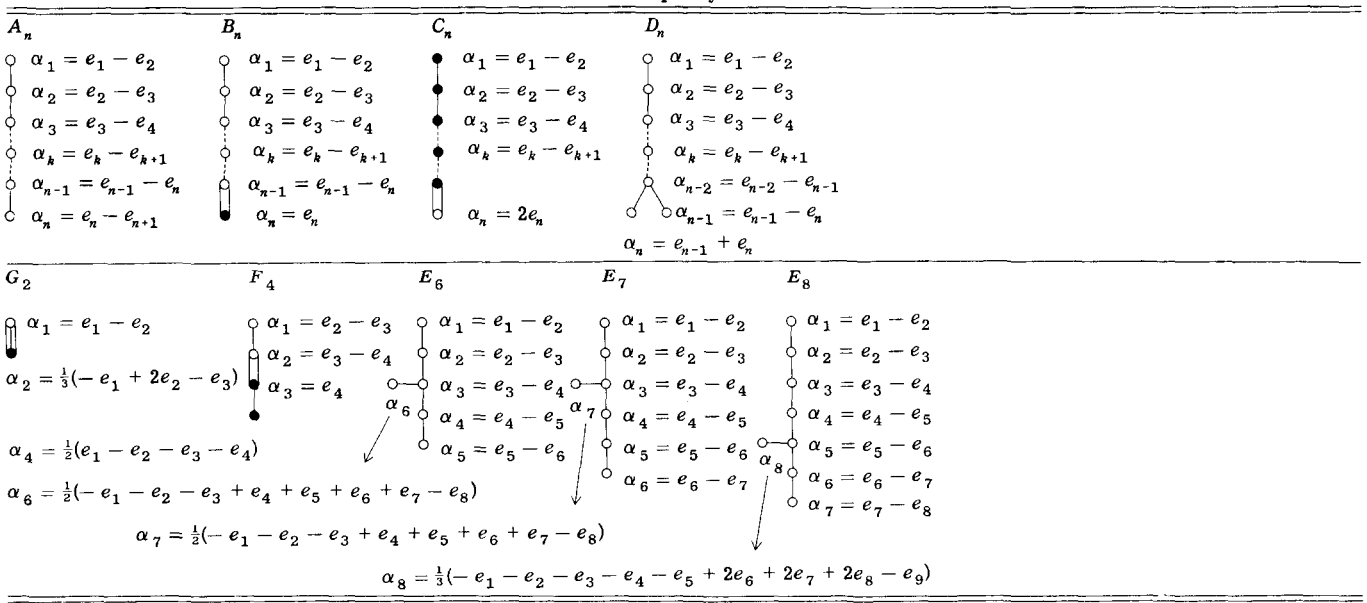
$$(\rho(H), \sigma(H)) = (H_\rho, H_\sigma), \tag{2.7}$$

where H_ρ, H_σ are the elements of K which correspond to the elements $\rho(H), \sigma(H)$ of K^*. From the automorphism of the spaces K and K^* it also follows that we may regard an element $\rho(H) \in K^*$ as an element ρ of the space K by means of the formula

$$\rho(H) = (\rho, H),$$

which expresses the linear form $\rho(H)$ as scalar pro-

TABLE I. Simple systems of roots.



duct between the elements $\rho, H \in K$. Subsequently $\rho(H)$ will be considered both as an element of K^* as well as an element $\rho \in K$.

The roots and weights belong to the space K^* and can be expanded with respect to some basis (for example, a π-system of simple roots[15]). Having chosen some basis, an ordering of the elements of K^* can be introduced with respect to this basis. If $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ represents the basis chosen, the vector

$$\rho = r_i \alpha_i$$

is called positive, and we write $\rho > 0$, if the first nonvanishing component r_i is positive. In particular the elements of the basis are positive. [With this definition one of the following alternatives holds for any element ρ of K^*: Either ρ is positive or $(-\rho)$ is positive or ρ is zero.] We say the element σ is higher than the element ρ, and write $\sigma > \rho$, if $\sigma - \rho > 0$. The element ρ is then said to be lower than the element σ.

The scalar product in K is defined up to an arbitrary constant. Generally this constant is fixed with the condition that the square of the length of the maximal root be equal to 2. With this normalization and using an orthonormal basis for K and K^*, we have

$$(\rho, \sigma) = (H_\rho, H_\sigma) = \sum_{i=1}^{n} r_i s_i, \tag{2.8}$$

where r_i and s_i are the cartesian components of ρ, σ and H_ρ, H_σ, respectively. For the algebra C_n, the scalar product is defined as (the square of the length of a maximal root for C_n is 4)

$$(\rho, \sigma) = (H_\rho, H_\sigma) = \tfrac{1}{2} \sum_{i=1}^{n} r_i s_i . \tag{2.9}$$

Throughout this article we use cartesian coordinates for the roots and weights. In Table I a representation is given for the systems of simple roots.

3. EMBEDDING OF A SUBALGEBRA

A faithful embedding f of an algebra $\tilde{G}$ in an algebra G is defined by an isomorphic mapping f of $\tilde{G}$ into G,

$$\tilde{X} \to f(\tilde{X}) \in G, \quad \text{for every} \quad \tilde{X} \in \tilde{G},$$
$$\text{such that} \quad f([\tilde{X}, \tilde{Y}]) = [f(\tilde{X}), f(\tilde{Y})]. \tag{3.1}$$

The image of $\tilde{X}$ can be expressed in terms of the generators of G

$$f(\tilde{X}) = \sum_{k=1}^{n} b_k H_k + \sum_{\alpha \in \Sigma} c_\alpha E_\alpha, \tag{3.2}$$

where the H_k denote an orthonormal basis in the Cartan subspace of G and E_α the root vectors of the subspaces G^α.

Two embeddings f_1 and f_2 of the algebra $\tilde{G}$ in G are called equivalent if there is an inner automorphism U of the algebra G such that

$$f_2(\tilde{X}) = U f_1(\tilde{X}), \quad \tilde{X} \in \tilde{G}. \tag{3.3}$$

In order to classify the semisimple Lie subalgebras of the simple Lie algebras, all the (inequivalent) classes of faithful embeddings of the subalgebras $\tilde{G}$ into the algebra G have to be found.

Given some embedding f of $\tilde{G}$ in G, both simple algebras, the relation

$$(f(\tilde{X}), f(\tilde{Y})) = j_f(\tilde{X}, \tilde{Y}), \quad \tilde{X}, \tilde{Y} \in \tilde{G}, \tag{3.4}$$

determines a scalar factor j_f *independent* of $\tilde{X}, \tilde{Y}$. Under an inner automorphism of G this factor is invariant, and thus is the same for all the equivalent embeddings of the algebra $\tilde{G}$ in G. For this reason this factor can be used to label the different classes of inequivalent embeddings. The factor j_f is called the *index of the simple subalgebra* $\tilde{G}$ *in the simple* algebra G. Some of the properties of this index are[16]

(i) j_f is an integer number.

(ii) If $G_1 \subseteq G_2 \subseteq G_3$ are simple algebras, the index of G_1 in G_3 is the product of the indices of G_1 in G_2 and G_2 in G_3.

(iii) If $f_1, f_2, \ldots, f_s$ are embeddings of a simple algebra $\tilde{G}$ in the simple algebra G and if

$$[f_i(\tilde{X}), f_j(\tilde{Y})] = 0 \quad \text{for} \quad i \neq j, \ \tilde{X}, \tilde{Y} \in \tilde{G},$$

then $f_1 + f_2 + \cdots + f_s$ is again an embedding and

$$j_f = j_{f_1} + j_{f_2} + \cdots + j_{f_s}.$$

(iv) Given some embedding f of the algebra $\tilde{G}$ in G, some linear representation ϕ of the algebra G and the representation $\tilde{\phi} = \phi f$ of $\tilde{G}$ which is induced by ϕ on the subalgebra $\tilde{G}$, the index is

$$j_f = l_{\phi f} / l_\phi, \tag{3.5}$$

where $l_{\phi f}, l_\phi$ are the indices of the representation $\phi f, \phi$, respectively.

From a theorem by Gantmacher[17] it follows that for any embedding of $\tilde{G}$ in G it is always possible to find a Cartan subalgebra K of G such that $f(\tilde{K}) \subset K$, where $\tilde{K}$ is a Cartan subalgebra of $\tilde{G}$. In particular

$$f(\tilde{H}_i) = \sum_{k=1}^{n} f_{ik} H_k, \quad i = 1, 2, \ldots, n', \ k = 1, 2, \ldots, n, \tag{3.6}$$

where $\tilde{H}_i$ and H_k are a basis for $\tilde{K}$ and K. Since this basis is orthonormal, we have

$$\sum_{k=1}^{n} f_{ik} f_{mk} = j_f \delta_{im}. \tag{3.7}$$

If G is restricted to a subalgebra $\tilde{G}$, the elements of the space K will be mapped through an orthogonal projection onto elements of the subspace $\tilde{K}$,

$$H \to f^*(H) \in \tilde{K}.$$

For two arbitrary elements $H \in K$ and $\tilde{H} \in \tilde{K}$, it holds that[18]

$$(H, f(\tilde{H})) = (f^*(H), \tilde{H}), \tag{3.8}$$

with

$$f^*(H_i) = \sum_{k=1}^{n'} f_{ki} \tilde{H}_k . \tag{3.9}$$

From the isomorphism between the Cartan space and its dual space it follows that

$$(\rho, f(\rho')) = (f^*(\rho), \rho') \tag{3.10}$$

for two arbitrary elements $\rho \in \tilde{K}^*$ and $\rho' \in \tilde{K}^*$. If ρ and ρ' are expressed in Cartesian coordinates, (3.10) reads

$$\sum_{ik} r_k (r_i' f_{ik}) = \sum_{ik} (f_{ik} r_k) r_i', \tag{3.11}$$

where f_{ik} is given by (3.6). A similar expression has been obtained in Ref. 19 for the relationship of the parameters of the toroid of G and its subalgebra $\tilde{G}$.

From a theorem by Dynkin[20] it follows that if $\Gamma_{\alpha'}$ is the set of roots of G which project onto a root α' of $\tilde{G}$, the embedding of the elements $\tilde{E}_{\alpha'}$ of $\tilde{G}$ in G is given as

$$f(\tilde{E}_{\alpha'}) = \sum_{\alpha \in \Gamma_{\alpha'}} c_{\alpha'\alpha} E_\alpha, \qquad E_\alpha \in G_\alpha. \tag{3.12}$$

Thereby the following properties hold:

(i) For two roots α', β' of $\tilde{G}$, $\Gamma_{\alpha'} \cap \Gamma_{\beta'} = 0$.

(ii) For a root α' and its negative $-\alpha'$ we have $\Gamma_{-\alpha'} = -\Gamma_{\alpha'}$.

(iii) $\bar{c}_{\alpha'\alpha} = c_{-\alpha',-\alpha}$ for arbitrary α', α, where $\bar{c}$ is the complex conjugate of c.

(iv) $f(\alpha') = \sum_{\alpha \in \Gamma_{\alpha'}} |c_{\alpha'\alpha}|^2 \alpha$.

(v) $j_f = \sum_{\alpha \in \Gamma_{\alpha'}} |c_{\alpha\alpha'}|^2$.

Collecting the coefficients f_{ik} and $c_{\alpha\alpha'}$ in a matrix U representing the embedding of $\tilde{G}$ in G, it holds

$$UU^\dagger = j_f 1.$$

4. CONDITIONS FOR A SUBALGEBRA

Let G be a semisimple algebra with Cartan subspace K, and let $H_{\alpha'}, H_{\beta'}, \ldots, H_{\delta'}$ be a set of linearly independent elements of K. Suppose that some elements $E_{\alpha'}, E_{-\alpha'}, \ldots, E_{\delta'}, E_{-\delta'}$ of the algebra G exist which satisfy the relations

(i) $[H_{\alpha'}, H_{\beta'}] = 0$,
(ii) $[H_{\alpha'}, E_{\pm\beta'}] = \pm (\alpha', \beta') E_{\pm\beta'}$
(iii) $[E_{\alpha'}, E_{-\alpha'}] = H_{\alpha'}$,
(iv) $[E_{\alpha'}, E_{-\beta'}] = 0, \ \alpha' \neq \beta'$.

} for all $\alpha', \beta', \ldots, \delta'$.

Then the minimal subalgebra $\tilde{G}$ of G which contains these elements is semisimple and the set of roots $\alpha', \beta', \ldots, \delta'$ forms a possible system of simple roots for $\tilde{G}$ with $E_{\alpha'}, E_{-\alpha'}, \ldots, E_{\delta'}, E_{-\delta'}$ as the corresponding root vectors.[21]

The conditions for a subalgebra as given above are equivalent to:

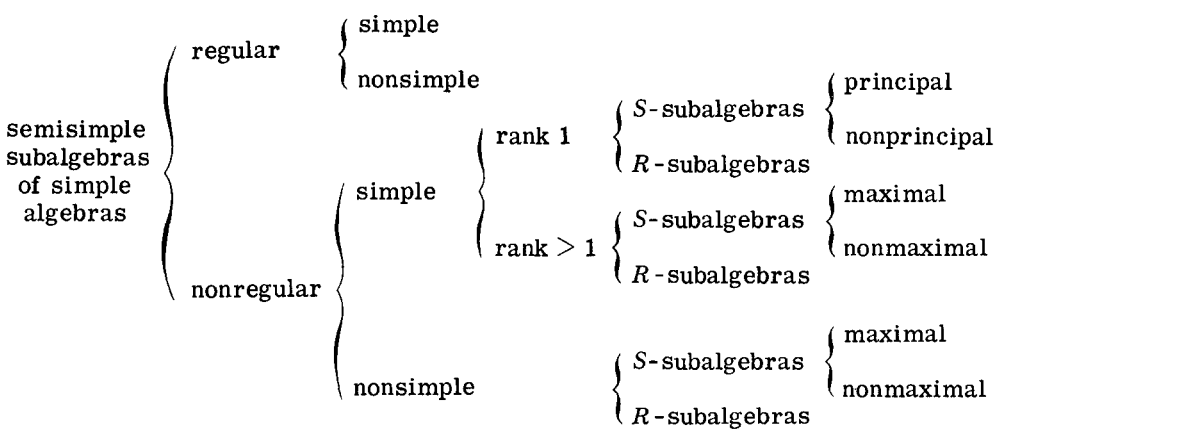
(i) There exists at least one representation ϕ of G whose weights project onto the weights of some representation $\tilde{\phi}$ of $\tilde{G}$.

(ii) For each simple root α' of $\tilde{G}$ there exists one or several roots α of G which project onto α'.

(iii) The root α' must be expressible as a linear combination of those roots α which project onto α'. The commutator $[E_{\alpha'}, E_{-\alpha'}]$ must belong to the Cartan space $\tilde{K}$ for all α'.

5. CLASSIFICATION OF SUBALGEBRAS

(a) Let G be a *simple* algebra. According to Dynkin's theory,[22] all the semisimple subalgebras of G belong to one of the following types of subalgebras:



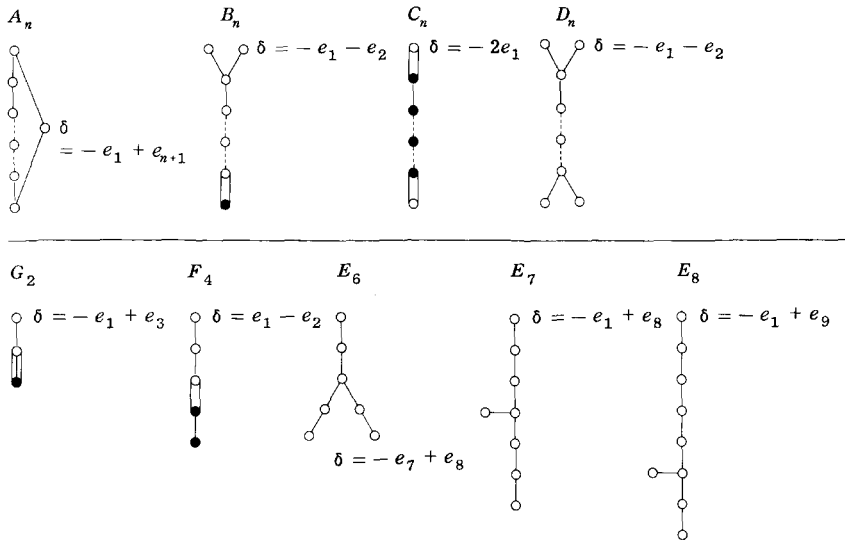
(b) Let G be a *semisimple* algebra and let

$$G = G_1 + G_2 + \cdots + G_s$$

be a decomposition of G into a direct sum of simple ideals. Suppose we know already all maximal semisimple subalgebras—regular as well as nonregular—of each simple algebra G_k, $k = 1, 2, \ldots, s$. Then all maximal *semisimple* subalgebras of the *semisimple* algebra G can be found following rules given in Sec. 11. Now let $\tilde{G}$ be one of these maximal *semisimple* subalgebras of the algebra G. In general $\tilde{G}$ will not be simple and in turn its semisimple maximal subalgebras will be found by following the rules given in Sec. 11. Obviously, these maximal subalgebras of $\tilde{G}$ are subalgebras of G, though not maximal subalgebras. Following this procedure all semisimple subalgebras of G will be obtained.

From the above it follows that all semisimple subalgebras of a semisimple algebra can be determined once all semisimple subalgebras of the simple algebras are known. Therefore, this article is limited to the determination of all semisimple subalgebras of the simple algebras.

TABLE II. Extension of the simple systems of roots.



6. REGULAR SUBALGEBRAS

Let G be a semisimple algebra. A subalgebra $\tilde{G}$ of the algebra G is called regular if there is a Cartan subspace K of G, such that for

$$\tilde{G} = \tilde{K} + \sum_{\alpha' \in \tilde{\Sigma}} \tilde{G}_{\alpha'},$$

the relations $\tilde{K} \subseteq K$ and $\tilde{\Sigma} \subseteq \Sigma$ hold.

The subalgebra $\tilde{G}$ is semisimple if the following conditions are satisfied:

(i) If $\alpha \in \tilde{\Sigma}$ then $-\alpha \in \tilde{\Sigma}$;
(ii) $\tilde{K}$ is the linear closure of $\tilde{\Sigma}$.

In order to construct the regular semisimple subalgebras it is more convenient to work with the systems of simple roots[15] (π-systems) due to the following theorem:[23]

Let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be a π-subsystem of roots of a semisimple algebra G. Let $\tilde{G}$ be the minimal subalgebra of G which contains the root vectors E_{α_1}, $E_{\alpha_2}, \ldots, E_{\alpha_m}, E_{-\alpha_1}, \ldots, E_{-\alpha_m}$. Then the subalgebra $\tilde{G}$ is a regular semisimple subalgebra and the system $\alpha_1, \alpha_2, \ldots, \alpha_m$ is a possible system of simple roots for $\tilde{G}$.

In order to find all the possible subsystems of Σ which are π-systems the following method can be used [24]:

(1) Let G be a simple algebra of rank n; its system of simple roots will be equivalent to one of the diagrams of Table I. Let us adjoin the lowest root (with respect to the ordering in K^*, Sec. 2) to this system. If the representation of the simple roots is that of Table I, the coordinates of the lowest root δ will be those given in Table II.

(2) Remove arbitrarily one of the roots $\alpha_i (i = 1, 2, \ldots, n)$ from an extended diagram. We will obtain at most n different diagrams corresponding to π-systems, which may split into mutually orthogonal subsystems.

(3) For each of the diagrams obtained in (2) apply step (1) to its nonsplitting *subsystems* of systems of simple roots. Repeat the same process until no new π-systems of n elements are obtained. The π-systems obtained in this manner are called maximal π-systems.

(4) From each maximal π-system remove arbitrarily $m < n$ roots in order to obtain π-systems of $n - m$ elements. All the possible π-systems which are subsystems of the system of roots Σ of G are exhausted by this method. Among the π-systems obtained in this manner there may be some which are transformed into each other by the Weyl group; in this case the corresponding subalgebras will be conjugate.

The *maximal* π-systems for the classical algebras up to rank 6 obtained in the manner described above are given in Table III. The maximal π-systems for the exceptional algebras are given in Table 10 of Ref. 13.

All the regular semisimple subalgebras for the classical algebras up to rank 6 are obtained by applying step (4) to the maximal π-systems. The result is listed in Table IV. To each type of subalgebra listed in this table corresponds one class of conjugate regular subalgebra, except in the case of D_n and for the case of subalgebras of the type $2A_1$ and A_3 in D_4; $3A_1, A_1 + A_3$, and A_5 in D_6: To every such type correspond two classes of conjugate subalgebras, transformable into each other by an outer automorphism of D_n. For the exceptional algebras, the regular subalgebras are given in Table 11 of Ref. 13. (In Table 4 we use the notation $2A_1$ for $A_1 + A_1$, $3A_1$ for $A_1 + A_1 + A_1$, etc. The index of the regular subalgebra is given for each simple ideal with respect to the simple algebra; this index is always 1, except when indicated explicitely in the form of a raised index, for example A_1^2).

The embeddings of the regular subalgebras $\tilde{G}$ in a simple algebra G are given in Table V. Also in Table V are listed the mappings f^* of the weights m of the algebras G onto weights m' of their regular subalgebras $\tilde{G}$.

TABLE III. Maximal π-systems of simple algebras.

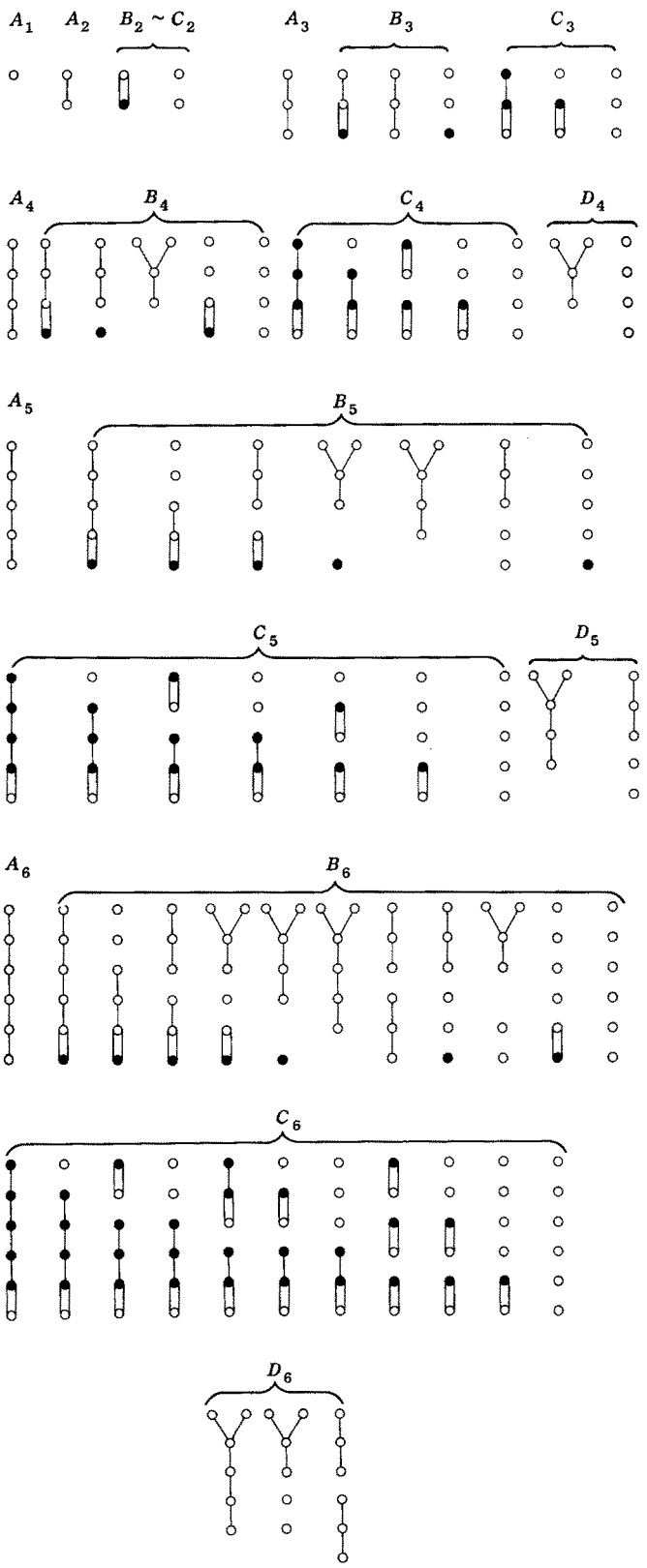


TABLE IV.
Regular semisimple subalgebras of classical algebras up to rank 6.

A_2	$B_2 \sim C_2$	$A_3 \sim D_3$	B_3	C_3
A_1	$2A_1$	A_2	A_3	$C_2 + A_1$
	A_1	$2A_1$	$2A_1 + A_1^2$	$3A_1$
	A_1^2	A_1	A_2	A_2^2
			B_2	C_2
			$2A_1$	$2A_1$
			$A_1 + A_1^2$	$A_1 + A_1^2$
			A_1	A_1
			A_1^2	A_1^2

A_4	B_4	C_4	D_4
A_3	D_4	$C_3 + A_1$	$4A_1$
$A_2 + A_1$	$A_3 + A_1^2$	$2C_2$	A_3
A_2	$B_2 + 2A_1$	$C_2 + 2A_1$	$3A_1$
$2A_1$	$4A_1$	$4A_1$	A_2
A_1	A_3	A_3^2	$2A_1$
	B_3	C_3	A_1
	$A_2 + A_1^2$	$A_2^2 + A_1$	
	$B_2 + A_1$	$C_2 + A_1$	
	$3A_1$	$C_2 + A_1^2$	
	$2A_1 + A_1^2$	$3A_1$	
	A_2	$2A_1 + A_1^2$	
	B_2	A_2	
	$2A_1$	C_2	
	$A_1 + A_1^2$	$2A_1$	
	A_1	$A_1 + A_1^2$	
	A_1^2	$2A_1^2$	
		A_1	
		A_1^2	

A_5	B_5		
A_4	D_5	$A_3 + A_1^2$	$A_2 + A_1^2$
$A_3 + A_1$	$D_4 + A_1^2$	$B_3 + A_1$	$B_2 + A_1$
$2A_2$	$A_3 + B_2$	$A_2 + B_2$	$3A_1$
A_3	$A_3 + 2A_1$	$A_2 + 2A_1$	$2A_1 + A_1^2$
$A_2 + A_1$	$B_3 + 2A_1$	$B_2 + 2A_1$	A_2
$3A_1$	$4A_1 + A_1^2$	$4A_1$	B_2
A_2	A_4	$3A_1 + A_1^2$	$2A_1$
$2A_1$	B_4	A_3	$A_1 + A_1^2$
A_1	D_4	B_3	A_1
	$A_3 + A_1$	$A_2 + A_1$	A_1^2

	C_5		D_5
$C_4 + A_1$	$A_2^2 + 2A_1$	$3A_1$	$A_3 + 2A_1$
$C_3 + C_2$	$2C_2$	$2A_1 + A_1^2$	A_4
$C_3 + 2A_1$	$C_2 + 2A_1$	$A_1 + 2A_1^2$	D_4
$2C_2 + A_1$	$C_2 + A_1 + A_1^2$	A_2^2	$A_3 + A_1$
$C_2 + 3A_1$	$4A_1$	C_2	$A_2 + 2A_1$
$5A_1$	$3A_1 + A_1^2$	$2A_1$	$4A_1$
A_4^2	A_3^2	$A_1 + A_1^2$	A_3
C_4	C_3	$2A_1^2$	$A_2 + A_1$
$A_3^2 + A_1$	$A_2^2 + A_1$	A_1	$3A_1$
$C_3 + A_1$	$A_2^2 + A_1^2$	A_1^2	A_2
$C_3 + A_1^2$	$C_2 + A_1$		$2A_1$
$A_2^2 + C_2$	$C_2 + A_1^2$		A_1

TABLE IV contd

A_6	B_6		
A_5	D_6	$A_3 + B_2$	$A_2 + A_1 + A_1^2$
$A_4 + A_1$	$D_5 + A_1^2$	$A_3 + 2A_1$	$B_2 + 2A_1$
$A_3 + A_2$	$B_4 + 2A_1$	$A_3 + A_1 + A_1^2$	$4A_1$
A_4	$D_4 + B_2$	$B_3 + A_2$	$3A_1 + A_1^2$
$A_3 + A_1$	$D_4 + 2A_1$	$B_3 + 2A_1$	A_3
$2A_2$	$2A_3$	$A_2 + 2A_1 + A_1^2$	B_3
$A_2 + 2A_1$	$A_3 + B_3$	$B_2 + 3A_1$	$A_2 + A_1$
A_3	$A_3 + 2A_1 + A_1^2$	$4A_1 + A_1^2$	$A_2 + A_1^2$
$A_2 + A_1$	$B_2 + 4A_1$	$5A_1$	$B_2 + A_1$
$3A_1$	$6A_1$	A_4	$3A_1$
A_2	A_5	B_4	$2A_1 + A_1^2$
$2A_1$	B_5	D_4	A_2
A_1	D_5	$A_3 + A_1$	B_2
	$A_4 + A_1^2$	$A_3 + A_1^2$	$2A_1$
	$B_4 + A_1$	$B_3 + A_1$	$A_1 + A_1^2$
	$D_4 + A_1$	$2A_2$	A_1
	$D_4 + A_1^2$	$A_2 + B_2$	A_1^2
	$A_3 + A_2$	$A_2 + 2A_1$	

C_6			D_6
$C_5 + A_1$	$A_2^2 + C_2 + A_1$	$C_2 + 2A_1^2$	$D_4 + 2A_1$
$C_4 + C_2$	$A_2^2 + 3A_1$	$4A_1$	$2A_3$
$C_4 + 2A_1$	$2C_2 + A_1$	$3A_1 + A_1^2$	A_5
$2C_3$	$2C_2 + A_1^2$	$2A_1 + 2A_1^2$	D_5
$C_3 + C_2 + A_1$	$C_2 + 3A_1$	A_3^2	$D_4 + A_1$
$C_3 + 3A_1$	$C_2 + 2A_1 + A_1^2$	C_3	$A_3 + A_2$
$3C_2$	$5A_1$	$A_2^2 + A_1$	$A_3 + 2A_1$
$2C_2 + 2A_1$	$4A_1 + A_1^2$	$A_2^2 + A_1^2$	$5A_1$
$C_2 + 4A_1$	A_4^2	$C_2 + A_1$	A_4
$6A_1$	C_4	$C_2 + A_1^2$	D_4
A_5^2	$A_3^2 + A_1$	$3A_1$	$A_3 + A_1$
C_5	$A_3^2 + A_1^2$	$2A_1 + A_1^2$	$2A_2$
$A_4^2 + A_1$	$C_3 + A_1$	$A_1 + 2A_1^2$	$A_2 + 2A_1$
$C_4 + A_1$	$C_3 + A_1^2$	$3A_1^2$	$4A_1$
$C_4 + A_1^2$	$2A_2^2$	A_2^2	A_3
$A_3^2 + C_2$	$A_2^2 + C_2$	C_2	$A_2 + A_1$
$A_3^2 + 2A_1$	$A_2^2 + 2A_1$	$2A_1$	$3A_1$
$C_3 + A_2^2$	$A_2^2 + A_1 + A_1^2$	$A_1 + A_1^2$	A_2
$C_3 + C_2$	$2C_2$	$2A_1^2$	$2A_1$
$C_3 + 2A_1$	$C_2 + 2A_1$	A_1	A_1
$C_3 + A_1 + A_1^2$	$C_2 + A_1 + A_1^2$	A_1^2	

TABLE Va.
Embeddings f of the regular subalgebras in a simple algebra.

$f(\tilde{E}_\alpha) = E_\alpha, \alpha \in \tilde{\Sigma}$

$f(\tilde{H}_i) = H_i, i = 1, 2, \ldots, n'$, for $G = B_n, C_n, D_n$
$\tilde{G} = B_{n'}, C_{n'}, D_{n'}$

$f(\tilde{H}_i) = H_i - [1/(n'+1)](H_1 + H_2 + \cdots + H_{n'+1}), i = 1, 2, \ldots, n'+1,$
for $G = B_n, C_n, D_n; \tilde{G} = A_{n'}$

$f(\tilde{H}_i) = H_i + [1/(n'+1)](H_{n'+2} + H_{n'+3} + \cdots + H_{n'+1}), i = 1, 2, \ldots,$
$n'+1$, for $G = A_n; \tilde{G} = A_{n'}$

with $n' \leq n$.

TABLE Vb. Mappings $f^*(m) = m'$ for subalgebras of Table Va.

$G = B_n, C_n, D_n; \tilde{G} = B_{n'}, C_{n'}, D_{n'}$:
$m_i' = m_i, i = 1, 2, \ldots, n',$

$G = B_n, C_n, D_n; \tilde{G} = A_{n'}$
$m_i' = m_i - [1/(n'+1)](m_1 + m_2 + \cdots + m_{n'+1}), i = 1, 2, \ldots, n'+1$

$G = A_n; \tilde{G} = A_{n'}$:
$m_i' = m_i + [1/(n'+1)](m_{n'+2} + m_{n'+3} + \cdots + m_{n+1}),$
$i = 1, 2, \ldots, n'+1,$

with $n' \leq n$.

7. THREE-DIMENSIONAL SUBALGEBRAS

The three algebras A_1, B_1, and C_1 are isomorphic. Their generators $\tilde{H}, \tilde{E}_+$, and $\tilde{E}_-$ satisfy the following commutation relations:

$$[\tilde{H}, \tilde{E}_+] = 2\tilde{E}_+, \quad [\tilde{H}, \tilde{E}_-] = -2\tilde{E}_-, \quad [\tilde{E}_+, \tilde{E}_-] = \tilde{H}. \tag{7.1}$$

We assume the algebra $\tilde{G}$, given by these commutation relations, to be a subalgebra of a semisimple algebra G. Let f denote the embedding of $\tilde{G}$ in G. Then, if $f(\tilde{H}) \in K$,

$$f(\tilde{H}) = \sum_{k=1}^{n} f_k H_k, \quad H_k \in K,$$

where the n elements H_k are assumed to form a basis for the vector space K. The vector

$$f = (f_1, f_2, \ldots, f_n), \quad f \in K,$$

is called the *defining vector* of the three-dimensional (3d-) subalgebra $\tilde{G}$. Considering the linear forms of the dual space K^* as elements of the space K (Sec. 2), the only positive root α of the 3d-subalgebra is in the space K given as

$$\alpha = [2f/(f,f)] \tag{7.2}$$

and vice versa

$$f = [2\alpha/(\alpha,\alpha)]. \tag{7.3}$$

The name defining vector stems from the following property[25]: Let G_1, G_2 be 3d-subalgebras of a semisimple algebra G, and let f_1, f_2 be the corresponding defining vectors. A necessary and sufficient condition that there exists an inner automorphism of G transforming G_1 into G_2 is that there exists an inner automorphism of G transforming f_1 into f_2. Thus the problem of finding all classes of conjugate 3d-algebras is equivalent to finding all defining vectors which are inequivalent with respect to the Weyl group.

Let us consider the Dynkin diagram (Table I) for the simple roots α_i of an algebra G and attach to each dot of the diagram (representing one of the simple roots) the value (f, α_i), where f is the *highest* defining vector with respect to some ordering in the Cartan subspace K of G. The resulting diagram (with numbers attached) is called the characteristic diagram of the 3d-subalgebra corresponding to f. A necessary and sufficient condition that two 3d-subalgebras of a semisimple algebra are conjugate is that their characteristic diagrams coincide.[26]

The defining vector f yields the following information:

(i) The embedding of the generator H of $\tilde{G}$ in G:

$$f(\tilde{H}) = \sum_{k=1}^{n} f_k H_k, \quad H_k \in G. \tag{7.4}$$

(ii) The mapping of the weights of a representation ϕ of G into the Cartan subspace $\tilde{K}$ of the 3d-subalgebra $\tilde{G}$:

$$f^*(m) = \frac{1}{2} \sum_{k=1}^{n} f_k m_k \tag{7.5}$$

(iii) The index j_f of the algebra $\tilde{G}$ in G:

$$j_f = \frac{1}{2} \sum_{k=1}^{n} f_k f_k,$$

$$= \sum_{k=1}^{n} f_k f_k, \quad \text{if } G = C_n. \tag{7.6}$$

(iv) The embedding of the positive root α' of $\tilde{G}$ in G:

$$f(\alpha') = f.$$

(v) The characteristic diagram of $\tilde{G}$ with respect to some ordering in G.

In deriving the properties listed above the fact has been used that $(\alpha', \alpha') = (\tilde{H}, \tilde{H}) = 2$. Moreover, the factor $\frac{1}{2}$ arising for the case $G = C_n$ due to Eq. (2.9), has been absorbed into the defining vector f.

For the classification of the nonregular semisimple subalgebras it is necessary to define the *R-subalgebras* and *S-subalgebras*. Let G be a semisimple algebra. An R-system is a subset of the elements of G such that these elements are contained in a proper regular subalgebra of G, that is in a regular subalgebra which is neither the entire algebra nor the null algebra. An S-system is a subset of the elements of G which is not a R-system. An R-subalgebra is an algebra which is an R-system. An S-subalgebra is an algebra which is an S-system. Every semisimple R-subalgebra is contained in some semisimple proper regular subalgebra G' of the algebra G.

Three-Dimensional S-Subalgebras[27]

(a) *Principal:* The highest defining vector satisfies $(f, \alpha_i) = 2$ for all simple roots α_i of G. If f is expressed in terms of the simple roots of G,

$$f = \sum_{\alpha \in \pi} |c_\alpha|^2 \alpha, \tag{7.7}$$

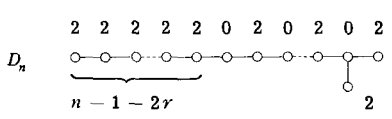
the embedding of the generators of $\tilde{G}$ in G is given by

$$f(\tilde{H}) = \sum_{\alpha \in \pi} |c_\alpha|^2 H_\alpha, \quad f(\tilde{E}_+) = \sum_{\alpha \in \pi} c_\alpha E_\alpha,$$

$$f(\tilde{E}_-) = \sum_{\alpha \in \pi} \bar{c}_\alpha E_{-\alpha}. \tag{7.8}$$

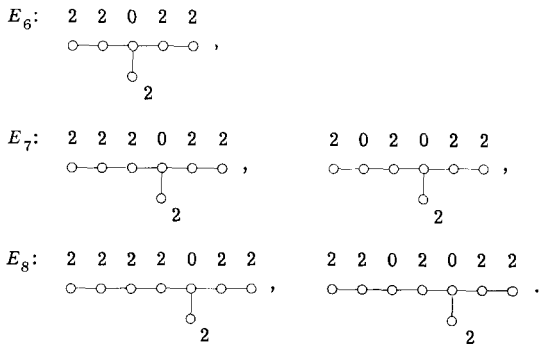
The elements $f(\tilde{H}), f(\tilde{E}_+), f(\tilde{E}_-)$ as given above satisfy the commutation relations (7.1).

(b) *Nonprincipal:* For the algebra D_n, besides the principal 3d-subalgebras, there are $[(n-2)/2]$ pairwise nonconjugate 3d-S-subalgebras; their characteristic diagram has the form



$(r = 1, 2, \ldots, [\frac{1}{2}(n-2)])$.

For the exceptional algebras there are, besides the principal S-subalgebras, the following S-subalgebras, given by the characteristics



In order to find the embedding of the nonprincipal 3d-subalgebras, we express the defining vector f in terms of the roots α of G which project onto the positive root α' of $\tilde{G}$. The set of roots α for which $f^*(\alpha) = \alpha'$ is denoted by $\Gamma_{\alpha'}$. Then

$$f = f(\alpha') = \sum_{\alpha \in \Gamma_{\alpha'}} |c_\alpha|^2 \alpha \tag{7.9}$$

and

$$f(\tilde{H}) = \sum_{\alpha \in \Gamma_{\alpha'}} |c_\alpha|^2 H_\alpha, \quad f(\tilde{E}_+) = \sum_{\alpha \in \Gamma_{\alpha'}} c_\alpha E_\alpha,$$

$$f(\tilde{E}_-) = \sum_{\alpha \in \Gamma_{\alpha'}} \bar{c}_\alpha E_{-\alpha}. \tag{7.10}$$

Three-Dimensional R-Subalgebras

Let G be a simple algebra and

$$G' = G'_1 + G'_2 + \cdots + G'_s$$

a proper regular subalgebra of G, with G'_k, $k = 1, 2, \ldots, s$, simple ideals. For each simple ideal G'_k $(k = 1, 2, \ldots, s)$, we take one 3d-S-subalgebra $\tilde{G}_k$. Then

$$\tilde{G} = \tilde{G}_1 + \tilde{G}_2 + \cdots + \tilde{G}_s$$

is a 3d-R-subalgebra with respect to G. If f_k is the defining vector of the subalgebra $\tilde{G}_k$, the sum

$$f = f_1 + f_2 + \cdots + f_s$$

is the defining vector of the subalgebra $\tilde{G}$.

This construction yields all the 3d-subalgebras of the simple algebras, because we know all the regular subalgebras as well as the 3d-S-subalgebras. Every 3d-subalgebra of a simple algebra G is either a regular subalgebra of G, an S-subalgebra of G, or an S-subalgebra of one of its proper regular subalgebras (an R-algebra of G).

Table VI gives all the 3d-subalgebras of the classical algebras up to rank 6. The 3d-subalgebras of the exceptional algebras are given in Tables 16–20 of Ref. 13.

TABLE VI. Three-dimensional subalgebras of classical algebras up to rank 6. The second column shows the minimal including regular subalgebra; when the S-subalgebra is not principal the symbol (a_r) is written after the regular subalgebra, r being the number of zeros in the characteristic diagram. The third column gives the index of the 3d-subalgebra; if there are several subalgebras with the same index, the first will be distinguished with one prime, the second one with two primes, and so on, corresponding to the order of the defining vectors $f_1 < f_2 < f_3 \cdots$. The fourth column gives the defining vector in cartesian coordinates. For the algebras $B_2 \sim C_2$ and $A_3 \sim D_3$ two sets of defining vectors are given corresponding to two different representations of these algebras.

Algebra	Regular subalgebra	Index	Defining vector	
A_2	A_1	1	(1, 0, −1)	
	A_2	4	(2, 0, −2)	
$B_2 \sim C_2$	A_1	1	(1, 1)	(1, 0)
	$2A_1; A_1^2$	2	(2, 0)	(1, 1)
	B_2	10	(4, 2)	(3, 1)
$A_3 \sim D_3$	A_1	1	(1, 0, 0, −1)	(1, 1, 0)
	$2A_1$	2	(1, 1, −1, −1)	(2, 0, 0)
	A_2	4	(2, 0, 0, −2)	(2, 2, 0)
	A_3	10	(3, 1, −1, −3)	(4, 2, 0)
B_3	A_1	1	(1, 1, 0)	
	$2A_1; A_1^2$	2	(2, 0, 0)	
	$A_1 + A_1^2$	3	(2, 1, 1)	
	$A_2; 2A_1 + A_1^2$	4	(2, 2, 0)	
	$A_3; B_2$	10	(4, 2, 0)	
	B_3	28	(6, 4, 2)	
C_3	A_1	1	(1, 0, 0)	
	$2A_1; A_1^2$	2	(1, 1, 0)	
	$3A_1; A_1 + A_1^2$	3	(1, 1, 1)	
	A_2^2	8	(2, 2, 0)	
	C_2	10	(3, 1, 0)	
	$C_2 + A_1$	11	(3, 1, 1)	
	C_3	35	(5, 3, 1)	
A_4	A_1	1	(1, 0, 0, 0, −1)	
	$2A_1$	2	(1, 1, 0, −1, −1)	
	A_2	4	(2, 0, 0, 0, −2)	
	$A_2 + A_1$	5	(2, 1, 0, −1, −2)	
	A_3	10	(3, 1, 0, −1, −3)	
	A_4	20	(4, 2, 0, −2, −4)	
B_4	A_1	1	(1, 1, 0, 0)	
	$2A_1$	2′	(1, 1, 1, 1)	
	$2A_1; A_1^2$	2″	(2, 0, 0, 0)	
	$3A_1; A_1 + A_1^2$	3	(2, 1, 1, 0)	
	$A_2; 4A_1; 2A_1 + A_1^2$	4	(2, 2, 0, 0)	
	$A_2 + A_1^2$	6	(2, 2, 2, 0)	
	A_3	10′	(3, 3, 1, 1)	
	$A_3; B_2$	10″	(4, 2, 0, 0)	
	$B_2 + A_1$	11	(4, 2, 1, 1)	
	$B_2 + 2A_1; A_3 + A_1^2; D_4(a_1)$	12	(4, 2, 2, 0)	
	$B_3; D_4$	28	(6, 4, 2, 0)	
	B_4	60	(8, 6, 4, 2)	
C_4	A_1	1	(1, 0, 0, 0)	
	$2A_1; A_1^2$	2	(1, 1, 0, 0)	
	$3A_1; A_1 + A_1^2$	3	(1, 1, 1, 0)	
	$4A_1; 2A_1 + A_1^2; 2A_1^2$	4	(1, 1, 1, 1)	
	A_2^2	8	(2, 2, 0, 0)	
	$A_2^2 + A_1$	9	(2, 2, 1, 0)	
	C_2	10	(3, 1, 0, 0)	
	$C_2 + A_1$	11	(3, 1, 1, 0)	
	$C_2 + 2A_1; C_2 + A_1^2$	12	(3, 1, 1, 1)	
	$2C_2; A_3^2$	20	(3, 3, 1, 1)	
	C_3	35	(5, 3, 1, 0)	
	$C_3 + A_1$	36	(5, 3, 1, 1)	
	C_4	84	(7, 5, 3, 1)	

Algebra	Regular subalgebra	Index	Defining vector
D_4	A_1	1	(1, 1, 0, 0)
	$2A_1$	2′	(1, 1, 1, ±1)
	$2A_1$	2″	(2, 0, 0, 0)
	$3A_1$	3	(2, 1, 1, 0)
	$4A_1; A_2$	4	(2, 2, 0, 0)
	A_3	10′	(3, 3, 1, ±1)
	A_3	10″	(4, 2, 0, 0)
	$D_4(a_1)$	12	(4, 2, 2, 0)
	D_4	28	(6, 4, 2, 0)
A_5	A_1	1	(1, 0, 0, 0, 0, −1)
	$2A_1$	2	(1, 1, 0, 0, −1, −1)
	$3A_1$	3	(1, 1, 1, −1, −1, −1)
	A_2	4	(2, 0, 0, 0, 0, −2)
	$A_2 + A_1$	5	(2, 1, 0, 0, −1, −2)
	$2A_2$	8	(2, 2, 0, 0, −2, −2)
	A_3	10	(3, 1, 0, 0, −1, −3)
	$A_3 + A_1$	11	(3, 1, 1, −1, −1, −3)
	A_4	20	(4, 2, 0, 0, −2, −4)
	A_5	35	(5, 3, 1, −1, −3, −5)
B_5	A_1	1	(1, 1, 0, 0, 0)
	$2A_1$	2′	(1, 1, 1, 1, 0)
	$2A_1; A_1^2$	2″	(2, 0, 0, 0, 0)
	$3A_1; A_1 + A_1^2$	3	(2, 1, 1, 0, 0)
	$2A_1 + A_1^2$	4′	(2, 1, 1, 1, 1)
	$A_2; 2A_1 + A_1^2; 4A_1$	4″	(2, 2, 0, 0, 0)
	$A_2 + A_1; 3A_1 + A_1^2$	5	(2, 2, 1, 1, 0)
	$A_2 + 2A_1; A_2 + A_1^2; 4A_1 + A_1^2$	6	(2, 2, 2, 0, 0)
	A_3	10′	(3, 3, 1, 1, 0)
	$A_3; B_2$	10″	(4, 2, 0, 0, 0)
	$A_3 + A_1; B_2 + A_1$	11	(4, 2, 1, 1, 0)
	$A_3 + A_1^2$	12′	(3, 3, 2, 1, 1)
	$A_3 + 2A_1; A_3 + A_1^2; B_2 + 2A_1; D_4(a_1)$	12″	(4, 2, 2, 0, 0)
	$A_2 + B_2; D_4(a_1) + A_1^2$	14	(4, 2, 2, 2, 0)
	$A_4; A_3 + B_2$	20	(4, 4, 2, 2, 0)
	$D_4; B_3$	28	(6, 4, 2, 0, 0)
	$B_3 + A_1$	29	(6, 4, 2, 1, 1)
	$D_4 + A_1^2; B_3 + 2A_1; D_5(a_1)$	30	(6, 4, 2, 2, 0)
	$D_5; B_4$	60	(8, 6, 4, 2, 0)
	B_5	110	(10, 8, 6, 4, 2)
C_5	A_1	1	(1, 0, 0, 0, 0)
	$2A_1; A_1^2$	2	(1, 1, 0, 0, 0)
	$3A_1; A_1 + A_1^2$	3	(1, 1, 1, 0, 0)
	$4A_1; 2A_1 + A_1^2; 2A_1^2$	4	(1, 1, 1, 1, 0)
	$5A_1; 3A_1 + A_1^2; A_1 + 2A_1^2$	5	(1, 1, 1, 1, 1)
	A_2^2	8	(2, 2, 0, 0, 0)
	$A_2^2 + A_1$	9	(2, 2, 1, 0, 0)
	$A_2^2 + 2A_1; A_2^2 + A_1^2$	10′	(2, 2, 1, 1, 0)
	C_2	10″	(3, 1, 0, 0, 0)
	$C_2 + A_1$	11	(3, 1, 1, 0, 0)

TABLE VI contd

Algebra	Regular subalgebra	Index	Defining vector
C_5	$C_2 + 2A_1; C_2 + A_1^2$	12	(3, 1, 1, 1, 0)
	$C_2 + 3A_1; C_2 + A_1 + A_1^2$	13	(3, 1, 1, 1, 1)
	$A_2^2 + C_2$	18	(3, 2, 2, 1, 0)
	$A_3^2; 2C_2$	20	(3, 3, 1, 1, 0)
	$A_3^2 + A_1; 2C_2 + A_1$	21	(3, 3, 1, 1, 1)
	C_3	35	(5, 3, 1, 0, 0)
	$C_3 + A_1$	36	(5, 3, 1, 1, 0)
	$C_3 + 2A_1; C_3 + A_1^2$	37	(5, 3, 1, 1, 1)
	A_4^2	40	(4, 4, 2, 2, 0)
	$C_3 + C_2$	45	(5, 3, 3, 1, 1)
	C_4	84	(7, 5, 3, 1, 0)
	$C_4 + A_1$	85	(7, 5, 3, 1, 1)
	C_5	165	(9, 7, 5, 3, 1)
D_5	A_1	1	(1, 1, 0, 0, 0)
	$2A_1$	2′	(1, 1, 1, 1, 0)
	$2A_1$	2″	(2, 0, 0, 0, 0)
	$3A_1$	3	(2, 1, 1, 0, 0)
	$A_2; 4A_1$	4	(2, 2, 0, 0, 0)
	$A_2 + A_1$	5	(2, 2, 1, 1, 0)
	$A_2 + 2A_1$	6	(2, 2, 2, 0, 0)
	A_3	10′	(3, 3, 1, 1, 0)
	A_3	10″	(4, 2, 0, 0, 0)
	$A_3 + A_1$	11	(4, 2, 1, 1, 0)
	$A_3 + 2A_1; D_4(a_1)$	12	(4, 2, 2, 0, 0)
	A_4	20	(4, 4, 2, 2, 0)
	D_4	28	(6, 4, 2, 0, 0)
	$D_5(a_1)$	30	(6, 4, 2, 2, 0)
	D_5	60	(8, 6, 4, 2, 0)
A_6	A_1	1	(1, 0, 0, 0, 0, 0, −1)
	$2A_1$	2	(1, 1, 0, 0, 0, −1, −1)
	$3A_1$	3	(1, 1, 1, 0, −1, −1, −1)
	A_2	4	(2, 0, 0, 0, 0, 0, −2)
	$A_2 + A_1$	5	(2, 1, 0, 0, 0, −1, −2)
	$A_2 + 2A_1$	6	(2, 1, 1, 0, −1, −1, −2)
	$2A_2$	8	(2, 2, 0, 0, 0, −2, −2)
	A_3	10	(3, 1, 0, 0, 0, −1, −3)
	$A_3 + A_1$	11	(3, 1, 1, 0, −1, −1, −3)
	$A_3 + A_2$	14	(3, 2, 1, 0, −1, −2, −3)
	A_4	20	(4, 2, 0, 0, 0, −2, −4)
	$A_4 + A_1$	21	(4, 2, 1, 0, −1, −2, −4)
	A_5	35	(5, 3, 1, 0, −1, −3, −5)
	A_6	56	(6, 4, 2, 0, −2, −4, −6)
B_6	A_1	1	(1, 1, 0, 0, 0, 0)
	$2A_1$	2′	(1, 1, 1, 1, 0, 0)
	$2A_1; A_1^2$	2″	(2, 0, 0, 0, 0, 0)
	$3A_1$	3′	(1, 1, 1, 1, 1, 1)
	$3A_1; A_1 + A_1^2$	3″	(2, 1, 1, 0, 0, 0)
	$4A_1; 2A_1 + A_1^2$	4′	(2, 1, 1, 1, 1, 0)
	$4A_1; 2A_1 + A_1^2; A_2$	4″	(2, 2, 0, 0, 0, 0)
	$5A_1; 3A_1 + A_1^2; A_2 + A_1$	5	(2, 2, 1, 1, 0, 0)
	$6A_1; 4A_1 + A_1^2; A_2 + 2A_1; A_2 + A_1^2$	6	(2, 2, 2, 0, 0, 0)
	$A_2 + A_1 + A_1^2$	7	(2, 2, 2, 1, 1, 0)
	$A_2 + 2A_1 + A_1^2; 2A_2$	8	(2, 2, 2, 2, 0, 0)

Algebra	Regular subalgebra	Index	Defining vector
B_6	A_3	10′	(3, 3, 1, 1, 0, 0)
	$A_3; B_2$	10″	(4, 2, 0, 0, 0, 0)
	$A_3 + A_1$	11′	(3, 3, 1, 1, 1, 1)
	$A_3 + A_1; B_2 + A_1$	11″	(4, 2, 1, 1, 0, 0)
	$A_3 + 2A_1; A_3 + A_1^2$	12′	(3, 3, 2, 1, 1, 0)
	$B_2 + 2A_1$	12″	(4, 2, 1, 1, 1, 1)
	$B_2 + 2A_1; A_3 + A_1^2;$ $A_3 + 2A_1; D_4(a_1)$	12‴	(4, 2, 2, 0, 0, 0)
	$A_3 + A_1 + A_1^2; B_2 + 3A_1;$ $D_4 + A_1(a_1)$	13	(4, 2, 2, 1, 1, 0)
	$A_3 + 2A_1 + A_1^2; B_2 + 4A_1;$ $D_4 + 2A_1(a_1);$ $D_4 + A_1^2(a_1); A_3 + A_2;$ $A_2 + B_2$	14	(4, 2, 2, 2, 0, 0)
	$A_3 + B_2$	20′	(4, 3, 3, 2, 1, 1)
	$A_3 + B_2; 2A_3; A_4$	20″	(4, 4, 2, 2, 0, 0)
	$D_4 + B_2(a_1); A_4 + A_1^2;$ $A_3 + B_2 + A_1^2$	22	(4, 4, 2, 2, 2, 0)
	$D_4; B_3$	28	(6, 4, 2, 0, 0, 0)
	$D_4 + A_1; B_3 + A_1$	29	(6, 4, 2, 1, 1, 0)
	$D_4 + 2A_1; D_5(a_1); D_4 + A_1^2;$ $B_3 + 2A_1$	30	(6, 4, 2, 2, 0, 0)
	$D_5 + A_1^2(a_1); B_3 + A_2$	32	(6, 4, 2, 2, 2, 0)
	A_5	35	(5, 5, 3, 3, 1, 1)
	$D_6(a_2); D_4 + B_2; A_3 + B_2$	38	(6, 4, 4, 2, 2, 0)
	$D_5; B_4$	60	(8, 6, 4, 2, 0, 0)
	$B_4 + A_1$	61	(8, 6, 4, 2, 1, 1)
	$D_6(a_1); D_5 + A_1^2; B_4 + 2A_1$	62	(8, 6, 4, 2, 2, 0)
	$D_6; B_5$	110	(10, 8, 6, 4, 2, 0)
	B_6	182	(12, 10, 8, 6, 4, 2)
C_6	A_1	1	(1, 0, 0, 0, 0, 0)
	$2A_1; A_1^2$	2	(1, 1, 0, 0, 0, 0)
	$3A_1; A_1 + A_1^2$	3	(1, 1, 1, 0, 0, 0)
	$4A_1; 2A_1 + A_1^2; 2A_1^2$	4	(1, 1, 1, 1, 0, 0)
	$5A_1; 3A_1 + A_1^2; A_1 + 2A_1^2$	5	(1, 1, 1, 1, 1, 0)
	$6A_1; 4A_1 + A_1^2; 2A_1 + 2A_1^2;$ $3A_1^2$	6	(1, 1, 1, 1, 1, 1)
	A_2^2	8	(2, 2, 0, 0, 0, 0)
	$A_2^2 + A_1$	9	(2, 2, 1, 0, 0, 0)
	$A_2^2 + 2A_1; A_2^2 + A_1^2$	10′	(2, 2, 1, 1, 0, 0)
	C_2	10″	(3, 1, 0, 0, 0, 0)
	$A_2^2 + 3A_1; A_2^2 + A_1 + A_1^2$	11′	(2, 2, 1, 1, 1, 0)
	$C_2 + A_1$	11″	(3, 1, 1, 0, 0, 0)
	$C_2 + 2A_1; C_2 + A_1^2$	12	(3, 1, 1, 1, 0, 0)
	$C_2 + 3A_1; C_2 + A_1 + A_1^2$	13	(3, 1, 1, 1, 1, 0)
	$C_2 + 4A_1; C_2 + 2A_1 + A_1^2;$ $C_2 + 2A_1^2$	14	(3, 1, 1, 1, 1, 1)
	$2A_2^2$	16	(2, 2, 2, 2, 0, 0)
	$A_2^2 + C_2$	18	(3, 2, 2, 1, 0, 0)
	$A_2^2 + C_2 + A_1$	19	(3, 2, 2, 1, 1, 0)
	$A_3^2; 2C_2$	20	(3, 3, 1, 1, 0, 0)
	$A_3^2 + A_1; 2C_2 + A_1$	21	(3, 3, 1, 1, 1, 0)
	$A_3^2 + 2A_1; A_3^2 + A_1^2$		
	$2C_2 + 2A_1; 2C_2 + A_1^2$	22	(3, 3, 1, 1, 1, 1)
	$A_3^2 + C_2; 3C_2$	30′	(3, 3, 3, 1, 1, 1)
	$A_3^2 + C_2$	30″	(4, 3, 2, 1, 0, 0)
	C_3	35	(5, 3, 1, 0, 0, 0)

TABLE VI contd

Algebra	Regular subalgebra	Index	Defining vector
C_6	$C_3 + A_1$	36	(5, 3, 1, 1, 0, 0)
	$C_3 + 2A_1; C_3 + A_1^2$	37	(5, 3, 1, 1, 1, 0)
	$C_3 + 3A_1; C_3 + A_1 + A_1^2$	38	(5, 3, 1, 1, 1, 1)
	A_4^2	40	(4, 4, 2, 2, 0, 0)
	$A_4^2 + A_1$	41	(4, 4, 2, 2, 1, 0)
	$C_3 + A_2^2$	43	(5, 3, 2, 2, 1, 0)
	$C_3 + C_2$	45	(5, 3, 3, 1, 1, 0)
	$C_3 + C_2 + A_1$	46	(5, 3, 3, 1, 1, 1)
	$A_5^2; 2C_3$	70	(5, 5, 3, 3, 1, 1)
	C_4	84	(7, 5, 3, 1, 0, 0)
	$C_4 + A_1$	85	(7, 5, 3, 1, 1, 0)
	$C_4 + 2A_1; C_4 + A_1^2$	86	(7, 5, 3, 1, 1, 1)
	$C_4 + C_2$	94	(7, 5, 3, 3, 1, 1)
	C_5	165	(9, 7, 5, 3, 1, 0)
	$C_5 + A_1$	166	(9, 7, 5, 3, 1, 1)
	C_6	286	(11, 9, 7, 5, 3, 1)
D_6	A_1	1	(1, 1, 0, 0, 0, 0)
	$2A_1$	2′	(1, 1, 1, 1, 0, 0)
	$2A_1$	2″	(2, 0, 0, 0, 0, 0)
	$3A_1$	3′	(1, 1, 1, 1, 1, ±1)
	$3A_1$	3″	(2, 1, 1, 0, 0, 0)
	$4A_1$	4′	(2, 1, 1, 1, 1, 0)
	$4A_1; A_2$	4″	(2, 2, 0, 0, 0, 0)
	$5A_1; A_2 + A_1$	5	(2, 2, 1, 1, 0, 0)
	$A_2 + 2A_1$	6	(2, 2, 2, 0, 0, 0)
	$2A_2$	8	(2, 2, 2, 2, 0, 0)
	A_3	10′	(3, 3, 1, 1, 0, 0)
	A_3	10″	(4, 2, 0, 0, 0, 0)
	$A_3 + A_1$	11′	(3, 3, 1, 1, 1, ±1)
	$A_3 + A_1$	11″	(4, 2, 1, 1, 0, 0)
	$A_3 + 2A_1$	12′	(3, 3, 2, 1, 1, 0)
	$A_3 + 2A_1; D_4(a_1)$	12″	(4, 2, 2, 0, 0, 0)
	$D_4(a_1) + A_1$	13	(4, 2, 2, 1, 1, 0)
	$D_4(a_1) + 2A_1; A_3 + A_2$	14	(4, 2, 2, 2, 0, 0)
	$A_4; 2A_3$	20	(4, 4, 2, 2, 0, 0)
	D_4	28	(6, 4, 2, 0, 0, 0)
	$D_4 + A_1$	29	(6, 4, 2, 1, 1, 0)
	$D_4 + 2A_1; D_5(a_1)$	30	(6, 4, 2, 2, 0, 0)
	A_5	35	(5, 5, 3, 3, 1, ±1)
	$D_6(a_2)$	38	(6, 4, 4, 2, 2, 0)
	D_5	60	(8, 6, 4, 2, 0, 0)
	$D_6(a_1)$	62	(8, 6, 4, 2, 2, 0)
	D_6	110	(10, 8, 6, 4, 2, 0)

8. THE DEFINING MATRIX

Let $\tilde{G}$ be a semisimple subalgebra of a simple algebra G. Let us choose an embedding f of $\tilde{G}$ in G such that the corresponding Cartan spaces satisfy $\tilde{K} \subseteq K$. Then the embedding f is given as

$$f(\tilde{H}_i) = \sum_{k=1}^{n} f_{ik} H_k, \qquad i = 1, 2, \ldots, n' \leq n, \tag{8.1}$$

where n and n' are the dimension of the Cartan subspaces K and $\tilde{K}$, respectively. We call the set of numbers f_{ik} the *defining matrix* of the embedding of $\tilde{G}$ in G. The defining matrix is a natural generalization of the defining vector. Since the $f(H_i)$ are elements of the Cartan space K of G, an inner automorphism of the group G will relate a given defining matrix (f_{ik}) into a defining matrix of a conjugate subalgebra. Two defining matrices related by an inner automorphism of the algebra G are called equivalent.

As in the case of 3d-subalgebras, the defining matrix characterizes the embedding of the algebra $\tilde{G}$ in G, because two semisimple subalgebras of G are conjugate if and only if the corresponding defining matrices are equivalent.

From the defining matrix of a subalgebra $\tilde{G}$ in G the following information can be obtained:

(i) Embedding of the generators $\tilde{H}_i$ of $\tilde{G}$ in G as given by Eq. (8.1).

(ii) Mapping of the weights m of any representation ϕ of G:

$$[f^*(m)]_i = m_i' = \sum_{k=1}^{n} f_{ik} m_k, \qquad m' \in \tilde{K},$$
$$i = 1, 2, \ldots, n' \leq n, \ 2 \leq n'.$$

(iii) Embedding of the roots α' of $\tilde{G}$ in G:

$$[f(\alpha')]_k = \sum_{i=1}^{n'} \alpha_i' f_{ik}.$$

(iv) Index j_f of embedding of $\tilde{G}$ in G:

$$\sum_{k=1}^{n} f_{ik} f_{jk} = \alpha_{ij} j_f,$$

where for

(a) $G = A_n, B_n, D_n, G_2, F_4, E_6, E_7, E_8$,
$\tilde{G} = B_{n'}, D_{n'}, F_4$ $(2 \leq n' \leq n)$, and for
$G = C_n, \tilde{G} = C_{n'}$ $(2 \leq n' < n)$,
$\alpha_{ij} = \delta_{ij}$;

(b) $G = A_n, B_n, D_n, G_2, F_4, E_6, E_7, E_8$,
$\tilde{G} = C_{n'}$ $(2 \leq n' \leq n)$,
$\alpha_{ij} = 2\delta_{ij}$;

(c) $G = C_n$,
$\tilde{G} = B_{n'}, D_{n'}, F_4$ $(2 \leq n' \leq n)$,
$\alpha_{ij} = \frac{1}{2}\delta_{ij}$;

(d) $G = A_n, B_n, D_n, G_2, F_4, E_6, E_7, E_8$,
$\tilde{G} = A_{n'}, G_2, E_6, E_7, E_8$ $(2 \leq n' \leq n)$,
$\alpha_{ij} = n'/(n' + 1)$, for $i = j$,
$\alpha_{ij} = -1/(n' + 1)$, for $i \neq j$;

(e) $G = C_n$,
$\tilde{G} = A_{n'}, G_2, E_6, E_7, E_8$ $(2 \leq n' \leq n)$,
$\alpha_{ij} = [n'/2(n' + 1)]$, for $i = j$,
$\alpha_{ij} = [-1/2(n' + 1)]$, for $i \neq j$.

(v) The defining matrix satisfies the following relations: For

(a) $G = B_n, C_n, D_n, F_4$,

$\tilde{G} = A_{n'}, G_2, E_6, E_7, E_8$ $(2 \leq n' \leq n)$,

$$\sum_{i=1}^{n'+1} f_{ik} = 0;$$

(b) $G = A_n, G_2, E_6, E_7, E_8$,

$\tilde{G} = A_{n'}, G_2, E_6, E_7, E_8$ $(2 \leq n' \leq n)$,

$$\sum_{i=1}^{n'+1} f_{ik} = c, \quad \sum_{k=1}^{n+1} f_{ik} = c\,\frac{n+1}{n'+1},$$

where the constant c has been set equal to zero.

(c) $G = A_n, G_2, E_6, E_1, E_8$,

$\tilde{G} = B_{n'}, C_{n'}, D_{n'}, F_4$,

$$\sum_{k=1}^{n} f_{ik} = 0.$$

The various values of α_{ij} arise from Eq. (2.8) and Eq. (2.9) as well as from the fact that the Cartan subspace of the algebras A_n, G_2, E_6, E_7, E_8 has been embedded into a space with one more dimension. (Thus the base is not orthonormal, but satisfies $(H_i, H_j) = [n/(n+1)]$, for $i = j$, and $(H_i, H_j) = [-1\,(n+1)]$, for $i \neq j$, $\sum_{i=1}^{n+1} H_i = 0$, n the rank of the algebra).

(vi) If $G_1 \subseteq G_2 \subseteq G_3$ are three semisimple algebras, then the defining matrix of G_1 in G_3 is the (matrix) product of the defining matrices of G_1 in G_2 and of G_2 in G_3.

(vii) If $f_{ik}^{(1)}, f_{ik}^{(2)}, \ldots, f_{ik}^{(s)}$ are defining matrices of a subalgebra $\tilde{G}$ with respect to different ideals of G, then

$$f_{ik} = f_{ik}^{(1)} + f_{ik}^{(2)} + \cdots + f_{ik}^{(s)}$$

is the defining matrix of $\tilde{G}$ with respect to G.

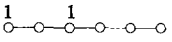
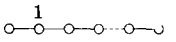
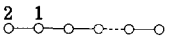
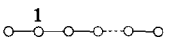
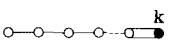
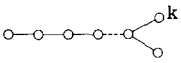
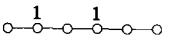
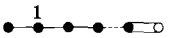
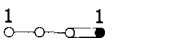
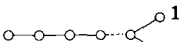
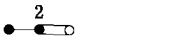
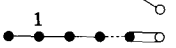
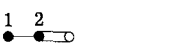
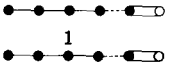
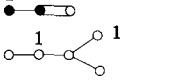
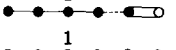
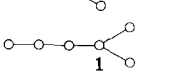
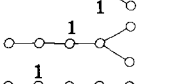
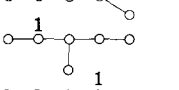
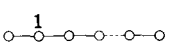
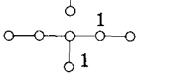
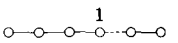
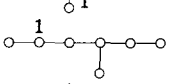
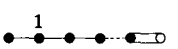
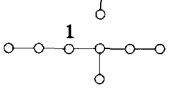
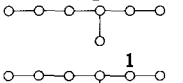
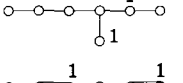
9. SIMPLE MAXIMAL S-SUBALGEBRAS

Let $SL(N)$ be the group of unimodular matrices of dimension N, $O(N)$ the group of orthogonal matrices of dimension N, and $Sp(N)$ the group of symplectic matrices of dimension N. The algebra A_n is obtained from the group $SL(n+1)$, the algebra B_n from the group $O(2n+1)$, the algebra C_n from the group $Sp(2n)$, and the algebra D_n from the group $O(2n)$.

A simple subalgebra $\tilde{G}$ of a simple algebra G is called *maximal* if for every subalgebra G' of G satisfying the inclusion $G \subseteq G' \subseteq G$ holds that either $G' \sim \tilde{G}$ or $G' \sim G$.

The problem of finding all maximal subalgebras of the classical algebras was solved by Dynkin.[28] This problem is equivalent to the problem of finding all maximal connected subgroups for the classical groups $\mathcal{G}$ (i.e., those connected subgroups of $\mathcal{G}$ which are not contained in any connected *proper* subgroup of $\mathcal{G}$). Dynkin was able to demonstrate that the set of all maximal subgroups of one of the classical groups essentially coincides with the set of simple *irreducible* subgroups. That is, with those simple *transformation subgroups* of a classical group, defined on an N-dimensional complex space $R^{(N)}$, which leave no subspace of $R^{(N)}$ invariant. The following theorem holds[29]:

TABLE VII. Semisimple S-subalgebras of the classical algebras which are maximal in G^* instead of in G.

$\tilde{G} \subset G^* \subset G$	Irred. represent. of $\tilde{G}$	Irred. represent. of G^*
$A_n^{n-1} \subset A_{(1/2)(n-1)(n+2)}$ $n > 3$		
$A_n^{n+3} \subset A_{(1/2)n(n+3)}$ $n > 1$		
$B_{2n+1}^1 \subset D_{2n+2}$ $n > 0$, $k \geq 1$		
$G_2^1 \subset B_3$ $k > 1$		
$A_1^{28} \subset G_2 \subset B_3$		
$A_5^6 \subset C_{10} \subset B_{94}$		
$B_4^2 \subset D_8 \subset D_{64}$		
$C_3^5 \subset C_7 \subset D_{45}$		
$C_3^5 \subset C_7 \subset C_{175}$		
$D_5^4 \subset A_{15} \subset A_{559}$		
$D_6^8 \subset C_{16} \subset B_{247}$		
$D_6^8 \subset C_{16} \subset A_{4927}$		
$E_6^6 \subset A_{26} \subset A_{350}$		
$E_6^6 \subset A_{26} \subset A_{17549}$		
$E_7^{12} \subset C_{28} \subset B_{769}$		
$E_7^{12} \subset C_{28} \subset C_{13832}$		
$E_7^{12} \subset C_{28} \subset D_{182875}$		
$E_7^{12} \subset C_{28} \subset D_{1896048}$		
$B_{n'} + B_{n''} \subset D_{n'+n''+1}$ $\begin{cases} n' \geq n'' \geq 1 \\ n' + n'' \geq 4 \end{cases}$		

Every *irreducible* group of unimodular linear transformations of the N-dimensional complex space $R^{(N)}$ (i.e., a group of transformations which does not leave invariant a proper subspace of $R^{(N)}$) is maximal either in $SL(N)$ (if the group does not have a bilinear invariant), or in $Sp(N)$ (if it has a skew-symmetric bilinear invariant), or in $O(N)$ (if it has a symmetric bilinear invariant). Exceptions to this rule are listed in Table VII.

Below, rules will be given for obtaining all maximal S-subalgebras of the classical algebras. These rules have been deduced by utilizing the theorem quoted above. In the first two steps a characterization for the irreducible representations $\tilde{\phi}$ with dimensionality N for a simple algebra $\tilde{G}$ is given (we write $\tilde{\phi}$ and $\tilde{G}$, because this algebra is going to be a subalgebra). For given algebra $\tilde{G}$ and given irreducible representation $\tilde{\phi}$ of this algebra, having dimensionality N, steps (3)–(5) of the rules consist in finding out which of the classical algebras $G = A_n, N = n+1, B_n, N = 2n+1, C_n, N = 2n$, and $D_n, N = 2n$ admits the algebra $\tilde{G}$ as maximal subalgebra. This depends on whether the representation $\tilde{\phi}$ of $\tilde{G}$ admits a symmetric bilinear form, an antisymmetric bilinear form, or none

at all, as well as on the fact whether N is even or odd. These conditions are mutually exclusive and follow from the fact that $R^{(N)}$ is the space on which both groups $\tilde{\mathfrak{G}}$ and $\mathfrak{G}$ act, namely $\tilde{\mathfrak{G}}$ (with algebra $\tilde{G}$) in its representation $\tilde{\phi}$ and the group $\mathfrak{G}$ (with algebra G) in its fundamental (defining) representation.

The rules for obtaining all maximal S-subalgebras of the classical algebras are:

(1) Given some irreducible representation $\tilde{\phi}$ of a simple algebra $\tilde{G}$ and the highest weight M of this representation, attach to every simple root α of its Dynkin diagram (to every dot of the Dynkin diagram) the number

$$M^{\alpha} = [2(M,\alpha)/(\alpha,\alpha)]. \tag{9.1}$$

These numbers are always nonnegative integers and are called the contravariant coordinates of M. All the irreducible representations of the algebra $\tilde{G}$ are obtained if we attach arbitrarily nonnegative integers to the Dynkin diagram of $\tilde{G}$. The dimension of the irreducible representation characterized in this manner is denoted by N.

(2) The bilinear invariant of an irreducible representation $\tilde{\phi}$ of $\tilde{G}$ is found, if it exists, from the Dynkin diagram as prepared in (1). All the irreducible representations of the simple algebras $B_n, C_n, D_{2k}, G_2, F_4, E_7, E_8$ have bilinear invariants. In order that a representation of the algebras A_n, D_{2k+1}, E_6 has a bilinear invariant, the numbers M^{α} attached to the corresponding diagrams must be symmetrically placed, as indicated in Table VIIIa.

Suppose an irreducible representation of the simple algebra $\tilde{G}$ has a bilinear invariant. We multiply each number M^{α} by the corresponding coefficient indicated in Table VIIIb. We add all the products obtained. Then the bilinear invariant is symmetric, if the sum is even, and is antisymmetric, if the sum is odd.

(3) If some N-dimensional representation $\tilde{\phi}$ of the algebra $\tilde{G}$ has a symmetric bilinear invariant, then $\tilde{G}$ is contained as maximal S-subalgebra in the algebra B_n if $N = 2n + 1$ or in D_n if $N = 2n$. If some N-dimensional representation $\tilde{\phi}$ of $\tilde{G}$ has an antisymmetric bilinear invariant, then $\tilde{G}$ is contained as maximal S-subalgebra in the algebra C_n, if $N = 2n$. If some N-dimensional representation of the simple algebra $\tilde{G}$ does not have any bilinear invariant, then $\tilde{G}$ is contained as maximal S-subalgebra in the algebra A_n with $N = n + 1$.

(4) If for some algebra $\tilde{G}$ and some representation $\tilde{\phi}$ of $\tilde{G}$, having dimensionality N, the algebra G obtained by step (3) if $\tilde{G}$ itself, then the algebra for which $\tilde{G}$ is a maximal S-subalgebra is A_n with $N = n + 1$. (In all cases the algebra A_n, $N = n + 1$, contains $\tilde{G}$ as a subalgebra. However, only in these exceptional cases and when the representation does not have a bilinear invariant, are they maximal S-subalgebra.)

(5) The subalgebras obtained in this manner which are *not* maximal S-subalgebras of the algebra G, as defined in step (3), but are maximal S-subalgebras of some subalgebra G^* of G are listed in Table VII. These algebras are irreducible in G but not maximal. (For the case of these subalgebras exists a proper subalgebra G^* of G such that $\tilde{G} \subset G^* \subset G$. All three algebras $\tilde{G}, G^*$, and G are listed in Table VII, except for those algebras G for which the expression for their rank is too bulky.)

(6) To each of the types of the simple subalgebras described in (3), (4), and (5), there corresponds one

TABLE VIIIa.
Irreducible representations with bilinear invariants.

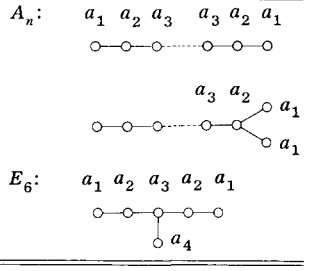


TABLE IX.
Defining Matrices for the maximal simple S-subalgebras.

(a) $G = B_n, C_n, D_n$:
$f_{ik} = m_i^{(k)}, k = 1, 2, \ldots, n; i = 1, 2, \ldots, n'$

(b) $G = A_n; \tilde{G} = B_{n'}, C_{n'}, D_{n'}, F_4$:
$f_{ik} = m_i^{(k)}, k = 1, 2, \ldots, n+1; i = 1, 2, \ldots, n'$

(c) $G = A_n; \tilde{G} = A_{n'}, G_2, E_6, E_7, E_8$:
$f_{ik} = m_i^{(k)} + [1/(n'+1)], k = 1, 2, \ldots, n+1; i = 1, 2, \ldots, n'+1.$

TABLE VIIIb. Coefficients for the calculation of the bilinear invariants of irreducible representations of the simple algebras.

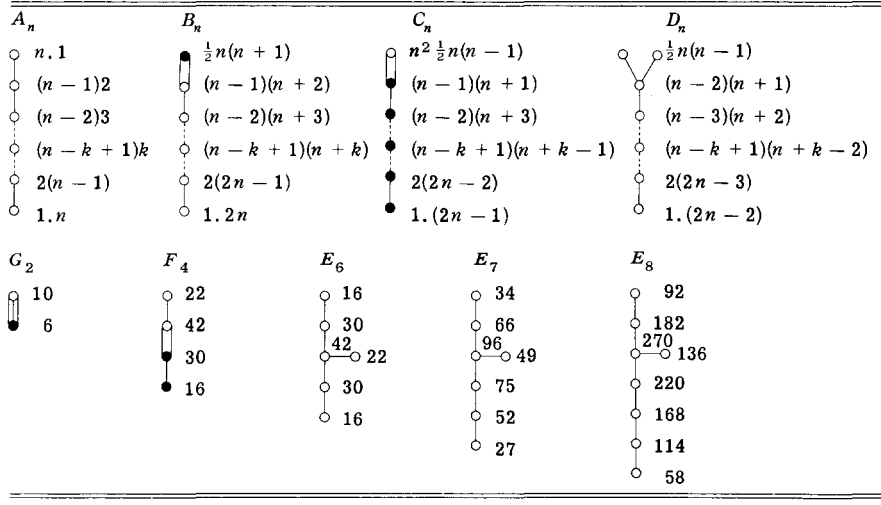


TABLE X. Defining matrix of semisimple maximal S-subalgebras of Table VII.

$A_n^{n-1} \subset A_{(1/2)(n-1)(n+2)}$ $n > 3$	$\frac{1}{n+1}$	$\begin{pmatrix} n & n & n & \cdots & -1 & -1 & -1 & \cdots & -1 & -1 & \cdots & -1 \\ n & -1 & -1 & \cdots & n & n & n & \cdots & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & n & -1 & -1 & \cdots & n & n & \cdots & -1 \\ -1 & -1 & n & \cdots & -1 & n & -1 & \cdots & n & -1 & \cdots & -1 \\ -1 & -1 & -1 & \cdots & -1 & -1 & n & \cdots & -1 & n & \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & -1 & \cdots & -1 & -1 & -1 & \cdots & -1 & -1 & \cdots & n \\ -1 & -1 & -1 & \cdots & -1 & -1 & -1 & \cdots & -1 & -1 & \cdots & n \end{pmatrix}$
$A_n^{n+3} \subset A_{(1/2)\cdot n(n+3)}$ $n \geq 2$	$\frac{1}{n+1}$	$\begin{pmatrix} 2n+1 & -1 & -1 & \cdots & -1 & n & n & \cdots & n & -1 & \cdots & -1 \\ -1 & 2n+1 & -1 & \cdots & -1 & n & -1 & \cdots & -1 & n & \cdots & -1 \\ -1 & -1 & 2n+1 & \cdots & -1 & -1 & n & \cdots & -1 & n & \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & -1 & \cdots & 2n+1 & -1 & -1 & \cdots & n & -1 & \cdots & n \end{pmatrix}$
$B_{2n+1}^1 \subset D_{2n+2}$		$\left.\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}\right\} 2n+2$
$G_2^1 \subset B_3$	$\frac{1}{3}$	$\begin{pmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \\ -2 & -1 & -1 \end{pmatrix}$
$A_1^{28} \subset G_2$		$(2 \quad 1 \quad -3)$
$A_5^6 \subset C_{10}$	$\frac{1}{2}$	$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \end{pmatrix}$
$B_4^2 \subset D_8$	$\frac{1}{2}$	$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$
$C_3^5 \subset C_7$		$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$
$D_5^4 \subset A_{15}$	$\frac{1}{2}$	$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix}$
$D_6^8 \subset C_{16}$	$\frac{1}{2}$	$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$
$E_7^{12} \subset C_{28}$	$\frac{1}{8}$	$\begin{pmatrix} 6 & 6 & 6 & 6 & 6 & 6 & 6 & -2 & -2 & -2 & -2 & -2 & -2 & \cdots & -2 \\ 6 & -2 & -2 & -2 & -2 & -2 & -2 & 6 & 6 & 6 & 6 & 6 & 6 & \cdots & -2 \\ -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & \cdots & -2 \\ -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & \cdots & -2 \\ -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & \cdots & -2 \\ -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & \cdots & -2 \\ -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & \cdots & 6 \\ -2 & -2 & -2 & -2 & -2 & -2 & 6 & -2 & -2 & -2 & -2 & -2 & 6 & \cdots & 6 \end{pmatrix}$
$E_6^6 \subset A_{26}$	$\frac{1}{6}$	$\begin{pmatrix} 5 & 5 & -1 & -1 & \cdots & -1 & -1 & 1 & 1 & 1 & 1 & \cdots & -5 & 0 & 0 & 0 \\ -1 & -1 & 5 & 5 & \cdots & -1 & -1 & 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & \cdots & -1 & -1 & 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & \cdots & -1 & -1 & 1 & 1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & \cdots & -1 & -1 & 1 & 1 & -5 & -5 & \cdots & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & \cdots & 5 & 5 & 5 & 5 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ 3 & -3 & 3 & -3 & \cdots & 3 & -3 & 3 & -3 & 3 & -3 & \cdots & -3 & 6 & -6 & 0 \\ -3 & 3 & -3 & 3 & \cdots & -3 & 3 & -3 & 3 & -3 & 3 & \cdots & 3 & -6 & 6 & 0 \end{pmatrix}$
$B_{n'}^1 + B_{n''}^1 \subset D_{n'+n''+1}$ $n' \geq n'' \geq 1$ $n' + n'' \geq 4$		$\left.\begin{pmatrix} 1 & 0 & 0 & \cdots & & & & & & & 0 & 0 \\ 0 & 1 & 0 & \cdots & & & & & & & 0 & 0 \\ 0 & 0 & 1 & \cdots & & & & & & & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & & & & & & & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & & 0 & 0 \\ \hline 0 & 0 & 0 & \cdots & & & & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & & & & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & & & & 0 & 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & & & & & & & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & & & & & & & 1 & 0 \end{pmatrix}\right\} \begin{matrix} n' \\ \\ n'' \end{matrix}$

class of conjugate subalgebras. An exception occurs in the case of subalgebras of D_n, when the representation induced on the subalgebra by the fundamental representation of $D_n(1, 0, \ldots, 0)$ does not split into two irreducible representations.[30] To every such type there corresponds two classes of nonequivalent subalgebras conjugated by an outer automorphism of D_n.

These nonequivalent but outer conjugate subalgebras of D_n will be linearly equivalent (L-equivalent) if and only if the corresponding defining matrices are equivalent.[31] The necessary and sufficient condition for that is the presence of the zero vector as a weight in the representation induced on the subalgebra by the fundamental representation of D_n.

(7) Given a maximal S-subalgebra $\tilde{G}$ of a classical algebra G, the N weights $M^{(k)}, k = 1, 2, \ldots, N$ of the fundamental (defining) representation ϕ of G project onto the N weights $m^{(k)}, k = 1, 2, \ldots, N$ of the representation $\tilde{\phi}$ of $\tilde{G}$. We have (Sec. 3)

$$m_i^{(k)} = \sum_{s=1}^{n} f_{is} M_s^{(k)}, \qquad i = 1, 2, 3, \ldots, n',$$

where n' and n are the rank of $\tilde{G}$ and G, respectively Choosing from the set of weights $M^{(k)}$ the n linearly independent weights of the form

$$(0, \ldots, 0, 1, 0, \ldots, 0),$$

we obtain the defining matrix, expressed in terms of the Cartesian coordinates $m_i^{(k)}$ (with a minor modification for A_n, due to the embedding of its subspace K in $R^{(n+1)}$, as given in Table IX).

The defining matrices of the exceptional cases are given in Table X.

The classification of the three-dimensional S-subalgebras of the classical algebras has been done previously in Sec. 7, using different theorems. The method as described in this section can obviously also be applied to classify them. It turns out that all the three-dimensional S-subalgebras of the algebras B_n and C_n are always maximal, except in the case $A_1^{28} \subset G_2^1 \subset B_3$; all the three-dimensional S-subalgebras of the algebras A_n and D_n are always nonmaximal, except in the case $A_1^4 \subset A_2$.

The simple maximal S-subalgebras of the exceptional algebras are

$$G_2: A_1^{28},$$
$$F_4: A_1^{156},$$
$$E_6: A_2^9, G_2^3, C_4^1, F_4^1,$$
$$E_7: A_1^{231}, A_1^{399}, A_2^{21},$$
$$E_8: A_1^{520}, A_1^{760}, A_1^{1240}, B_2^{12}.$$

To each type there corresponds one class of conjugate simple subalgebras, except in the cases A_2 and G_2 in E_6, to which there correspond two classes of conjugate subalgebras, transformable one into the other by an automorphism of E_6. The embedding of the simple roots and generators of all the maximal S-subalgebras of the exceptional algebras are given in Tables 14, 15, and 24 of Ref. 13.

10. NONSIMPLE MAXIMAL S-SUBALGEBRAS

Let U', U'' be two irreducible groups of dimension N', N'', respectively (i.e., defined and irreducible on $R^{(N')}$ and $R^{(N'')}$, respectively). By $U' \times U''$ we denote the direct product of these two groups. In the product space $R^{(N')} \times R^{(N'')}$ the algebra of this group is given by

$$G' \times 1'' + 1' \times G'',$$

TABLE XI. Embedding of the Generators I_1' of the Cartan subspace of a nonsimple maximal S-subalgebra in the Cartan subspace of a simple algebra.

Subalgebras of the type I:	I_1'	$= H_1 + H_2 + \ldots + H_{n''+1}$
	I_2'	$= H_{n''+2} + H_{n''+3} + \ldots + H_{2n''+2}$
	⋮	
	$I_{n'+1}'$	$= H_{(n''+1)n'+1} + \ldots + H_{(n'+1)(n''+1)}$
	I_1''	$= H_1 + H_{n''+2} + \ldots + H_{n'(n''+1)+1}$
	I_2''	$= H_2 + H_{n''+3} + \ldots + H_{n'(n''+2)+2}$
	⋮	
	$I_{n''+1}''$	$= H_{n''+1} + H_{2n''+2} + \ldots + H_{(n'+1)(n''+1)}$
Subalgebras of the type II. a, III, IV. a:	I_1'	$= H_1 + H_2 + \ldots + H_{2n''}$
	I_2'	$= H_{2n''+1} + H_{2n''+2} + \ldots + H_{4n''}$
	⋮	
	$I_{n'}'$	$= H_{2n''(n'-1)+1} + \ldots + H_{2n'n''}$
	I_1''	$= H_2 - H_{2n''} + H_{2n''+1} - H_{4n''} + \ldots + H_{2n''(n'-1)+1} - H_{2n'n''}$
	I_2''	$= H_2 - H_{2n''-1} + H_{2n''+2} - H_{4n''-1} + \ldots + H_{2n''(n'-1)+2} - H_{2n'n''-1}$
	⋮	
	$I_{n''}''$	$= H_{n''} - H_{n''+1} + H_{3n''} - H_{3n''+1} + \ldots + H_{n''(2n'-1)} - H_{n''(2n'-1)+1}$
Subalgebras of the type II. b, IV. b:	I_1'	$= H_1 + H_2 + \ldots + H_{2n''+1}$
	I_2'	$= H_{2n''+2} + H_{2n''+3} + \ldots + H_{4n''+2}$
	⋮	
	$I_{n'}'$	$= H_{(n'-1)(2n''+1)+1} + \ldots + H_{n'(2n''+1)}$
	I_1''	$= H_1 - H_{2n''+1} + H_{2n''+2} - H_{4n''+2} + \ldots + H_{(n'-1)(2n''+1)+1} - H_{n'(2n''+1)}$
	I_2''	$= H_2 - H_{2n''} + H_{2n''+3} - H_{4n''+1} + \ldots + H_{(n'-1)(2n''+1)+2} - H_{n'(2n''+1)-1}$
	⋮	
	$I_{n''}''$	$= H_{n''} - H_{n''+2} + H_{3n''+1} - H_{3n''+3} + \ldots + H_{(n'-1)(2n''+1)+n''} - H_{(n'-1)(2n''+1)+n''}$
Subalgebras of the type IV. c:	I_1'	$= H_1 + H_2 + \ldots + H_{2n''+1}$
	I_2'	$= H_{2n''+2} + H_{2n''+3} + \ldots + H_{4n''+2}$
	⋮	
	$I_{n'}'$	$= H_{(n'-1)(2n''+1)+1} + \ldots + H_{n'(2n''+1)}$
	I_1''	$= H_1 - H_{2n''+1} + H_{2n''+2} - H_{4n''+2} + \ldots + H_{2n'n''+n'-2n''} - H_{n'(2n''+1)}$
	I_2''	$= H_2 - H_{2n''} + H_{2n''+3} - H_{4n''+1} + \ldots + H_{2n'n''+n'-2n''+2} - H_{2n'n''+n'-1}$
	⋮	
	$I_{n''}''$	$= H_{n''} - H_{n''+2} + H_{3n''+1} - H_{3n''+3} \cdots + H_{2n'n''+n'+n''} - H_{2n'n''+n'+n''+2}$

where G' and G'' are the algebras of U' and U'', and $1'$ and $1''$ are unit matrices of dimension N' and N''.

The corresponding nonsimple maximal S-subalgebras with their indices are given by (with one exception)[32]

I.: $A_{n'}^{n''+1} + A_{n''}^{n'+1} \subset A_{(n'+1)(n''+1)-1}, \quad n' \geq n'' \geq 1;$

II. a: $C_{n'}^{2n''} + D_{n''}^{4n'} \subset C_{2n'n''}, \quad n' \geq 1, n'' \geq 3,$

b: $C_{n'}^{3} + A_1^{8n'} \subset C_{3n'}, \quad n' \geq 1,$

$$C_{n'}^{2n''+1} + B_{n''}^{4n'} \subset C_{n'(2n''+1)}, \quad n' \geq 1, n'' \geq 2;$$

$$A_1^4 + A_1^4 + A_1^4 \subset C_4$$

III.: $C_{n'}^{n''} + C_{n''}^{n'} \subset D_{2n'n''}, \quad n' \geq n'' \geq 1;$

IV. a: $D_{n'}^{2n''} + D_{n''}^{2n'} \subset D_{2n'n''}, \quad n' \geq n'' \geq 3,$

b: $D_{n'}^{2n''+1} + B_{n''}^{2n'} \subset D_{n'(2n''+1)}, \quad n' \geq 3, n'' \geq 2,$

$$D_{n'}^3 + A_1^{4n'} \subset D_{3n'}, \quad n' \geq 3,$$

c: $B_{n'}^3 + A_1^{2(2n'+1)} \subset B_{3n'+1}, \quad n' \geq 2,$

$$B_{n'}^{2n'+1} + B_{n''}^{2n'+1} \subset B_{2n'n+n'+n''}, \quad n' \geq n'' \geq 2.$$

The only exception is given at the end of Table VII.

The nonsimple maximal S-subalgebras of the exceptional algebras are

F_4: $G_2^1 + A_1^8$,
E_6: $G_2^1 + A_2^2$,
E_7: $G_2^1 + G_3^1$, $F_4^1 + A_1^3$, $G_2^1 + A_1^7$, $A_1^{24} + A_1^{15}$,
E_8: $G_2^1 + F_4^1$, $A_2^6 + A_1^{16}$.

The embedding of these subalgebras is given in Table 35 of Ref. 13.

The defining matrices of the nonsimple maximal S-subalgebras of the classical algebras can be obtained by expressing the generators $H' \times 1'' = I''$ and $1' \times H'' = I''$ of the subalgebra $G' + G''$ in terms of the generators H of G. This information is contained in Table XI.[33]

11. SEMISIMPLE NONMAXIMAL S-SUBALGEBRAS

The construction of all the semisimple S-subalgebras of a simple algebra is carried out by the following method[34]:

Let G be a simple algebra. We find all its maximal S-subalgebras (simple and nonsimple). These subalgebras are called *subalgebras of the first stage*. For each of the subalgebras of the first stage we find all the semisimple maximal subalgebras (regular as well as nonregular) and eliminate those which are R-subalgebras in G. The remaining set of algebras is called the *subalgebras of the second stage*. In general, if the *subalgebras of the* k*'th stage* have been obtained, the *subalgebras of the* $(k+1)$*th stage* are obtained by determining all semisimple maximal subalgebras in each subalgebra of the k'th stage, eliminating those subalgebras which are R-algebras in G and keeping the rest. (The schemes, given in Table XII, will serve to demonstrate the procedure. The subgroups listed along a horizontal line are subalgebras of the same stage. It should be noted, however, that there are inclusion relations between the subalgebras, i.e., a certain subalgebra may appear in different stages. The inclusion relations between the subalgebras are also given in Table XII.)

In order to carry out the construction described, it is essential to be able to determine all maximal subalgebras of a semisimple algebra. The following theorem of Dynkin solves this problem: Let

$$G = G_1 + G_2 + \cdots + G_s \tag{11.1}$$

be a decomposition of a semisimple algebra G into a direct sum of simple ideals. The collection of all maximal subalgebras of G is given by the formulas

$$G_1 + G_2 + \cdots + \tilde{G}_i + \cdots + G_s \tag{11.2}$$

and

$$\sum_{\substack{k=1 \\ k \neq i,j}}^{s} G_k + \{G_i + P_j G_i\}, \ (i < j). \tag{11.3}$$

In Eq. (11.2) the index i runs through the values $1, 2, \ldots, s$, while $\tilde{G}_i$ runs through all maximal subalgebras (regular and nonregular) of the subalgebra G_i of G. In formula (11.3) the indices i,j $(i < j)$ run through all possible pairs of indices for which G_i and G_j are isomorphic. P_j denotes an arbitrary isomorphic mapping of G_i on G_j, and $\{G_i + P_j G_i\}$ denotes the collection of all elements of the form

$$X + P_j X, \quad X \in G_i.$$

A necessary and sufficient condition that two algebras of the form (11.2) are conjugate in G is that two subalgebras $\tilde{G}_i$ and $\tilde{G}_i'$ of the subalgebra G_i of G are conjugate in G_i. A necessary and sufficient condition that two subalgebras of the form (11.3) are conjugate is that $i_1 = i_2$ and $j_1 = j_2$ [the subindices 1, 2 refer to the two subalgebras of form (11.3)] and that P_{j_1} and P_{j_2} are transformed into each other by some inner automorphism of G. The subalgebras (11.2) and (11.3) cannot be conjugate.

For subalgebras of the form (11.3) the index and the defining matrix of the algebra $\{G_i + P_j G_i\}$ is obtained by adding the indices and defining matrices of the two subalgebras G_i and G_j.

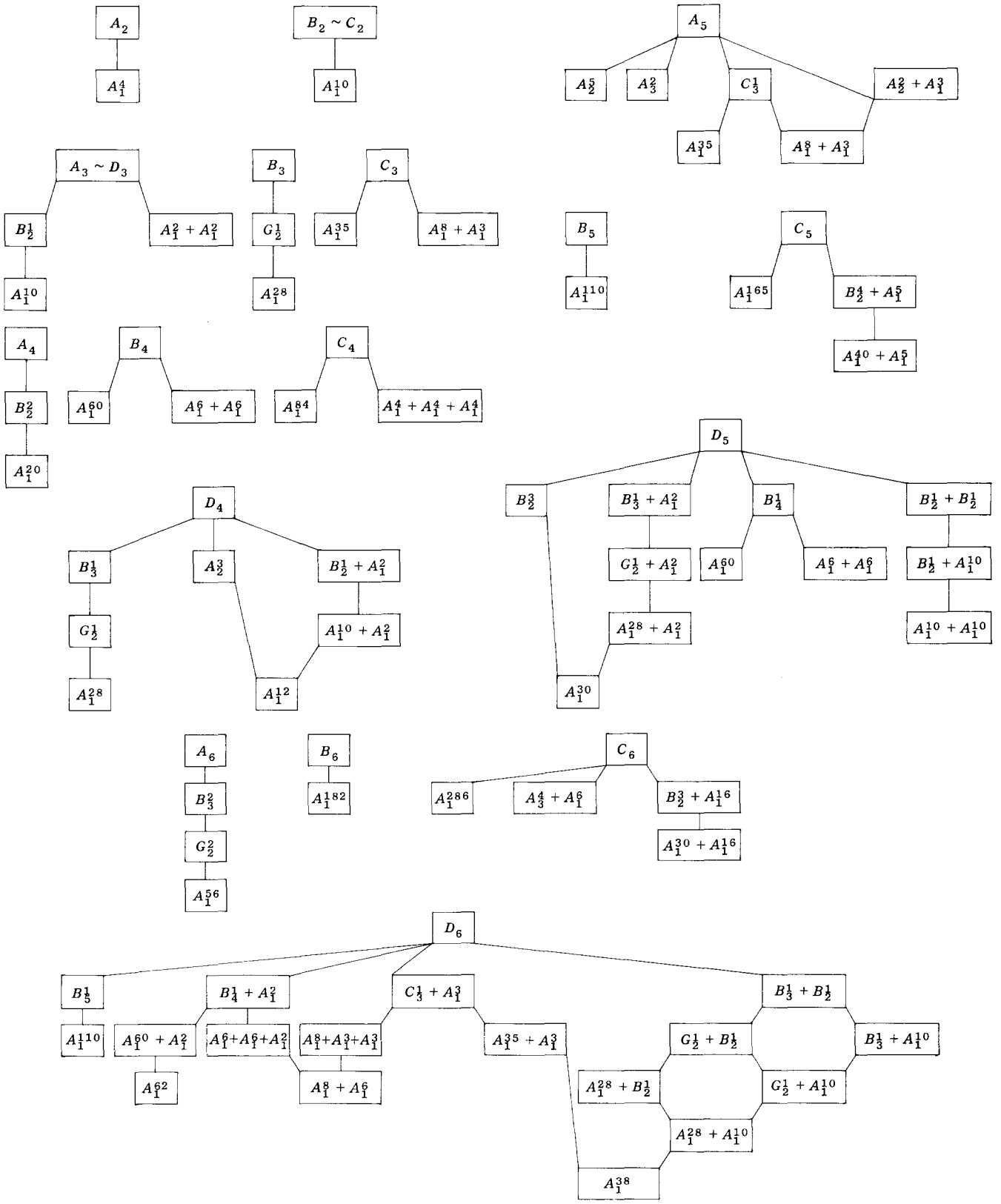
In Table XII all the semisimple S-subalgebras of the classical algebras up to rank 6 are given, as well as the inclusion relations among them. The theorem, quoted above, has been used in the derivation of this table. Thus, examples can be found in Table XII which will help to illustrate the content of this theorem. The semisimple S-subalgebras of the exceptional algebras and the inclusion relations among them are given in Table 39 of Ref. 13.

Table XIII contains all simple S-algebras of rank exceeding one for the classical algebras up to rank 6 (all rank 1 subalgebras, S-algebras, and non-S-algebras, are given in Table VI). In this table the embedding of the generators of the simple subalgebra $\tilde{G}$ in the algebra G is given explicitly, as well as the embedding of the simple roots of $\tilde{G}$ in G. Moreover the decomposition (branching) of the fundamental (defining) and adjoint representation of the simple algebra G into irreducible representations of its subalgebra $\tilde{G}$ is given. The notation used to denote an irreducible representation is $D^N(m)$, where m is the highest weight of the representation and N its dimension. The symbol $E_1, E_2, \ldots$, etc. denote the root vectors corresponding to the simple roots $\alpha_1, \alpha_2 \cdots$ according to the canonical ordering of Table I. The definition

$$E_{k_1 k_2 \cdots k_s} = [E_{k_1}, E_{k_2}], E_{k_s}], \cdots], E_{k_s}]$$

is used, where $E_{k_1 k_2 \cdots k_s}$ is the root vector corresponding to the root $\alpha_1 + \alpha_2 + \cdots + \alpha_s$, with α_i simple roots.

TABLE XII. Semisimple S-subalgebras of the classical algebras up to rank 6.



To each of the types of simple subalgebras given in Table XIII, there corresponds one class of conjugate subalgebras, except in the cases $A_2^3 \subset D_4$ where there are six classes of nonequivalent subalgebras conjugate by an outer automorphism of D_4; $B_2^3 \subset D_5$, where there are two classes of nonequivalent subalgebras conjugate by an automorphism of D_5; and $B_3 \subset D_4$, where there are three classes of nonequivalent subalgebras conjugate by an outer automorphism of D_4.

In Table XIV all nonsimple S-subalgebras of the classical algebras up to rank 6 are listed. The defining matrices of the subalgebras with respect to the algebra containing them, is given explicitly. A horizontal line in a defining matrix is used to distinguish

TABLE XIII.
Simple S-subalgebras of the classical algebras up to rank 6.

	$f(\tilde{H}_1) = H_1 + H_2$
	$f(\tilde{H}_2) = -H_2 + H_3$
	$f(\tilde{H}_3) = -H_1 - H_3$
	$f(\tilde{E}_1) = E_{12} + E_{23} + E_{24}$
$A_2^3 \subset D_4$	$f(\tilde{E}_2) = E_1 + \frac{1}{2}(1 \mp i\sqrt{3})E_3 + \frac{1}{2}(1 \pm i\sqrt{3})E_4$
	$f(1,-1,0) = (1,0,-1,0) + (0,1,0,-1) + (0,1,0,1)$
	$f(0,1,-1) = (1,-1,0,0) + (0,0,1,-1) + (0,0,1,1)$
	$D^{28}(1,1,0,0) \to D^8(1,0,-1) + D^{10}(2,-1,-1) + D^{10}(1,1,-2)$
	$D^8(1,0,0,0) \to D^8(1,0,-1)$
	$f(\tilde{H}_1) = H_1 + H_2 + H_3 - H_6$
	$f(\tilde{H}_2) = H_1 - H_3 + H_4 + H_6$
	$f(\tilde{H}_3) = -H_1 + H_3 + H_5 + H_6$
	$f(\tilde{E}_1) = \sqrt{2}(E_1 + E_{23}) + E_{34}$
	$f(\tilde{E}_2) = E_2 + \sqrt{2}(E_4 + E_5)$
$A_2^5 \subset A_5$	$f(1,-1,0) = 2(1,-1,0,0,0,0) + 2(0,1,0,-1,0,0) + (0,0,1,0,-1)$
	$f(0,1,-1) = (0,1,-1,0,0,0) + 2(0,0,0,1,-1,0) + 2(0,0,0,0,1,-1)$
	$D^{35}(1,0,0,0,0,-1) \to D^8(1,0,-1) + D^{27}(2,0,-2)$
	$D^6(\frac{5}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6}) \to D^6(\frac{2}{3},\frac{2}{3},-\frac{4}{3})$
	$f(\tilde{H}_1) = \frac{1}{2}(H_1 + H_2 - H_3 - H_4)$
	$f(\tilde{H}_2) = \frac{1}{2}(H_1 - H_2 + H_3 - H_4)$
	$f(\tilde{E}_1) = E_2$
$B_2^1 \subset A_3$	$f(\tilde{E}_2) = (1/\sqrt{2})(E_1 + E_3)$
	$f(1,-1) = (0,1,-1,0)$
	$f(0,1) = \frac{1}{2}(1,-1,0,0) + \frac{1}{2}(0,0,1,-1)$
	$D^{15}(1,0,0,-1) \to D^{10}(1,1) + D^5(1,0)$
	$D^4(\frac{3}{4},-\frac{1}{4},-\frac{1}{4},-\frac{1}{4}) \to D^4(\frac{1}{2},\frac{1}{2})$
	$f(\tilde{H}_1) = H_1 - H_5$
	$f(\tilde{H}_2) = H_2 - H_4$
	$f(\tilde{E}_1) = E_1 + E_4$
	$f(\tilde{E}_2) = E_2 + E_3$
$B_2^2 \subset A_4$	$f(1,-1) = (1,-1,0,0,0) + (0,0,0,1,-1)$
	$f(0,1) = (0,1,-1,0,0) + (0,0,1,-1,0)$
	$D^{24}(1,0,0,0,-1) \to D^{10}(1,1) + D^{14}(2,0)$
	$D^5(\frac{4}{5},-\frac{1}{5},-\frac{1}{5},-\frac{1}{5},-\frac{1}{5}) \to D^5(1,0)$
	$f(\tilde{H}_1) = H_1 + H_2 + H_3$
	$f(\tilde{H}_2) = H_1 - H_3 + H_4$
	$f(\tilde{E}_1) = E_{23} + E_{34} + E_{35}$
	$f(\tilde{E}_2) = E_1 + E_2 + \frac{1}{2}(1 \pm i)E_4 + \frac{1}{2}(1 \mp i)E_5$
$B_2^3 \subset D_5$	$f(1,-1) = (0,1,0,-1,0) + (0,0,1,0,-1) + (0,0,1,0,1)$
	$f(0,1) = (1,-1,0,0,0) + (0,1,-1,0,0) + \frac{1}{2}(0,0,0,1,-1) + \frac{1}{2}(0,0,0,1,1)$
	$D^{45}(1,1,0,0,0) \to D^{10}(1,1) + D^{35}(2,1)$
	$D^{10}(1,0,0,0,0) \to D^{10}(1,1)$
	$f(\tilde{H}_1) = \frac{1}{3}(H_1 + 2H_2 - H_3)$
	$f(\tilde{H}_2) = \frac{1}{3}(H_1 - H_2 + 2H_3)$
	$f(\tilde{H}_3) = \frac{1}{3}(-2H_1 - H_2 - H_3)$
	$f(\tilde{E}_1) = E_2$
$G_2^1 \subset B_3$	$f(\tilde{E}_2) = (1/\sqrt{3})E_1 + \sqrt{2/3}\,E_3$
	$f(1,-1,0) = (0,1,-1)$
	$f(-\frac{1}{3},\frac{2}{3},-\frac{1}{3}) = \frac{1}{3}(1,-1,0) + \frac{2}{3}(0,0,1)$
	$D^{21}(1,1,0) \to D^{14}(1,0,-1) + D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3})$
	$D^7(1,0,0) \to D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3})$
	$f(\tilde{H}_1) = \frac{1}{3}(H_1 + 2H_2 - H_3)$
	$f(\tilde{H}_2) = \frac{1}{3}(H_1 - H_2 + 2H_3)$
	$f(\tilde{H}_3) = \frac{1}{3}(-2H_1 - H_2 - H_3)$
	$f(\tilde{E}_1) = E_2$
	$f(\tilde{E}_2) = (1/\sqrt{3})(E_1 + E_3 + E_4)$
$G_2^1 \subset D_4$	$f(1,-1,0) = (0,1,-1,0)$
	$f(-\frac{1}{3},\frac{2}{3},-\frac{1}{3}) = \frac{1}{3}(1,-1,0,0) + \frac{1}{3}(0,0,1,-1) + \frac{1}{3}(0,0,1,1)$
	$D^{28}(1,1,0,0) \to D^{14}(1,0,-1) + 2D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3})$
	$D^8(1,0,0,0) \to D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3}) + D^1(0,0,0)$
	$f(\tilde{H}_1) = \frac{1}{3}(2H_1 + 3H_2 + H_4 + 2H_5 - H_6)$
	$f(\tilde{H}_2) = \frac{1}{3}(2H_1 + 3H_3 + H_4 - H_5 + 2H_6)$
	$f(\tilde{H}_3) = \frac{1}{3}(-H_1 + H_4 + 2H_5 + 2H_6 + 3H_7)$
	$f(\tilde{E}_1) = E_2 + E_5$
	$f(\tilde{E}_2) = (1/\sqrt{3})(E_1 + E_6) + \sqrt{2/3}(E_3 + E_4)$
$G_2^2 \subset A_6$	$f(1,-1,0) = (0,1,-1,0,0,0,0) + (0,0,0,0,1,-1,0)$
	$f(-\frac{1}{3},\frac{2}{3},-\frac{1}{3}) = \frac{1}{3}(1,-1,0,0,0,0,0) + \frac{1}{3}(0,0,0,0,0,1,-1) + \frac{2}{3}(0,0,1,-1,0,0,0) + \frac{2}{3}(0,0,0,1,-1,0,0)$
	$D^{48}(1,0,0,0,0,0,-1) \to D^{14}(1,0,-1) + D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3}) + D^{27}(\frac{2}{3},\frac{2}{3},-\frac{4}{3})$
	$D^7(\frac{6}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7}) \to D^7(\frac{1}{3},\frac{1}{3},-\frac{2}{3})$
	$f(\tilde{H}_1) = \frac{1}{4}(3H_1 + 3H_2 + 3H_3 - H_4 - H_5 - H_6)$
	$f(\tilde{H}_2) = \frac{1}{4}(3H_1 - H_2 - H_3 + 3H_4 + 3H_5 - H_6)$
	$f(\tilde{H}_3) = \frac{1}{4}(-H_1 + 3H_2 - H_3 + 3H_4 - H_5 + 3H_6)$
	$f(\tilde{H}_4) = \frac{1}{4}(-H_1 - H_2 + 3H_3 - H_4 + 3H_5 + 3H_6)$
	$f(\tilde{E}_1) = E_{23} + E_{34}$
	$f(\tilde{E}_2) = E_1 + E_5$
$A_3^2 \subset A_5$	$f(\tilde{E}_3) = E_2 + E_4$
	$f(1,-1,0,0) = (0,1,0,-1,0,0) + (0,0,1,0,-1,0)$
	$f(0,1,-1,0) = (1,-1,0,0,0,0) + (0,0,0,0,1,-1)$
	$f(0,0,1,-1) = (0,1,-1,0,0,0) + (0,0,0,1,-1,0)$
	$D^{35}(1,0,0,0,0,-1) \to D^{15}(1,0,0,-1) + D^{20}(1,1,-1,-1)$
	$D^6(\frac{6}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6}) \to D^6(\frac{1}{2},\frac{1}{2},-\frac{1}{2},-\frac{1}{2})$
	$f(\tilde{H}_1) = H_1$
	$f(\tilde{H}_2) = H_2$
	$f(\tilde{H}_3) = H_3$
	$f(\tilde{E}_1) = E_1$
	$f(\tilde{E}_2) = E_2$
$B_3^1 \subset D_4$	$f(\tilde{E}_3) = (1/\sqrt{2})(E_3 + E_4)$
	$f(1,-1,0) = (1,-1,0,0)$
	$f(0,1,-1) = (0,1,-1,0)$
	$f(0,0,1) = \frac{1}{2}(0,0,1,-1) + \frac{1}{2}(0,0,1,1)$
	$D^{28}(1,1,0,0) \to D^{21}(1,1,0) + D^7(1,0,0)$
	$D^8(1,0,0,0) \to D^7(1,0,0) + D^1(0,0,0)$
	$f(\tilde{H}_1) = H_1 - H_7$
	$f(\tilde{H}_2) = H_2 - H_6$
	$f(\tilde{H}_3) = H_3 - H_5$
	$f(\tilde{E}_1) = E_1 + E_6$

TABLE XIII contd

$B_3^2 \subset A_6$	$f(\tilde{E}_2) = E_2 + E_5$ $f(\tilde{E}_3) = E_3 + E_4$ $f(1,-1,0) = (1,-1,0,0,0,0,0) + (0,0,0,0,0,1,-1)$ $f(0,1,-1) = (0,1,-1,0,0,0,0) + (0,0,0,0,1,-1,0)$ $f(0,0,1) = (0,0,1,-1,0,0,0) + (0,0,0,1,-1,0,0)$ $D^{48}(1,0,0,0,0,0,-1) \to D^{21}(1,1,0) + D^{27}(2,0,0)$ $D^{7}(\frac{6}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},-\frac{1}{7}) \to D^{7}(1,0,0)$
$C_3^1 \subset A_5$	$f(\tilde{H}_1) = H_1 - H_6$ $f(\tilde{H}_2) = H_2 - H_5$ $f(\tilde{H}_3) = H_3 - H_4$ $f(\tilde{E}_1) = E_1 + E_5$ $f(\tilde{E}_2) = E_2 + E_4$ $f(\tilde{E}_3) = \sqrt{2}\,E_3$ $f(1,-1,0) = (1,-1,0,0,0,0) + (0,0,0,0,1,-1)$ $f(0,1,-1) = (0,1,-1,0,0,0) + (0,0,0,1,-1,0)$ $f(0,0,2) = 2(0,0,1,-1,0,0)$ $D^{35}(1,0,0,0,0,-1) \to D^{21}(2,0,0) + D^{14}(1,1,0)$ $D^{6}(\frac{5}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6}) \to D^{6}(1,0,0)$
	$f(\tilde{H}_1) = H_1$ $f(\tilde{H}_2) = H_2$ $f(\tilde{H}_3) = H_3$ $f(\tilde{H}_4) = H_4$ $f(\tilde{E}_1) = E_1$ $f(\tilde{E}_2) = E_2$ $f(\tilde{E}_3) = E_3$ $f(\tilde{E}_4) = (1/\sqrt{2})(E_4 + E_5)$
$B_4^1 \subset D_5$	$f(1,-1,0,0) = (1,-1,0,0,0)$ $f(0,1,-1,0) = (0,1,-1,0,0)$ $f(0,0,1,-1) = (0,0,1,-1,0)$ $f(0,0,0,1) = \frac{1}{2}(0,0,0,1,-1) + \frac{1}{2}(0,0,0,1,1)$ $D^{45}(1,1,0,0,0) \to D^{36}(1,1,0,0,) + D^{9}(1,0,0,0)$ $D^{10}(1,0,0,0,0) \to D^{9}(1,0,0,0) + D^{1}(0,0,0,0)$
$B_5 \subset D_6$	$f(\tilde{H}_1) = H_1$ $f(\tilde{H}_2) = H_2$ $f(\tilde{H}_3) = H_3$ $f(\tilde{H}_4) = H_4$ $f(\tilde{H}_5) = H_5$ $f(\tilde{E}_1) = E_1$ $f(\tilde{E}_2) = E_2$ $f(\tilde{E}_3) = E_3$ $f(\tilde{E}_4) = E_4$ $f(\tilde{E}_5) = (1/\sqrt{2})(E_5 + E_6)$
	$f(1,-1,0,0,0) = (1,-1,0,0,0,0)$ $f(0,1,-1,0,0) = (0,1,-1,0,0,0)$ $f(0,0,1,-1,0) = (0,0,1,-1,0,0)$ $f(0,0,0,1,-1) = (0,0,0,1,-1,0)$ $f(0,0,0,0,1) = \frac{1}{2}(0,0,0,0,1,-1) + \frac{1}{2}(0,0,0,0,1,1)$ $D^{66}(1,1,0,0,0,0) \to D^{55}(1,1,0,0,0) + D^{11}(1,0,0,0,0)$ $D^{12}(1,0,0,0,0,0) \to D^{11}(1,0,0,0,0) + D^{1}(0,0,0,0,0)$

TABLE XIV. Nonsimple S-subalgebras of the classical algebras up to rank 6.

$A_1^2 + A_1^2 \subset A_3$	$\frac{1}{2}\begin{pmatrix}1&1&-1&-1\\1&-1&1&-1\end{pmatrix}$	$D^4(\frac{3}{4},-\frac{1}{4},-\frac{1}{4},-\frac{1}{4}) \to D^{2\times 2}(\frac{1}{2};\frac{1}{2})$
$A_1^8 + A_1^3 \subset C_3$	$\frac{1}{2}\begin{pmatrix}2&0&-2\\1&1&1\end{pmatrix}$	$D^6(1,0,0) \to D^{3\times 2}(1;\frac{1}{2})$
$A_1^4 + A_1^4 + A_1^4 \subset C_4$	$\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&1&-1\end{pmatrix}$	$D^8(1,0,0,0) \to D^{2\times 2\times 2}(\frac{1}{2};\frac{1}{2};\frac{1}{2})$
$A_1^6 + A_1^6 \subset B_4$	$\begin{pmatrix}1&1&1&0\\1&0&-1&1\end{pmatrix}$	$D^9(1,0,0,0) \to D^{3\times 3}(1;1)$
$A_1^{10} + A_1^2 \subset D_4$	$\frac{1}{2}\begin{pmatrix}3&1&-1&-3\\1&1&1&1\end{pmatrix}$	$D^8(1,0,0,0) \to D^{4\times 2}(\frac{3}{2};\frac{1}{2})$
$A_1^8 + A_1^3 \subset A_5$	$\frac{1}{2}\begin{pmatrix}2&0&-2&2&0&-2\\1&1&1&1&1&1\end{pmatrix}$	$D^6(\frac{5}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6}) \to D^{3\times 2}(1;\frac{1}{2})$
$A_1^{10} + A_1^2 \subset D_4$	$\begin{pmatrix}2&1&0&0\\0&0&1&0\end{pmatrix}$	$D^8(1,0,0,0) \to D^{5\times 1}(2;0) + D^{1\times 3}(0;1)$
$A_1^{40} + A_1^5 \subset C_5$	$\frac{1}{2}\begin{pmatrix}4&2&0&-2&-4\\1&1&1&1&1\end{pmatrix}$	$D^{10}(1,0,0,0,0) \to D^{5\times 2}(3;\frac{1}{2})$
$A_1^6 + A_1^6 \subset D_5$	$\begin{pmatrix}1&1&1&0&0\\1&0&-1&1&0\end{pmatrix}$	$D^{10}(1,0,0,0,0) \to D^{3\times 3}(1;1) + D^{1\times 1}(0;0)$
$A_1^{10} + A_1^{10} \subset D_5$	$\begin{pmatrix}2&1&0&0&0\\0&0&2&1&0\end{pmatrix}$	$D^{10}(1,0,0,0,0) \to D^{5\times 1}(2;0) + D^{1\times 5}(0;2)$
$A_1^{30} + A_1^{16} \subset C_6$	$\frac{1}{2}\begin{pmatrix}3&3&3&1&1&1\\2&0&-2&2&0&-2\end{pmatrix}$	$D^{12}(1,0,0,0,0,0) \to D^{4\times 3}(\frac{3}{2};1)$
$A_1^8 + A_1^3 + A_1^3 \subset D_6$	$\frac{1}{2}\begin{pmatrix}2&0&-2&2&0&-2\\1&1&1&-1&-1&-1\\1&1&1&1&1&1\end{pmatrix}$	$D^{12}(1,0,0,0,0,0) \to D^{3\times 2\times 2}(1;\frac{1}{2};\frac{1}{2})$
$A_1^8 + A_1^6 \subset D_6$	$\begin{pmatrix}1&0&-1&1&0&-1\\1&1&1&0&0&0\end{pmatrix}$	$D^{12}(1,0,0,0,0,0) \to D^{3\times 3}(1;1) + D^{3\times 1}(1;0)$
$A_1^{35} + A_1^3 \subset D_6$	$\frac{1}{2}\begin{pmatrix}5&3&1&-1&-3&-5\\1&1&1&1&1&1\end{pmatrix}$	$D^{12}(1,0,0,0,0,0) \to D^{6\times 2}(\frac{5}{2};\frac{1}{2})$
$A_1^{28} + A_1^{10} \subset D_6$	$\begin{pmatrix}3&2&1&0&0&0\\0&0&0&2&1&0\end{pmatrix}$	$D^{12}(1,0,0,0,0,0) \to D^{7\times 1}(3;0) + D^{1\times 5}(0;2)$
$A_2^2 + A_1^3 \subset A_5$	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\\\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\end{pmatrix}$	$D^6(\frac{5}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6},-\frac{1}{6}) \to D^{3\times 2}(\frac{2}{3},-\frac{1}{3},-\frac{1}{3};\frac{1}{2})$

TABLE XIV contd

$B_2^1 + A_1^2 \subset D_4$	$\left(\begin{array}{cccc} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ \hline \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right)$	$D^8(1,0,0,0) \to D^{4\times 2}(1,0;\frac{1}{2})$
$B_2^1 + A_1^2 \subset D_4$	$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array}\right)$	$D^8(1,0,0,0) \to D^{5\times 1}(1,0;0) + D^{1\times 3}(0,0;1)$
$B_2^4 + A_1^5 \subset C_5$	$\left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ \hline \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{5\times 2}(1,0;\frac{1}{2})$
$B_2^1 + A_1^{10} \subset D_5$	$\left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 & 0 \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{5\times 1}(1,0;0) + D^{1\times 5}(0,0;2)$
$B_2^3 + A_1^{16} \subset C_6$	$\left(\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & -1 & 1 & 0 & -1 \end{array}\right)$	$D^{12}(1,0,0,0,0,0,0) \to D^{4\times 3}(1,0;1)$
$B_2^1 + A_1^{28} \subset D_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 3 & 2 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{5\times 1}(1,0;0) + D^{1\times 7}(0,0;3)$
$B_2^1 + B_2^1 \subset D_5$	$\left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{5\times 1}(1,0;0,0) + D^{1\times 5}(0,0;1,0)$
$G_2^1 + B_2^1 \subset D_6$	$\frac{1}{3}\left(\begin{array}{cccccc} 1 & 2 & -1 & 0 & 0 & 0 \\ 1 & -1 & 2 & 0 & 0 & 0 \\ -2 & -1 & -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{1\times 5}(0,0,0;1,0) + D^{7\times 1}(\frac{1}{3},\frac{1}{3},-\frac{2}{3};0,0)$
$G_2^1 + A_1^{10} \subset D_6$	$\frac{1}{3}\left(\begin{array}{cccccc} 1 & 2 & -1 & 0 & 0 & 0 \\ 1 & -1 & 2 & 0 & 0 & 0 \\ -2 & -1 & -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 6 & 3 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{1\times 5}(0,0,0;2) + D^{7\times 1}(\frac{1}{3},\frac{1}{3},-\frac{2}{3};0)$
$A_3^4 + A_1^6 \subset C_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ \hline \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right)$	$D^{12}(1.0,0,0,0,0) \to D^{6\times 2}(1,0,0;\frac{1}{2})$
$B_3^1 + B_2^1 \subset D_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{7\times 1}(1,0,0;0,0) + D^{1\times 5}(0,0,0;1,0)$
$C_3^1 + A_1^3 \subset D_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ \hline \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{6\times 2}(1,0,0;\frac{1}{2})$
$B_3^1 + A_1^{10} \subset D_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{7\times 1}(1,0,0;0) + D^{1\times 5}(0,0,0;2)$
$B_3^1 + A_1^2 \subset D_5$	$\left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{7\times 1}(1,0,0;0) + D^{1\times 3}(0,0,0;1)$
$G_2^1 + A_1^2 \subset D_5$	$\frac{1}{3}\left(\begin{array}{ccccc} 1 & 2 & -1 & 0 & 0 \\ 1 & -1 & 2 & 0 & 0 \\ -2 & -1 & -1 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 0 \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{7\times 1}(\frac{1}{3},\frac{1}{3},-\frac{2}{3};0) + D^{1\times 3}(0,0,0;1)$
$A_1^{28} + A_1^2 \subset D_5$	$\left(\begin{array}{ccccc} 3 & 2 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{10}(1,0,0,0,0) \to D^{7\times 1}(3;0) + D^{1\times 3}(0;1)$
$B_4^1 + A_1^2 \subset D_6$	$\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{9\times 1}(1,0,0,0;0) + D^{1\times 3}(0,0,0,0;1)$
$A_1^{60} + A_1^2 \subset D_6$	$\left(\begin{array}{cccccc} 4 & 3 & 2 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{9\times 1}(4;0) + D^{1\times 3}(0;1)$
$A_1^6 + A_1^6 + A_1^2 \subset D_6$	$\left(\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$	$D^{12}(1,0,0,0,0,0) \to D^{3\times 3\times 1}(1;1;0) + D^{1\times 1\times 3}(0;0;1)$

the parts of a defining matrix referring to the various simple ideals of a semisimple subalgebra. In addition, the decomposition of the fundamental representation of the simple algebra into irreducible representations of its semisimple subalgebra is listed.

12. SEMISIMPLE R-SUBALGEBRAS

The R-subalgebras of the simple algebras are obtained in a way similar to the case of the 3d-subalgebras.

TABLE XV. Simple subalgebras of the type A_2.

Algebra	Regular subalgebra	Index	Defining matrix
A_3	A_2	1'	$\frac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \end{pmatrix}$
	A_2	1"	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & -2 \\ -1 & -2 & 1 \end{pmatrix}$
B_3	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & -2 \\ -1 & -2 & 1 \end{pmatrix}$
C_3	A_2^2	2	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & -2 \\ -1 & -2 & 1 \end{pmatrix}$
A_4	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 \end{pmatrix}$
B_4	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 \\ -1 & 1 & -2 & 0 \\ -1 & -2 & 1 & 0 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 \end{pmatrix}$
C_4	A_2^2	2	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 \\ -1 & 1 & -2 & 0 \\ -1 & -2 & 1 & 0 \end{pmatrix}$
D_4	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 \\ -1 & 1 & -2 & 0 \\ -1 & -2 & 1 & 0 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 0 \end{pmatrix}$
A_5	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \end{pmatrix}$
	$2A_2$	2	$\frac{1}{3}\begin{pmatrix} 2 & 2 & -1 & -1 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & -1 & -1 & 2 & 2 \end{pmatrix}$
	A_5	5	$\frac{1}{3}\begin{pmatrix} 2 & 2 & 2 & -1 & -1 & -4 \\ 2 & -1 & -4 & 2 & -1 & 2 \\ -4 & -1 & 2 & -1 & 2 & 2 \end{pmatrix}$
B_5	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 \end{pmatrix}$
C_5	A_2^2	2	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 \end{pmatrix}$
D_5	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 \end{pmatrix}$
A_6	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}$
	$2A_2$	2	$\frac{1}{3}\begin{pmatrix} 2 & 2 & -1 & -1 & -1 & -1 & 0 \\ -1 & -1 & 2 & 2 & -1 & -1 & 0 \\ -1 & -1 & -1 & -1 & 2 & 2 & 0 \end{pmatrix}$
	A_5	5	$\frac{1}{3}\begin{pmatrix} 2 & 2 & 2 & 0 & -1 & -1 & -4 \\ 2 & -1 & -4 & 0 & 2 & -1 & 2 \\ -4 & -1 & 2 & 0 & -1 & 2 & 2 \end{pmatrix}$
B_6	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 & 0 \end{pmatrix}$
	$2A_2$	2	$\frac{1}{3}\begin{pmatrix} 2 & 2 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -2 & -2 \\ -1 & -1 & -2 & -2 & 1 & 1 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}$
	A_5	5	$\frac{1}{3}\begin{pmatrix} 4 & 2 & 2 & 2 & 1 & 1 \\ -2 & 2 & -1 & -4 & 1 & -2 \\ -2 & -4 & -1 & 2 & -2 & 1 \end{pmatrix}$
C_6	A_2^2	2	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 & 0 \end{pmatrix}$
	$2A_2^2$	4	$\frac{1}{3}\begin{pmatrix} 2 & 2 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -2 & -2 \\ -1 & -1 & -2 & -2 & 1 & 1 \end{pmatrix}$
	A_5^2	10	$\frac{1}{3}\begin{pmatrix} 4 & 2 & 2 & 2 & 1 & 1 \\ -2 & 2 & -1 & -4 & 1 & -2 \\ -2 & -4 & -1 & 2 & -2 & 1 \end{pmatrix}$
D_6	A_2	1	$\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 & 0 & 0 \\ -1 & -2 & 1 & 0 & 0 & 0 \end{pmatrix}$
	$2A_2$	2	$\frac{1}{3}\begin{pmatrix} 2 & 2 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -2 & -2 \\ -1 & -1 & -2 & -2 & 1 & 1 \end{pmatrix}$
	D_4	3	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}$
	A_5	5	$\frac{1}{3}\begin{pmatrix} 4 & 2 & 2 & 2 & 1 & \pm 1 \\ -2 & 2 & -1 & -4 & 1 & \mp 1 \\ -2 & -4 & -1 & 2 & -2 & \pm 1 \end{pmatrix}$

TABLE XVI. Simple subalgebras of the type $B_2 \sim C_2$.

Algebra	Regular subalgebra	Index	Defining matrix $\tilde{G} = (e_1 - e_2; e_2)$	Defining matrix $\tilde{G} = (e_1 - e_2; 2e_2)$
A_3	A_3	1	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix}$
B_3	$A_3; B_2$	1	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$
C_3	C_2	1	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
A_4	A_3	1	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix}$
	A_4	2	$\begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 & -1 \end{pmatrix}$
B_4	$A_3; B_2$	1'	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}$
	A_3	1"	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$
C_4	C_2	1	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

TABLE XVI contd

	$A_3^2; 2C_2$	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&1&-1&-1\end{pmatrix}$	$\begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix}$
D_4	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&\pm1\\1&1&-1&\mp1\end{pmatrix}$	$\begin{pmatrix}1&1&0&0\\0&0&1&\pm1\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0\\1&-1&0&0\end{pmatrix}$
A_5	A_3	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&-1&-1\\1&-1&0&0&1&-1\end{pmatrix}$	$\begin{pmatrix}1&0&0&0&0&-1\\0&1&0&0&-1&0\end{pmatrix}$
	A_4	2	$\begin{pmatrix}1&0&0&0&0&-1\\0&1&0&0&-1&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&-1&-1\\1&-1&0&0&1&-1\end{pmatrix}$
B_5	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$
	$A_3; B_2$	1"	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0\\1&-1&0&0&0\end{pmatrix}$
	$A_4; A_3 + B_2$	2	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$
	D_5	3	$\begin{pmatrix}1&1&1&0&0\\1&0&-1&1&0\end{pmatrix}$	$\begin{pmatrix}2&1&1&0&0\\0&1&-1&2&0\end{pmatrix}$
C_5	C_2	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&0\\1&-1&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\end{pmatrix}$
	$A_3^2; 2C_2$	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$
	A_4^2	4	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$
D_5	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0\\1&-1&0&0&0\end{pmatrix}$
	A_4	2	$\begin{pmatrix}1&1&0&0&0\\0&0&1&1&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\end{pmatrix}$
	D_5	3	$\begin{pmatrix}1&1&1&0&0\\1&0&-1&1&0\end{pmatrix}$	$\begin{pmatrix}2&1&1&0&0\\0&1&-1&2&0\end{pmatrix}$
A_6	A_3	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&0&-1&-1\\1&-1&0&0&0&1&-1\end{pmatrix}$	$\begin{pmatrix}1&0&0&0&0&0&1\\0&1&0&0&0&-1&0\end{pmatrix}$
	A_4	2	$\begin{pmatrix}1&0&0&0&0&0&-1\\0&1&0&0&0&-1&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&-1&-1\\1&-1&0&0&0&1&-1\end{pmatrix}$
B_6	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$
	$A_3; B_2$	1"	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&0\\1&-1&0&0&0&0\end{pmatrix}$
	$A_3 + B_2$	2'	$\begin{pmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0\\0&\frac{1}{2}&\frac{1}{2}&-\frac{1}{2}&-\frac{1}{2}&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0&0\\1&-1&0&0&1&1\end{pmatrix}$
	$A_3 + B_2; A_4; 2A_3$	2"	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$
	D_5	3	$\begin{pmatrix}1&1&1&0&0&0\\1&0&-1&1&0&0\end{pmatrix}$	$\begin{pmatrix}2&1&1&0&0&0\\0&1&-1&2&0&0\end{pmatrix}$
C_6	C_2	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&0&0\\1&-1&0&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\end{pmatrix}$
	$A_3^2; 2C_2$	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$
	$A_3^2 + C_2; 3C_2$	3	$\frac{1}{2}\begin{pmatrix}1&1&1&1&1&1\\1&1&1&-1&-1&-1\end{pmatrix}$	$\begin{pmatrix}1&1&1&0&0&0\\0&0&0&1&1&1\end{pmatrix}$
	A_4^2	4	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$
D_6	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&0&0&0&0\\1&-1&0&0&0&0\end{pmatrix}$
	$A_4; 2A_3$	2	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\end{pmatrix}$
	D_5	3	$\begin{pmatrix}1&1&1&0&0&0\\1&0&-1&1&0&0\end{pmatrix}$	$\begin{pmatrix}2&1&1&0&0&0\\0&1&-1&2&0&0\end{pmatrix}$

Let G be an arbitrary simple algebra. We take a regular subalgebra of G from each class of conjugate regular subalgebras. Suppose such a regular subalgebra is the direct sum of simple ideals

$$G' = G_1' + G_2' + \cdots + G_s'.$$

In this subalgebra G' we take a complete set of nonconjugate semisimple S-subalgebras. If $\tilde{G}_i$ is a semi-

TABLE XVII. Subalgebras of the type G_2.

Algebra	Regular subalgebra	Index	Defining matrix
B_3	B_3	1	$\frac{1}{3}\begin{pmatrix}1&2&-1\\1&-1&2\\-2&-1&-1\end{pmatrix}$
B_4	$B_3;D_4$	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0\\1&-1&2&0\\-2&-1&-1&0\end{pmatrix}$
D_4	D_4	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0\\1&-1&2&0\\-2&-1&-1&0\end{pmatrix}$
B_5	$B_3;D_4$	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0&0\\1&-1&2&0&0\\-2&-1&-1&0&0\end{pmatrix}$
D_5	D_4	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0&0\\1&-1&2&0&0\\-2&-1&-1&0&0\end{pmatrix}$
A_6	A_6	2	$\frac{1}{3}\begin{pmatrix}1&2&-1&0&1&-2&-1\\1&-1&2&0&-2&1&-1\\-2&-1&-1&0&1&1&2\end{pmatrix}$
B_6	$B_3;D_4$	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0&0&0&0\\1&-1&2&0&0&0&0\\-2&-1&-1&0&0&0&0\end{pmatrix}$
D_6	D_4	1	$\frac{1}{3}\begin{pmatrix}1&2&-1&0&0&0&0\\1&-1&2&0&0&0&0\\-2&-1&-1&0&0&0&0\end{pmatrix}$

simple S-subalgebra of the algebra G_i', then

$$\tilde{G} = \tilde{G}_1 + \tilde{G}_2 + \cdots + \tilde{G}_s$$

is a semisimple S-subalgebra of the algebra G' and, therefore, is an R-subalgebra of the algebra G. If $\tilde{f}_{ik}^{(r)}$ is the defining matrix of the subalgebra $\tilde{G}_r$ with respect to G_r', then

$$\tilde{f}_{ik} = \sum_{r=1}^{s} \tilde{f}_{ik}^{(r)}$$

is the defining matrix of $\tilde{G}$ with respect to G' and

$$f_{im} = \sum_{k=1}^{n'} \tilde{f}_{ik} f'_{km},$$

the defining matrix of $\tilde{G}$ with respect to G, with f'_{km} the defining matrix of G' with respect to G. Applying the Weyl group of G simultaneously to the rows of the defining matrix, all the equivalent defining matrices corresponding to all conjugate subalgebras are obtained.

In Tables XV–XX all simple subalgebras of rank exceeding 1 for the classical algebras up to rank 6

TABLE XVIII. Simple subalgebras of the type $A_3 \sim D_3$.

Algebra	Regular Subalgebra	Index	Defining matrix $\tilde{G} = (e_2 + e_3, e_1 - e_2, e_2 - e_3)$	Defining matrix $\tilde{G} = (e_1 - e_2, e_2 - e_3, e_3 - e_4)$
B_3	A_3	1	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1\\1&-1&-1\\-1&1&-1\\-1&-1&1\end{pmatrix}$
A_4	A_3	1	$\frac{1}{2}\begin{pmatrix}1&1&0&-1&-1\\1&-1&0&1&-1\\1&-1&0&-1&1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&-1&-1&-1&0\\-1&3&-1&-1&0\\-1&-1&3&-1&0\\-1&-1&-1&3&0\end{pmatrix}$
B_4	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&1&-1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1\\-1&1&1&-3\\-1&1&-3&1\\-1&-3&1&1\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0\\1&-1&-1&0\\-1&1&-1&0\\-1&-1&1&0\end{pmatrix}$
C_4	A_3^2	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&1&-1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1\\-1&1&1&-3\\-1&1&1&-3\\-1&-3&1&1\end{pmatrix}$
D_4	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&\pm1\\1&1&-1&\pm1\\1&-1&1&\mp1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&\pm1\\-1&1&1&\mp1\\-1&1&-3&\pm1\\-1&-3&1&\pm1\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0\\1&-1&-1&0\\-1&1&-1&0\\-1&-1&1&0\end{pmatrix}$
A_5	A_3	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&-1&-1\\1&-1&0&0&1&-1\\1&-1&0&0&-1&1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&-1&-1&-1&0&0\\-1&3&-1&-1&0&0\\-1&-1&3&-1&0&0\\-1&-1&-1&3&0&0\end{pmatrix}$
	A_5	2	$\begin{pmatrix}1&0&0&0&0&-1\\0&1&0&0&-1&0\\0&0&1&-1&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&-1&-1&-1\\1&-1&-1&1&1&-1\\-1&1&-1&1&-1&1\\-1&-1&1&-1&1&1\end{pmatrix}$
B_5	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\\1&-1&1&-1&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0\\-1&1&1&-3&0\\-1&1&-3&1&0\\-1&-3&1&1&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0&0\\1&-1&-1&0&0\\-1&1&-1&0&0\\-1&-1&1&0&0\end{pmatrix}$

TABLE XVIII contd

C_5	A_3^2	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\\1&-1&1&-1&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0\\-1&1&1&-3&0\\-1&1&-3&1&0\\-1&-3&1&1&0\end{pmatrix}$
D_5	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0\\1&1&-1&-1&0\\1&-1&1&-1&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0\\-1&1&1&-3&0\\-1&1&-3&1&0\\-1&-3&1&1&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0&0\\1&-1&-1&0&0\\-1&1&-1&0&0\\-1&-1&1&0&0\end{pmatrix}$
A_6	A_3	1	$\frac{1}{2}\begin{pmatrix}1&1&0&0&0&-1&-1\\1&-1&0&0&0&1&-1\\1&-1&0&0&0&-1&1\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&-1&-1&-1&0&0&0\\-1&3&-1&-1&0&0&0\\-1&-1&3&-1&0&0&0\\-1&-1&-1&3&0&0&0\end{pmatrix}$
	A_5	2	$\begin{pmatrix}1&0&0&0&0&0&-1\\0&1&0&0&0&-1&0\\0&0&1&0&-1&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0&-1&-1&-1\\1&-1&-1&0&1&1&-1\\-1&1&-1&0&1&-1&1\\-1&-1&1&0&-1&1&1\end{pmatrix}$
B_6	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\\1&-1&1&-1&0&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0&0\\-1&1&1&-3&0&0\\-1&1&-3&1&0&0\\-1&-3&1&1&0&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0&0&0\\1&-1&-1&0&0&0\\-1&1&-1&0&0&0\\-1&-1&1&0&0&0\end{pmatrix}$
	$A_5;2A_3$	2	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&1&1&1\\1&1&-1&-1&-1&-1\\-1&-1&1&1&-1&-1\\-1&-1&-1&-1&1&1\end{pmatrix}$
C_6	A_3^2	2	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\\1&-1&1&-1&0&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0&0\\-1&1&1&-3&0&0\\-1&1&-3&1&0&0\\-1&-3&1&1&0&0\end{pmatrix}$
	A_5^2	4	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&1&1&1\\1&1&-1&-1&-1&-1\\-1&-1&1&1&-1&-1\\-1&-1&-1&-1&1&1\end{pmatrix}$
D_6	A_3	1'	$\frac{1}{2}\begin{pmatrix}1&1&1&1&0&0\\1&1&-1&-1&0&0\\1&-1&1&-1&0&0\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&1&1&1&0&0\\-1&1&1&-3&0&0\\-1&1&-3&1&0&0\\-1&-3&1&1&0&0\end{pmatrix}$
	A_3	1"	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&0&0&0\\1&-1&-1&0&0&0\\-1&1&-1&0&0&0\\-1&-1&1&0&0&0\end{pmatrix}$
	$A_5;2A_3$	2	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&\pm1\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}1&1&1&1&1&\pm1\\1&1&-1&-1&-1&\mp1\\-1&-1&1&1&-1&\mp1\\-1&-1&-1&-1&1&\pm1\end{pmatrix}$

TABLE XIX. Simple subalgebras of type B_3.

Algebra	Regular subalgebra	Index	Defining matrix
B_4	$B_3; D_4$	1	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$
D_4	D_4	1	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$
B_5	$B_3; D_4$	1	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\end{pmatrix}$
D_5	D_4	1	$\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\end{pmatrix}$
A_6	A_6	2	$\begin{pmatrix}1&0&0&0&0&0&-1\\0&1&0&0&0&-1&0\\0&0&1&0&-1&0&0\end{pmatrix}$
B_6	$B_3; D_4$	1	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$
D_6	D_4	1	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$

TABLE XX. Subalgebras of the type C_3.

Algebra	Regular subalgebra	Index	Defining matrix
C_4	C_3	1	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$
A_5	A_5	1	$\begin{pmatrix}1&0&0&0&0&-1\\0&1&0&0&-1&0\\0&0&1&-1&0&0\end{pmatrix}$
C_5	C_3	1	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$
A_6	A_5	1	$\begin{pmatrix}1&0&0&0&0&0&-1\\0&1&0&0&0&-1&0\\0&0&1&0&-1&0&0\end{pmatrix}$
B_6	A_5	1	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}$
C_6	C_3	1	$\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}$
	$A_5^2; 2C_3$	2	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}$
D_6	A_5	1	$\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&\pm1\end{pmatrix}$

(except the trivial cases $A_4 - D_4, A_5 - D_5$) are given. These tables contain the minimal including regular subalgebra, the index of the simple algebra with respect to the simple algebra and the highest defining matrix. If the same index corresponds to different subalgebras $B_2 \sim C_2$ and $A_3 \sim D_3$, two kinds of defining matrices are given according to the different system of simple roots which can be used for the simple subalgebras.

All the simple subalgebras of rank exceeding one of the exceptional algebras are given in Table 25 of Ref. 13.

* This work has been supported in part by the National Science Foundation, Grant GP-9632.

† The co-author (M.L.) wishes to express his gratitude to the Dean of the Facultad de Ciencias of the Universidad de Madrid for granting him leave of absence.

‡ This work contains part of the dissertation of M.L. which was written at Saint Louis University. (Tesis Doctoral, Universidad de Madrid, Spain, 1972)

1 *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1968).
2 P. Kramer and M. Moshinsky, in Ref. 1.
3 M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964).
4 B. R. Judd, in Ref. 1; B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley-Interscience, New York, 1970).
5 L. O'Raifeartaigh in Ref. 1; R. E. Behrends, in Ref. 1.
6 B. T. Feld, *Models of Elementary Particles* (Ginn/Blaisdell, Boston, Mass., 1969).
7 G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
8 A. U. Klimyk, "On Decomposition of Irreducible Representations of Semisimple Lie Group into Representations of its Semisimple Subgroup," Academy of Science, Ukranian SSR, Kiev Institute for Theoretical Physics, Preprint No. 67-17 (1967).
9 A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).
10 R. M. Delaney and B. Gruber, J. Math. Phys. **10**, 252 (1969).
11 M. K. F. Wong, J. Math. Phys. **11**, 1489 (1970).
12 A. P. Stone, J. Math. Phys. **11**, 29 (1970).
13 E. Dynkin, Mat. Sb. **30**, 349 (1952); Amer. Math. Soc. Transl., Ser. 2, **6**, 111 (1957), Theorem 2. 5.
14 Reference 13, Table 5.
15 N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 120.
16 Reference 13, Chap. 1, Parag. 2.
17 F. R. Gantmacher, Mat. Sb. N.S. **5** (47), 101 (1939), Theorem 11.
18 Reference 13, formula (1. 6).
19 Reference 10, formula (10); Ref. 12, p. 33.
20 Reference 13, Theorem 4. 2.
21 Reference 13, Theorem 4. 1.
22 Reference 13, Chap. VI, Parag. 15.
23 Reference 13, Theorem 5. 1.
24 Reference 13, Theorems 5. 2 and 5. 3.
25 Reference 13, Theorem 8. 1.
26 Reference 13, Theorem 8. 2.
27 Reference 13, Sec. 9.
28 E. Dynkin, Tr. Mosk. Mat. Obshchestva. **1**, 39 (1952) [Amer. Math. Soc. Transl. Ser. 2 **6**, 245 (1957)].
29 Reference 28, Theorem 1. 5.
30 A. I. Mal'cev, Isv. Akad. Nauk SSSR. Ser. Mat. **8**, 143 (1944) [Amer. Math. Soc. Transl. No. 33 (1950)], pp. 154 and 169.
31 Reference 13, pp. 121 and 127.
32 Reference 28, Theorems 1. 3 and 1. 4.
33 This mapping was given independently by A. U. Klimyk, Ref. 8, p. 16 and M. Lorente, thesis, Saint Louis University, 1968 (unpublished), p. 18.
34 Reference 13, Theorems 14. 2 and 15. 1.

Errata: S Matrix in the Heisenberg Representation

[J. Math. Phys. **11**, 3487 (1970)]

Edith Borie
National Bureau of Standards, Washington, D.C. 20234

(Received 15 May 1972)

Equation (2.14) should read

$$f^{\pm} = (1 + G_0{}^{\pm}X^{\pm})(1 + G_0{}^{\mp}X^{\mp})^{-1}f^{\mp} = f^{+} \pm \tilde{G}Uu.$$

Equation (3.9) should read:

$$\langle 0, +\infty | 0, -\infty \rangle = C_0^* \equiv e^{iW[\phi]}.$$

Equation (3.19) should read

$$W_{,j} = \tfrac{1}{2} i S_{,jkl} G^{(+)kl} - \tfrac{1}{8}(S_{,lmn} G^{lp} S_{,pab} G^{(+)mn} G^{(+)ab})_{,j}$$
$$- \tfrac{1}{8}(S_{,klmn} G^{(+)kl} G^{(+)mn})_{,j}$$
$$- \tfrac{1}{12}(S_{,lmn}(G^{np} S_{,pab} G^{(+)ma} + G^{-np} S_{,pab}$$
$$\times (G^{(+)ma} + G^{(+)am})) G^{(+)lb})_{,j};$$

only the second term is affected.

The following changes should be made in Appendix A:

$$\tilde{G} = G^{+} - G^{-} = (1 + G_0{}^{\pm}X^{\pm})\tilde{G}_0(1 + X^{\mp}G_0{}^{\mp}); \quad \text{(A8d)}$$

$$G^{(\pm)} = (1 + G_0 X) G_0{}^{(\pm)}(1 + X^{\pm} G_0{}^{\pm}); \quad \text{(A8e)}$$

$$G^{(+)ij} = -G^{(-)ji}. \quad \text{(A8j)}$$