Considerations on the Classical Spinning Electron

JACK COHN
Department of Physics, University of Oklahoma, Norman, Oklahoma

(Received 25 November 1968)

The present work is concerned with the asymptotic properties of the field radiated by a classical spinning (and nonspinning) point charge in arbitrary motion. It is demonstrated that the expression found for the radiated angular momentum is satisfactory in that it is a 4-tensor which is spacelike surface-independent. The angular-momentum radiation rate is then expressed in terms of the retarded kinematic properties of the particle. In addition, a "center-of-energy" theorem is proved for nonspinning charges in arbitrary motion—a result which follows from the demonstrated result that nonspinning charges also radiate angular momentum. Finally, it is demonstrated that the (linear) energy–momentum radiation rate is independent of the spin of the involved charge. The discussion is manifestly covariant throughout and many mathematical details are deferred to appendices.

I. INTRODUCTION

The present work is concerned with certain asymptotic properties of the field radiated by a charged particle. Specifically, we are concerned with finding the asymptotic expressions for the angular and linear momentum radiated by a spinning or nonspinning classical point charge undergoing arbitrary motion. By a classical point charge we mean the following: The charge is considered to be classical in that we shall consider it to be composed of an essentially arbitrary distribution of charge and the exterior field of this charge will then be calculated as that arising from this distribution. The charge is considered as a point charge in that, among other things, retardation effects arising from the charge's size are neglected. However, the particle is still assumed to have a nonzero moment of inertia about its spin axis. These points will be elaborated later on in the presentation.

The format of the paper is as follows:

In the next section we derive the field produced by a spinning point charge.

In Sec. III we give the definition of the angular momentum radiated by a spinning point charge and demonstrate that the definition is satisfactory.

In Sec. IV the angular-momentum radiation rate is then expressed in terms of the charge's retarded kinematical properties.

In Sec. V we prove the "center-of-energy" theorem which is a consequence of the demonstrated result that even *nonspinning* charged particles radiate angular momentum.

Finally, in Sec. VI, it is demonstrated that the emission of (linear) energy–momentum is independent of the spin of the charge involved.

Several appendices are included where many of the mathematical details are considered.

II. FIELD OF A SPINNING POINT CHARGE

The model of the electron assumed in this paper is that of a spherical charge of radius r_0 (in the momentary rest system) spinning about an axis through the center of the charge. The electron is further assumed to have a cylindrically symmetric charge and mass distribution about the spin axis (again in the rest system). After a certain stage of the calculation, r_0 will be allowed to go to zero.

For the moment, it is assumed that $r_0 \doteq 0$ and we consider the charge to have its center of mass at rest in an inertial frame and to be spinning at a constant rate about a fixed axis. The charge can then be thought of as composed of many small parallel current loops. The vector potential $\Delta\mathbf{A}$, due to a very small current loop of radius ξ carrying a constant current Δi, is approximately given by the expression

$$\Delta\mathbf{A} = \Delta i \frac{\pi\xi^2}{cr^3}\,\mathbf{n}\times\mathbf{r}, \tag{1}$$

where $\mathbf{n}$ is a unit vector in the direction of the spin axis and $\mathbf{r}$ is the vector going from the current loop to the field point.

Assuming then that the spinning charge is made up of such current loops and that the charge density $\rho(x)$ depends only on the distance x from the spin axis, we have for the vector potential $\mathbf{A}_\uparrow$, due to the spinning electron, the approximate relation

$$\mathbf{A}_\uparrow = \frac{\omega}{2cr^3}\,\mathbf{n}\times\mathbf{r}\int_{\text{electron}} x^2\rho(x)\,d^3x, \tag{2}$$

where ω is the spin in rad/sec.

Now, we assume that

$$\rho(x) = \frac{e}{m}\,\rho_{\text{mass}}(x), \tag{3}$$

where e and m are the charge and rest mass of the electron, respectively, and $\rho_{\text{mass}}(x)$ refers to the mass density within the electron.

Therefore, for such a spinning electron we have, as $r_0 \to 0$, the relation

$$\mathbf{A}_\uparrow = \frac{eI}{2mcr^3}\boldsymbol{\omega} \times \mathbf{r}, \quad \boldsymbol{\omega} \equiv \omega \mathbf{n}, \tag{4}$$

where I is the moment of inertia about the spin axis. This expression is now exact since we have let r_0 go to zero. Note, however, that I remains finite. This means, of course, that, as $r_0 \to 0$, $\rho_{\text{mass}}(x) \to \infty$.

Finally, note that the vector potential referred to so far pertains only to that part arising from the electron spin—denoted by a vertical arrow as subscript.

We shall now write the above expression in a Lorentz covariant manner as follows: In the following, a superscript 0 placed above a quantity indicates it is evaluated in the electron rest system, otherwise the quantity is to be evaluated in an arbitrary Lorentz frame; Latin indices go from 1 to 3, while Greek indices go from 1 to 4; and the space–time metric $g_{\mu\nu}$ is taken as the diagonal matrix $(1, 1, 1, -1)$, where the invariant arc length is given by $-c^2\,d\tau^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu$; finally, let $U^{0\mu} \sim (\mathbf{r}/r, 0)$ denote a contravariant 4-vector. Then,

$$\begin{aligned} U^{0\mu}U^0_\mu &= U^\mu U_\mu = 1, \\ U^{0\mu}V^0_\mu &= U^\mu V_\mu = 0, \end{aligned} \tag{5}$$

where V^μ is the world velocity $dx^\mu/d\tau$.

We may now write Eq. (4) in the form

$$A^{0i} = \frac{e}{2mc\rho^2} M^{0i\beta}U^0_\beta, \tag{6}$$

where ρ (not to be confused with charge or mass density) $\equiv r$ and

$$M^{0\mu\beta} = -M^{0\beta\mu} \equiv c^{-1}\delta^{\mu\alpha\beta\gamma} I\omega^0_\alpha V^0_\gamma \tag{7}$$

and where $\delta^{\mu\alpha\beta\gamma}$ is the Levi-Civita tensor density and ω^0_α is the 4-angular velocity vector $(\omega^i, 0)$. It is instructive to display the tensor $M^{0\alpha\beta}$; it has the form

$$M^{0\mu\nu} \sim \begin{pmatrix} 0 & I\omega_3 & -I\omega_2 & 0 \\ -I\omega_3 & 0 & I\omega_1 & 0 \\ I\omega_2 & -I\omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8}$$

Therefore,

$$M^{0\mu\nu}V^0_\nu = M^{\mu\nu}V_\nu = 0. \tag{9}$$

Now, the contribution to the vector potential coming from the charge itself arises only in the 4-component, and is given by the relation[1]

$$A^{04} = \frac{eV^{04}}{c\rho}. \tag{10}$$

Therefore, using Eqs. (6) and (10), we have for the *entire* vector potential the expression

$$A^{0\mu}_t = \frac{e}{2mc\rho^2} M^{0\mu\beta}U^0_\beta + \frac{eV^{0\mu}}{c\rho} \tag{11}$$

in the rest frame of an electron undergoing constant linear and angular motion. The subscript t is used here to denote the entire (total) vector potential.

We now consider the case where the electron's linear and rotational velocities may be arbitrary. A little reflection shows that for a particle of *zero* size the above relation must again be valid in the electron's momentary rest system. Therefore, we have for an electron in arbitrary motion (in an arbitrary Lorentz frame) the relation

$$A^\mu_t = \frac{e}{2mc\rho^2} M^{\mu\beta}U_\beta + \frac{eV^\mu}{c\rho} \equiv A^\mu_\uparrow + A^\mu. \tag{12}$$

We are now able to evaluate the field tensor $F^{\mu\nu}_t = F^{\mu\nu}_\uparrow + F^{\mu\nu}$, where

$$\begin{aligned} F^{\mu\nu}_t &= \partial^\mu A^\nu_t - \partial^\nu A^\mu_t, \\ F^{\mu\nu}_\uparrow &= \partial^\mu A^\nu_\uparrow - \partial^\nu A^\mu_\uparrow, \\ F^{\mu\nu} &= \partial^\mu A^\nu - \partial^\nu A^\mu. \end{aligned} \tag{13}$$

We first evaluate $F^{\mu\nu}_\uparrow$ as follows:

Using the relations[2]

$$\begin{aligned} \partial^\mu \rho &= U^\mu + A_u R^\mu/c^2, \\ \partial^\mu f(\tau) &= -\frac{1}{c}\frac{df}{d\tau}\frac{R^\mu}{\rho}, \end{aligned} \tag{14}$$

where all quantities are at their retarded values, $f(\tau)$ is any function of the proper time, $R = \rho(U^\mu + V^\mu/c)$ is the null vector going from the retarded location of the charge to the field point, and $A_u \equiv A_\sigma U^\sigma$, we obtain the result

$$\begin{aligned} \partial^\mu A^\nu_\uparrow = &-\frac{3e}{2mc\rho^3}\left(U^\mu + \frac{A_u R^\mu}{c^2}\right)M^{\nu\beta}U_\beta \\ &- \frac{e}{2mc^2\rho^3}\dot{M}^{\nu\beta}U_\beta R^\mu \\ &+ \frac{e}{2mc^3\rho^3}M^{\nu\beta}A_\beta R^\mu + \frac{e}{2mc\rho^3}M^{\nu\mu}, \end{aligned} \tag{15}$$

[1] See, for example, F. Rohrlich, *Classical Charged Particles* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965), p. 83. The notation in this work will, for the most part, follow that in this reference.

[2] Reference 1, pp. 83–84

where $A_\beta \equiv \dot{V}_\beta = dV_\beta/d\tau$. From this expression we then have $F^{\mu\nu}_{\uparrow}$. We also have the familiar expression[3]

$$F^{\mu\nu} = \frac{2e}{c\rho^2} V^{[\mu}U^{\nu]} + \frac{2e}{\rho c^2}\{A^{[\mu}V^{\nu]} - U^{[\mu}(V^{\nu]}A_u + A^{\nu]})\}, \tag{16}$$

where the kinematic quantities are retarded and

$$A^{[\mu}b^{\nu]} \equiv \tfrac{1}{2}(A^\mu b^\nu - A^\nu b^\mu). \tag{17}$$

Therefore, from Eqs. (15) and (16) we have the expression for $F^{\mu\nu}_t = F^{\mu\nu}_{\uparrow} + F^{\mu\nu}$.

III. DEFINITION OF RADIATED ANGULAR MOMENTUM

In this section we wish to construct a suitable definition of the radiated angular momentum from an arbitrarily moving (possibly spinning) point charge.

In the case of free radiation fields one defines the energy–momentum vector of the field P^μ as

$$P^\mu = \frac{1}{c}\int \theta^{\mu\nu}\, d\sigma_\nu, \tag{18}$$

where $\theta^{\mu\nu}$ is the stress–momentum tensor given by the expression

$$\theta^{\mu\nu} = \frac{1}{4\pi}(F^{\mu\alpha}F^\nu_\alpha + \tfrac{1}{4}g^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}) \tag{19}$$

and where the above integration can be over *any* (plane) spacelike surface, i.e., it is spacelike surface-independent.

Again for free fields one defines the angular momentum of the field as

$$J^{\mu\nu} = \frac{1}{c}\int J^{\alpha\mu\nu}\, d\sigma_\alpha, \tag{20}$$

where the integration is again over any spacelike surface and

$$J^{\alpha\mu\nu} \equiv -(\theta^{\alpha\mu}x^\nu - \theta^{\alpha\nu}x^\mu). \tag{21}$$

This definition is satisfactory as long as the entire radiation field is contained within a finite region.

Now, in the case of an arbitrarily moving charge one defines the energy–momentum radiated by it—in a proper time $d\tau$—by modifying Eq. (18) to give

$$dP^\mu_{rad} = \lim_{\rho\to\infty}\frac{1}{c}\int_{(\Delta\sigma)} \theta^{\mu\nu}\, d\sigma_\nu, \tag{22}$$

where $\Delta\sigma$ is a spacelike surface segment (dependent on $d\tau$) that is infinitely far away from the location of the charge during $d\tau$, as depicted in Fig. 1. This definition is satisfactory since it can be shown that the energy–momentum radiation rate so defined is a 4-vector independent of which spacelike surface segment is selected at $\rho = \infty$.[4]

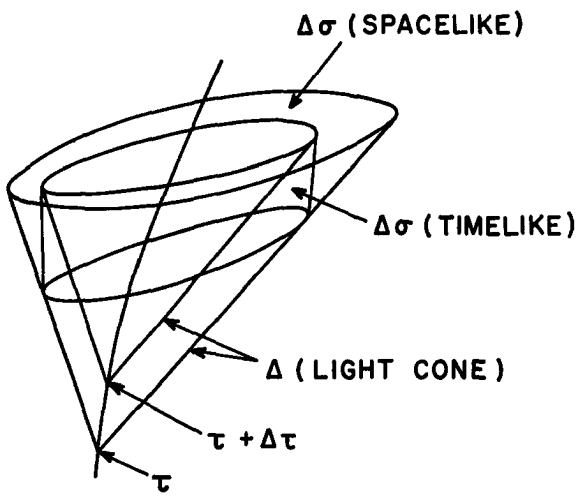


FIG. 1. Particle world line and associated surfaces.

We shall now consider a similar procedure to define the rate of emission of angular momentum by a (possibly spinning) charged particle in arbitrary motion, where the entire radiation field is assumed to be contained within a finite region. We tentatively define then the angular momentum radiated by the charge in the proper time $d\tau$ as

$$dJ^{\mu\nu}_t = \lim_{\rho\to\infty}\frac{1}{c}\int_{(\Delta\sigma)} J^{\alpha\mu\nu}_t\, d\sigma_\alpha, \tag{23}$$

where $\Delta\sigma$ is a limiting surface segment determined by $d\tau$ in the same manner as in the energy–momentum case.[5] Note the subscript t here, which indicates that the total field is being used. To demonstrate that this definition is satisfactory we presently show that the above quantity is a tensor which is independent of the spacelike surface segment $\Delta\sigma$. For this to be so, it is sufficient that (see Fig. 1):

$$\lim_{\rho\to\infty}\int_{\Delta(\text{light cone})} J^{\alpha\mu\nu}_t\, d\sigma_\alpha = 0, \tag{24}$$

since $J^{\alpha\mu\nu}_{t,\alpha} = 0$ everywhere outside the charge. This statement is now proved as follows:

On the light cone $d\sigma_\alpha$ is of order ρ, viz., $d\sigma_\alpha = R_\alpha d\omega$, where $d\omega$ is an element of surface area. The only nonvanishing contributions to the above integral must then come from the terms in $\theta^{\mu\nu}_t$ of order ρ^{-2}.

[3] Reference 1, p. 106.

[4] Reference 1, p. 108.

[5] This definition has also been recently considered by J. N. Goldberg [*Perspectives in Geometry and Relativity* (Indiana University Press, Bloomington, Indiana, 1966), p. 167; J. Math. Phys. **5**, 172 (1964)] in a somewhat more general context than that involved here.

$\theta_t^{\mu\nu}$ is composed of $F_t^{\mu\nu}$, which, in turn, is made up of $F^{\mu\nu}$ and $F_\uparrow^{\mu\nu}$. As such, $\theta_t^{\mu\nu}$ contains terms ranging in ρ from ρ^{-6} to ρ^{-2}. The terms involving ρ^{-n} $(n > 2)$ will not contribute in the limit as $\rho \to \infty$. The only terms in $\theta_t^{\mu\nu}$ involving ρ^{-2} come from the "radiation" term in $\theta^{\mu\nu}$, i.e., from the term

$$\frac{e^2}{4\pi c^4 \rho^2}(A_u^2 - A^2)\frac{R^\mu R^\nu}{\rho^2} \equiv \lambda R^\mu R^\nu. \tag{25}$$

Thus,

$$\lim_{\rho\to\infty} \int_{\Delta(\text{light cone})} J_t^{\alpha\mu\nu}\, d\sigma_\alpha = \frac{1}{c}\int \lambda(R^\alpha R^\mu x^\nu - R^\alpha R^\nu R^\mu)R_\alpha\, d\omega = 0, \tag{26}$$

since $R^\alpha R_\alpha = 0$. Therefore, expression (23) is a 4-tensor which is spacelike-surface-independent and can be taken as the proper expression for the angular-momentum emission. Because the integral over the light cone vanishes, we may also write (see Fig. 1):

$$dJ_t^{\mu\nu} = \lim_{\rho\to\infty}\frac{1}{c}\int_{\Delta\sigma(\text{spacelike})} J_t^{\alpha\mu\nu}\, d\sigma_\alpha = \lim_{\rho\to\infty} -\frac{1}{c}\int_{\Delta\sigma(\text{timelike})} J_t^{\alpha\mu\nu}\, d\sigma_\alpha, \tag{27}$$

where $d\sigma_\alpha = U_\alpha \rho^2\, d\Omega c\, d\tau$ for the timelike surface and Gauss' theorem has been used. In this latter form, the emission rate is more easily calculated.

IV. KINEMATIC EXPRESSION FOR THE EMISSION RATE

Using the above expression for the angular-momentum emission, we shall now calculate the angular-momentum-emission *rate* from an arbitrarily moving spinning (and nonspinning) point charge. This expression will be formulated entirely in terms of the retarded kinematical properties of the emitting charge.

Now, since $d\sigma_\alpha$, for a timelike surface, is of order ρ^2, the only contributions to $dJ_t^{\mu\nu}$ will come from terms in $J_t^{\alpha\mu\nu}$ of order ρ^{-2} and higher, that is, from terms in $\theta_t^{\mu\nu}$ of order ρ^{-3} and ρ^{-2}. The terms in $\theta_t^{\mu\nu}$ of order ρ^{-2} will give a contribution of order ρ, which, when $\rho \to \infty$, might diverge. It is shown in Appendix A that this, in fact, is not so, the contribution being zero. We next turn to the contributions to $dJ_t^{\mu\nu}$ coming from terms in $\theta_t^{\mu\nu}$ of order ρ^{-3}. Such terms come from two sources: that depending on spin terms and that independent of spin terms. That is, $\theta_t^{\mu\nu}$ can be decomposed as follows:

$$\theta_t^{\mu\nu} = \theta^{\mu\nu} + \frac{1}{4\pi}(F_\uparrow^{\mu\alpha}F^{\sigma\nu}g_{\sigma\alpha} + \tfrac{1}{4}g^{\mu\nu}F_{\alpha\beta}^\uparrow F_\uparrow^{\alpha\beta}) + \frac{1}{4\pi}(F_\uparrow^{\mu\alpha}F^{\sigma\nu}g_{\sigma\alpha} + \tfrac{1}{4}g^{\mu\nu}F_{\alpha\beta}^\uparrow F^{\alpha\beta}) + \frac{1}{4\pi}(F^{\mu\alpha}F_\uparrow^{\sigma\nu}g_{\sigma\alpha} + \tfrac{1}{4}g^{\mu\nu}F_{\alpha\beta}F_\uparrow^{\alpha\beta}) \equiv \theta^{\mu\nu} + \theta_\uparrow^{\mu\nu}. \tag{28}$$

We consider the spin contribution first. The only terms of order ρ^{-3} come from a term of order ρ^{-1} in $F^{\mu\nu}$ and one of order ρ^{-2} in $F_\uparrow^{\mu\nu}$ and a product of terms in $F^{\mu\nu}$ of order ρ^{-1} and ρ^{-2}. The term of order ρ^{-2} coming from $F_\uparrow^{\mu\nu}$ will be denoted by $F_\uparrow^{\mu\nu}\{O(\rho^{-2})\}$ and is given by the expression

$$F_\uparrow^{\mu\nu}\{O(\rho^{-2})\} = -\frac{e}{mc^2\rho^4}\left\{R_\beta R^{[\mu}\left(2\dot{M}^{\nu]\beta} + \frac{3A_u}{c}M^{\nu]\beta}\right)\right\} \tag{29}$$

and the terms in $F^{\mu\nu}$ of order ρ^{-1} are given by

$$F^{\mu\nu}\{O(\rho^{-1})\} = \frac{2e}{\rho c^3}\{A^{[\mu}V^{\nu]} - cU^{[\mu}A^{\nu]} - A_u U^{[\mu}V^{\nu]}\} \tag{30}$$

and

$$F_\alpha^\nu\{O(\rho^{-1})\} = \frac{2e}{\rho c^3}\{A_{[\alpha}V^{\nu]} - cU_{[\alpha}A^{\nu]} - A_u U_{[\alpha}V^{\nu]}\}, \tag{31}$$

where

$$A_{[\alpha}B^{\beta]} \equiv \tfrac{1}{2}(A_\alpha B^\beta - A^\beta B_\alpha). \tag{32}$$

Now, let $\theta_\uparrow^{\mu\nu}\{O(\rho^{-3})\}$ denote the term of order ρ^{-3} in $\theta_\uparrow^{\mu\nu}$. Then we have

$$\theta^{\mu\nu}\{O(\rho^{-3})\} = \frac{1}{4\pi}F_\uparrow^{\mu\alpha}\{O(\rho^{-1})\}F_\alpha^\nu\{O(\rho^{-1})\} + \frac{1}{4\pi}F^{\mu\alpha}\{O(\rho^{-1})\}F_{\alpha\uparrow}^\nu\{O(\rho^{-1})\} + \frac{1}{8\pi}g^{\mu\nu}F_{\alpha\beta}\{O(\rho^{-1})\}F_\uparrow^{\alpha\beta}\{O(\rho^{-1})\}. \tag{33}$$

After a straightforward but very lengthy calculation (see Appendix B) we finally obtain the result

$$\int(\theta_\uparrow^{\mu i}\{O(\rho^{-3})\}x^j - \theta_\uparrow^{\mu j}\{O(\rho^{-3})\}x^i)U_\mu\rho^2\, d\Omega\big|_{\text{r.s.}} = -\frac{2e^2}{3mc^5}A_\alpha(\dot{M}^{\alpha j}V^i - \dot{M}^{\alpha i}V^j)\big|_{\text{r.s.}}, \tag{34}$$

where r.s. refers to the rest system. Therefore, the contribution to the rate of emission of (space–space) angular momentum due to the spin is, in the electron rest system,

$$\frac{dJ^{ij}}{d\tau}(\text{spin})\big|_{\text{r.s.}} = -\frac{2e^2}{3mc^5}A_\alpha(\dot{M}^{\alpha j}V^i - \dot{M}^{\alpha i}V^j)\big|_{\text{r.s.}} = 0. \tag{35}$$

We now evaluate the space-time contribution to the emission rate due to the spin. After a lengthy calculation (see Appendix C), we obtain the expression

$$\frac{dJ^{i4}}{d\tau}(\text{spin})\Big|_{\text{r.s.}} = -\frac{4e^2}{3mc^4}\dot{M}^{\alpha i}A_\alpha\Big|_{\text{r.s.}}. \tag{36}$$

We may write the results of Eqs. (35) and (36) covariantly as

$$\frac{dJ^{\alpha\beta}}{d\tau}(\text{spin}) = -\frac{4e^2}{3mc^5}A_\alpha(\dot{M}^{\mu\beta}V^\alpha - \dot{M}^{\mu\alpha}V^\beta). \tag{37}$$

We now turn to the contribution to the emission of angular momentum coming from the nonspin terms, i.e., from $\theta^{\mu\nu}$. We are interested, then, in evaluating the quantity

$$\frac{dJ^{\mu\nu}}{d\tau}(\text{no spin}) = -\lim_{\rho\to\infty}\int J^{\alpha\mu\nu}U_\alpha\rho^2\,d\Omega, \tag{38}$$

where we are only interested in terms of order ρ^{-3} in $\theta^{\mu\nu}$ contained in the above integrand. We then take

$$J^{\alpha\mu\nu} = \theta^{\alpha\mu}\{O(\rho^{-3})\}x^\nu - \theta^{\alpha\nu}\{O(\rho^{-3})\}x^\mu,$$

where $\theta^{\alpha\beta}\{O(\rho^{-3})\}$ is given by the expression

$$\theta^{\alpha\beta}\{O(\rho^{-3})\} = \frac{e^2}{2\pi c^2\rho^3}\left[A_u\frac{R^\alpha R^\beta}{\rho^2} - \left(\frac{V^{(\alpha}A_u}{c} + A^{(\alpha}\right)\frac{R^{\beta)}}{\rho}\right], \tag{39}$$

where

$$A^{(\alpha}B^{\beta)} \equiv \tfrac{1}{2}(A^\alpha B^\beta + A^\beta B^\alpha). \tag{40}$$

Inserting this into Eq. (38) then gives the result

$$\frac{dJ^{\mu\nu}}{d\tau}(\text{no spin}) = \frac{e^2}{2\pi c^2}\int(A_uU^\mu x^\nu - A_uU^\nu x^\mu - A^\mu x^\nu + A^\nu x^\mu)\tfrac{1}{2}\,d\Omega. \tag{41}$$

Therefore,

$$\frac{dJ^{ij}}{d\tau}(\text{no spin})\Big|_{\text{r.s.}} = 0 \tag{42}$$

and

$$\frac{dJ^{4i}}{d\tau}(\text{no spin})\Big|_{\text{r.s.}} = -\frac{4e^2}{3c^2}A^i\Big|_{\text{r.s.}}. \tag{43}$$

Expressing these results in covariant form, we then have

$$\frac{dJ^{\mu\nu}}{d\tau}(\text{no spin}) = \frac{4e^2}{3c^3}(A^\mu V^\nu - A^\nu V^\mu). \tag{44}$$

Therefore, even a *nonspinning* charge emits angular momentum in a form given by a covariant "Thomas precession" term.

Combining Eqs. (44) and (37), we then have as the expression for the total asymptotic angular-momentum rate of emission by a classical charged particle undergoing arbitrary spin and linear motion

$$\frac{dJ_i^{\alpha\beta}}{d\tau} = \frac{4e^2}{3mc^5}A_\mu(\dot{M}^{\mu\alpha}V^\beta - \dot{M}^{\mu\beta}V^\alpha) + \frac{4e^2}{3c^3}(A^\alpha V^\beta - A^\beta V^\alpha), \tag{45}$$

where all quantities have their retarded values.

V. CENTER-OF-ENERGY THEOREM

The fact that even a nonspinning charged particle may emit angular momentum leads to an interesting theorem.

Consider the angular momentum dJ^{4i} (no spin), radiated by the moving charge in the proper time $d\tau$, as given by the expression

$$dJ^{4i}(\text{no spin})\Big|_{\text{r.s.}} = \lim_{\rho\to\infty}\frac{1}{c}\int_{(\Delta\sigma)}(\theta^{44}x^i - \theta^{4i}x^4)\,d^3x\Big|_{\text{r.s.}}, \tag{46}$$

where we have selected a spacelike surface segment (dependent on $d\tau$ as in Fig. 1) which is orthogonal to the retarded world velocity of the particle (recalling that the choice of surface segment is unimportant).

Now,

$$\lim_{\rho\to\infty}\int_{(\Delta\sigma)}\theta^{44}\,d^3x\Big|_{\text{r.s.}} = -dW\Big|_{\text{r.s.}},$$
$$\lim_{\rho\to\infty}\int_{(\Delta\sigma)}\theta^{4i}\,d^3x\Big|_{\text{r.s.}} = -c\,dP^i\Big|_{\text{r.s.}} = 0, \tag{47}$$

where dW and dP^i are the energy and linear momentum radiated by the charge during $d\tau$.

We now define the "center of radiated energy" by the relation

$$R^i \equiv \lim_{\rho\to\infty}\left.\frac{\int_{\Delta\sigma}x^i\theta^{44}\,d^3x}{\int_{\Delta\sigma}\theta^{44}\,d^3x}\right|_{\text{r.s.}}. \tag{48}$$

Therefore,

$$dJ^{4i}(\text{no spin})\Big|_{\text{r.s.}} = -\frac{R^i}{c}\,dW\Big|_{\text{r.s.}}. \tag{49}$$

It then follows that

$$R^i = -c\left.\frac{dJ^{4i}}{dW}\right|_{\text{r.s.}} = -c\left.\frac{dJ^{4i}/d\tau}{dW/d\tau}\right|_{\text{r.s.}}. \tag{50}$$

Now,

$$\frac{dW}{d\tau} = \frac{2}{3}\frac{e^2}{c^3}\mathbf{A}^2\Big|_{\text{r.s.}}. \tag{51}$$

From Eqs. (51) and (43) we then obtain

$$\mathbf{R} = \frac{2c^2\mathbf{A}}{\mathbf{A}^2}\bigg|_{\text{r.s.}}. \tag{52}$$

That is, if we follow the radiation out from an accelerating charge, its "center" as defined above has a fixed location relative to the charge at the (retarded) time of emission which depends only on the charge's *intrinsic* acceleration A (retarded).[6]

VI. ENERGY-MOMENTUM EMISSION

It is also interesting to calculate the asymptotic rate of energy–momentum emission from an arbitrarily moving spinning charge.

As with the nonspinning charge, we shall define the energy–momentum emission for a spinning charged particle by the relation

$$dP_t^\mu = \lim_{\rho\to\infty} \frac{1}{c}\int_{(\Delta\sigma)} \theta_t^{\mu\nu}\, d\sigma_\nu, \tag{53}$$

where $\Delta\sigma$ is again the customary limiting spacelike surface segment.

Since

$$\int_{\Delta(\text{light cone})} \theta_\uparrow^{\mu\nu}\, d\sigma_\nu \to 0, \quad \text{as} \quad \rho\to\infty, \tag{54}$$

we again have that dP_t^μ is surface-independent as it must be.

Then we also have the relation

$$\begin{aligned} dP_t^\mu &= \lim_{\rho\to\infty} \frac{1}{c}\int_{\Delta\sigma(\text{spacelike})} \theta_t^{\mu\nu}\, d\sigma_\nu \\ &= \lim_{\rho\to\infty} -\frac{1}{c}\int_{\Delta\sigma(\text{timelike})} \theta_t^{\mu\nu}\, d\sigma_\nu. \end{aligned} \tag{55}$$

Now, $\theta_\uparrow^{\mu\nu}$ is composed of terms of orders ρ^{-6}, ρ^{-5}, ρ^{-4}, and ρ^{-3}. Therefore, in the limit as $\rho\to\infty$, $\theta_\uparrow^{\mu\nu}$ gives no contribution to the above integral and we have

$$\frac{dP_t^\mu}{d\tau} = \frac{2}{3}\frac{e^2}{c^5} A^2 V^\mu. \tag{56}$$

That is, the asymptotic value for the rate of emission of energy–momentum is *independent* of the spin of the charged particle.

APPENDIX A: CONTRIBUTION TO ANGULAR MOMENTUM FROM TERMS OF ORDER ρ^{-2}

Here it is shown that the terms in $\theta_t^{\mu\nu}$ of order ρ^{-2} give a contribution of zero to the angular-momentum-emission rate. For the order being considered, then, the term we must consider is

$$I^{\mu\nu} \equiv \lim_{\rho\to\infty} \frac{-e^2}{4\pi c^5 \rho^2} \times \int (A_u^2 - A^2)(R^\alpha R^\mu x^\nu - R^\alpha R^\nu x^\mu) U_\alpha c\, d\Omega. \tag{A1}$$

It will presently be demonstrated that this integral is zero. Using the relation $R^\mu = \rho(U^\mu + V^\mu/c)$, the above expression becomes

$$I^{\mu\nu} = \lim_{\rho\to\infty} \frac{-e^2}{4\pi c^4} \int (A_u^2 - A^2)\left\{\left(U^\mu + \frac{V^\mu}{c}\right)x^\nu - \left(U^\nu + \frac{V^\nu}{c}\right)x^\mu\right\} c\, d\Omega. \tag{A2}$$

There are various types of terms to evaluate here. If these are evaluated in the rest system, we obtain

$$\begin{aligned} \int (U^\mu x^\nu - U^\nu x^\mu)\, d\Omega\big|_{\text{r.s.}} &= 0, \\ \int (V^\mu x^\nu - V^\nu x^\mu)\, d\Omega\big|_{\text{r.s.}} &= 0, \\ \int A_u^2 (U^\mu x^\nu - U^\nu x^\mu)\, d\Omega\big|_{\text{r.s.}} &= 0, \\ \int A_u^2 (V^\mu x^\nu - V^\nu x^\mu)\, d\Omega\big|_{\text{r.s.}} &= 0. \end{aligned} \tag{A3}$$

Therefore, $I^{\mu\nu} = 0$ in the rest system and, since $I^{\mu\nu}$ is a tensor, $I^{\mu\nu} = 0$ in any Lorentz frame.

APPENDIX B: SPACE–SPACE CONTRIBUTION TO ANGULAR MOMENTUM

Here we calculate the space–space contribution to the emitted angular momentum due to the spin portion $\theta_\uparrow^{\mu\nu}$. As discussed before, we only need consider terms of order ρ^{-3}. We wish then to evaluate the quantity

$$\frac{dJ^{ij}}{d\tau}(\text{spin}) = -\lim_{\rho\to\infty} \int [\theta_\uparrow^{\mu i}\{O(\rho^{-3})\}x^j - \theta_\uparrow^{\mu j}\{O(\rho^{-3})\}x^i] U_\mu \rho^2\, d\Omega. \tag{B1}$$

Referring to Eq. (31), we see that

$$\begin{aligned} \theta_\uparrow^{\mu\nu}\{O(\rho^{-3})\} = {} & \frac{1}{4\pi} F_\uparrow^{\mu\alpha}\{O(\rho^{-2})\} F_\alpha^\nu\{O(\rho^{-1})\} \\ & + \frac{1}{4\pi} F^{\mu\alpha}\{O(\rho^{-1})\} F_{\alpha\uparrow}^\nu\{O(\rho^{-2})\} \\ & + \frac{1}{8\pi} g^{\mu\nu} F_{\alpha\beta}\{O(\rho^{-1})\} F_\uparrow^{\alpha\beta}\{O(\rho^{-1})\}. \end{aligned} \tag{B2}$$

[6] This may be regarded as a generalization of the "center-of-mass" theorem for free radiation fields. See Ref. 1, p. 99.

The above terms (on the right-hand side) are given as follows:

$$\text{first term} = -\frac{e^2}{2\pi c^5 m\rho^5} R_\beta R^{[\mu}\Lambda^{\alpha]\beta} \times (A_{[\alpha}V^{\nu]} - cU_{[\alpha}A^{\nu]} - A_u U_{[\alpha}V^{\nu]}),$$

$$\text{second term} = -\frac{e^2}{2\pi c^5 m\rho^5} R_\beta R_{[\alpha}\Lambda^{\nu]\beta} \times (A^{[\mu}V^{\alpha]} - cU^{[\mu}A^{\alpha]} - A_u U^{[\mu}V^{\alpha]}), \tag{B3}$$

$$\text{third term} = -\frac{e^2}{4\pi c^5 m\rho^5} g^{\mu\nu} R_\sigma R^{[\alpha}\Lambda^{\beta]\sigma} \times (A_{[\alpha}V_{\beta]} - cU_{[\alpha}A_{\beta]} - A_u U_{[\alpha}V_{\beta]}),$$

where

$$\Lambda^{\nu\beta} \equiv 2\dot{M}^{\nu\beta} + \frac{3A_u}{c} M^{\nu\beta}. \tag{B4}$$

In the rest system, one obtains the following results (where $m \neq 4$):

$$\int (\text{first term}) U_\mu x^m \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\{-\tfrac{1}{15}(M^{\alpha m}A_\alpha A^\nu + M^{\alpha\nu}A^m A_\alpha) - \tfrac{2}{3}\dot{M}^{m\beta}V_\beta A^\nu + \tfrac{2}{3}\dot{M}^{\alpha m}(A_\alpha V^\nu - A^\nu V_\alpha)\}, \tag{B5}$$

$$\int (\text{second term}) U_\beta x^m \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\{-\tfrac{2}{15}(\dot{M}^{m\beta}V_\beta A^\nu + \dot{M}^{\nu\beta}V_\beta A^m) + \tfrac{2}{3}\dot{M}^{\alpha m}A_\alpha V^\nu + \tfrac{1}{5}(M^{\alpha\nu}A^m A_\alpha + M^{\alpha m}A^\nu A_\alpha)\}, \tag{B6}$$

$$\int (\text{third term}) U_\mu x^m \rho^2 \, d\Omega = -\frac{e^2}{4mc^5}\{-\tfrac{2}{15}(\dot{M}^{m\sigma}V_\sigma A^\nu + \dot{M}^{\nu\sigma}V_\sigma A^m) - \tfrac{2}{5}(M^{\beta m}A_\beta A^\nu + M^{\beta\nu}A_\beta A^m)\}. \tag{B7}$$

Inserting these expressions into Eq. (B1) then leads to Eq. (35) for $dJ^{ij}/d\tau$ (spin).

APPENDIX C: SPACE-TIME CONTRIBUTION TO ANGULAR MOMENTUM

Here we calculate the space-time contribution to the emitted angular momentum due to the spin portion $\theta^{\mu\nu}_{\uparrow}$. We desire, then, to evaluate the quantity

$$\frac{dJ^{i4}}{d\tau}(\text{spin}) = -\lim_{\rho\to\infty}\int [\theta^{\mu i}_{\uparrow}\{O(\rho^{-3})\}x^4 - \theta^{\mu 4}_{\uparrow}\{O(\rho^{-3})\}x^i]U_\mu \rho^2 \, d\Omega. \tag{C1}$$

Again, in the rest frame, we obtain the following results (for $\nu = i$):

$$\int (\text{first term}) U_\mu x^4 \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\left\{\tfrac{2}{3}c\dot{M}^{\alpha\nu}A_\alpha + \frac{4}{3c}\dot{M}^{\alpha\beta}V_\beta A_\alpha V^\nu\right\}, \tag{C2}$$

$$\int (\text{second term}) U_\mu x^4 \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\left\{\frac{2}{c}\dot{M}^{\alpha\beta}V_\beta V^\nu A_\alpha - \frac{2}{3c}\dot{M}^{\alpha\beta}V_\beta A_\alpha V^\nu + \tfrac{2}{3}c\dot{M}^{\alpha\nu}A_\alpha\right\} \tag{C3}$$

$$\int (\text{third term}) U_\mu x^4 \rho^2 \, d\Omega = 0. \tag{C4}$$

Further, with $\nu = 4$, we have

$$\int (\text{first term}) U_\mu x^i \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\{\tfrac{2}{3}cA_\alpha \dot{M}^{\alpha i}\}, \tag{C5}$$

$$\int (\text{second term}) U_\mu x^i \rho^2 \, d\Omega = -\frac{e^2}{2mc^5}\{\tfrac{2}{3}cA_\alpha \dot{M}^{\alpha i}\}, \tag{C6}$$

$$\int (\text{third term}) U_\mu x^i \rho^2 \, d\Omega = 0. \tag{C7}$$

Inserting these results into Eq. (C1) then yields

$$\frac{dJ^{i4}}{d\tau}(\text{spin})\Big|_{\text{r.s.}} = -\frac{4e^2}{3mc^5}\cdot c\dot{M}^{\alpha i}A_\alpha\Big|_{\text{r.s.}}. \tag{C8}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Nonsaturation of Gravitational Forces

JEAN-MARC LÉVY-LEBLOND
*Physique Théorique, Nice**†

(Received 30 March 1968)

Rigorous inequalities are derived for the ground-state energy of a nonrelativistic quantum-mechanical system of N particles in gravitational interaction. It is shown that gravitational forces do not saturate, the binding energy per particle increasing with N, like N^2 for a Bose system, like $(N^{4/3})$ for a Fermi system. As a by-product, we obtain a generally valid Heisenberg-like inequality for N-fermion systems, expressing very simply the effect of the Pauli exclusion principle. These results are extended to the case of a system of oppositely charged particles which is shown to behave, with respect to gravitational forces, as a Fermi system as soon as particles with one sign of charge only are identical fermions. This explains quantitatively how and when gravitational forces finally predominate over Coulomb forces for large enough bodies (planets). A further extension to the case where relativistic effects enter only at the kinematical level permits us to derive rigorously from first principles the existence and an estimate of the Chandrasekhar mass limit, above which no collection of cold matter is stable (white dwarf stars).

INTRODUCTION

It has been emphasized by Fisher and Ruelle[1] that, in order to establish a rigorous basis for the statistical mechanics of an infinite system of interacting particles, one must be sure that the relevant forces have a saturating character: The total energy of a (finite) system ought to possess a lower bound which is extensive, i.e., proportional to the number of particles. When this is not true, the binding energy *per particle* increases indefinitely with the number of particles, so that it becomes obviously impossible to define the usual thermodynamic variables for infinite systems; the ground state of the system is a single bound state, more and more condensed as its mass increases. Fisher and Ruelle have given general criteria which guarantee the saturation property for not-too-singular forces.[1] However, there exists in nature only a very limited number of interactions, and it is for these, in the last resort, that we are interested in the saturation problem.

In the case of Coulomb forces, it has been proved by Dyson and Lenard,[2] in a splendid analysis, that there is saturation if particles with charges of a given sign belong to a finite number of fermion species. We thus understand the stability of ordinary matter. It is clearly impossible at the present stage to refine this understanding by including the relativistic features of electromagnetic interactions (retardation effects, radiation, etc.) or to discuss the saturation problem for strong or weak interactions. There still remains to be considered the universal gravitational interaction, which can in fact be discussed much more easily than the Coulomb case, in the nonrelativistic limit at least. Such considerations are relevant for the study of matter in bulk on an astronomical scale.

As can be easily conjectured, and is proved below, gravitational forces, because of their long-range and purely attractive character, *do not* saturate. This already results from Theorem III of Fisher and Ruelle,[1] which, however, does not give the exact degree of nonsaturation. The nonextensive nature of the thermodynamic functions has also been noted by Salzberg[3] in an investigation of classical systems with gravitational interactions in one and two dimensions. A more detailed investigation seems interesting as it touches upon topics of general interest in astrophysics. Obviously, we need to state our problem in quantum-mechanical terms since, classically, the simplest two-body gravitational system already has an energy spectrum unbounded from below. The crushing of matter under its own gravitational pull can only be prevented by quantum effects. We thus consider the Hamiltonian of an N-particle system interacting via gravitational forces and want to study the ground-state energy as a function of N.

We start by giving in Sec. I a heuristic and elementary derivation of our results. In Sec. II, upper and lower bounds are rigorously derived for the ground-state energy of a gravitationally interacting N-particle system with a special emphasis on the role of the exclusion principle. In the last section, we consider realistic systems where Coulomb and Newton forces are simultaneously operating, and comment on the physical meaning and applications of our results.

I. HEURISTIC DISCUSSION

Consider a system of N particles with a common mass m. Let p be the average momentum of a particle in the system and r be some average distance between

* Equipe de Recherche Associée au C.N.R.S.
† Postal address: Physique Théorique, Faculté des Sciences, Parc Valrose, 06 - Nice, France.

[1] M. E. Fisher and D. Ruelle, J. Math. Phys. **7**, 260 (1966).
[2] F. J. Dyson and A. Lenard, J. Math. Phys. **8**, 423 (1967).
[3] A. M. Salzberg, J. Math. Phys. **6**, 158 (1965).

any two particles. We may estimate the total energy of the gravitationally interacting system by

$$E(N) \simeq N(p^2/2m) - \tfrac{1}{2}N^2(Gm^2/r), \tag{1}$$

G being Newton's constant, and $\frac{1}{2}N^2 \simeq \frac{1}{2}N(N-1)$ the number of interacting pairs. According to Heisenberg's uncertainty relation, the position of a particle cannot be defined to better than $\hbar/p$, neither can then the distance between two particles, so that quantum mechanics requires $r \geqslant \hbar/p$. We obtain

$$E(N) \geqslant N(p^2/2m) - \tfrac{1}{2}N^2(Gm^2p/\hbar) \tag{2}$$

and minimize this estimate with respect to the free parameter p. For

$$p_0 \simeq N(Gm^3/2\hbar), \tag{3}$$

we have the lowest energy

$$E_0(N) \simeq -N^3(G^2m^5/8\hbar^2). \tag{4}$$

Without further constraints we are led to predict a cubic variation of $E_0(N)$ with N.

Things, however, are different if the N particles are identical fermions. Indeed, we may interpret the exclusion principle as forbidding the presence of more than one particle within any region with linear dimensions of the order of the average de Broglie wavelength $\hbar/p$.[4] Accordingly, N particles occupy a total volume $N(\hbar/p)^3$, so that the average separation of a pair is of order $N^{\frac{1}{3}}(\hbar/p)$. We now have:

$$E^{\text{fer}}(N) \geqslant N\frac{p^2}{2m} - \frac{N^2}{2}\frac{Gm^2}{N^{\frac{1}{3}}\hbar/p} = N\frac{p^2}{2m} - N^{\frac{5}{3}}\frac{Gm^2}{2\hbar}p. \tag{5}$$

The minimal energy is obtained for

$$p_0^{\text{fer}} \simeq N^{\frac{2}{3}}(Gm^3/2\hbar), \tag{6}$$

and its value is:

$$E_0^{\text{fer}}(N) \simeq -N^{\frac{7}{3}}(G^2m^5/8\hbar^2). \tag{7}$$

The exclusion principle, as expected, reduces the degree of unstability, but is by far insufficient to guarantee the saturation of gravitational forces.[5]

II. PURELY GRAVITATIONAL FORCES

Theorem 1: For a system of N particles with common mass[6] m in gravitational interaction, when no exclusion principle operates, the ground-state energy obeys

$$-AN(N-1)^2(G^2m^5/\hbar^2) \leq E_0(N) \leq -BN(N-1)^2(G^2m^5/\hbar^2), \tag{8}$$

A and B being positive constants.

Proof:

(a) Lower bound: We write the Hamiltonian for the system as

$$H = \sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m} - \sum\sum_{1\leq i<j\leq N}\frac{Gm^2}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{9}$$

where $\mathbf{p}_i$ and $\mathbf{r}_i$ are the momentum and position of the ith particle. Using a trick of Ruelle and Fisher,[1] as reformulated by Dyson and Lenard,[2] one may write

$$H = \sum\sum_{i<j}\left[\frac{\mathbf{p}_i^2}{2(N-1)m} + \frac{\mathbf{p}_j^2}{2(N-1)m} - \frac{Gm^2}{|\mathbf{r}_i - \mathbf{r}_j|}\right] = \sum\sum_{i<j}h_{ij}. \tag{10}$$

The Hamiltonian H thus appears as a sum of $\frac{1}{2}N(N-1)$ two-body Hamiltonians h_{ij} of hydrogenic type with known states:

$$\epsilon_0 = \inf\langle h_{ij}\rangle = -\tfrac{1}{4}(N-1)(G^2m^5/\hbar^2). \tag{11}$$

Then:

$$E_0(N) = \inf\langle H\rangle \geq \sum\sum_{i<j}\inf\langle h_{ij}\rangle, \tag{12}$$

$$E_0(N) \geq -\tfrac{1}{8}N(N-1)^2(G^2m^5/\hbar^2). \tag{13}$$

(b) Upper bound: We use the Rayleigh–Ritz variational principle to obtain an upper bound for the ground-state energy by computing the expectation value of the Hamiltonian with the completely symmetrical wavefunction

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \prod_{i=1}^{N}\phi(\lambda\mathbf{r}_i), \tag{14}$$

where the scale factor λ and the function ϕ are to be varied. The result is

$$(\Psi, H\Psi) = \lambda^2N(\hbar^2/2m)\alpha - \lambda\tfrac{1}{2}N(N-1)Gm^2\beta, \tag{15}$$

with

$$\alpha = \int|\nabla\phi|^2\,d^3r$$

$$\beta = \iint\frac{|\phi(\mathbf{r})|^2\,|\phi(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}\,d^3r\,d^3r'. \tag{16}$$

Minimizing the expectation value (15) with respect to λ, one obtains

$$E_0(N) \leq -\tfrac{1}{8}(\beta^2/\alpha)N(N-1)^2(G^2m^5/\hbar^2). \tag{17}$$

The largest value for the coefficient β^2/α might be obtained by solving the corresponding variational problem for the wavefunction ϕ, very similar in form

[4] It may be amusing to note that we do in fact obtain a proof of an inequality justifying this interpretation of the Pauli principle, following a rigorous treatment of the gravitational Hamiltonian (see Sec. II).

[5] Note that in a d-dimensional space, (7) would be replaced by the following estimate: $E_0^{\text{fer}}(N) \simeq -N^{3-2/d}(G^2m^5/8\hbar^2)$ so that gravitational forces just saturate for a one-dimensional fermion system.

[6] The same results remain true for particles with different masses if m is interpreted as the largest mass. The assumption of a common mass merely simplifies the writing of our equations.

to Hartree's equation for a 2-electron atom. However, we do not try here to optimize our bounds and will be content to say that with a simple exponential trial function, one may obtain the value $\beta^2/\alpha = (\frac{5}{8})^2$.

Theorem 2: For a system of N identical fermions with common mass m, the ground-state energy obeys

$$-A'N(N-1)^{\frac{4}{3}}(G^2m^5/\hbar^2) \leq E_0^f(N) \leq -B'N^{\frac{1}{3}}(N-1)^2(G^2m^5/\hbar^2), \quad (18)$$

A' and B' being positive constants.

Proof:

(a) Lower bound: We just need to refine the technique already used and partition the Hamiltonian (9) in the following way:

$$H = \sum_{i=1}^{N}\left[\sum_{j\neq i}\left(\frac{\mathbf{p}_j^2}{2(N-1)m} - \frac{G}{2}\frac{m^2}{|\mathbf{r}_i - \mathbf{r}_j|}\right)\right] = \sum_{i=1}^{N} h_i. \quad (19)$$

H now appears as a sum of N similar Hamiltonians h_i, each representing $(N-1)$ independent particles $(j \neq i)$ in the field of a fixed one (the ith particle). The ground state of h_i is obtained by distributing $(N-1)$ fermions over the first $(N-1)$ levels of this hydrogenic atom. These levels are given by

$$\eta_n = -n^{-2}\tfrac{1}{4}(N-1)(G^2m^5/\hbar^2), \quad n = 1, 2, \cdots, \quad (20)$$

with a degree of degeneracy $g_n = n^2$. The last level to be completely filled is η_ν such that

$$\sum_{n=1}^{\nu} g_n \leq N - 1 < \sum_{n=1}^{\nu+1} g_n, \quad (21)$$

$$\tfrac{1}{3}\nu(\nu+\tfrac{1}{2})(\nu+1) \leq N-1 < \tfrac{1}{3}(\nu+1)(\nu+\tfrac{3}{2})(\nu+2), \quad (22)$$

and one has

$$\langle h_i\rangle \geq \sum_{n=1}^{\nu+1} g_\nu\eta_\nu = -(\nu+1)\tfrac{1}{4}(N-1)(G^2m^5/\hbar^2), \quad (23)$$

$$\langle h_i\rangle \geq -\tfrac{1}{2}(N-1)^{\frac{4}{3}}(G^2m^5/\hbar^2), \quad (24)$$

since (22) implies $\nu + 1 \leq 2(N-1)^{\frac{1}{3}}$, for $N \geq 2$.[7]

Finally,

$$E_0^f(N) \geq N\langle h_i\rangle = -\tfrac{1}{2}N(N-1)^{\frac{4}{3}}(G^2m^5/\hbar^2). \quad (25)$$

(b) Upper bound: We use the variational principle, computing the expectation value of H for the antisymmetrical Slater wavefunction

$$\Psi^f(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = (N!)^{-\frac{1}{2}}\det[\psi_k(\lambda\mathbf{r}_l)], \quad (26)$$

where λ is a scale parameter and $\{\psi_k\}$ a set of localized wavefunctions

$$\psi_k(\mathbf{r}) = \psi(\mathbf{r} - \mathbf{a}_k). \quad (27)$$

[7] Asymptotically, for N large enough, this can be improved into $\nu + 1 \leq 3^{\frac{1}{3}}(N-1)^{\frac{1}{3}}$.

The $\{\mathbf{a}_k\}$ are N fixed points, chosen so that their minimal separation is kept fixed:

$$|\mathbf{a}_k - \mathbf{a}_l| \geq 1, \quad \forall k, l, \quad (28)$$

and ψ is a continuous function vanishing outside of a sphere with radius $\frac{1}{2}$. Now,

$$(\Psi^f, H\Psi^f) = \lambda^2 N(\hbar^2/2m)\alpha - \lambda(\tfrac{1}{2}Gm^2)\sum_{k\neq l}\beta_{kl}, \quad (29)$$

where

$$\alpha = \int |\nabla\psi|^2\, d^3r,$$

$$\beta_{kl} = \iint [|\psi_k(\mathbf{r})\psi_l(\mathbf{r}')|^2 - \overline{\psi_k(\mathbf{r})}\psi_k(\mathbf{r}')\psi_l(\mathbf{r})\overline{\psi_l(\mathbf{r}')}]\frac{d^3r\, d^3r'}{|\mathbf{r}-\mathbf{r}'|}. \quad (30)$$

Due to the support property of ψ, however, the exchange term in β_{kl} vanishes. On the other hand, α is maximized for $\phi^{(m)}(\mathbf{r}) \propto \sin(2\pi r)$, giving

$$\alpha^{(m)} = 4\pi^2,$$
$$\beta_{kl}^{(m)} = |\mathbf{a}_k - \mathbf{a}_l|^{-1}, \quad (31)$$

because of the spherical symmetry of $\phi^{(m)}$. Consider now a cubic lattice with period 1. The N points $\{a_k\}$ may be chosen to lie on this lattice within a cube with side $(\mu - 1)$, μ being an integer, such that

$$(\mu-1)^3 \leq N \leq \mu^3. \quad (32)$$

Then, the largest distance between any two points being along the diagonal,

$$|\mathbf{a}_k - \mathbf{a}_l| \leq \sqrt{3}(\mu-1) \leq \sqrt{3}\,N^{\frac{1}{3}}, \quad \forall k, l. \quad (33)$$

From (29), (31), and (33), it follows that:

$$(\Psi^f, H\Psi^f) \leq \lambda^2 2\pi^2 N(\hbar^2/m) - \lambda(2\sqrt{3})^{-1}N^{\frac{2}{3}}(N-1)Gm^2. \quad (34)$$

Minimizing with respect to λ, we obtain

$$E_0^f(N) \leq -(3.2^5\pi^2)^{-1}N^{\frac{1}{3}}(N-1)^2(G^2m^5/\hbar^2). \quad (35)$$

Corollary: Consider an arbitrary system of N identical fermions. The average squared momentum $\langle\mathbf{p}^2\rangle$ of a particle and the average inverse distance between two particles $\langle|\mathbf{r}-\mathbf{r}'|^{-1}\rangle$ obey the generalized uncertainty relation

$$(\langle\mathbf{p}^2\rangle)^{\frac{1}{2}}(\langle|\mathbf{r}-\mathbf{r}'|^{-1}\rangle)^{-1} \geq \tfrac{1}{2}(N-1)^{\frac{1}{3}}\hbar. \quad (36)$$

This expresses the physical nature of the exclusion principle as forbidding the simultaneous presence of more than one fermion in a volume with dimensions comparable to the de Broglie wavelength (see Sec. II).[8]

[8] A similar uncertainty relation dealing with the average squared two-particle distance may be derived from a study of an N-particle system with two-body harmonic forces. See J.-M. Lévy-Leblond, Phys. Letters **26A**, 540 (1968).

Proof: We define

$$\langle \mathbf{p}^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle \mathbf{p}_i^2 \rangle, \tag{37}$$

$$\left\langle \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right\rangle = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \sum \left\langle \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right\rangle. \tag{38}$$

Consider the fictitious Hamiltonian

$$H_{\text{fict}} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2} - g \sum_{1 \le i < j \le N} \sum \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{39}$$

From the lower bound (25), we have

$$\frac{N}{2} \langle \mathbf{p}^2 \rangle - \tfrac{1}{2}[N(N-1)] g \left\langle \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right\rangle \geq -\tfrac{1}{2} N(N-1)^{\frac{4}{3}} (g^2/\hbar^2). \tag{40}$$

For this inequality to hold true for any g, one must have (36).[9]

Theorem 3: For a system of N different particles with maximal mass m and belonging to q species separately obeying the exclusion principle, the ground-state energy obeys

$$-Aq^{\frac{4}{3}} N^{\frac{7}{3}} (G^2 m^5/\hbar^2) \leq E_0^{(q)}(N) \leq -Bq^{\frac{4}{3}} N^{\frac{7}{3}} (G^2 m^5/\hbar^2). \tag{41}$$

This general theorem may be proved in a manner so similar to Theorem 2, as a particular case of it, that we shall skip its proof.

To conclude this section, we note that our results are incomplete in two respects:

(a) We did not attempt to optimize the bounds obtained, that is, to find the best possible constants. This might be interesting for (31), since this general inequality hopefully may find some applications.

(b) We did not succeed in proving that the power laws obtained in Theorems 1, 2, 3 represent the real asymptotic behavior of the ground-state energy. From Theorem 1, for instance, it is natural to conjecture that

$$\lim_{N \to \infty} N^{-3} E_0(N) = C^{\text{st}} \tag{42}$$

for bosons, and similarly for Theorems 2 and 3.

III. GRAVITATIONAL AND ELECTROSTATIC FORCES

A. Heuristic Arguments

The preceding theorems do not apply to the real world, since ordinary matter in bulk consists of charged particles and its behavior depends on electrostatic forces in an essential way. We now extend our considerations to systems of particles where gravitational and electrostatic forces operate simultaneously. Let us first qualitatively discuss the situation. Ordinary matter consists of electrons, light negatively-charged identical fermions, and nuclei, heavy positively-charged particles with either statistics. The main effect of Coulomb forces is to ensure the most precise local neutrality of the system. Any deviation from an exact balance between both types of charge in any region of space would produce huge repulsive forces. As a result, the spatial distribution of the nuclei is much the same as that of the electrons. In particular, even when the nuclei obey Bose statistics, the exclusion principle operating on the electrons to limit their density is "transmitted" to the nuclei by the interplay of Coulomb forces. Due to this adjustment of the spatial distributions for both types of particles, electrons and nuclei also have the same momentum distribution. The essential contribution to the kinetic energy then is furnished by the electrons, because of their much smaller mass. On the other hand, the gravitational potential energy comes essentially from the mutual interaction between the heavier particles, the nuclei.

Finally, at least for a large enough number of particles, the system practically behaves as a collection of N neutral identical fermions in gravitational interaction with the electron mass as their inertial mass and the average nuclear mass as their gravitational mass. We are then led to expect an $N^{\frac{7}{5}}$ behavior of the ground-state energy, despite the possible bosonic nature of the heavy particles. Of course, if N is sufficiently small, the system is dominated by the Coulomb forces and the ground-state energy is roughly a linear function of N[2]. We comment on the transition from Coulomb-like to Newton-like behavior after a precise study of the problem.

B. A Rigorous Result

Theorem 4: For a system consisting of N identical fermions with mass m and charge $(-e)$, and N particles with mass M, charge e, and unspecified statistics, the ground-state energy obeys

$$-CN \frac{me^4}{2\hbar^2} \left(1 + cN^{\frac{2}{3}} \frac{GM^2}{e^2}\right)^2 \leq E_0(N) \leq -DN \frac{me^4}{2\hbar^2} \left(1 + dN^{\frac{2}{3}} \frac{GM^2}{e^2}\right)^2, \tag{43}$$

C, c, D, d being positive constants.

[9] Corresponding to Footnote 7, for $N \to \infty$, the coefficient $\frac{1}{2}$ in (36) may be asymptotically improved in $2^{-\frac{1}{3}} 3^{-\frac{1}{6}} = 0.59 \cdots$.

Proof:

(a) Lower bound: The Hamiltonian of the system is

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m} + \sum_{k=1}^{N} \frac{\mathbf{P}_k^2}{2M} + \sum\sum_{1 \le i < j \le N} \frac{e^2 - Gm^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i=1}^{N}\sum_{k=1}^{N} \frac{e^2 + GmM}{|\mathbf{r}_i - \mathbf{R}_k|} + \sum\sum_{1 \le k < l \le N} \frac{e^2 - GM^2}{|\mathbf{R}_k - \mathbf{R}_l|} \quad (44)$$

with rather obvious notations. This may be written as

$$H = H_e + H_G, \quad (45)$$

where

$$H_e = \lambda \sum \frac{\mathbf{p}_i^2}{2m} + \sum \frac{\mathbf{P}_k^2}{2M} + \sum\sum \frac{e'^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum\sum \frac{e'E'}{|\mathbf{r}_i - \mathbf{R}_k|} + \sum\sum \frac{E'^2}{|\mathbf{R}_k - \mathbf{R}_l|}, \quad (46)$$

$$H_G = (1 - \lambda) \sum \frac{\mathbf{p}_i^2}{2m} - \sum\sum \frac{GM'^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (47)$$

In these definitions, λ is a real number, $0 < \lambda < 1$, and we have introduced fictitious charges (e', E') and mass (M') by the formulas:

$$\begin{aligned} E'^2 &= e^2 - GM^2, \\ e'E' &= e^2 + GmM, \\ e'^2 - GM'^2 &= e^2 - Gm^2, \end{aligned} \quad (48)$$

or

$$\begin{aligned} E' &= e(1 - GM^2/e^2)^{\frac{1}{2}}, \\ e' &= e(1 - GM^2/e^2)^{-\frac{1}{2}}(1 + GmM/e^2), \\ M' &= (M + m)(1 - GM^2/e^2)^{-\frac{1}{2}}. \end{aligned}$$

These definitions require $GM^2 < e^2$, which is true, by and large, for ordinary matter, in which case we have, in effect,

$$e' \simeq E' \simeq e, \quad M' \simeq M + m. \quad (49)$$

The physical interpretation is straightforward. H_e is the Hamiltonian for a system of N fictitious fermions with mass $\lambda^{-1}m$ and charge $-e'$, and N particles with mass M and charge E', interacting via pure Coulomb forces. We know from the work by Dyson and Lenard,[2] that

$$\langle H_e \rangle \ge -C'\lambda^{-1}N(me'^4/2\hbar^2) = -C\lambda^{-1}N(me^4/2\hbar^2) \quad (50)$$

for some positive constant C'. [We then define $C = C'(e'/e)^4$.] H_G is the Hamiltonian for a system of N fictitious neutral fermions with inertial mass $(1 - \lambda)^{-1}m$, interacting via pure Newton forces with gravitational masses M'. Theorem 2, adapted to the present case with two types of masses, tells us that

$$\langle H_G \rangle \ge -A'(1 - \lambda)^{-1}N^{\frac{7}{3}}(G^2mM'^4/\hbar^2) = -A(1 - \lambda)^{-1}N^{\frac{7}{3}}(GmM^4/\hbar^2). \quad (51)$$

The sum of the lower bounds for $\langle H_e \rangle$, Eq. (50), and $\langle H_G \rangle$, Eq. (51), may now be maximized with respect to λ. The best lower bound for $\langle H \rangle = \langle H_e \rangle + \langle H_G \rangle$ is obtained for

$$\lambda_0 = [1 + cN^{\frac{2}{3}}(GM^2/e^2)]^{-1}, \quad (52)$$

where $c = (2A/a)^{\frac{1}{2}}$, and yields the lower bound in (43).

(b) Upper bound: We apply the variational principle to a wavefunction very similar to the one used in the proof of Theorem 2.

C. Newton vs Coulomb

Theorem 4 proves that for ordinary matter, the ground-state energy is linear in the number N of particles for N small enough, when Coulomb forces dominate, but varies as $N^{\frac{7}{3}}$ for N so large that Newton forces become more important. The transition from the Coulomb zone to the Newton zone occurs around a critical number of particles

$$N_c \simeq (e^2/GM^2)^{\frac{3}{2}} \quad (53)$$

within a numerical coefficient, the value of which we ignore. It is clear *a priori* that N_c must be a function of e^2/GM^2, the dimensionless number which characterizes the relative intensity of Coulomb and Newton forces. However, to obtain the exponent $\frac{3}{2}$, it was necessary to solve the saturation problem for gravitational forces. This result can also be interpreted in a very naive way as follows: Let r_0 be the average nearest-neighbor distance between the N particles with mass M and charge e. The mean distance between any two particles then is of order $N^{\frac{1}{3}}r_0$ and the gravitational energy is estimated as

$$E_G \simeq \tfrac{1}{2}N^2(GM^2/N^{\frac{1}{3}}r_0), \quad (54)$$

while the Coulomb energy, due to the self-screening property of electrostatic interactions,[2] is of order

$$E_e \simeq N(e^2/r_0). \quad (55)$$

The two contributions become comparable for the critical number of particles N_c, Eq. (53), already obtained. This argument also shows that the estimate (53) remains valid for ordinary matter, consisting of atoms with arbitrary charge and mass numbers (Z, A), provided M is interpreted as the total atom mass. N_c should then be viewed as the number of atoms above which the cohesive properties of matter

are essentially due to the gravitational pull rather than to Coulomb forces. At that stage, it is only the interatomic bonds which are disrupted by the gravitational forces. A second critical number of atoms N_c' may be defined which corresponds to the gravitational potential energy overcoming the intra-atomic Coulomb energy of matter. For (Z, A) atoms with mass M, an argument similar to those already presented yields

$$N_c' \simeq Z^{\frac{7}{2}}(e^2/GM^2)^{\frac{3}{2}} = Z^{\frac{2}{2}}N_c. \tag{56}$$

For most solid matter, the composition is ironlike, with $Z \simeq 25$, $M \simeq 10^{-25}$ kg. This gives critical masses $\mathcal{M}_c \simeq 10^{23} - 10^{24}$ kg and $\mathcal{M}_c' \simeq 10^{28} - 10^{29}$ kg. It is gratifying to check that the masses of Earth-like planets (composed of such material) are in the vicinity of 10^{24} kg. Indeed, we know gravitational and Coulomb cohesive forces both to be important for their internal structure. Similar remarks could be made for hydrogen-composed planets like Jupiter.

D. The Semirelativistic Case: White Dwarf Stars

We have shown that for N large enough, a system consisting of N electrons and N protons essentially behaves as a system of N identical fermions with inertial mass m and gravitational mass M, governed by an effective Hamiltonian

$$H_G = \sum \frac{\mathbf{p}_i^2}{2m} - \sum\sum \frac{GM^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{57}$$

In the ground state, the average momentum of these particles is of order

$$\bar{p} = N^{\frac{2}{3}}(GM^2m/\hbar). \tag{58}$$

For N sufficiently large so that $\bar{p} \simeq mc$, the electrons can no longer be treated as nonrelativistic and must be endowed with the correct relativistic energy–momentum relationship. No question arises at that stage for the positive particles, much heavier, so that, with the same average momentum, they still behave nonrelativistically. Nor does the potential energy need to be modified since it is mostly due, by and large, to the mutual interaction of these heavy particles. Only the kinetic energy of the electrons, that is the first term of the total Hamiltonian (44), must be changed. As before, the Coulomb energy essentially ensures the local neutrality of the system, for which a semirelativistic effective Hamiltonian may be written similarly to, and in place of, (57):

$$H_G' = \sum (\mathbf{p}_i^2c^2 + m^2c^4)^{\frac{1}{2}} - \sum\sum GM^2/|\mathbf{r}_i - \mathbf{r}_j|. \tag{59}$$

A heuristic evaluation of the ground-state energy may be done as in Sec. I. Calling p the average momentum of a particle, one has

$$E'(N) \simeq N(p^2c^2 + m^2c^4)^{\frac{1}{2}} - \tfrac{1}{2}N^{\frac{5}{3}}(GM^2p/\hbar). \tag{60}$$

The minimal value is reached for

$$p_0 \simeq mc(N/N_r)^{\frac{2}{3}}[1 - (N/N_r)^{\frac{4}{3}}]^{-\frac{1}{2}} \tag{61}$$

and is given by

$$E_0'(N) \simeq Nmc^2[1 - (N/N_r)^{\frac{4}{3}}]^{\frac{1}{2}}, \tag{62}$$

where

$$N_r = (2\hbar c/GM^2)^{\frac{3}{2}}. \tag{63}$$

We see that the binding energy, in this semirelativistic region, increases even faster with N. Above the critical number of particles N_r, the Hamiltonian (59) is no longer bounded from below and the system faces an unescapable collapse. This catastrophy is readily perceived on the ultrarelativistic approximation for $E'(N)$; when $p \simeq mc$, the estimate (60) may be rewritten as

$$E'(N) \simeq N[1 - (N/N_r)^{\frac{2}{3}}]pc. \tag{64}$$

For $N > N_r$, this negative quantity can be made as large as desired by choosing a high enough momentum p. These results may be derived in a rigorous way by using the variational principle to obtain an upper bound for the energy of the considered system. We thus prove:

Theorem 5: The Hamiltonian for a system consisting of N negative "light" fermions and N positive particles with mass M, interacting via Newton and Coulomb forces, is not bounded from below for $N > a(2\hbar c/GM^2)^{\frac{3}{2}}$, where a is some numerical constant.

This situation is illustrated by the white dwarf stars, where the pressure of the degenerate electron gas (the kinetic energy of the electrons) cannot balance the gravitational pull if the total mass is higher than the so-called "Chandrasekhar limit," which is precisely given, within a numerical coefficient, by Eq. (63).[10] This limit, which holds for any system of cold matter, has been derived and estimated here right from the first principles, without using any statistical or thermodynamic argument. In fact, this

[10] S. Chandrasekhar, Monthly Notices Roy. Astron. Soc. **91**, 456 (1931). See, for instance, S. Chandrasekhar, *Introduction to the Study of Stellar Structure* (The University of Chicago Press, Chicago, 1939), or E. Schatzman, *White Dwarfs* (Interscience Publishers, Inc., New York, 1958).

approach merely is a rigorous form of the original discussion by Landau, who first showed the existence of such a limit.[11]

The present considerations, of course, do not apply to ordinary stars where radiation pressure, a typically relativistic electromagnetic effect, holds the system in equilibrium against gravitational collapse. They cannot be generalized either to systems where the heavy particles in turn become degenerate and where other forces (strong, weak) enter the picture, such as in neutron stars, and/or where a relativistic treatment of the gravitational interaction itself becomes necessary.[12]

ACKNOWLEDGMENTS

It is a pleasure to thank S. Deser, F. Dyson, E. Lieb, E. Montroll, and D. Ruelle for useful remarks on topics included in or related to this paper.

[11] L. Landau, Phys. Z. Sowjetunion **1**, 285 (1932), reprinted in *Collected Papers of L. D. Landau*, D. ter Haar, Ed. (Pergamon Press, Inc., New York, 1965), p. 60.

[12] The nonlinear character of a completely relativistic theory of gravitation drastically modifies the saturation problem. For a very simple-minded discussion of this point, see J.-M. Lévy-Leblond and P. Thurnauer, Am. J. Phys. **34**, 1110 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Zero-Mass Representation of Poincaré Group and Conformal Invariance*

S. K. Bose and R. Parker
Department of Physics, University of Notre Dame, Notre Dame, Indiana

(Received 5 September 1968)

By explicit construction it is shown how to extend zero-mass, discrete spin representations of the Poincaré group to corresponding representations of the conformal group of Minkowski space.

It is well known that relativistic wave equations for zero mass are invariant under transformation of the conformal group $SO(4, 2)$ of Minkowski space. This result, first discovered for the Maxwell's equations,[1] was subsequently extended to other zero-mass relativistic wave equations.[2] Some time ago, Gross[3] analyzed this situation in light of the Wigner theorem,[4] which states the complete equivalence between the Lorentz invariance of a quantum-mechanical system and the existence of a unitary, irreducible representation of Poincaré group on the Hilbert space of that quantum system. Specifically, Gross proved that solutions of Maxwell's equations provide unitary representations of the conformal group of Minkowski space and then extended this result to other massless relativistic equations using the Bargmann–Wigner[5] description of particles of zero mass and discrete spin. In other words, the zero-mass, discrete spin representation of the Poincaré group can be extended to provide a representation of the conformal group. We have re-examined this problem and found a rather elementary proof of this result. This is achieved by avoiding the explicit use of the Bargmann–Wigner method which for the massless case introduces a certain redundancy of description (which is to be then eliminated by imposing subsidiary conditions) and using instead an explicit realization of the infinitesimal generators of Poincaré group via one-component wavefunctions given by Lomont and Moses.[6]

We want to prove the following theorem: Any zero-mass, discrete spin representation of Poincaré group automatically admits a unitary representation of conformal group $SO(4, 2)$. We proceed as follows. First, the commutation relations of the 15 infinitesimal generators of the Lie algebra of conformal group are given by[7]

$$[M_{\mu\nu}, M_{\lambda\sigma}] = i(g_{\nu\lambda}M_{\mu\sigma} - g_{\nu\sigma}M_{\mu\lambda} - g_{\mu\lambda}M_{\nu\sigma} + g_{\mu\sigma}M_{\nu\lambda}), \quad (1a)$$

$$[M_{\mu\nu}, P_\lambda] = i(g_{\nu\lambda}P_\mu - g_{\mu\lambda}P_\nu), \quad (1b)$$

$$[P_\mu, P_\nu] = 0, \quad (1c)$$

$$[M_{\mu\nu}, K_\lambda] = i(g_{\nu\lambda}K_\mu - g_{\mu\lambda}K_\nu), \quad (1d)$$

$$[K_\mu, K_\nu] = 0, \quad (1e)$$

* Work supported in part by the U.S. Atomic Energy Commission.

[1] H. Bateman, Proc. London Math. Soc. **8**, 223 (1910); E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910).

[2] J. A. McLennan, Nuovo Cimento **3**, 1360 (1956); J. S. Lomont, Nuovo Cimento **22**, 673 (1961).

[3] L. Gross, J. Math. Phys. **5**, 687 (1964).

[4] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[5] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. US **34**, 211 (1949).

[6] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962). See also Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **33**, 1208 (1958) [Sov. Phys.—JETP **6**, 929 (1958)].

[7] See, for instance, Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953); also H. A. Kastrup, Ann. Physik **9**, 388 (1962).

approach merely is a rigorous form of the original discussion by Landau, who first showed the existence of such a limit.[11]

The present considerations, of course, do not apply to ordinary stars where radiation pressure, a typically relativistic electromagnetic effect, holds the system in equilibrium against gravitational collapse. They cannot be generalized either to systems where the heavy particles in turn become degenerate and where other forces (strong, weak) enter the picture, such as in neutron stars, and/or where a relativistic treatment of the gravitational interaction itself becomes necessary.[12]

ACKNOWLEDGMENTS

It is a pleasure to thank S. Deser, F. Dyson, E. Lieb, E. Montroll, and D. Ruelle for useful remarks on topics included in or related to this paper.

[11] L. Landau, Phys. Z. Sowjetunion **1**, 285 (1932), reprinted in *Collected Papers of L. D. Landau*, D. ter Haar, Ed. (Pergamon Press, Inc., New York, 1965), p. 60.

[12] The nonlinear character of a completely relativistic theory of gravitation drastically modifies the saturation problem. For a very simple-minded discussion of this point, see J.-M. Lévy-Leblond and P. Thurnauer, Am. J. Phys. **34**, 1110 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Zero-Mass Representation of Poincaré Group and Conformal Invariance*

S. K. Bose and R. Parker
Department of Physics, University of Notre Dame, Notre Dame, Indiana

(Received 5 September 1968)

By explicit construction it is shown how to extend zero-mass, discrete spin representations of the Poincaré group to corresponding representations of the conformal group of Minkowski space.

It is well known that relativistic wave equations for zero mass are invariant under transformation of the conformal group $SO(4, 2)$ of Minkowski space. This result, first discovered for the Maxwell's equations,[1] was subsequently extended to other zero-mass relativistic wave equations.[2] Some time ago, Gross[3] analyzed this situation in light of the Wigner theorem,[4] which states the complete equivalence between the Lorentz invariance of a quantum-mechanical system and the existence of a unitary, irreducible representation of Poincaré group on the Hilbert space of that quantum system. Specifically, Gross proved that solutions of Maxwell's equations provide unitary representations of the conformal group of Minkowski space and then extended this result to other massless relativistic equations using the Bargmann–Wigner[5] description of particles of zero mass and discrete spin. In other words, the zero-mass, discrete spin representation of the Poincaré group can be extended to provide a representation of the conformal group. We have re-examined this problem and found a rather elementary proof of this result. This is achieved by avoiding the explicit use of the Bargmann–Wigner method which for the massless case introduces a certain redundancy of description (which is to be then eliminated by imposing subsidiary conditions) and using instead an explicit realization of the infinitesimal generators of Poincaré group via one-component wavefunctions given by Lomont and Moses.[6]

We want to prove the following theorem: Any zero-mass, discrete spin representation of Poincaré group automatically admits a unitary representation of conformal group $SO(4, 2)$. We proceed as follows. First, the commutation relations of the 15 infinitesimal generators of the Lie algebra of conformal group are given by[7]

$$[M_{\mu\nu}, M_{\lambda\sigma}] = i(g_{\nu\lambda}M_{\mu\sigma} - g_{\nu\sigma}M_{\mu\lambda} - g_{\mu\lambda}M_{\nu\sigma} + g_{\mu\sigma}M_{\nu\lambda}), \quad (1a)$$

$$[M_{\mu\nu}, P_\lambda] = i(g_{\nu\lambda}P_\mu - g_{\mu\lambda}P_\nu), \quad (1b)$$

$$[P_\mu, P_\nu] = 0, \quad (1c)$$

$$[M_{\mu\nu}, K_\lambda] = i(g_{\nu\lambda}K_\mu - g_{\mu\lambda}K_\nu), \quad (1d)$$

$$[K_\mu, K_\nu] = 0, \quad (1e)$$

* Work supported in part by the U.S. Atomic Energy Commission.

[1] H. Bateman, Proc. London Math. Soc. **8**, 223 (1910); E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910).

[2] J. A. McLennan, Nuovo Cimento **3**, 1360 (1956); J. S. Lomont, Nuovo Cimento **22**, 673 (1961).

[3] L. Gross, J. Math. Phys. **5**, 687 (1964).

[4] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[5] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. US **34**, 211 (1949).

[6] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962). See also Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **33**, 1208 (1958) [Sov. Phys.—JETP **6**, 929 (1958)].

[7] See, for instance, Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953); also H. A. Kastrup, Ann. Physik **9**, 388 (1962).

$$[K_\mu, P_\nu] = 2i(g_{\mu\nu}D - M_{\mu\nu}), \tag{1f}$$

$$[M_{\mu\nu}, D] = 0, \tag{1g}$$

$$[D, P_\mu] = iP_\mu, \tag{1h}$$

$$[D, K_\mu] = -iK_\mu. \tag{1i}$$

The metric chosen is $g_{00} = 1$, $g_{11} = g_{22} = g_{33} = -1$, $g_{ij} = 0$ otherwise. The 10 generators P_μ and $M_{\mu\nu}$ $(= -M_{\nu\mu})$ correspond to the Poincaré group, which is a subgroup of $SO(4, 2)$. Let us now consider the zero-mass, discrete spin representation of Poincaré group which is characterized by $P_\mu^2 = 0$, $W_\mu = \lambda P_\mu$; W_μ $(= -\frac{1}{2}\epsilon_{\mu\nu\lambda\sigma}M_{\nu\lambda}P_\sigma)$ being the Pauli–Lubanski operator and λ the helicity. We do not concern ourselves with the zero-mass, continuous spin representations $P_\mu^2 = 0$, $W_\mu^2 > 0$ in this paper. An explicit form of realizations of the infinitesimal generators for massless, discrete spin is given in Ref. 6. Consider a space V of complex functions $f(\mathbf{p}, \lambda)$, where $\mathbf{p} = (p_1, p_2, p_3)$ and the range of each variable p_i extends over the entire real axis. The value of λ, which includes the sign, is helicity for spin $|\lambda|$ and is fixed. Then the form of infinitesimal generators in V is given by[6]

$$P_0 f(\mathbf{p}, \lambda) = pf(\mathbf{p}, \lambda),$$

$$P_i f(\mathbf{p}, \lambda) = p_i f(\mathbf{p}, \lambda); \quad p = (p_1^2 + p_2^2 + p_3^2)^{\frac{1}{2}}, \tag{2a}$$

$$M_1 = -i\epsilon_{1jk}p_j\partial_k + \frac{\lambda p_1}{p + p_3}, \tag{2b}$$

$$M_2 = -i\epsilon_{2jk}p_j\partial_k + \frac{\lambda p_2}{p + p_3}, \tag{2c}$$

$$M_3 = -i\epsilon_{3jk}p_j\partial_k + \lambda, \tag{2d}$$

$$N_1 = -ip\partial_1 - \frac{\lambda p_2}{p + p_3}, \tag{2e}$$

$$N_2 = -ip\partial_2 + \frac{\lambda p_1}{p + p_3}, \tag{2f}$$

$$N_3 = -ip\partial_3. \tag{2g}$$

In the above, the usual notations $M_{ij} = \epsilon_{ijk}M_k$ and $N_i = M_{0i}$ $(i = 1, 2, 3)$ have been used and, furthermore, ∂_i denotes $\partial/\partial p_i$ and repeated indices implies summation. The scalar product in V is defined as

$$(f^{(1)}, f^{(2)}) = \int \frac{d^3p}{p} f^{(1)*}(\mathbf{p}, \lambda) f^{(2)}(\mathbf{p}, \lambda). \tag{3}$$

With respect to the above scalar product operators, $M_{\mu\nu}$ and P_μ are Hermitian. We are now in a position to formulate our problem mathematically. To prove the desired theorem we have to be able to show that there exist linear operators in space V with correct commutation properties which can be identified with generators K_μ and D. The proof is by explicit construction. We write down the form of these operators in V:

$$D = i(p \cdot \partial + 1), \tag{4a}$$

$$K_0 = -p\partial^2 + 2i\lambda\left(\frac{p_2}{p + p_3}\partial_1 - \frac{p_1}{p + p_3}\partial_2\right) + \frac{2\lambda^2}{p + p_3}, \tag{4b}$$

$$K_1 = 2\partial_1 + 2(p \cdot \partial)\partial_1 - p_1\partial^2 + 2i\lambda\left(\frac{p_2}{p + p_3}\partial_3 - \partial_2\right), \tag{4c}$$

$$K_2 = 2\partial_2 + 2(p \cdot \partial)\partial_2 - p_2\partial^2 + 2i\lambda\left(\partial_1 - \frac{p_1}{p + p_3}\partial_3\right), \tag{4d}$$

$$K_3 = 2\partial_3 + 2(p \cdot \partial)\partial_3 - p_3\partial^2 + 2i\lambda\left(\frac{p_1}{p + p_3}\partial_2 - \frac{p_2}{p + p_3}\partial_1\right) - \frac{2\lambda^2}{p + p_3}. \tag{4e}$$

In Eq. (4), we have used the abbreviations $\partial^2 = \partial_1^2 + \partial_2^2 + \partial_3^2$ and $p \cdot \partial = p_1\partial_1 + p_2\partial_2 + p_3\partial_3$. It may be verified by direct calculation that commutation relations (1d)–(1i) are satisfied by the operators given by Eq. (4). With respect to the scalar product (3), these operators are Hermitian. This completes the proof. We may finally note that the limiting case $\lambda = 0$ of Eq. (4) yields the known form[8] of operators D and K_μ for that special case.

ACKNOWLEDGMENTS

We are thankful to Professor J. T. Cushing and Professor W. D. McGlinn for many helpful conversations.

[8] H. A. Kastrup, Phys. Rev. **142**, 1060 (1966).

Best Error Bounds for Padé Approximants to Convergent Series of Stieltjes*

GEORGE A. BAKER, JR.
Applied Mathematics Department, Brookhaven National Laboratory, Upton, Long Island, New York

(Received 2 August 1968)

We consider series of Stieltjes with a nonzero radius of convergence R. We establish by way of Padé approximants the allowed range of values for such functions at any point in the cut $(-\infty < z \leq -R)$ complex plane when a finite number of Taylor-series coefficients are known. The previous results for z real and positive are sharpened. We investigate the fitting problem and again give error bounds throughout the cut complex plane and we give necessary and sufficient conditions that the set of values fitted be values of a series of Stieltjes with radius of convergence at least R.

1. INTRODUCTION

Common[1] has recently proposed the following interesting problem: Suppose $f(z)$ is a member of the class of functions called series of Stieltjes (defined in Sec. 2) with nonzero radius of convergence. From a finite number only of known Taylor-series coefficients, what bounds can be placed on the error of the $[N, M]$ Padé approximant at any point in the complex z plane? The $[N, M]$ Padé approximant to a function[2]

$$f(z) = \sum_{j=0}^{\infty} a_j z^j \tag{1.1}$$

is defined by the equations

$$f(z)P_M(z) - Q_N(z) = O(z^{M+N+1}),$$
$$Q_N(z) = 1.0, \tag{1.2}$$

where P_M and Q_N are polynomials of degree at most M and N, respectively. The totality of Padé approximants is arranged in a 2-dimensional, semi-infinite table. The entries in the first row comprise the partial sums of the Taylor series and are denoted $[0, 0]$, $[0, 1]$, $[0, 2], \cdots$. The second row has a linear denominator and is denoted $[1, 0]$, $[1, 1]$, $[1, 2]$, $[1, 3], \cdots$, and so forth. The genesis of the recent interest in Common's problem lies in the application of Padé-approximant techniques to partial-wave integral equations. Mason[3] has shown that the s-wave scattering amplitude for two spinless particles, when considered as a function of the strength of the left-hand cut discontinuity, is related to a series of Stieltjes and hence the Padé can be used as an alternative to the N/D method.[4] His result is particularly desirable because, for series of Stieltjes, one can prove when z is real and positive that

$$[N, N-1] \leq f(z) \leq [N, N]. \tag{1.3}$$

For $z < 0$, all the Padé approximants form lower bounds. Common[1] points out that it is also desirable to have error bounds on the negative real axis as well [and more generally throughout the cut $(-\infty < z \leq -R)$ complex plane] and does obtain bounds on the negative real axis and in part of the complex plane.

In this paper we give the solution to the Common[1] problem. Through the use of Padé approximants we prove that given any point z in the cut $(-\infty < z \leq -R)$ complex plane, the first p coefficients determine a lens-shape range of possible values for $f(z)$. These results are the best possible in the sense that any point of this range can actually be attained by some member of the class of functions considered.

We show that the knowledge that $f(z)$ has a radius of convergence of at least R allows us to sharpen (1.3) for positive real z.

We have also investigated the fitting problem, i.e., determining Padé approximants from the values of $f(z)$ at a set of points rather than from a sequence of derivatives. We derive necessary and sufficient conditions that the $\{f(z_j)\}$ are values of a series of Stieltjes with a radius of convergence at least R and value ranges (error bounds) for $f(z)$ with the values $\{f(z_j), j = 0, 1, \cdots, p\}$ for z in the cut $(-\infty < z \leq -R)$ complex plane.

2. RANGE OF A RESTRICTED SERIES OF STIELTJES

In this section we show how to compute the possible values $f(z)$ of any series of Stieltjes at any point z in the cut $(-\infty \leq z \leq -R)$ complex z plane, when we are given the first n terms of the Taylor-series development of $f(z)$ about $z = 0$ and the information that this series has a radius of convergence of at least R. This range for $f(z)$ is simply obtained in terms of adjacent Padé approximants in the Padé table.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] A. K. Common, J. Math. Phys. **9**, 32 (1968).
[2] G. A. Baker, Jr., Advan. Theoret. Phys. **1**, 1 (1965).
[3] D. Mason, J. Math. Phys. **8**, 512 (1967).
[4] G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).

By definition,

$$f(z) = \sum_{j=0}^{\infty} f_j(-z)^j \tag{2.1}$$

is a series of Stieltjes, if and only if there is a bounded, nondecreasing function $\varphi(u)$ taking on infinitely many values in the interval $0 \leq u < \infty$, such that

$$f_j = \int_0^{\infty} u^j \, d\varphi(u). \tag{2.2}$$

If we introduce the determinants

$$D(m, n) = \det \begin{vmatrix} f_m & f_{m+1} & \cdots & f_{m+n} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n+1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_{m+n} & f_{m+n+1} & \cdots & f_{m+2n} \end{vmatrix}, \tag{2.3}$$

then this definition is equivalent[2] to the conditions

$$D(0, n) > 0, \quad D(1, n) > 0, \quad n = 0, 1, 2, \cdots. \tag{2.4}$$

All the $D(m, n)$ are actually positive,[2] but we only require $m = 0, 1$ to prove form (2.2).

The first step in our procedure is to use the known result that, if $f(z)$ is a series of Stieltjes of radius of convergence at least R, then $g(z)$, defined by

$$f(z) = \frac{f_0}{1 + zg(z)}, \tag{2.5}$$

is also. That this is so can be seen formally through the use of Hadamard's[5] theorem on determinants. If $D_g(m, n)$ are the determinants for $g(z)$ analogous to (2.3) for $f(z)$, then

$$\begin{aligned} D_g(0, p) &= D(1, p)/f_0^{2p+4} > 0, \\ D_g(1, p) &= D(0, p + 1)/f_0^{2p+5} > 0, \end{aligned} \tag{2.6}$$

which via (2.4) establishes that $g(z)$ is a series of Stieltjes. As $f(z)$ is assumed to have a radius of convergence R, it has the integral representation[2]

$$f(z) = \int_0^{1/R} \frac{d\varphi(u)}{1 + uz}. \tag{2.7}$$

From this representation it is clear that $f(z)$ is regular in the cut plane $(-\infty \leq z \leq -R)$. Except for possible polar singularities, $g(z)$ must be also, by (2.5). However, if $g(z)$ had a pole, then $f(z)$ would vanish; but it follows easily from (2.7) that $f(z) \neq 0$ in the cut z plane. Thus $g(z)$ also has a radius of convergence of at least R. As

$$\frac{-[f(z) - f_0 + f_1 z - f_2 z^2 + \cdots - f_{n-1}(-z)^{n-1}]}{(-z)^n} = \int_0^{1/R} \frac{u^n \, d\varphi(u)}{1 + uz} \tag{2.8}$$

is a series of Stieltjes, then $g(z)$, defined by

$$f(z) = f_0 - f_1 z + f_2 z^2 + \cdots + f_{n-1}(-z)^{n-1} + \frac{f_n(-z)^n}{1 + zg(z)}, \tag{2.9}$$

is also a series of Stieltjes.

As $g(z)$ in Eq. (2.5) is a member of the same class as $f(z)$, namely, a series of Stieltjes of radius of convergence at least R, we are free to iterate that form, starting either with (2.5) or (2.9). When we do, we obtain

$$f(z) = \cfrac{f_0}{1 + \cfrac{za_1}{1 + \cfrac{za_2}{1 + \ddots \cfrac{}{1 + \cfrac{za_p}{1 + zh_p(z)}}}}}, \tag{2.10}$$

where $h_p(z)$ is again a series of Stieltjes with radius of convergence at least R. Equation (2.10) may be re-expressed as

$$f(z) = \frac{A_p(z) + zh_p(z)A_{p-1}(z)C_p}{B_p(z) + zh_p(z)B_{p-1}(z)C_p}, \tag{2.11}$$

where the A's and B's are polynomials in z, and the C_p are constants. That the form is correct can be seen by setting $h_p = 0$ and ∞ in (2.10) and noting that we obtain the pth and $(p - 1)$th expressions, respectively. Wall[6] shows (Theorems 96.1 and 97.1) that the fractions $A_p(z)/B_p(z)$ are in fact Padé approximants and fill a stairlike sequence in the Padé table, i.e., [0, 0], [1, 0], [1, 1], [2, 1], [2, 2], $\cdots$. If we start with (2.9) instead of (2.5), we obtain the sequence $[0, n]$, $[1, n]$, $[1, n + 1]$, $[2, n + 1]$, $[2, n + 2]$, $\cdots$, in an exactly analogous manner.

Returning to Eq. (2.5), we note one further restriction that $g(z)$ must satisfy, namely, as $g(z)$ is monotonic in the range $-R \leq z \leq 0$, we must have

$$\lim_{z \to -R} zg(z) \leq 1, \tag{2.12}$$

[5] J. Hadamard, J. Math. (4) 8, 101 (1892).

[6] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand, Inc., Princeton, N.J., 1948).

in order that $f(z)$ be free of singularities in this range. Therefore,

$$g(-R) \leq R^{-1}. \tag{2.13}$$

As we iterate (2.5), the $h_p(z)$ obtained will be similarly restricted, although the bound is now a function of all the constants in (2.10) which may be defined from[6] the first p coefficients of the power series for $f(z)$. We choose to redefine $h_p(z)$, so that

$$h_p(-R) \leq R^{-1}, \tag{2.14}$$

and absorb the change in normalization in the C_p. We may now easily solve for C_p from (2.11), using the critical equation which makes $z = -R$ a pole of $f(z)$. Thus,

$$C_p = B_p(-R)/B_{p-1}(-R). \tag{2.15}$$

Having established relation (2.11) subject to (2.14) and (2.15), we see that the range of $f(z)$, given the first p coefficients and given that it is a series of Stieltjes with radius of convergence of at least R, is just the linear fractional transformation of the range of $h_p(z)$. The reader will note that this portion of our discussion has followed Gordon's[7] adaption of the ideas of Hamburger,[8] although with different emphasis.

We now turn our attention to the problem of computing the range of $h_p(z)$, subject to (2.14). Now we can represent

$$h_p(z) = \int_0^{1/R} \frac{d\psi_p(u)}{1 + uz} \tag{2.16}$$

which evaluated at $z = -R$ is

$$h_p(-R) = \int_0^{1/R} \frac{d\psi_p(u)}{1 - Ru} \leq \frac{1}{R}. \tag{2.14'}$$

By (2.14'),

$$d\omega_p(u) = \frac{R\, d\psi_p(u)}{1 - Ru} \tag{2.17}$$

is also an allowable measure with

$$\int_0^{1/R} d\omega_p(u) \leq 1. \tag{2.18}$$

Hence we may rewrite

$$h_p(z) = \frac{1}{R} \int_0^{1/R} \frac{(1 - Ru)\, d\omega_p(u)}{1 + uz}, \tag{2.19}$$

with $d\omega_p$ an arbitrary, non-negative-definite, normalized measure. It follows at once from (2.19) that

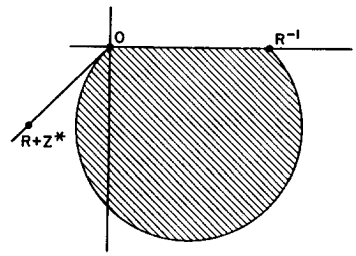


FIG. 1. The shaded region is the allowed range of values of $h_p(z)$ for the case $z = -2R + iR$. Note that the circle is tangent to the line $\overline{0, R + z^*}$.

if H_1 and H_2 are possible values of $h_p(z)$ then $\alpha H_1 + (1 - \alpha)H_2$ $(0 \leq \alpha \leq 1)$ are also, as if ω and ω' are allowed measures in (2.19) then so is $\alpha\omega + (1 - \alpha)\omega'$. Consequently, the range of $h_p(z)$ is a convex region. The integrand of (2.19) is a weighted sum of

$$\frac{1}{R}\left(\frac{1 - Ru}{1 + uz}\right), \quad 0 \leq u \leq \frac{1}{R}, \tag{2.20}$$

which is a linear fractional transformation of the segment of the u axis. By a well-known property of linear fractional transformations,[9] the result is the arc of a circle. Hence the range of $h_p(z)$ is the convex hull of this arc. We have illustrated (Fig. 1) the sample case $z = -2R + iR$. It is tangent, at the origin to a line through $R + z^*$. The vertical height can be easily calculated to be

$$0 \leq -\operatorname{Im} h_p(z) \leq \frac{y}{2R\{[(R + x)^2 + y^2]^{\frac{1}{2}} + (R + x)\}}, \tag{2.21}$$

where x and y are the real and imaginary parts of z. The range is the complete, convex, lens-shaped region described above, since any point on the circular arc can be obtained for $d\omega_p(u) = \delta(u - u_0)\, du$, where u_0 is appropriately selected. Any point on the real-axis portion of the boundary can be obtained as $d\omega(u) = h_p(z)\delta(u)\, du$. All points in the interior can be obtained in an infinite number of ways as linear combinations of boundary points. The range of $f(z)$ is therefore the map of this lens-shaped region under (2.11). The resulting range $F_p(z)$ will again be a lens-shaped[9] region, although not necessarily convex. By the method of construction, we have

$$f(z) \in F_p(z) \subset F_{p-1}(z) \subset \cdots \subset F_1(z), \tag{2.22}$$

where $\subset$ means "contained in." We expect the sequence of ranges based on (2.5) to be superior (for the same number of coefficients) to those based on (2.9) with $n > 0$, as they maintain an additional provable requirement that $f(z)$ be bounded at infinity (away

[7] R. G. Gordon, J. Math. Phys. **9**, 1087 (1968).

[8] H. Hamburger, Math. Ann. **81**, 235 (1920); **82**, 120, 168 (1921), as reviewed by J. A. Shahat and J. D. Tamarkin, *Mathematical Surveys, Number I: The Problem of Moments* (American Mathematical Society, New York, 1943).

[9] A linear fractional transformation maps the family of circles and straight lines into itself. See, for example, E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University Press, London, 1948), p. 11, Ex. 6.

from the cut). We now see that this is indeed the case and, hence, $F_p(z)$ is the best possible bound for the values of $f(z)$ based on p coefficients and the assumption that the radius of convergence is at least R.

In order to discuss the relations between the possible ranges for $f(z)$ based on (2.9) for different values of n, we introduce a range for every nearest-neighbor pair in the (upper-half) Padé table. Following the notation of Chap. II, Sec. B of our review article[2] for the numerators and denominators of Padé approximants

$$[N, N+j](z) = \frac{P_N^{(j)}(z)}{Q_N^{(j)}(z)}, \tag{2.23}$$

we may write out explicitly, for the even

$$f(z) = \frac{P_m^{(n)}(z)Q_m^{(n-1)}(-R) + zh_{2m,n}(z)P_m^{(n-1)}(z)Q_m^{(n)}(-R)}{Q_m^{(n)}(z)Q_m^{(n-1)}(-R) + zh_{2m,n}(z)Q_m^{(n-1)}(z)Q_m^{(n)}(-R)} \tag{2.24}$$

and odd

$$f(z) = \frac{P_{m+1}^{(n-1)}(z)Q_m^{(n)}(-R) + zh_{2m+1,n}(z)P_m^{(n)}(z)Q_{m+1}^{(n-1)}(-R)}{Q_{m+1}^{(n-1)}(z)Q_m^{(n)}(-R) + zh_{2m+1,n}(z)Q_m^{(n)}(z)Q_{m+1}^{(n-1)}(-R)} \tag{2.25}$$

values of p, the formulas relating $f(z)$ to $h(z)$. The map of the range of the $h(z)$ found by considering (2.20) defines the set of ranges $F_{p,n}(z)$. Equation (2.22) becomes, when we start with (2.9), for $n = 0, 1, 2, \cdots$,

$$f(z) \in F_{p,n}(z) \subset F_{p-1,n}(z) \subset \cdots \subset F_{1,n}(z), \qquad n = 0, 1, 2, \cdots. \tag{2.26}$$

These relations generalize the inequalities (II.19a) and (II.19c) of Theorem 6 of our review article[2] to arbitrary values of z in the cut complex plane. To complete the picture we need to examine the relations between F for different values of n. We first consider the possible relation

$$F_{2m+1,n}(z) \subset F_{2m,n+1}(z). \tag{2.27}$$

To this end we define

$$\tau = zh_{2m+1,n}(z)Q_{m+1}^{(n-1)}(-R)/Q_m^{(n)}(-R),$$
$$\sigma = zh_{2m,n+1}(z)Q_m^{(n+1)}(-R)/Q_m^{(n)}(-R). \tag{2.28}$$

If we now solve for τ in terms of σ from Eqs. (2.24) and (2.25) we get

$$\tau = \frac{[P_{m+1}^{(n-1)}(z)Q_m^{(n+1)}(z) - P_m^{(n+1)}(z)Q_m^{(n-1)}(z)] + [P_{m+1}^{(n-1)}(z)Q_m^{(n)}(z) - P_m^{(n)}(z)Q_{m+1}^{(n-1)}(z)]\sigma}{[P_m^{(n+1)}(z)Q_m^{(n)}(z) - Q_m^{(n+1)}(z)P_m^{(n)}(z)]}. \tag{2.29}$$

In a manner exactly analogous to the derivation of Eqs. (II.18) and (II.20) of Ref. 2, we may show that

$$P_{m+1}^{(n-1)}(z)Q_m^{(n+1)}(z) - P_m^{(n+1)}(z)Q_{m+1}^{(n-1)}(z) = (-z)^{2m+n+2}D^2(1+n, m+1), \tag{2.30a}$$

$$P_{m+1}^{(n-1)}(z)Q_m^{(n)}(z) - P_m^{(n)}(z)Q_{m+1}^{(n-1)}(z) = (-z)^{2m+n+1}D(n, m)D(1+n, m), \tag{2.30b}$$

$$P_m^{(n+1)}(z)Q_m^{(n)}(z) - P_m^{(n)}(z)Q_m^{(n+1)}(z) = (-z)^{2m+n+1}D(2+n, m-1)D(1+n, m). \tag{2.30c}$$

If we use (2.30) to simplify (2.29) we obtain

$$\begin{pmatrix}\tau\\ z\end{pmatrix} = -\frac{[D(1+n, 1+m)]^2}{D(2+n, m-1)D(1+n, m)} + \frac{D(n, m)}{D(2+n, m-1)}\begin{pmatrix}\sigma\\ z\end{pmatrix}, \tag{2.31}$$

which is a *linear* relationship between the points of $F_{2m,n+1}$ and $F_{2m+1,n}$. Since (2.31) is a linear transformation, it is automatically a conformal one. As the constants are real, the real-axis portion of the boundary to the range of (σ/z) maps into a portion of the real axis of the (τ/z) plane. The origin in the (σ/z) plane maps into a real negative point in the (τ/z) plane. As Eq. (2.29) is equivalent to (2.31), we may obtain the value of τ which corresponds to the value of σ when $h_{2m,n+1}(z) = 1/R$ by substituting into (2.29) for the particular value of $z = -R$. This yields

$$\tau = -Q_m^{(n)}(-R)/Q_{m+1}^{(n-1)}(-R), \tag{2.32}$$

which implies the corresponding mapped value

$$h_{2m+1,n}(-R) = 1/R, \tag{2.33}$$

which is the same maximum value that we have previously derived. As the map of the (σ/z)-plane, straight-line portion of the boundary of the range includes all of the straight-line portion of the boundary of the range in the (τ/z) plane, and as the transformation is a linear, shape-preserving one, we see that the curved boundary lies outside the boundary for the

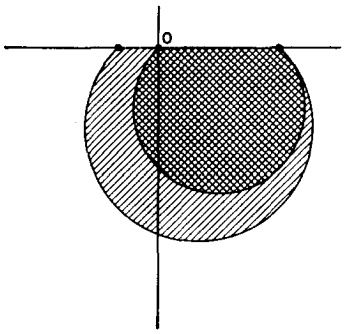


FIG. 2. The shaded area is the map of the allowed region in the (σ/z) plane. The cross-hatched area is the region in the (τ/z) plane.

region in the (τ/z) plane. (See Fig. 2). Thus we have proved (2.27). The similar relation for $F_{2m,n}$ can be proven in an exactly analogous manner, except that the map of the (σ/z)-plane range has the curved side in common with the (τ/z)-plane range instead of the straight side. We may thus supplement the inclusion relations (2.26) with the inclusion relations

$$F_{p,n}(z) \subset F_{p-1,n+1}(z). \tag{2.34}$$

Taken together these imply that

$$f(z) \in F_{p,0}(z) \tag{2.35}$$

is the best possible bound for $f(z)$ which we may form from the coefficients through z^p, where the $F_{p,0}(z)$ are defined by (2.24) and (2.25) with the aforementioned permissible range for the h's as illustrated in Fig. 1, and defined as the convex hull of Eq. (2.20).

The best bounds in the cut z plane $(-\infty < z \leq 0)$ for the problem (2.2) can be obtained as a special case of our results. This represents the limiting case where $R \to 0$. The formulas (2.24) and (2.25) remain valid. By definition $Q_m^{(n)}(0) = D(1+n, m-1)$ for all m, n. The range for $h_{p,n}$ follows directly as the limit as $R \to 0$ of (2.20). It is the complete angular wedge bounded by the real axis and the ray through $R + z^*$. The sequence with $n = 0$ is again the best one. As a wedge is merely a lens-shaped region with one vertex at infinity, the ranges $F_{p,n}$ for the value of $f(z)$ are again lens-shaped regions.

The problem where

$$f_j = \int_{-\infty}^{+\infty} u^j \, d\varphi(u) \tag{2.36}$$

was completely treated by Hamburger[8] and Gordon.[7] Here the range of values becomes the entire half-plane Im $(h(z))$ Im $(z) < 0$. The map of this is a circle, which is a degenerate lens in which the two sides meet at a straight angle.

We consider one further case.

If the problem is defined by

$$f_j = \int_{-1/S}^{1/R} u^j \, d\varphi(u) \tag{2.37}$$

and neither R nor S is zero, then we may show (Chap. IV, Sec. A in Ref. 2) that both

$$\frac{R+S}{S+R(1+\xi)} f\left(\frac{RS\xi}{S+R(1+\xi)}\right) \tag{2.38a}$$

$$\frac{R+S}{R+S(1+\xi)} f\left(\frac{-RS\xi}{R+S(1+\xi)}\right) \tag{2.38b}$$

are series of Stieltjes with a radius of convergence of unity. Thus the results of this section apply to them. The limits $R \to 0$ and $S \to 0$ may be treated by letting $\eta = R\xi$ in (2.38a) and $\eta = S\xi$ in (2.38b), respectively, which yields

$$\frac{S}{S+\eta} f\left(\frac{S\eta}{S+\eta}\right) \tag{2.39a}$$

$$\frac{R}{R+\eta} f\left(\frac{-R\eta}{R+\eta}\right) \tag{2.39b}$$

as series of Stieltjes of form (2.2). Hence the results discussed above apply to these cases also.

3. SPECIALIZATION TO REAL ARGUMENT

When the best possible bounds derived in the previous section are specialized to real values of z, the lens-shaped region collapses into a real interval. The restriction that $f(z)$ has a radius of convergence of at least R enables us to derive sharper bounds than has been possible heretofore. For real z, it follows from (2.20) that

$$0 \leq h(z) \leq 1/R. \tag{3.1}$$

Hence, for $z > 0$, the inclusion relation (2.35) becomes the inequalities

$$\frac{P_m^{(0)}(z)Q_m^{(-1)}(-R) + (z/R)P_m^{(-1)}(z)Q_m^{(0)}(-R)}{Q_m^{(0)}(z)Q_m^{(-1)}(-R) + (z/R)P_m^{(-1)}(z)Q_m^{(0)}(-R)}$$

or

$$\leq f(z) \leq [m, n] \tag{3.2}$$

$$[m+1, m] \leq f(z)$$

$$\leq \frac{P_{m+1}^{(-1)}(z)Q_m^{(0)}(-R) + (z/R)P_m^{(0)}(z)Q_{m+1}^{(-1)}(-R)}{Q_{m+1}^{(-1)}(z)Q_m^{(0)}(-R) + (z/R)Q_m^{(0)}(z)Q_{m+1}^{(-1)}(-R)}. \tag{3.3}$$

Equation (3.2) is applicable when the last available coefficient is for z^{2m} and (3.3) is applicable for z^{2m+1}. When $-R < z < 0$, the sense of the inequality signs in (3.2) reverses, and that in (3.3) remains unchanged.

These bounds $(z > 0)$ reduce to the previous results [(II.19c) of Ref. 2]:

$$[N, N-1] \leq f(z) \leq [N, N] \tag{3.4}$$

in the limit as $(z/R) \to \infty$ $(R \to 0)$. For (z/R) finite, a better bound is provided, either upper, if the last available coefficient is for z^{2m+1}, or lower, if it is for z^{2m}.

For, $z < 0$, as has been pointed out by Common,[1] both the $[N, N]$ and $[N, N-1]$ Padé approximants

TABLE I. $f(z)$ and its approximants (values are multiplied by 10^2).

z	[3, 2]	$f(z)$	max $\{F_{5,0}(z)\}$	(3, 2)
−0.6	8.8036	8.8051	8.8072	8.8081
−0.8	12.168	12.216	12.289	12.448
−0.9	15.829	16.213	16.881	19.770

form lower bounds to $f(z)$. He has introduced a set of modified approximants which provide upper bounds in this region $(-R \leq z \leq 0)$. They are defined as follows: Let $f(z)$ be a series of Stieltjes

$$f(z) = \sum_{j=0}^{\infty} f_j(-z)^j \tag{3.5}$$

and define

$$k_j = \frac{1}{j+1}\left\{\frac{f_0}{R^{j+1}} - f_{j+1}\right\}, \tag{3.6}$$

$$K(z) = \sum_{j=0}^{\infty} k_j(-z)^j, \tag{3.7}$$

then

$$f(z) = [(f_0R)/(R+z) + z[zK'(z) + K(z)]. \tag{3.8}$$

Common proves that $K(z)$ is a series of Stieltjes with radius of convergence of at least R. Common's approximants, which he denotes by (M, N), consist of replacing $K(z)$ by its $[M, N]$ Padé approximant. He proves, among other things, that

$$(N, N-1) \geq (N, N) \geq (N+1, N) \geq f(z) \geq [N+1, N] \geq [N, N] \geq [N, N-1] \tag{3.9}$$

over the range $-R \leq z \leq 0$. He has illustrated his results for the example

$$\begin{aligned} f(z) &= \frac{1}{4\pi}\int_0^1 \frac{u^{\frac{1}{2}}\,du}{1+zu} \\ &= \frac{1}{2\pi z}\left\{1 - \frac{1}{2(-z)^{\frac{1}{2}}}\log\left[\frac{1+(-z)^{\frac{1}{2}}}{1-(-z)^{\frac{1}{2}}}\right]\right\} \\ &= \frac{1}{2\pi}\left[\frac{1}{3} - \frac{z}{5} + \frac{z^2}{7} - \frac{z^3}{9} + \cdots\right]. \end{aligned} \tag{3.10}$$

It is interesting to compare them with our best possible results. We have done this briefly in Table I. The values of Common's approximants are taken from his paper.[1] It will be observed that the best possible upper bounds are significantly closer than Common's approximants and for this example the function value lies in the central third of the allowed range.

4. THE FITTING PROBLEM FOR A SERIES OF STIELTJES

We consider here the problem of constructing the possible range of values of a restricted series of Stieltjes (radius of convergence of at least R) when we are given, not a set of derivatives at a single point, but a set of values at various points $\{f(z_j);\ j = 1, 2, \cdots, n\}$. We are able to treat this problem in a manner quite similar to our treatment for the sequence-of-derivatives problem. What emerges are error bounds for an interpolation formula closely similar to Thiele's reciprocal-difference interpolation formula.[10] We first make the following observation. Let

$$f(z) = \int_0^{1/R} \frac{d\varphi(u)}{1+uz}; \tag{4.1}$$

then, by the change of variables of integration (w real):

$$v = u/(1+uw), \tag{4.2}$$

we have

$$f(z+w) = \int_0^{1/(R+w)} \frac{(1-vw)\,d\varphi[v/(1-vw)]}{1+zv}, \tag{4.3}$$

which is, by its form, a series of Stieltjes in z, provided $w > -R$, with a radius of convergence of at least $R + w$. Consequently, by arguments exactly analogous to those of Sec. 2, if $f(z)$ is a series of Stieltjes and $\{z_j;\ j = 1, 2, \cdots, p\}$ are points in the real line $-R < z < +\infty$, then

$$f(z) = \cfrac{a_0}{1 + \cfrac{(z-z_0)a_1}{1 + \cfrac{(z-z_1)a_2}{1 + \ddots \cfrac{(z-z_{p-1})a_p}{1 + (z-z_p)g_{p+1}(z)}}}} \tag{4.4}$$

defines $g_p(z)$ as a series of Stieltjes, provided the a_p are selected so as to fit $f(z)$ at $z = z_0, z_1, \cdots, z_p$. This is conveniently done through the relations

$$\begin{aligned} g_0(z) &= f(z), \\ g_p(z) &= \frac{g_{p-1}(z_{p-1}) - g_{p-1}(z)}{(z-z_{p-1})g_{p-1}(z)}, \quad p \geq 1, \end{aligned} \tag{4.5}$$

then

$$a_p = g_p(z_p). \tag{4.6}$$

Equation (4.4) may be re-expressed as

$$f(z) = \frac{A_p(z) + (z-z_p)g_{p+1}(z)A_{p-1}(z)C_p}{B_p(z) + (z-z_p)g_{p+1}(z)B_{p-1}(z)C_p} \tag{4.7}$$

where the A's and B's are polynomials in z which result in fitting A_p/B_p through (4.4)–(4.6) to the first 0 to p points, and the C_p are constants. That this form is correct may be seen by setting $g_{p+1}(z) = 0$ and ∞ in (4.4), and noting that we obtain the pth and $(p-1)$th

[10] See, for example, L. M. Milne-Thomson, *The Calculus of Finite Differences* (The Macmillan Company, New York, 1951).

expressions respectively. Again, the function $A_p(z)/B_p(z)$ are rational fractions and the sequence of degrees of the numerator and denominator is the same as in the stair-step sequence [0, 0], [1, 0], [1, 1], [2, 1], [2, 2], $\cdots$. Let us change the normalization of $g_{p+1}(z)$, absorbing this change by changing C_p correspondingly. We normalize $g_{p+1}(z)$ as

$$(R + z_p)g_{p+1}(-R) \leq 1. \tag{4.8}$$

That $g_{p+1}(-R)$ is bounded follows, as it did in Sec. 2 for $h_p(-R)$. If we consider $g_{p+1}(w + z_p) = k_{p+1}(w)$, then we have shown $k_{p+1}(w)$ to be a series of Stieltjes with radius of convergence at least $R + z_p$. Hence the bounds on its range given by the convex hull of (2.20), where $(R + z_p)$ replaces R, are valid. We may evaluate C_p similarly, as before, obtaining

$$f(z) = \frac{B_{p-1}(-R)A_p(z) + (z - z_p)g_{p+1}(z)B_p(-R)A_{p-1}(z)}{B_{p-1}(-R)B_p(z) + (z - z_p)g_{p+1}(z)B_{p-1}(z)B_p(-R)}. \tag{4.9}$$

The range of $f(z)$ at any point in the cut $(-\infty < z \leq -R)$ complex plane is a lens-shaped region which is the map under (4.9) of

$$\text{Range}\{g_{p+1}(z)\} = \underset{0 \leq u \leq 1/(R+z_p)}{\text{convex hull}} \left[\frac{1}{R + z_p}\left(\frac{1 - (R + z_p)u}{1 + u(z - z_p)}\right)\right]. \tag{4.10}$$

This lens-shaped region reduces to an interval for z real. This bound on the value of $f(z)$ is the best possible, because we have shown in Sec. 2 that we can construct a series of Stieltjes of the required type to take on any value in the range given by (4.10).

When the set $\{z_j\}$ processes a limit point in the real line $-R < z < +\infty$, then the necessary and sufficient conditions that a set of values $\{f(z_j)\}$ are values of a series of Stieltjes are that

$$f(z_{p+1}) \in F_p(z_{p+1}), \quad p = 0, 1, \cdots, \tag{4.11}$$

where $F_p(z)$ is the allowable range of $f(z)$ defined by (4.9) and (4.10), using the points $z_0, z_1, \cdots, z_p$. To see that this result is so we note that by the above derivation, if $f(z)$ is a series of Stieltjes, then we have already proved that (4.11) must hold. On the other hand, if (4.11) holds we may construct a sequence of series of Stieltjes $[S_p(z) = A_p(z)/B_p(z)]$ which agree with $f(z)$ at the points $(z_0, z_1, \cdots, z_p)$. Since by arguments analogous to those of Sec. 2 we must have

$$f(z) \in F_p(z) \subset F_{p-1}(z) \subset \cdots \subset F_1(z) \subset F_0(z); \tag{4.12}$$

as $F_0(z)$ is uniformly bounded over any closed domain interior to the cut $(-\infty < z \leq -R)$ complex plane, so is every $S_p(z)$, uniformly on p. We may therefore apply the Osgood–Vitali convergence theorem[11] to the sequence of analytic functions $S_p(z)$. It implies that the sequence $S_p(z)$ constructed tends uniformly to an analytic function in any closed domain interior to the cut complex plane. That this function is a series of Stieltjes can be seen by first noting that all the derivatives at a point (e.g., the origin) also converge by Weirstrass's theorem.[12] Hence the determinantal conditions

$$D_p(m, n) \geq 0 \tag{4.13}$$

satisfied by the $S_p(z)$ also converge and hold in the limit as p tends to infinity. Thus the limiting function satisfies (4.13) as well and hence is a series of Stieltjes,[2] as was to be shown.

We remark that, in spite of the seemingly unsymmetric method of construction of the function $[A_p(x)/B_p(x)]$, it does not, in fact, depend on the order, but only on which points are chosen at which to fit $f(z)$. This result is demonstrated by the solution to the fitting problem given by Muir,[13] where he shows that the solution is always expressible in terms of the determinants

$$V_q = \det \begin{vmatrix} 1 & z_0 & z_0^2 & \cdots & z_0^{p-1} & z_0^q f(z_0) \\ 1 & z_1 & z_1^2 & \cdots & z_1^{p-1} & z_1^q f(z_1) \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 1 & z_p & z_p^2 & \cdots & z_p^{p-1} & z_1^q f(z_1) \end{vmatrix} \tag{4.14}$$

and the van der Monde determinant, which is manifestly symmetric in the fitted points.[10]

We remark that it is clearly possible to combine the results of Sec. 2 and this section, so that a combination of values and successive derivatives at various points can be treated.

Note Added in Proof: We wish to thank Dr. J. Holderman for drawing to our attention the work by Henrici *et al.*[14,15] They have obtained a special case of our results for problem (2.2) ($R = 0$). They also show that the regions $F_p(z)$ are convex. This result is also valid for our problem and has, as they point out, the important consequence that the diameter of the allowed region is bounded by a function $\kappa(z)$ times the difference of the two Padé approximants.

[11] See, for example, H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, London, 1950), Sec. 11.21.

[12] See, for example, Ref. 9, Sec. 5.2.

[13] T. Muir, *A Treatise on the Theory of Determinants* (revised by W. H. Metzler) (Dover Publications, Inc., New York, 1960), Sec. 454.

[14] A. Pfluger and P. Henrici, in *International Colloquium: Theory of Analytic Functions*, M. A. Lavrentév, Ed. (Erivan, Russia, 1965), p. 257.

[15] J. Gargantini and P. Henrici, Math. Computation **21**, 18 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Criteria for Slater Determinants and Quasiparticle Vacuum States

A. NAVON*
Institute of Theoretical Physics, McGill University, Montreal, Canada

(Received 13 December 1968)

New criteria characterizing Slater determinants and quasiparticle vacuum states are obtained. These criteria are expressed as quadratic homogeneous equations in the coefficients of the development of the trial wavefunction in a basis of Slater determinants. A redefinition of quasiparticle vacuum states permits the introduction of quasiparticle transformations which are more general than the generalized Bogoliubov transformations.

1. INTRODUCTION

In Hartree–Fock and Hartree–Bogoliubov methods applied to nuclear calculations, the expectation value $\langle\Phi| H |\Phi\rangle/\langle\Phi \mid \Phi\rangle$ of the Hamiltonian is expressed in terms of certain variational parameters which specify indirectly the trial wavefunction $|\Phi\rangle$. The stationarity condition imposed on the expectation value is then supplemented by criteria imposed on the variational parameters which ensure that $|\Phi\rangle$ is a Slater determinant (SD) or a quasiparticle vacuum state (QPVS).[1,2]

Two sets of variational parameters and corresponding criteria were used until now in the literature. The first set of parameters is formed by the coefficients in the development of single-particle (quasiparticle) operators b_i^+, b_i specifying $|\Phi\rangle$, in terms of fixed single-particle operators a_i^+, a_i. The corresponding criterion for $|\Phi\rangle$ being an SD (QPVS) is expressed by orthonormality (pseudo-orthonormality) conditions imposed on these coefficients. The second set of parameters is formed by the elements $\rho_{ij} \equiv \langle\Phi| a_j^+ a_i |\Phi\rangle$ of the one-particle density matrix and $\chi_{ij} \equiv \langle\Phi| a_i a_j |\Phi\rangle$ of the pairing tensor. For this set, the criterion ensuring that $|\Phi\rangle$ is an SD is expressed by the matrix equation $\rho^2 = \rho$, and the criterion ensuring that $|\Phi\rangle$ is a QPVS is expressed by the two matrix equations $\rho^2 - \chi\chi^* = \rho$ and $\rho\chi = \chi\rho^*$. The relations between the two sets of variational parameters and between the corresponding criteria were discussed by several authors.[1–3]

In this paper, we develop new criteria for characterizing a Slater determinant or a quasiparticle vacuum state in terms of a third set of possible variational parameters. This set consists of the coefficients in the development of the trial wavefunction in a basis of Slater determinants. The Slater determinants are constructed from a finite number of given single-fermion states. The new criteria are expressed as sets of quadratic homogeneous equations in the coefficients of the trial wavefunction. To derive the criterion characterizing a QPVS we need to generalize the usual definitions of a QPVS. This is done by introducing quasiparticle transformations which are more general than the generalized Bogoliubov transformations. The new transformations conserve fermion anticommutation relations, but may violate the relations of adjointness between creation and annihilation operators. The new criteria may prove useful in problems where the explicit expression of the trial wavefunction is needed, instead of intermediate quantities such as the density matrix and the pairing tensor. This is the case, for instance, in some problems of neutron–proton correlations.[4,5]

2. CRITERION FOR SLATER DETERMINANTS

In this section, we derive the criterion characterizing a Slater determinant. We consider a finite number of given single-fermion states

$$a_1^+ |0\rangle, a_2^+ |0\rangle, \cdots, a_n^+ |0\rangle. \tag{2.1}$$

An arbitrary N-particle wavefunction $|\Phi\rangle$ may be developed in the basis of Slater determinants constructed from the states (2.1) as follows:

$$|\Phi\rangle = \sum_{i_1<i_2<\cdots<i_N}^{n} C_{i_1 i_2 \cdots i_N} a_{i_1}^+ a_{i_2}^+ \cdots a_{i_N}^+ |0\rangle. \tag{2.2}$$

We prove in the following that the criterion ensuring that $|\Phi\rangle$ is a SD consists in the set of equations

$$C_{[i_1 i_2 \ldots i_N} C_{k]k_1 \ldots k_{N-1}} = 0, \tag{2.3}$$

where the coefficients $C_{i_1 \ldots i_N}$ are antisymmetric functions of their indices, and the product of two such coefficients is to be antisymmetrized furthermore with respect to the indices enclosed in square brackets; in the case $n = 4$, $N = 2$, for example, the set (2.3)

* Postdoctoral Fellow of the National Research Council of Canada.

[1] M. Baranger, *1962 Cargese Lectures in Theoretical Physics* (W. A. Benjamin, Inc., New York, 1963).

[2] M. Baranger, Phys. Rev. **122**, 992 (1961).

[3] W. H. Young, Phys. Rev. **131**, 476 (1963).

[4] J. Flores and P. A. Mello, Nucl. Phys. **88**, 609 (1966).

[5] M. Ichimura, Progr. Theoret. Phys. (Kyoto) **31**, 575 (1964).

reduces to the single equation

$$C_{[12}C_{3]4} \equiv C_{12}C_{34} - C_{13}C_{24} + C_{14}C_{23} = 0. \quad (2.4)$$

We first show that the criterion (2.3) is a necessary condition. If $|\Phi\rangle$ is an SD, then it can be written

$$|\Phi\rangle = \prod_{i=1}^{N} b_i^+ |0\rangle, \quad (2.5)$$

where

$$b_i^+ = \sum_{j=1}^{n} U_{ij} a_j^+, \quad (2.6)$$

U being some unitary transformation. Inserting Eqs. (2.6) into Eq. (2.5), $|\Phi\rangle$ takes the general form (2.2), where the coefficients $C_{i_1 \cdots i_N}$ are given as the determinants:

$$C_{i_1 \cdots i_N} = \begin{vmatrix} U_{1i_1} \cdots U_{1i_N} \\ \cdots \\ U_{Ni_1} \cdots U_{Ni_N} \end{vmatrix}. \quad (2.7)$$

We introduce now the creation operators

$$d^+(i_1 \cdots i_{s-1} i_{s+1} \cdots i_N) = \sum_{k=1}^{n} C_{i_1 \cdots i_{s-1} k i_{s+1} \cdots i_N} a_k^+. \quad (2.8)$$

We note that, for each choice $(i_1 \cdots i_N)$ of N indices, it is possible to construct N operators d^+. However, these N operators are not necessarily linearly independent. By replacing the coefficients $C_{i_1 \cdots i_N}$ in Eqs. (2.8) by their expressions (2.7) in the form of determinants and by expanding the determinants with respect to the k column, we see that the operators d^+ have the form

$$d^+(i_1 \cdots i_{s-1} i_{s+1} \cdots i_N) = x_1 b_1^+ + x_2 b_2^+ + \cdots + x_N b_N^+. \quad (2.9)$$

Then, from Eqs. (2.5) and (2.9), it follows that

$$d^+(i_1 \cdots i_{s-1} i_{s+1} \cdots i_N) |\Phi\rangle = 0. \quad (2.10)$$

We obtain without difficulty the quadratic equations (2.3) from the relations (2.10), by replacing the operators d^+ and the wavefunction $|\Phi\rangle$, respectively, by their expressions (2.8) and (2.2) in terms of the coefficients $C_{i_1 \cdots i_N}$.

To show that criterion (2.3) is also a sufficient condition, let us suppose that $|\Phi\rangle$ has the general form (2.2) with $C_{12 \cdots N} \neq 0$. If $|\Phi\rangle$ satisfies Eqs. (2.3), then it satisfies Eqs. (2.10), the two sets of equations being equivalent. Let us consider now the N operators

$$d_i^+ = \sum_{k=1}^{n} C_{1 \cdots i-1 k i+1 \cdots N} a_k^+, \quad i = 1, \cdots, N. \quad (2.11)$$

The $N \times N$ determinant formed by the components of the operators $d_1^+, \cdots, d_N^+$ on the operators $a_1^+, \cdots, a_N^+$, is equal to $(C_{1 \cdots N})^N \neq 0$. The N operators (2.11) are therefore linearly independent and the wavefunction $|\Phi\rangle$ may be written, up to a multiplicative constant, as a Slater determinant

$$|\Phi\rangle = \prod_{i=1}^{N} d_i^+ |0\rangle. \quad (2.12)$$

3. QUASIPARTICLE TRANSFORMATIONS

In this section, we introduce quasiparticle transformations more general than the usual generalized Bogoliubov transformations.[1,2] We shall need these new transformations for the derivation of the criterion characterizing a QPVS. Let us consider the creation and annihilation operators

$$a_1^+, \cdots, a_n^+, \quad a_1, \cdots, a_n \quad (3.1)$$

corresponding to the given single-fermion states (2.1). These $2n$ operators satisfy the anticommutation relations

$$[a_j^+, a_k^+]_+ = [a_j, a_k]_+ = 0, \quad [a_j^+, a_k]_+ = \delta_{jk} \quad (3.2)$$

and the relations of adjointness

$$a_j^+ = (a_j)^+, \quad a_j = (a_j^+)^+. \quad (3.3)$$

It will be convenient to introduce a $2n$-dimensional vector space V formed by all complex linear combinations of the operators (3.1). These operators form a possible basis for the space V, and the anticommutation relations (3.2) can be interpreted as defining a bilinear symmetric form on this basis.[6] The $2n$ operators

$$h_j = (2)^{-\frac{1}{2}}(a_j^+ + a_j), \quad h_{n+j} = i(2)^{-\frac{1}{2}}(a_j^+ - a_j), \quad j = 1, \cdots, n, \quad (3.4)$$

whose anticommutation relations read

$$[h_j, h_k]_+ = \delta_{jk}, \quad (3.5)$$

form an orthonormal basis of V. The relations (3.3) become, in the basis h_j,

$$h_j^+ = h_j. \quad (3.6)$$

We are now interested in finding the explicit form of transformations T which conserve only fermion anticommutation relations, and, also, of generalized Bogoliubov transformations T_B which conserve both of the relations (3.2) and (3.3). Notice first that the transformations T or T_B act in two spaces: (i) in the operator space V following the rule $a_j \rightarrow Ta_jT^{-1}$, and (ii) in the space of wavefunctions, which we shall denote by W, following the rule $|\Phi\rangle \rightarrow T|\Phi\rangle$. The group G formed by the transformations T leaves

[6] A. K. Bose and A. Navon, Phys. Letters 17, 112 (1965).

invariant the bilinear symmetric form of the complex space V; it must, therefore, be isomorphic to the group $O(2n, C)$ of complex orthogonal matrices. On the other hand, it was shown already[6,7] that the group of generalized Bogoliubov transformations which we denote by G_B is isomorphic to the group $O(2n, R)$ of real orthogonal matrices.

We determine first the form of the transformations in the connected subgroups G' and G'_B of G and G_B, isomorphic, respectively, to the proper orthogonal subgroups $SO(2n, C)$ and $SO(2n, R)$. It is known[8] that any transformation of a connected group can be written as one term or as a product of terms of the form $\exp t$, where t is a Lie algebra element. A Lie algebra element t will act in the space V following the rule

$$a_j \to [t, a_j]_- . \tag{3.7}$$

Let us consider the $n(2n-1)$ elements $h_j h_k$ $(j < k)$. These elements are linearly independent, generate by commutation a closed Lie algebra, and map any element of V, according to the rule (3.7), into another element of V. An element t of the Lie algebra of G' must be antisymmetric with respect to transposition, while an element t_B of the Lie algebra of G'_B must be both antisymmetric and real. The operation of transposition is to be defined for elements expressed in second quantization with respect to the bilinear form represented in V by the anticommutator. One obtains that the transpose of a product of operators h_j or a_j^+, a_j, is the product of the same operators written in reversed order. Thus one has

$$(h_j h_k)^t = h_k h_j = -h_j h_k, \tag{3.8}$$

and the operators $h_j h_k$ form a basis for the Lie algebras of the groups G' and G'_B. Arbitrary elements t and t_B may now be written in the form

$$\sum_{j<k} m_{jk} h_j h_k, \tag{3.9}$$

where the coefficients m_{jk} are complex for t and real for t_B. In terms of the operators a_j^+, a_j, the elements t and t_B may be expressed as follows:

$$t = \sum_{j<k} (m_{jk} a_j^+ a_k^+ + n_{jk} a_j a_k) + \sum_{j \neq k} p_{jk} a_j^+ a_k + \sum_j r_j (2a_j^+ a_j - 1), \tag{3.10}$$

$$t_B = \sum_{j<k} [m'_{jk}(a_j^+ a_k^+ - a_k a_j) + in'_{jk}(a_j^+ a_k^+ + a_k a_j) + p'_{jk}(a_j^+ a_k - a_k^+ a_j) + iq'_{jk}(a_j^+ a_k + a_k^+ a_j)] + \sum_j ir'_j (2a_j^+ a_j - 1), \tag{3.11}$$

where the primed coefficients are restricted to be real.

From the isomorphism between the groups G, G_B and the groups $O(2n, C)$, $O(2n, R)$, we can assert now that transformations T or T_B are either proper or may be written as products of a chosen reflection by proper transformations. In order to know the general expression of transformations T or T_B, it will be sufficient to determine the form of some reflection. If we put

$$R = (2)^{\frac{1}{2}} h_i = a_i^+ + a_i \tag{3.12}$$

and notice that

$$R^{-1} = R, \tag{3.13}$$

it follows that the basis vectors h_j of V transform like

$$h_i \to R h_i R^{-1} = h_i, \quad h_j \to R h_j R^{-1} = -h_j, \quad j \neq i, \tag{3.14}$$

and R is thus clearly a reflection.

Let us discuss now briefly how the generalized quasiparticle transformations T' of the connected group G' act in the space W of wavefunctions. This space is of dimension 2^n, a general element of W being a mixture of 0-, 1-, $\cdots$, n-particle wavefunctions. W can be decomposed as direct sum of two 2^{n-1}-dimensional subspaces W^e and W^0, containing respectively the wavefunctions with an even number of particles and the wavefunctions with an odd number of particles. The transformations T', which can be written as one term or as a product of terms $\exp t$, are even operators [cf. Eq. (3.10)] and do not change the parity of the number of particles. Thus, T' acts inside each of the subspaces W^e and W^0. Furthermore, it was established in mathematical literature[9] that the two representations of the group G' in the subspaces W^e and W^0 are irreducible and isomorphic to the spinor representations of the group $SO(2n, C)$. This implies that there exists always a transformation T' mapping two given wavefunctions of W^e or of W^0 one into the other.

4. CRITERION FOR QUASIPARTICLE VACUUM STATES

A quasiparticle vacuum state $|\Phi\rangle$ is usually defined[1] in one of the two ways:

(i) there exist n annihilation quasiparticle operators $c_i = T_B a_i T_B^{-1}$, which satisfy the relations

$$c_i |\Phi\rangle = 0; \tag{4.1}$$

(ii) $|\Phi\rangle$ may be expressed in the form

$$|\Phi\rangle = T_B |0\rangle. \tag{4.2}$$

[7] C. Bloch and A. Messiah, Nucl. Phys. **39**, 95 (1962).

[8] R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc., New York, 1966).

[9] C. Chevalley, *The Algebraic Theory of Spinors* (Columbia University Press, New York, 1954).

These definitions can be generalized by replacing the Bogoliubov transformations T_B by the more general transformations T introduced in the previous section. This is possible because the set of quasiparticle operators (c_i) defining $|\Phi\rangle$ in (i) is characterized only by the anticommutation relations

$$[c_i, c_j]_+ = 0 \tag{4.3}$$

and do not imply relations of adjointness. Thus, any QPVS may be written in the form

$$|\Phi\rangle = T|0\rangle, \tag{4.4}$$

from which it is also seen that a QPVS is either odd or even with respect to the parity of the number of particles.

Let us consider now an arbitrary even wavefunction which we denote by $|\Phi\rangle$. Assuming that the number n of given single-particle states is also even, $|\Phi\rangle$ can be developed as follows:

$$|\Phi\rangle = \Big(C_0 + \sum_{i_1<i_2} C_{i_1 i_2} a_{i_1}^+ a_{i_2}^+ + \cdots + C_{12\ldots n} a_1^+ a_2^+ \cdots a_n^+\Big)|0\rangle \equiv S|0\rangle. \tag{4.5}$$

Here S stands for the expression multiplying the particle vacuum. We prove in the following that the criterion for $|\Phi\rangle$ to be a QPVS consists in the set of equations

$$\sum_{l,k}{}' (-1)^{\frac{1}{2}k(k-1)} C_{[i_1\cdots i_l} C_{j_1\cdots j_{k-c}]\cdots j_k} = 0, \tag{4.6}$$

where the prime on the sign of summation means that $l + k$ is to be kept constant. All equations of the set are obtained when c varies from 0 to n and $l + k$ varies from $2c + 2$ to $n + c$. In the case $n = 4$, the set (4.6) reduces to the single equation

$$C_0 C_{1234} - C_{12}C_{34} + C_{13}C_{24} - C_{14}C_{23} = 0, \tag{4.7}$$

considered already in problems of neutron–proton correlations.[4,5]

We prove first that Eqs. (4.6) are a necessary condition. If $|\Phi\rangle$ is a QPVS, then we can write it as

$$|\Phi\rangle \equiv S|0\rangle = T|0\rangle. \tag{4.8}$$

In the case of a finite number n of single-particle states, the product of all annihilation operators $\prod_{i=1}^n a_i$ has exactly the same properties as the particle vacuum $|0\rangle$. Equation (4.8) may thus be rewritten as

$$S\prod_{i=1}^n a_i = T\prod_{i=1}^n a_i. \tag{4.9}$$

Taking the transpose on both sides of Eq. (4.9) following the rule established in Sec. 3, and noting that $a_n \cdots a_2 a_1 = (-1)^{\frac{1}{2}n(n-1)} a_1 a_2 \cdots a_n$ and that $T^t = T^{-1}$, we obtain

$$\prod_{i=1}^n a_i S^t = \prod_{i=1}^n a_i T^{-1}. \tag{4.10}$$

From Eqs. (4.9) and (4.10) it follows now that

$$S\prod_{i=1}^n a_i S^t = T\prod_{i=1}^n a_i T^{-1}. \tag{4.11}$$

We note that A is the operator $S\prod_{j=1}^n a_i S^t$ arranged in normal order, i.e., creation operators to the left of annihilation operators. We remark that A is quadratic in the coefficients C of the wavefunction $|\Phi\rangle$. If $c_i = Ta_iT^{-1}$ are the quasiparticle operators generated by the transformation T from the n operators a_i, the right-hand side of Eq. (4.11) can be written as the product $\prod_{i=1}^n c_i$. This product, expanded as a sum of tensors in a_i^+, a_i, will not contain components of rank higher than n. The same condition will be valid for the operator A and will imply restrictions on the coefficients C. These restrictions, calculated in the Appendix, are found to be identical with Eqs. (4.6).

We assume now the equivalence between Eqs. (4.6) and the condition that the operator A has no components of rank higher than n, and proceed to the proof that criterion (4.6) is also a sufficient condition. Let us consider the operator $A' = S'\prod_{i=1}^n a_i S'^t$, corresponding to the wavefunction $|\Phi\rangle \equiv S'|0\rangle = T'|\Phi\rangle$, obtained by applying a quasiparticle transformation T' to the trial wavefunction $|\Phi\rangle \equiv S|0\rangle$. The operator A' may be written in terms of the operator A, corresponding to $|\Phi\rangle$, as

$$A' = T'AT'^{-1}. \tag{4.12}$$

It results from Eq. (4.12) that, if A has no components of rank higher than n, the same is true for A'. Thus, it will suffice to prove that Eqs. (4.6) are a sufficient condition for a wavefunction $|\Phi'\rangle$ obtained from the trial wavefunction $|\Phi\rangle$ by a quasiparticle transformation.

We shall first choose the wavefunction $|\Phi'\rangle$ and the transformation T' such that the coefficient C_0' of S' is nonzero. This is possible because the space W^e of even wavefunctions is irreducible with respect to the group G of quasiparticle transformations and, therefore, a transformation T' of G transforming a wavefunction with $C_0 = 0$ into one with $C_0' \neq 0$ must exist. We may set in fact $C_0' = 1$. We can assume also that S' has no component of rank two. For, if S' has such a component, e.g., $S_2' = \sum C'_{i_1 i_2} a_{i_1}^+ a_{i_2}^+$, we may consider instead of $|\Phi'\rangle$ the wavefunction

$$|\Phi''\rangle = \exp\Big(-\sum C'_{i_1 i_2} a_{i_1}^+ a_{i_2}^+\Big)|\Phi'\rangle \tag{4.13}$$

with $S_2'' = 0$. With the above choice of wavefunction, which we shall denote from now on simply by $|\Phi\rangle$, assuming that Eqs. (4.6) are satisfied, we shall prove that $|\Phi\rangle = |0\rangle$.

Suppose that S has a component of rank $h \geq 4$:

$$S_h = \sum C_{i_1 \cdots i_h} a_{i_1}^+ \cdots a_{i_h}^+ . \tag{4.14}$$

We may assume that one of the coefficients, e.g., $C_{12\cdots h}$ for the sake of simplicity, is not equal to zero. Defining now the operator

$$\bar{S} = C_{12\cdots h}^{-1} a_{h+1}^+ \cdots a_n^+ , \tag{4.15}$$

we can write the following relations:

$$\bar{S} S_h = \prod_{i=1}^{n} a_i^+ , \tag{4.16}$$

$$\bar{S} S = \bar{S} + \prod_{i=1}^{n} a_i^+ . \tag{4.17}$$

If the operator A corresponding to $|\Phi\rangle$ has no components of rank higher than n, then the operator $a_i^+ A a_i^+$ will not have such components because the operator a_i^+ on the right either contracts by crossing over A or gives zero by multiplying the a_i^+ on the left. From this it follows that the operator

$$\bar{A} = \bar{S} A \bar{S}^t = (\bar{S}S) \prod_{i=1}^{n} a_i (\bar{S}S)^t \tag{4.18}$$

must not have components of rank higher than n. With the help of Eqs. (4.17) and (4.18), $\bar{A}$ can be written

$$\begin{aligned} \bar{A} = {} & (C_{1\cdots h}^{-1} a_{h+1}^+ \cdots a_n^+ + a_1^+ \cdots a_n^+) a_1 \cdots a_n \\ & \times (C_{1\cdots h}^{-1} a_{h+1}^+ \cdots a_n^+ + a_1^+ \cdots a_n^+)^t , \end{aligned} \tag{4.19}$$

in which form it is the sum of four terms. The two terms

$$C_{1\cdots h}^{-2} a_{h+1}^+ \cdots a_n^+ a_1 \cdots a_n (a_{h+1}^+ \cdots a_n^+)^t$$

and

$$a_1^+ \cdots a_n^+ a_1 \cdots a_n (a_1^+ \cdots a_n^+)^t$$

have no normal components of rank higher than n because the wavefunctions $a_{h+1}^+ \cdots a_n^+ |0\rangle$ and

$$a_1^+ \cdots a_n^+ |0\rangle$$

are obviously QPVS. The term

$$C_{1\cdots h}^{-1} a_1^+ \cdots a_n^+ a_1 \cdots a_n a_n^+ \cdots a_{h+1}^+$$

has no normal components of rank smaller than $n + h$. However, the term

$$C_{1\cdots h}^{-1} a_{h+1}^+ \cdots a_n^+ a_1 \cdots a_n a_n^+ \cdots a_1^+$$

has clearly nonzero components of rank $n + h - 2 > n$. Thus, $\bar{A}$ has nonzero components of rank higher than n, which is contradictory. This entails that $|\Phi\rangle = |0\rangle$, which proves also that Eqs. (4.6) are a sufficient condition.

APPENDIX

In this appendix we prove that the condition, that the operator A has no components of rank higher than n, gives Eqs. (4.6).

We evaluate first, by using Wick's theorem, the normal form of the operator

$$a_1 \cdots a_n a_{j_1}^+ \cdots a_{j_k}^+ . \tag{A1}$$

We consider a general term in (A1) obtained by the contraction of c operators $a_{j_{s_1}}^+, \cdots, a_{j_{s_c}}^+$, from the set $a_{j_1}^+, \cdots, a_{j_k}^+$. We denote by $(j_1', \cdots, j_{k-c}')$ and $(j_1'', \cdots, j_{n-c}'')$; respectively, the indices remaining in the sequences $(j)_k \equiv (j_1, \cdots, j_k)$ and $(1, \cdots, n)$ after the elimination of $j_{s_1}, \cdots, j_{s_c}$. To be contracted, the operator $a_{j_{s_1}}^+$ must cross $(s_1 - 1)$ creation and $(n - j_{s_1})$ annihilation operators, which gives a sign $(-1)^{n - j_{s_1} - s_1 - 1}$; the operator $a_{j_{s_2}}^+$ must cross $(s_2 - 2)$ creation and $(n - j_{s_2})$ annihilation operators, which gives a sign $(-1)^{n - j_{s_2} - s_2 - 2}$, etc. The contraction of all c operators introduces the sign $(-1)^{nc - \frac{1}{2} \cdot c(c+1) - \sum j_s + \sum s}$, where $\sum j_s = j_{s_1} + \cdots + j_{s_c}$ and $\sum s = s_1 + \cdots + s_c$. The normal product

$$N(a_{j_1''} \cdots a_{j_{n-c}''} a_{j_1'}^+ \cdots a_{j_{k-c}'}^+)$$

contributes also the sign $(-1)^{(n-c)(k-c)}$. Noting that n and k are even, the sign of the contracted term in (A1) will be $(-1)^{\frac{1}{2}c(c-1) + \sum j_s + \sum s}$. The normal form of the operator $a_1 \cdots a_n a_{j_1}^+ \cdots a_{j_k}^+$ is therefore

$$\sum_{0}^{k} \sum_{c} \sum_{(j_s)_c} (-1)^{\frac{1}{2}c(c-1) + \sum j_s + \sum s} a_{j_1'}^+ \cdots a_{j_{k-c}'}^+ a_{j_1''} \cdots a_{j_{n-c}''} , \tag{A2}$$

where $(j_s)_c$ denotes a sequence of c indices $j_{s_1}, \cdots, j_{s_c}$, chosen from the sequence $(j)_k$.

The operator A, whose detailed form is

$$\begin{aligned} A = {} & \left(C_0 + \cdots + \sum_{i_1 < \cdots < i_l} C_{i_1 \cdots i_l} a_{i_1}^+ \cdots a_{i_l}^+ + \cdots \right) \\ & \times a_1 \cdots a_n \left(C_0 + \cdots + (-1)^{\frac{1}{2}k(k-1)} \right. \\ & \left. \times \sum_{j_1 < \cdots < j_k} C_{j_1 \cdots j_k} a_{j_1}^+ \cdots a_{j_k}^+ + \cdots \right) , \end{aligned} \tag{A3}$$

can now be written with the help of expression (A2) as

$$\begin{aligned} A = {} & \sum_{l,k}^{n} \sum_{(i)_l} \sum_{(j)_k} \sum_{c}^{k} \sum_{(j_s)_c} (-1)^{\frac{1}{2}k(k-1)} (-1)^{\frac{1}{2}c(c-1) + \sum j_s + \sum s} \\ & \times C_{i_1 \cdots i_l} C_{j_1 \cdots j_k} a_{i_1}^+ \cdots a_{i_l}^+ a_{j_1'}^+ \cdots a_{j_{k-c}'}^+ a_{j_1''} \cdots a_{j_{n-c}''} . \end{aligned} \tag{A4}$$

We must equate to zero the coefficients of the products $a_{i_1}^+ \cdots a_{i_l}^+ a_{j_1'}^+ \cdots a_{j_{k-c}'}^+ a_{j_1''} \cdots a_{j_{n-c}''}$ whose degrees d are higher than n. One has that $d = l + k + n - 2c$, and then $d > n$ implies $l + k > 2c$; on the other

hand, the number $l + k - c$ of creation operators in a product cannot exceed n, which implies that $l + k \leq n + c$. Hence, the restriction

$$2c < l + k \leq n + c. \tag{A5}$$

It can be easily verified that the coefficient of a given product of degree d may be written

$$(-1)^{\frac{1}{2}c(c-1)+\Sigma j_s+\Sigma s} \sum_{l,k}{}' (-1)^{\frac{1}{2}k(k-1)} \times C_{[j_1 \cdots i_l} C_{j_1 \cdots j_{k-c}]j_{k-c+1} \cdots j_k}, \tag{A6}$$

where the prime on the summation sign means that $l + k$ is constant and equal to $d + 2c - n$, and the square bracket notation was already defined in Eq. (2.3). The set of equations obtained by equating to zero the coefficients of products with $d > n$ is finally

$$\sum_{l,k}{}' (-1)^{\frac{1}{2}k(k-1)} C_{[i_1 \cdots i_l} C_{j_1 \cdots j_{k-c}]j_{k-c+1} \cdots j_k} = 0, \tag{A7}$$

where $l + k$ is restricted by (A5) and c may take all values 0 to n.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Macroscopic Causality Conditions and Properties of Scattering Amplitudes*

COLSTON CHANDLER† AND HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 27 August 1967)

Two causality conditions that refer only to mass-shell quantities are formulated and their consequences explored. The first condition, called weak asymptotic causality, expresses the requirement that some interaction between the initial particles must occur before the last interaction from which final particles emerge. This condition is shown to imply that if a two-body scattering function is analytic except for singularities in the energy variable at normal thresholds, then (a) the physical scattering functions in two adjacent parts of the physical region separated by any normal threshold are parts of a single analytic function; (b) the path of continuation joining these two parts bypasses the singularity in the upper half-plane of the energy variable; and (c) the integral over the physical function can be represented as an integral over a contour that is distorted into the upper-half energy plane (hence not, for example, by a principal-value integral). Singularities possessing finite derivatives of all orders with respect to real variations of the energy are not encompassed by this result. The second causality condition, called strong asymptotic causality, expresses the requirement that, apart from contributions whose effects fall off faster than any inverse power of Euclidean distance, momentum-energy is carried over macroscopic distances only by stable physical particles. This condition implies that all n-particle scattering functions ($n \geq 4$) are analytic, apart from infinitely differentiable singularities, at physical points not lying on any positive-α Landau surface. Moreover, the scattering functions on the two sides of any such Landau surface are analytically connected by a path that passes around the singularity surface in a well defined manner, which is the same as in perturbation theory. Thus, apart from possible infinitely differentiable singularities, the physical region singularity structure is derived from a mass-shell causality requirement. Several properties of the set $\mathfrak{L}^+$ of physical region positive-α Landau surfaces are derived.

1. INTRODUCTION

By a causality condition we mean a requirement that events identified as effects occur later than events identified as their causes. Such conditions have led to important properties of the basic functions of classical electrodynamics,[1] nonrelativistic quantum mechanics,[2] and quantum field theory.[3] The aim of the present work is to formulate causality conditions within a mass-shell S-matrix theory and to derive from them certain properties of the physical-region scattering functions.[4,5]

The procedure is as follows. The momentum space wavefunctions representing the initial and final particles of a scattering experiment are chosen to be Schwartz test functions, and the scattering functions are shown to be Schwartz distributions. The mass-shell

* This work was done under the auspices of the United States Atomic Energy Commission.

† Present Address: Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico.

[1] H. A. Kramers, Atti Congr. Intern. Fisici **2**, 545 (1927); R. Kronig, J. Opt. Soc. Am. **12**, 547 (1926).

[2] N. G. van Kampen, Phys. Rev. **89**, 1072 (1953) and **91**, 1267 (1953); J. S. Toll, *ibid.* **104**, 1760 (1956).

[3] E. C. G. Stueckelberg and D. Rivier, Helv. Phys. Acta **23**, 215 (1950); M. Fierz, *ibid.* **23**, 731 (1950); M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. **95**, 1612 (1955).

[4] Causality requirements in S-matrix theory have been discussed also by the following authors: G. Wanders, Nuovo Cimento **14**, 168 (1959) and Helv. Phys. Acta **38**, 142 (1965); D. Branson, Phys. Rev. **135**, B1255 (1964); R. J. Eden and P. V. Landschoff, Ann. Phys. **31**, 370 (1965); A. Peres, *ibid.* **37**, 179 (1966); D. Iagolnitzer, J. Math. Phys. **6**, 1576 (1965); F. Pham, Ann. Inst. Henri Poincaré **6**, 89 (1967); R. Omnes, Phys. Rev. **146**, 1123 (1966).

[5] H. P. Stapp, Phys. Rev. **139**, B257 (1965).

hand, the number $l + k - c$ of creation operators in a product cannot exceed n, which implies that $l + k \leq n + c$. Hence, the restriction

$$2c < l + k \leq n + c. \tag{A5}$$

It can be easily verified that the coefficient of a given product of degree d may be written

$$(-1)^{\frac{1}{2}c(c-1)+\Sigma j_s+\Sigma s} \sum_{l,k}{}' (-1)^{\frac{1}{2}k(k-1)} \times C_{[j_1 \cdots i_l} C_{j_1 \cdots j_{k-c}]j_{k-c+1} \cdots j_k}, \tag{A6}$$

where the prime on the summation sign means that $l + k$ is constant and equal to $d + 2c - n$, and the square bracket notation was already defined in Eq. (2.3). The set of equations obtained by equating to zero the coefficients of products with $d > n$ is finally

$$\sum_{l,k}{}' (-1)^{\frac{1}{2}k(k-1)} C_{[i_1 \cdots i_l} C_{j_1 \cdots j_{k-c}]j_{k-c+1} \cdots j_k} = 0, \tag{A7}$$

where $l + k$ is restricted by (A5) and c may take all values 0 to n.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Macroscopic Causality Conditions and Properties of Scattering Amplitudes*

COLSTON CHANDLER† AND HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 27 August 1967)

Two causality conditions that refer only to mass-shell quantities are formulated and their consequences explored. The first condition, called weak asymptotic causality, expresses the requirement that some interaction between the initial particles must occur before the last interaction from which final particles emerge. This condition is shown to imply that if a two-body scattering function is analytic except for singularities in the energy variable at normal thresholds, then (a) the physical scattering functions in two adjacent parts of the physical region separated by any normal threshold are parts of a single analytic function; (b) the path of continuation joining these two parts bypasses the singularity in the upper half-plane of the energy variable; and (c) the integral over the physical function can be represented as an integral over a contour that is distorted into the upper-half energy plane (hence not, for example, by a principal-value integral). Singularities possessing finite derivatives of all orders with respect to real variations of the energy are not encompassed by this result. The second causality condition, called strong asymptotic causality, expresses the requirement that, apart from contributions whose effects fall off faster than any inverse power of Euclidean distance, momentum-energy is carried over macroscopic distances only by stable physical particles. This condition implies that all n-particle scattering functions ($n \geq 4$) are analytic, apart from infinitely differentiable singularities, at physical points not lying on any positive-α Landau surface. Moreover, the scattering functions on the two sides of any such Landau surface are analytically connected by a path that passes around the singularity surface in a well defined manner, which is the same as in perturbation theory. Thus, apart from possible infinitely differentiable singularities, the physical region singularity structure is derived from a mass-shell causality requirement. Several properties of the set $\mathfrak{L}^+$ of physical region positive-α Landau surfaces are derived.

1. INTRODUCTION

By a causality condition we mean a requirement that events identified as effects occur later than events identified as their causes. Such conditions have led to important properties of the basic functions of classical electrodynamics,[1] nonrelativistic quantum mechanics,[2] and quantum field theory.[3] The aim of the present work is to formulate causality conditions within a mass-shell S-matrix theory and to derive from them certain properties of the physical-region scattering functions.[4,5]

The procedure is as follows. The momentum space wavefunctions representing the initial and final particles of a scattering experiment are chosen to be Schwartz test functions, and the scattering functions are shown to be Schwartz distributions. The mass-shell

* This work was done under the auspices of the United States Atomic Energy Commission.

† Present Address: Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico.

[1] H. A. Kramers, Atti Congr. Intern. Fisici **2**, 545 (1927); R. Kronig, J. Opt. Soc. Am. **12**, 547 (1926).

[2] N. G. van Kampen, Phys. Rev. **89**, 1072 (1953) and **91**, 1267 (1953); J. S. Toll, *ibid.* **104**, 1760 (1956).

[3] E. C. G. Stueckelberg and D. Rivier, Helv. Phys. Acta **23**, 215 (1950); M. Fierz, *ibid.* **23**, 731 (1950); M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. **95**, 1612 (1955).

[4] Causality requirements in S-matrix theory have been discussed also by the following authors: G. Wanders, Nuovo Cimento **14**, 168 (1959) and Helv. Phys. Acta **38**, 142 (1965); D. Branson, Phys. Rev. **135**, B1255 (1964); R. J. Eden and P. V. Landschoff, Ann. Phys. **31**, 370 (1965); A. Peres, *ibid.* **37**, 179 (1966); D. Iagolnitzer, J. Math. Phys. **6**, 1576 (1965); F. Pham, Ann. Inst. Henri Poincaré **6**, 89 (1967); R. Omnes, Phys. Rev. **146**, 1123 (1966).

[5] H. P. Stapp, Phys. Rev. **139**, B257 (1965).

constraints on these wavefunctions imply that the space–time wavefunctions defined by Fourier transformation are solutions of the free-particle Klein–Gordon equation. Consequently the regions over which these space–time functions are nonzero cannot be bounded; these wavefunctions have appreciable values on cones, called velocity cones, running from the infinite past to the infinite future. It is argued in Sec. 2 that these velocity cones can be interpreted as the trajectory regions of the corresponding particles in the sense that the transition amplitude of a reaction will be small unless the velocity cones of appropriate particles intersect. These intersections are interpreted as the locations of the possible particle collisions. It is their space–time ordering that is restricted by the causality conditions.

The space–time wavefunctions are not strictly confined to their velocity cones, but have "tails" that extend over all space–time. This means that the locations of collisions are not sharply defined. This presents a difficulty that must be surmounted.

In Sec. 3 a condition called weak asymptotic causality (WAC) is formulated. This condition expresses the general idea that if a time t can be found such that none of the collisions between initial particles occur at times earlier than t, and none of the collisions from which final particles emerge occur at times later than t, then the corresponding transition amplitude should be small. In other words, the first collision between initial particles should occur no later than the last collision that produces final particles. The WAC condition is formulated so that it refers only to the asymptotic regions long before or long after the relevant collisions take place. Indeed, it is only in these regions that the free-particle wave functions should have physical significance. From the WAC condition we derive the $i\epsilon$ rule for continuation past any physical-region Landau singularity surface of the two-body scattering functions. This weak condition is not strong enough, however, to give the rule for continuation past an arbitrary Landau singularity surface of a general n-particle scattering function.

In Sec. 4 a stronger condition, called strong asymptotic causality (SAC), is formulated. It embodies the idea that energy–momentum is carried over macroscopic distances only by physical particles. More precisely, the probabilities of interactions having energy–momentum transfers that cannot be attributed to physical particles are required to fall off faster than any inverse power of the Euclidean distance, as the distances involved become infinite. The SAC condition is shown to imply that the scattering functions are infinitely differentiable at all physical points not lying on a positive-α Landau surface.

Points that do lie on some positive-α surface are classified as Type I points or Type II points. Points which lie on only one positive-α Landau surface are included among the Type I points. The only known examples of Type II points are points at which two initial or two final particle energy–momentum vectors are collinear. The SAC condition is shown to imply that in a neighborhood of a Type I point $\bar{K}$ a scattering function can be represented as a sum of a finite number of terms of which the first is infinitely differentiable, while the others are boundary values of holomorphic functions. Furthermore, these boundary values are themselves infinitely differentiable except on the relevant Landau surfaces. If $\bar{K}$ belongs to only one positive-α Landau surface, then there is only one of these boundary-value terms, and the $i\epsilon$ prescription that defines the boundary value agrees with that of perturbation theory. Similar results are derived for Type I points at which several positive-α surfaces intersect. No results are obtained for Type II points.

The results described above are useful in the following way. In analytic S-matrix theory, it is assumed that the only singularities of the scattering functions are those that arise from the unitarity equations. But even granting that the positions of the singularities are known, there is the question of how to continue around them. There is even the prior question of whether the physical scattering functions on the two sides of a singularity passing through the physical region are analytically connected at all. That these two functions can differ is a real possibility. For example, the K matrix, which also has singularities on the Landau surfaces, is not represented in sectors separated by these surfaces by the same analytic function. This property is a special feature of the scattering matrix. It has usually been assumed that one could accept the results of perturbation theory on this point and take the scattering function in the various sectors to be parts of a single analytic function, with the rule for continuation around singularities the same as in perturbation theory. The present work provides a physical basis for these assumptions. Infinitely differentiable singularities are not encompassed. Since, however, the singularities generated by the unitarity equations are apparently never infinitely differentiable, this omission is of no practical significance in this context.

As a by-product we obtain a number of useful results concerning the nature of the set $\mathfrak{L}^+$ of physical points lying on positive-α Landau surfaces. Let $\mathcal{M}$ be

the mass shell. This consists of points in energy–momentum space that satisfy the mass constraints and the conservation laws. Let $\mathcal{M}_0$ be the subset of $\mathcal{M}$ where two (or more) initial or two (or more) final energy–momentum vectors are collinear. Let $\mathfrak{L}^+[\mathfrak{D}]$ be the Landau surface in $\mathcal{M}$ associated with the Landau diagram $\mathfrak{D}$, and let $\mathfrak{L}_0^+[\mathfrak{D}]$ be the subset of $\mathfrak{L}^+[\mathfrak{D}]$ that excludes points lying on the $\mathfrak{L}^+[\mathfrak{D}']$ of any contraction $\mathfrak{D}'$ of $\mathfrak{D}$. Then $\mathfrak{L}^+$ is the union of points lying on the various $\mathfrak{L}_0^+[\mathfrak{D}]$. Each point $K \notin \mathcal{M}_0$ of $\mathfrak{L}_0^+[\mathfrak{D}]$ is shown to correspond to a unique (apart from scaling) point in the space of Feynman α's. Each surface $\mathfrak{L}_0^+[\mathfrak{D}]$ is shown to be an analytic submanifold of $\mathcal{M} - \mathcal{M}_0$ of codimension 1. It is shown that the $i\epsilon$ prescriptions associated with a set of intersecting Landau surfaces $\mathfrak{L}_0^+[\mathfrak{D}_i]$ associated with a set of $\mathfrak{D}_i$ that are all contractions of some single $\mathfrak{D}$ are necessarily compatible.

2. BASIC FORMALISM

A. Transition Amplitudes

The basic observables in scattering experiments can be considered to be the scattering amplitudes for transitions from initial systems of freely moving particles to final systems of freely moving particles. The general mathematical form of these transition amplitudes is dictated in the following way by physical requirements.

Consider an arbitrary reaction involving a total of n initial and final particles. Let the particles be labeled by an index i, $1 \leq i \leq n$. Each particle is represented by a complex-valued momentum space wavefunction ψ_i which, because the particles are freely moving, is a mapping $\psi_i: \mathcal{M}_i \to \mathbf{C}$ from the real manifold

$$\mathcal{M}_i = \{k_i \mid k_i^2 \equiv (k_{i0})^2 - \mathbf{k}_i^2 = \mu_i^2,\ \sigma_i k_{i0} > 0\} \quad (2.1)$$

into the space $\mathbf{C}$ of complex numbers. The vector k_i is the mathematical energy–momentum of the ith particle and is defined by $k_i = \sigma_i p_i$, where p_i is the physical energy–momentum of the particle, and

$$\sigma_i = \begin{cases} +1, & \text{for final particles,} \\ -1, & \text{for initial particles.} \end{cases} \quad (2.2)$$

The mass μ_i of each particle is assumed to be nonzero. Other quantum numbers such as spin, isospin, charge, etc., are unimportant in this discussion and are not indicated explicitly. The functions ψ_i can, for the present purposes, be assumed to belong to the spaces $\mathfrak{D}(\mathcal{M}_i)$ of functions that have compact support $\operatorname{supp} \psi_i \subset \mathcal{M}_i$ and continuous partial derivatives of all orders in $\mathcal{M}_i$.

The transition from the initial system of particles to the final system is represented by a functional $S[\psi_1, \cdots, \psi_n]$ which, when all of the wavefunctions ψ_i have unit norm

$$\|\psi_i\| = \left\{(2\pi)^{-3}\int d^4k\theta(\sigma_i k_0)\delta(k^2 - \mu_i^2)|\psi_i(k)|^2\right\}^{\frac{1}{2}}, \quad (2.3)$$

is a probability amplitude. The functional S is assumed to be linear in the wavefunctions of the initial particles and antilinear in the wavefunctions of the final particles. This linearity, together with the probability interpretation of S, implies the inequality

$$|S[\psi_1, \cdots, \psi_n]| \leq \prod_i \|\psi_i\|. \quad (2.4)$$

This inequality in turn implies the continuity of S in each variable ψ_i in the topology induced by the norm (2.3),[6] and hence also in the topology of $\mathfrak{D}(\mathcal{M}_i)$.[7] The functional S can, therefore, by virtue of the nuclear theorem,[8] be written $S[\psi_1, \cdots, \psi_n] = S[\psi]$, where ψ is the product wavefunction

$$\psi(k_1, \cdots, k_n) = \prod_{\text{initial}} \psi_i(k_i) \prod_{\text{final}} \psi_i^*(k_i), \quad (2.5)$$

and $S[\psi]$ is a continuous linear functional (Schwartz distribution) on the space $\mathfrak{D}(\otimes\mathcal{M}_i)$ of functions with compact support and continuous partial derivatives of all orders in the product space

$$\otimes\mathcal{M}_i = \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \cdots \otimes \mathcal{M}_n.$$

Conservation of energy and momentum requires S to be concentrated on the set

$$\mathcal{M} = \{K \mid K = (k_1, \cdots, k_n) \in \otimes\mathcal{M}_i,\ \Sigma k_i = 0\}. \quad (2.6)$$

The restricted real mass-shell $\mathfrak{W}$ is the subset of all points K of $\mathcal{M}$ at which at least two of the vectors k_i are linearly independent. The restriction of S to the set

$$\mathfrak{B}(\mathfrak{W}) = \{\psi \mid \psi \in \mathfrak{D}(\otimes\mathcal{M}_i),\ (\mathcal{M} \cap \operatorname{supp} \psi) \subset \mathfrak{W}\} \quad (2.7)$$

then has the representation[9]

$$S[\psi] = \int dK \psi(K) S(K), \quad (2.8)$$

where $S(K)$ is a Schwartz distribution and

$$dK = (2\pi)^{4-3n}\delta(\Sigma k_i)\Pi\, d^4k_i\delta(k_i^2 - \mu_i^2)\theta(\sigma_i k_{i0}) \quad (2.9)$$

is the (Lorentz invariant) volume element of $\mathfrak{W}$.

[6] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, New York, 1958), Vol. 1, p. 59.

[7] H. Bremermann, *Distributions, Complex Variables, and Fourier Transforms* (Addison-Wesley Publ. Co., Reading, Mass., 1965), p. 39; or R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin Co., Inc., 1964), Chap. 2.

[8] R. F. Streater and A. S. Wightman, Ref. 7, p. 43.

[9] D. Williams, J. Math. Phys. **8**, 1807 (1967).

It is convenient to use, instead of $S(K)$, the distribution

$$T(K) = S(K) - S_0(K), \tag{2.10}$$

where $S_0(K)$ is the no-scattering part of the S matrix. Our causality conditions will be formulated in terms of the corresponding functionals $T[\psi]$.

B. Infinite Differentiability

In the following sections the distribution T is sometimes said to be infinitely differentiable, and sometimes holomorphic, at a point $\bar{K}$ of $\mathfrak{W}$. These statements are given precise meaning in the following way.

The restricted real mass-shell $\mathfrak{W}$ is a subset of the restricted complex mass-shell $\mathfrak{W}_c$. The definition of $\mathfrak{W}_c$ is analogous to that of $\mathfrak{W}$: $\mathcal{M}_c$ is the set defined by

$$\mathcal{M}_c = \{K \mid k_i^2 = \mu_i^2, \Sigma k_i = 0\}, \tag{2.11}$$

where the components of the vectors k_i are now allowed to assume complex values, and $\mathfrak{W}_c$ is the set of all points $K = (k_1, \cdots, k_n)$ of $\mathcal{M}_c$ at which two or more of the vectors k_i are linearly independent. This set $\mathfrak{W}_c$ is a $(3n-4)$-dimensional submanifold of $\mathbf{C}^{4n}$, which means that at every point $\bar{K} \in \mathfrak{W}_c$ there is a (nonunique) local coordinate system.[10] This local coordinate system is defined as a triple $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ consisting of a neighborhood $\Delta_c(\bar{K}) \subset \mathbf{C}^{4n}$ of $\bar{K}$, a polydisk

$$D_c(\bar{K}) = \{z \mid z \in \mathbf{C}^{3n-4}, |z_\lambda - \bar{z}_\lambda| < r_\lambda, \quad \bar{z} \in \mathbf{C}^{3n-4}, r_\lambda > 0\}, \tag{2.12}$$

and a nonsingular holomorphic mapping $\Pi_{\bar{K}}: D_c(\bar{K}) \to \Delta_c(\bar{K})$ which is such that $\bar{K} = \Pi_{\bar{K}}(\bar{z})$ and

$$\mathfrak{W}_c \cap \Delta_c(\bar{K}) = \Pi_{\bar{K}}(D_c(\bar{K})).$$

At points $\bar{K}$ of $\mathfrak{W}$ this mapping can and will be chosen so that $\Delta_c(\bar{K}) \cap \mathfrak{W} = \Pi_{\bar{K}}(D_c(\bar{K}) \cap \mathbf{R}^{3n-4})$.

It is sometimes convenient to choose a local coordinate system in which the local coordinates z_λ are defined by the equations $z_\lambda = U_\lambda \cdot K$, where the $U_\lambda = (u_{\lambda 1}, \cdots, u_{\lambda n})$ are appropriately chosen n-tuples of four-vectors and

$$U_\lambda \cdot K = \sum_{i=1}^{n} u_{\lambda i} \cdot k_i = \sum_{i=1}^{n} \sum_{\nu=0}^{3} g^{\nu\nu} u_{\lambda i \nu} k_{i\nu}. \tag{2.13}$$

(The metric is $g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$.) Such a coordinate system will be called a *simple* coordinate system.

Infinite differentiability on $\mathfrak{W}$ can now be defined as follows.

Definition 1: Let $F(K)$ be a function defined on some open set $\mathcal{N} \subset \mathfrak{W}$. The function $F(K)$ is said to be *infinitely differentiable at* $\bar{K} \in \mathcal{N}$ if and only if for every choice $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ of local coordinates at $\bar{K}$, the function $F \circ \Pi_{\bar{K}}$ has continuous partial derivatives of all order in some neighborhood $D \subset \Pi_{\bar{K}}^{-1}(\mathcal{N} \cap \Delta_c(\bar{K}))$ of $\bar{z} = \Pi_{\bar{K}}^{-1}(\bar{K})$. If, in addition, the function $F \circ \Pi_{\bar{K}}$ can be represented by a convergent power series in a neighborhood of $\bar{z}$, the function F is said to be *holomorphic* at $\bar{K}$.

Definition 2: Let $T(K)$ be a Schwartz distribution defined on some open set $\mathcal{N} \subset \mathfrak{W}$. The distribution $T(K)$ is said to be infinitely differentiable (holomorphic) at $\bar{K} \in \mathcal{N}$ if on some neighborhood $\mathcal{N}' \subset \mathcal{N}$ of $\bar{K}$ there is defined an infinitely differentiable (holomorphic) function $F(K)$ that satisfies the equation

$$\int dK \psi(K)[T(K) - F(K)] = 0 \tag{2.14}$$

for all wavefunctions ψ in

$$\mathcal{B}(\mathcal{N}') = \{\psi \mid \psi \in \mathcal{D}(\otimes \mathcal{M}_i), \quad (\mathcal{M} \cap \operatorname{supp} \psi) \subset \mathcal{N}'\}. \tag{2.15}$$

Because the different possible local coordinate systems are holomorphically equivalent,[11] the conditions of the definitions are satisfied for all choices of local coordinate systems if they are satisfied for any particular choice.

C. Space–Time Wavefunctions

A preliminary problem of this paper is to develop some kind of space–time picture of a scattering process. To this end we introduce space–time wavefunctions

$$\tilde{\psi}_i(x) = (2\pi)^{-3} \int d^4k \delta(k^2 - \mu_i^2) \theta(\sigma_i k_0) e^{-i\sigma_i k \cdot x} \psi_i(k). \tag{2.16}$$

These functions $\tilde{\psi}_i(x)$ have the important property[12] that, for every positive integer N, the equation

$$\lim_{\tau \to \infty} \tau^N \tilde{\psi}_i(\hat{x}\tau) = 0 \tag{2.17}$$

is satisfied uniformly in $\hat{x}$ on compact subsets of the complement of

$$\hat{V}(\psi_i) = \{\hat{x} \mid \hat{x} = kt, k \in \operatorname{supp} \psi_i, t \text{ real}\}. \tag{2.18}$$

This property entails that for any fixed positive numbers ϵ, N, and δ there exists a τ_0 such that for all $\tau > \tau_0$ one has $|\tilde{\psi}_i(\hat{x}\tau)| < (|\hat{x}|\,\tau)^{-N}\delta$ for all $\hat{x}$ in

[10] The notation is essentially that of R. C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables* (Prentice-Hall, 1965). See also Appendix B.

[11] Reference 10, p. 17.

[12] D. Ruelle, Helv. Phys. Acta **35**, 147 (1962).

the complement of the set

$$\hat{V}_\epsilon(\psi_i) \equiv \{\hat{x} \mid \hat{x} = kt, |k - k'| \leq \epsilon, \quad k' \in \text{supp } \psi_i, \ t \text{ real}\} \bigcup \{\hat{x} \mid |\hat{x}| \leq \epsilon\}. \quad (2.19)$$

[The norm $|x|$ of any four-vector x is the Euclidean norm $|x| = (\Sigma x_\nu^2)^{\frac{1}{2}}$.] The rapid uniform collapse of $\tilde{\psi}_i(\hat{x}\tau)$ into $\hat{V}_\epsilon(\psi_i)$ as $\tau \to \infty$ suggests that the ith particle may in some limiting sense be regarded as confined to

$$V_\epsilon(\psi_i, \tau) = \{x \mid x = \hat{x}\tau, \hat{x} \in \hat{V}_\epsilon(\psi_i)\}. \quad (2.20)$$

This suggestion is supported by the following consideration. Let the various wavefunctions $\tilde{\psi}_i$ be displaced by the respective amounts $u_i\tau$. The displaced momentum–space wavefunctions are

$$\psi_i(k) \exp(i\sigma_i k \cdot u_i \tau),$$

and the corresponding transition amplitude is denoted by $T[\psi; U\tau]$. Thus, for product wavefunctions ψ in $\mathfrak{B}(\mathfrak{V})$, the amplitude $T[\psi; U\tau]$ has the representation

$$T[\psi; U\tau] = \int dK e^{-iU\cdot K\tau} \psi(K) T(K). \quad (2.21)$$

If $T(K)$ is essentially constant in the (perhaps very small) support of ψ, the approximation

$$T[\psi; U\tau] \approx \lambda \int d^4x \prod_{\text{initial}} \tilde{\psi}_i(x - u_i\tau) \prod_{\text{final}} \tilde{\psi}_i^*(x - u_i\tau) \quad (2.22a)$$

$$= \lambda \int d^4(\hat{x}\tau) \prod_{\text{initial}} \tilde{\psi}_i((\hat{x} - u_i)\tau) \prod_{\text{final}} \tilde{\psi}_i^*((\hat{x} - u_i)\tau) \quad (2.22b)$$

can be made. If an $\epsilon > 0$ can be found such that no point lies simultaneously in all of the displaced cones

$$\hat{V}_\epsilon(\psi_i; u_i) = \{\hat{x} \mid \hat{x} - u_i \in \hat{V}_\epsilon(\psi_i)\}, \quad (2.23)$$

then Eqs. (2.17) and (2.22) imply that

$$\lim_{\tau \to \infty} \tau^N T[\psi; U\tau] = 0 \quad (2.24)$$

for all positive integers N. [Henceforth, the notation $f(\tau) \Rightarrow 0$ will indicate the rapid decrease (2.24) of any function $f(\tau)$.] That is, if the intersection of all the sets $\hat{V}_\epsilon(\psi_i; u_i)$ is empty, then the probability that a reaction of the corresponding particles takes place decreases rapidly as τ becomes infinite.

This result provides a justification for considering the particles to be mainly confined to the space–time regions where the corresponding wavefunctions $\tilde{\psi}_i$ are not small. It also suggests that the image under $\hat{x} \to x = \hat{x}\tau$ of the region of intersection of the displaced cones $\hat{V}_\epsilon(\psi_i; u_i)$ should be interpretable as the location of the "collision" of the corresponding particles, in the limit $\tau \to \infty$. This idea has been discussed in detail in Ref. 5, and shown to be completely in accord with the nature of the one-particle exchange contribution to a scattering process.

This interpretation of overlap regions as the locations of the corresponding collisions is the basis of the present work. These collisions constitute the "events" of S-matrix theory, and causality conditions place restrictions on their space–time ordering.

3. WEAK ASYMPTOTIC CAUSALITY (WAC)

A. Formulation of WAC

If the particle trajectories (i.e., the displaced velocity cones) are such that all possible collisions involving two or more initial particles occur later than all possible collisions from which two or more final particle can emerge, the reaction is considered to be *acausal* and the corresponding transition amplitude is required to be small. This requirement is made precise in the following way. Let ψ be a product wavefunction and let $T[\psi]$ be the corresponding transition amplitude. Let the particles represented by ψ be displaced by amounts $u_i\tau$. The displaced particles are represented by the wavefunctions $\psi_i(k) \exp(i\sigma_i k \cdot u_i\tau)$, and the transition amplitude corresponding to them is denoted by $T[\psi; U\tau]$. For any fixed time $\hat{t}$ and positive number ϵ define the two sets

$$\hat{D}^\pm(\hat{t}, \epsilon) = \{\hat{x} \mid \pm(\hat{x}_0 - \hat{t}) \geq -\epsilon\}. \quad (3.1)$$

Finally, let $\mathcal{A}(\hat{t}, \epsilon, \psi)$ be the set of all n-particle displacements $U = (u_1, u_2, \cdots, u_n)$ such that (a) the Euclidean distance between points of $\hat{V}_\epsilon(\psi_i; u_i) \cap \hat{D}^-(\hat{t}, \epsilon)$ and points of $\hat{V}_\epsilon(\psi_j; u_j) \cap \hat{D}^-(\hat{t}, \epsilon)$ has a lower bound $d_{ij}^- > 0$ for all pairs $(i \neq j)$ of initial particles, and (b) the distance between points of $\hat{V}_\epsilon(\psi_i; u_i) \cap \hat{D}^+(\hat{t}, \epsilon)$ and points of $\hat{V}_\epsilon(\psi_j; u_j) \cap \hat{D}^+(\hat{t}, \epsilon)$ has a lower bound $d_{ij}^+ > 0$ for all pairs $(i \neq j)$ of final particles. A set $\mathcal{A}(\hat{t}, \epsilon, \psi)$ is called a set of acausal displacements. The weak causality condition is as follows:

Weak Asymptotic Causality (WAC): For any fixed product wavefunction ψ in $\mathfrak{B}(\mathfrak{V})$, fixed time $\hat{t}$, and fixed positive number ϵ, the condition $T[\psi; U\tau] \Rightarrow 0$ is satisfied uniformly in U on every compact subset of the set $\mathcal{A}(\hat{t}, \epsilon, \psi)$ of acausal displacements.

This causality condition is justified in Appendix A by proving that it holds in nonrelativistic quantum mechanics and in all classical models with finite-range interactions.

The WAC condition is also plausible within the framework of relativistic theories. If the set of particle

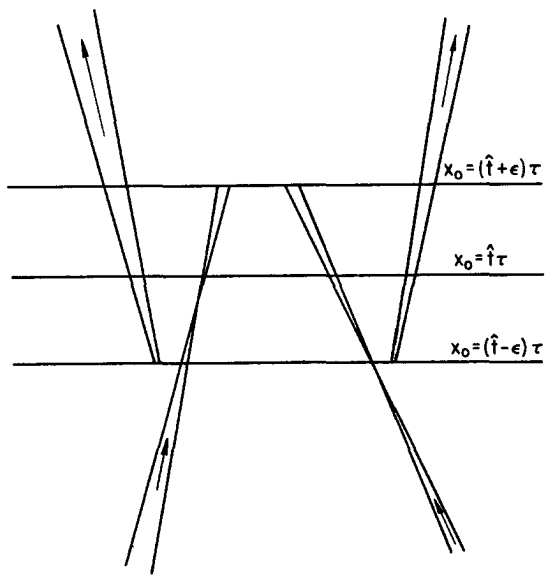


FIG. 1. Displaced velocity cones which are acausal with respect to WAC.

displacements U belongs to $\mathcal{A}(\hat{t}, \epsilon, \psi)$, then the displaced velocity cones

$$V_\epsilon(\psi_i; u_i, \tau) \equiv \{x = \hat{x}\tau \mid \hat{x} \in \hat{V}_\epsilon(\psi_i; u_i)\} \quad (3.2)$$

of the initial particles become increasingly far apart, as τ becomes infinite, for all times $x_0 < \hat{t}\tau + \epsilon\tau$, and the displaced velocity cones of the final particles become increasingly far apart for all times $x_0 > \hat{t}\tau - \epsilon\tau$. But if the initial particles become increasingly far apart in $x_0 < \hat{t}\tau + \epsilon\tau$, then the state generated near $x_0 = \hat{t}\tau$ by the initial particles should be represented with increasing precision, as $\tau \to \infty$, by the displaced initial free-particle state. Similarly, the state near $x_0 = \hat{t}\tau$ that develops into the final free-particle state should be represented with increasing precision by the displaced final free-particle state. (See Fig. 1.) Therefore, both these states near $x_0 = \hat{t}\tau$ are represented with increasing precision by the corresponding free-particle states, and the transition amplitude $T[\psi; U\tau]$ should approach its no-scattering value. This value is zero since the no-scattering part has been subtracted from T.

According to this argument, the amplitude $T[\psi; U\tau]$ would be expected to vanish as τ becomes infinite. But should it decrease faster than every inverse power of τ? This property means that for any fixed N, no matter how large, the amplitude decreases faster than τ^{-N}. Now, the overlap integrals

$$\mathcal{O}_{ij}^{\pm}(U, \tau) = \int_{D^{\pm}(\hat{t}, \epsilon, \tau)} d^4x \, |\tilde{\psi}_i(x - u_i\tau)\tilde{\psi}_j(x - u_j\tau)|, \quad (3.3)$$

where

$$D^{\pm}(\hat{t}, \epsilon, \tau) \equiv \{x = \hat{x}\tau \mid \hat{x} \in \hat{D}^{\pm}(\hat{t}, \epsilon)\}, \quad (3.4)$$

should provide a measure of the probability that interactions take place in $D^{\pm}(\hat{t}, \epsilon, \tau)$. If U belongs to $\mathcal{A}(\hat{t}, \epsilon, \psi)$, then (a) $\mathcal{O}_{ij}^{-}(U, \tau) \Rightarrow 0$ for all pairs $(i \neq j)$ of initial particles, and (b) $\mathcal{O}_{ij}^{+}(U, \tau) \Rightarrow 0$ for all pairs $(i \neq j)$ of final particles.[13] The overlap integrals therefore decrease faster than, say, τ^{-N^N}. If the propagation of dynamical effects is itself causal, at least up to terms that fall off faster than any inverse power of (Euclidean) distance, the fact that the initial and final overlaps fall off at a very large rate (τ^{-N^N}) should insure that the transition amplitude falls off at least at a relatively slow rate (τ^{-N}).

The discussion of the previous two paragraphs is based on the idea of a development of a system in time. It does not, however, require a fundamental quantity that represents the "state" of a system at an instant of time. As τ becomes infinite, the duration of the strip $\epsilon\tau \geq (x_0 - \hat{t}\tau) \geq -\epsilon\tau$ over which the initial and final particle states are compared becomes infinite. Therefore, the notion of a "state" of a system needs to become precise only when the time interval to which it refers becomes infinite. This is in accord with the general S-matrix philosophy.

B. Consequences of WAC

The weak asymptotic causality condition does not permit a complete specification of the singularity structure of $T(K)$, but it does have some useful consequences. Suppose that $\mathcal{N}$ is a connected open set in $\mathfrak{V}$ and that the set $\mathfrak{L}^+$ of points lying on positive-α Landau surfaces[14] passes through $\mathcal{N}$. Suppose also that $T(K)$ is holomorphic on $\mathcal{N} - \mathfrak{L}^+$. It is then of interest to know whether the functions that represent $T(K)$ in the various regions of $\mathcal{N} - \mathfrak{L}^+$ are holomorphic continuations of each other, and if so, to know the path that connects them.

In Sec. 5 it is shown that almost every point K on $\mathfrak{L}^+$ has a neighborhood $\mathcal{N}$ such that

$$\mathcal{N} \cap \mathfrak{L}^+ = \{K \mid K \in \mathcal{N}, \Lambda(K) = 0\} \quad (3.5)$$

where $\Lambda(K)$ is a real analytic function defined in a full $4n$-dimensional neighborhood of $\mathcal{N}$. The gradient $\nabla\Lambda(K) = (u_1, \cdots, u_n)$, where $u_{iv} = \partial\Lambda/\partial k_i^v$, is well defined and is nonzero in $\mathcal{N}$. This result motivates the following theorem.

Theorem 1: Suppose the following four conditions are satisfied:

(a) A real analytic function $\Lambda(K)$ is defined in a

[13] This is an easily proved consequence of (2.17) and the intersection properties of the sets $\hat{V}_\epsilon(\psi_i; u_i)$.

[14] L. D. Landau, Nuclear Phys. 13, 181 (1959). See also Ref. 17 p. 1556.

full neighborhood of a neighborhood $\mathcal{N} \subset \mathcal{W}$ of $\bar{K} \in \mathcal{W}$;

(b) there is a local coordinate system $(\Delta_c'(\bar{K}), \Pi_{\bar{K}}', D_c'(\bar{K}))$ with $z_1 = \Lambda(K)$,[15] such that the distribution $T = T \circ \Pi_{\bar{K}}'$ is a distribution in z_1 that is infinitely smooth in the variables $(z_2, \cdots, z_{3n-4})$. That is, for any test function $\psi(z)$ with support in $D_c'(\bar{K})$ the amplitude $T[\psi]$ has the representation

$$T[\psi] = \int dz F(z) \frac{d^m}{dz_1^m} [J(z)\psi(z)], \tag{3.6}$$

where m is an integer, $J(z)$ is the Jacobian appropriate to the transformation $\Pi_{\bar{K}}'$, and $F(z)$ is continuous in z_1 and has continuous derivatives of all orders in $(z_2, \cdots, z_{3n-4})$;

(c) for some fixed time $\hat{t}$, some fixed $\epsilon > 0$, and some fixed product wavefunction ϕ in $\mathcal{B}(\mathcal{W})$, with $\phi(\bar{K}) \neq 0$, the set $\mathcal{A}(\hat{t}, \epsilon, \phi)$ contains $-\nabla\Lambda(\bar{K})$;

(d) the WAC condition is valid.

Let $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ be any simple coordinate system at $\bar{K}$. Then for any α, $0 < \alpha < 1$, there exists a real neighborhood $\mathcal{N}' \subset (\mathcal{N} \cap \Delta_c(\bar{K}) \cap \text{supp } \phi)$ of $\bar{K}$ such that the restriction of the functional T to $\mathcal{B}(\mathcal{N}')$ can be written in the form

$$T[\psi] = \lim_{\substack{|\delta| \to 0 \\ \delta \in C^+(\alpha)}} \int dK \psi(K)[T^0(K) + T^1(K'(K, \delta))], \tag{3.7}$$

where

$$K'(K, \delta) \equiv \Pi_{\bar{K}}(\Pi_{\bar{K}}^{-1}(K) + i\delta) \tag{3.8}$$

and

$$C^+(\alpha) = \{\delta \mid \delta \in \mathbf{R}^{3n-4}, (\delta, \gamma) > |\delta|\,|\gamma|\,\alpha\}. \tag{3.9}$$

The vector γ in (3.9) is nonzero and is given by

$$\gamma_\lambda = [\partial\Lambda \circ \Pi_{\bar{K}}/(\partial z_\lambda)](\bar{z}), \quad 1 \leq \lambda \leq 3n - 4, \tag{3.10}$$

where $\Pi_{\bar{K}}(\bar{z}) = \bar{K}$. The function $T^0(K)$ is infinitely differentiable on $\mathcal{N}'$, and the function $T^1(K)$ is holomorphic (has a power series expansion in local coordinates[10]) on

$$\mathcal{E}_\alpha = \{K \mid K \in \mathcal{W}_c \cap \Delta_c(\bar{K}), \text{Im } [\Pi_{\bar{K}}^{-1}(K)] \in C^+(\alpha)\}. \tag{3.11}$$

This theorem is proved in Appendix C.

The specific form of the domain $\mathcal{E}_\alpha$ of Theorem 1 depends on the particular choice of simple coordinate system. A variation of Theorem 1 that does not refer to a particular simple coordinate system is the following Theorem 1A, which is also proved in Appendix C.

[15] The necessary condition that $\bar{\Lambda}(z) \equiv \Lambda \circ \Pi_{\bar{K}}'(z)$ have nonvanishing gradient follows from condition (c) of the theorem. This result is contained in the proof of the theorem. See also Sec. 5.

Theorem 1A: Suppose the assumptions of Theorem 1 are satisfied. For any $\epsilon > 0$ define

$$C_\epsilon^+(\bar{K}) = \{K \mid \text{Im } \{[\nabla\Lambda(\bar{K}) + U] \cdot K\} > 0 \quad \text{for all } U \in R_\epsilon\}, \tag{3.12}$$

where

$$R_\epsilon = \left\{U \mid U = (u_1, \cdots, u_n), \|U\| = \left[\sum_{i\nu} u_{i\nu}^2\right]^{\frac{1}{2}} \leq \epsilon\right\}, \tag{3.13}$$

the components $u_{i\nu}$ being real. Then for any $\epsilon > 0$ there exists a complex neighborhood $\mathcal{N}'_\epsilon \subset \mathcal{W}_c$ of $\bar{K}$ such that the restriction of T to $\mathcal{B}(\mathcal{W} \cap \mathcal{N}'_\epsilon)$ has the form

$$T[\psi] = \lim_{s \to 0} \int dK \psi(K)[T^0(K) + T^1(K'(K, s))], \tag{3.14}$$

where $K'(K, s)$ is any function uniformly continuous in $K \in \mathcal{N}'_\epsilon \cap \mathcal{W}$ and s, $0 < s < 1$, such that: $K'(K, s)$ is infinitely differentiable on $\mathcal{N}'_\epsilon \cap \mathcal{W}$ and all derivatives are continuous in both K and s; $K'(K, 0) = K$ for all $K \in [\mathcal{N}'_\epsilon \cap \mathcal{W}]$; and $K'(K, s) \in [\mathcal{N}'_\epsilon \cap C_\epsilon^+(\bar{K})]$ for all $s > 0$. The function $T^0(K)$ is infinitely differentiable on $\mathcal{N}'_\epsilon \cap \mathcal{W}$ and $T^1(K)$ is holomorphic on $\mathcal{N}'_\epsilon \cap C_\epsilon^+(\bar{K})$.

The content of Theorem 1A is this: At points K sufficiently near $\bar{K}$, the functional $T[\psi]$ is represented by a function that is, apart from infinitely differentiable singularities, holomorphic in a domain that is essentially the upper half-plane of the variable

$$\sigma(K; \bar{K}) = \nabla\Lambda(\bar{K}) \cdot K.$$

This theorem is applicable, for example, to the case of two-particle scattering [1 + 2 → 3 + 4]. The only (positive-α) Landau surfaces in the physical region are those corresponding to normal thresholds in $s = (k_3 + k_4)^2$. These surfaces are given by functions Λ of the form $\Lambda(K) = (k_3 + k_4)^2 - M^2$. Thus the displacement $\nabla\Lambda(\bar{K})$ has the form

$$\nabla\Lambda(\bar{K}) = (0, 0, u, u), \tag{3.15}$$

where $u = 2(\bar{k}_3 + \bar{k}_4)$. This displacement vector simply shifts the two final particles, 3 and 4, by twice the total energy–momentum vector of the reaction, as is illustrated in Fig. 2. If $\bar{k}_1$ and $\bar{k}_2$ are not collinear and if $\bar{k}_3$ and $\bar{k}_4$ are not collinear, then it is clear from the figure that for any product wavefunction ϕ with sufficiently small compact support centered at $\bar{K}$, there exists a $\hat{t}$ and an ϵ for which $-\nabla\Lambda(\bar{K})$ belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi)$. Indeed, because u is positive timelike (k_3 and k_4 are positive timelike), the displacement $-\nabla\Lambda(\bar{K})$ moves the regions of intersection of the final-particle velocity cones to a position earlier than

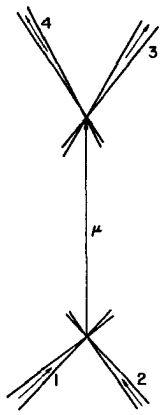
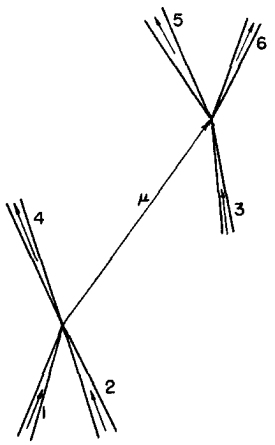


FIG. 2. Two-particle scattering ($1 + 2 \to 3 + 4$). The velocity cones are displaced by $U = (0, 0, u, u)$, where u is positive timelike.

FIG. 3. Three-particle scattering ($1 + 2 + 3 \to 4 + 5 + 6$). The velocity cones are displaced by $U = (0, 0, u, 0, u, u)$, where u is positive timelike.



that of the initial-particle cones. Thus condition (c) can be satisfied for any value of $\hat{t}$ lying between these two regions, for some sufficiently small ϵ. If $\tilde{T}(z)$ is analytic in the variables other than $z_1 = \Lambda(K)$, and if WAC is valid, then all the conditions of the theorem are met. The function $T^1(K)$ is then holomorphic in what is essentially the upper half-plane of the variable $\sigma(K; \bar{K})$. This upper half-plane is to lowest order in $(K - \bar{K})$, the upper half-plane of the variable s, so $T^1(K)$ is holomorphic in the intersection of a neighborhood of $\bar{K}$ with what is essentially the upper half-plane of the variable s.

Another application of Theorem 1 is to the pole contribution to the three-particle scattering amplitude. If in the vicinity of the pole at $\Lambda(K) = 0$ the amplitude is assumed to have the form $T(K) = R(K)D[\Lambda(K)] + H(K)$, where $R(K)$ and $H(K)$ are holomorphic and $D[\Lambda]$ is a distribution that is holomorphic for $\Lambda \neq 0$, then the conditions of the theorem on the structure of $T(K)$ are satisfied. The function $\Lambda(K)$ is given by $\Lambda(K) = (k_3 + k_4 + k_6)^2 - M^2$, and the displacement $\nabla\Lambda(\bar{K})$ is, therefore,

$$\nabla\Lambda(\bar{K}) = (0, 0, u, 0, u, u), \tag{3.16}$$

where $u = 2(\bar{k}_3 + \bar{k}_4 + \bar{k}_6)$. The result of this displacement is shown in Fig. 3. Suppose now that none of the initial-particle momenta are collinear and none of the final-particle momenta are collinear. Then inspection of Fig. 3 shows that for wavefunctions ϕ with sufficiently small compact support centered at $\bar{K}$, there exists a $\hat{t}$ and an ϵ for which $-\nabla\Lambda(\bar{K})$ belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi)$. Theorem 1 again prescribes a path of continuation of T^1 which involves infinitesimal detours into the upper half-plane of $\sigma(K; \bar{K})$.

The WAC condition does not give the $i\epsilon$ prescriptions for normal thresholds of all types of reactions. For example, if the case just considered were modified by adding one external line at each vertex in such a way that each subreaction involved two initial and two final particles, then the conditions of the theorem could not be satisfied. Indeed the conditions of the theorem provide, in such a case, no distinction between the two collisions that allows one to identify one collision as the cause and the other as the effect; the two vertices are completely equivalent so far as weak causality is concerned.

The two vertices are, of course, not completely equivalent. Positive energy is generally carried into one and out of the other by the external particles. This provides the necessary distinction between cause and effect, because energy–momentum is always transferred over macroscopic distances in a way such that positive energy flows forward in time. To proceed further, this energy-balance consideration must be incorporated into the causality condition.

The WAC condition can be augmented by an energy-balance condition so as to give the $i\epsilon$ prescriptions for all normal thresholds. Rather than dwelling on this point, we pass directly to the logical extension of this idea. Transmission of energy and momentum over macroscopic distances is, as far as we know, associated not only with the forward transmission of positive energy, but with transmission of just those amounts of energy and momentum that can be carried by physical particles. A formulation of this idea is given in the next section.

4. STRONG ASYMPTOTIC CAUSALITY (SAC)

A. Formulation of SAC

The condition of strong asymptotic causality (SAC) is a formulation of the notion that momentum–energy is transmitted over macroscopic distances only by stable physical particles: If a reaction requires a transfer of energy–momentum that cannot be carried by stable physical particles, then SAC requires the

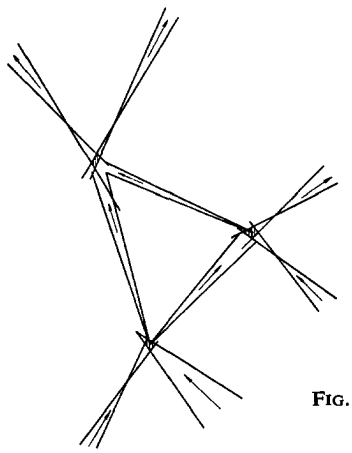


FIG. 4. A causal network.

probability of that reaction to fall off faster than any inverse power of the lower bound on the Euclidean distances over which such transfers must carry.

The central idea in the formulation of this requirement is that particle collisions are located in the intersections of the trajectory regions (i.e., displaced velocity cones) of the corresponding wavefunctions. From a collision involving two or more initial particles certain other stable physical particles may emerge. The momenta of these new particles must be consistent with conservation laws, and their trajectory regions must originate in the collision region where they are produced. These new trajectory regions may intersect other trajectory regions, defining new collision regions from which additional particles may emerge. In this fashion a causal network of collision regions connected by physical particle trajectories can be built up. (See Fig. 4.)

In order to formulate this idea more precisely the following definitions are introduced.

Definition 3: A causal space–time diagram $\mathfrak{D}$ is a triple $\mathfrak{D} = (V, L, \epsilon)$ consisting of a set $V = (v_1, \cdots, v_m)$ of space–time points (vertices), a set $L = (L_1, \cdots, L_s)$ of directed line segments of space–time points, and a matrix ϵ of structure constants. The following properties hold:

(a) Each line segment L_j has the representation

$$L_j = \{x \mid x = tl_j^+ + (1-t)l_j^-, 0 \leq t \leq 1\}, \quad (4.1)$$

where the end points l_j^σ are space–time points;

(b) the set V is the intersection of the end points:

$$V = \{x \mid x = l_i^\sigma = l_j^{\sigma'} \text{ for some } \sigma, \sigma' \text{ and } i \neq j\}. \quad (4.2)$$

Lines intersect effectively only at end points;

(c) the structure constants ϵ_{jr} $(1 \leq j \leq s, 1 \leq r \leq m)$ are defined by

$$\epsilon_{jr} = \begin{cases} +1, & \text{if } v_r = l_j^+, \\ -1, & \text{if } v_r = l_j^-, \\ 0, & \text{otherwise;} \end{cases} \quad (4.3)$$

(d) each line segment L_j is associated with a freely moving physical particle of nonzero mass μ_j and momentum–energy p_j. The real momentum–energy vector p_j satisfies $p_j^0 > 0$ and $p_j^2 = \mu_j^2$, and is related to L_j by

$$\Delta_j \equiv l_j^+ - l_j^- = \alpha_j p_j, \quad (4.4)$$

where α_j is some positive real number;

(e) momentum is conserved at each vertex:

$$\sum_j p_j \epsilon_{jr} = 0, \quad \text{all } r \quad (4.5)$$

(any other additively conserved quantum number must obey a similar conservation law);

(f) each v_r satisfies (4.2) with $\sigma = \sigma' = +1$ and also with $\sigma = \sigma' = -1$. (This condition can be imposed by virtue of the stability condition on the masses of physical particles.)

The line segments of $\mathfrak{D}$ are divided into two classes: internal and external. A line segment is internal if the set V contains both of its endpoints. Otherwise it is external. The vertices are similarly classified: a vertex is external if it is the end point of at least one external line. Otherwise it is internal. A $\mathfrak{D}$ with no internal lines is called trivial.

Definition 4: Let $\psi = \prod \psi_i$ be a product wavefunction. An n-particle displacement $U = (u_1, \cdots, u_n)$ belongs to the set $\mathcal{C}(\psi)$, and is called *causal with respect to* ψ, if and only if for each $\epsilon > 0$ there exists a causal space–time diagram $\mathfrak{D}_\epsilon$ such that:

(a) the diagram $\mathfrak{D}_\epsilon$ has n external lines that are associated (in the sense of Definition 3) in a one-to-one fashion with the n initial and final particles represented by ψ. In particular, the physical momentum–energy vectors associated with the external lines are $p_i = \sigma_i k_i$, where $K = (k_1, \cdots, k_n)$ belongs to the support of ψ;

(b) the vertex of $\mathfrak{D}_\epsilon$ that contains the endpoint of the ith external line is contained in $\hat{V}_\epsilon(\psi_i; u_i)$.

The sets of displacements that are not causal with respect to ψ are *acausal* with respect to ψ:

$$\mathcal{A}(\psi) = \{U \mid U \notin \mathcal{C}(\psi)\}. \quad (4.6)$$

The strong asymptotic causality condition analogous to WAC would be the requirement that for any fixed

product wavefunction $\psi \in \mathfrak{B}(\mathfrak{W})$ the relation

$$T[\psi; U\tau] \Rightarrow 0$$

be satisfied uniformly on compact subsets of $\mathcal{A}(\psi)$.

We shall, however, deal directly with the connected part $T_c[\psi]$ of $T[\psi]$. Only the connected causal space–time diagrams $\mathfrak{D}$ should be relevant to $T_c[\psi]$. (A connected diagram is one for which the point set $\bigcup L_j$ is connected.) Let $\mathcal{C}_c(\psi)$ be the subset of $\mathcal{C}(\psi)$ which is formed by requiring also that the space–time diagram $\mathfrak{D}$ of Definition 3 be connected. The corresponding acausal set is

$$\mathcal{A}_c(\psi) = \{U \mid U \notin \mathcal{C}_c(\psi)\}. \tag{4.7}$$

The SAC condition is then defined as follows:

Strong Asymptotic Causality. For any fixed product wavefunction $\psi \in \mathfrak{B}(\mathfrak{W})$ the condition $T_c[\psi; U\tau] \Rightarrow 0$ is satisfied uniformly on compact subsets of $\mathcal{A}_c(\psi)$.

B. Consequences of SAC

Consider displacements of the form

$$U_0(K) = (a + t_1k_1, a + t_2k_2, \cdots, a + t_nk_n), \tag{4.8}$$

where $K = (k_1, \cdots, k_n)$ is any point of supp ψ, a is any real four-vector, and the t_i are real scalars. If the momenta of the external lines of a diagram $\mathfrak{D}$ are given by K and the positions of these lines are specified by a set of displacements from a common origin of the form $U_0(K)$, then the external lines of $\mathfrak{D}$ all pass through a common point. The set $\mathcal{C}_0(\psi)$ of all displacements of the form (4.8) is then immediately seen to be a subset of $\mathcal{C}_c(\psi)$.

The sets $\mathcal{A}_c(K)$ and $\mathcal{C}_c(K)$ are defined to be the sets obtained by replacing supp ψ by K in the foregoing definitions.

The set $\mathfrak{L}^+$ is defined as the union of all positive-α Landau surfaces that are associated with connected nontrivial Landau diagrams.

Theorem 2: $\mathfrak{L}^+$ coincides with the set of all $K \in \mathcal{M}$ for which $\mathcal{C}_c(K) - \mathcal{C}_0(K)$ is nonempty.

Proof: The positive-α Landau loop equations associated with a diagram $\mathfrak{D}$ are precisely the statement that the set of vectors $\Delta_j \equiv \alpha_j p_j$ fit together to form a nontrivial causal diagram $\mathfrak{D}$. The conservation law constraints and mass-shell conditions are demanded both by the Landau equations and by the existence of $\mathfrak{D}$. Thus, the statement that there exists a nontrivial connected causal diagram $\mathfrak{D}$ satisfying $K(\mathfrak{D}) = \bar{K}$, where $K(\mathfrak{D})$ is the set of energy–momentum vectors associated with external lines of $\mathfrak{D}$, is equivalent to the statement that the Landau equations associated with diagram $\mathfrak{D}$ have a positive-α solution at $\bar{K}$.[16] At a point $\bar{K} \in (\mathcal{M} - \mathcal{M}_0)$, where $\mathcal{M}_0$ is the subset of the mass-shell $\mathcal{M}$ in which two or more initial-particle energy–momenta are collinear or two or more final-particle energy–momenta are collinear, the existence of a nontrivial connected causal diagram $\mathfrak{D}$, satisfying $K(\mathfrak{D}) = \bar{K}$, is equivalent to the fact that $\mathcal{C}_c(\bar{K}) - \mathcal{C}_0(\bar{K})$ is nonempty. This is because the trivial connected causal diagrams $\mathfrak{D}$ satisfying $K(\mathfrak{D}) = \bar{K}$ come only from $\mathcal{C}_0(\bar{K})$ and each nontrivial one is given by some U in $\mathcal{C}_c(\bar{K})$ that is not in $\mathcal{C}_0(\bar{K})$. At points $\bar{K}$ in $\mathcal{M}_0$ the set $\mathcal{C}_c(\bar{K}) - \mathcal{C}_0(\bar{K})$ is nonempty. (See Sec. 5, paragraph 2.) But all points $\bar{K} \in \mathcal{M}_0$ clearly lie on some positive-α Landau surface. This completes the proof.

This geometric interpretation of the Landau equations has been emphasized by Coleman and Norton.[16] We use it continually. In particular, the set of points lying on positive-α Landau surfaces is regarded as precisely the set of points K at which $K = K[\mathfrak{D}(K)]$ for some causal nontrivial $\mathfrak{D} = \mathfrak{D}(K)$.

We consider only connected diagrams, and by a Landau surface always mean a Landau surface associated with a nontrivial connected causal diagram.

A first consequence of SAC is Theorem 3.

Theorem 3: SAC implies that the scattering function $T_c(K)$ is infinitely differentiable at all points of $\mathfrak{W} - \mathfrak{L}^+$.

The proof is given in Appendix D. Theorems 2 and 3 combine to say that the singularities of $T_c(K)$ [or more precisely, the points at which $T_c(K)$ is not infinitely differentiable] are confined to the positive-α Landau surfaces.

We next turn to points that lie on $\mathfrak{L}^+$. Let $\bar{K}$ be a point of $\mathfrak{L}^+$. Let $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ be any set of $(3n-4)n$-particle displacements that define a simple local coordinate system at $\bar{K}$ through the equations $z_\lambda = U_\lambda \cdot K$. Define the set

$$\Gamma(\mathfrak{U}) = \{U \mid U = \Sigma t_\lambda U_\lambda, |t| = 1\}. \tag{4.9}$$

[The norm $|t|$ is the Euclidean norm of $t = (t_1, \cdots, t_{3n-4})$.] A *product neighborhood* $\mathcal{N}$ is a neighborhood such that for some product wavefunction χ, supp $\chi = \bar{\mathcal{N}}$. For any product neighborhood $\mathcal{N}$ define the closed set

$$\Gamma_c(\mathfrak{U}; \mathcal{N}) = \{U \mid U \in \mathcal{C}_c(\chi) \cap \Gamma(\mathfrak{U})\}, \tag{4.10}$$

where supp $\chi = \bar{\mathcal{N}}$ and the bar indicates closure.

Definition 5: A point $\bar{K}$ of $\mathfrak{L}^+$ is of Type I if and only if for every set $\mathfrak{U}$ that can be used to define a simple

[16] R. E. Norton, Phys. Rev. **135**, B1381 (1964); R. E. Norton and S. Coleman, Nuovo Cimento **38**, 438 (1965).

coordinate system $(\Delta_i(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ at $\bar{K}$ there exists a product neighborhood $\mathcal{N}$ of $\bar{K}$, $(\mathcal{N} \cap \mathcal{W}) \subset (\mathcal{W} \cap \Delta_c(\bar{K}))$, such that: (a) the set $\Gamma_c(\mathcal{U}; \mathcal{N})$ is contained in a finite number of closed disjoint subsets $\Gamma_c^j(\mathcal{U}; \mathcal{N})$; and (b) each of these sets $\Gamma_c^j(\mathcal{U}; \mathcal{N})$ can be contained in a corresponding set of the form

$$\Gamma^+(\mathcal{U}; e_j) = \{U \mid U = \Sigma t_\lambda U_\lambda, |t| = 1, (t, e_j) > 0\}, \quad (4.11)$$

where e_j is some vector in $\mathbf{R}^{3n-4}$. A point $\bar{K} \in \mathfrak{L}^+$ is of Type II if it is not of Type I.

The set $\mathcal{M}_0 \subset \mathfrak{L}^+$ of points $K = (k_1, \cdots, k_n)$ of $\mathcal{M}$ at which two initial or two final-particle momenta k_i are collinear consists entirely of Type II points. No other Type II points are known. The problem of showing that various points $\bar{K} \in \mathfrak{L}^+ - \mathcal{M}_0$ are of Type I is considered in the next section.

The structure of $T_c(K)$ near Type I points is intimately related to the geometric structure of the set $\Gamma_c(\mathcal{U}; \mathcal{N})$. Let Ω be the unit sphere

$$\Omega = \{t \mid t \in \mathbf{R}^{3n-4}, |t| = 1\} \quad (4.12)$$

and let

$$\Omega_c^j(\mathcal{U}; \mathcal{N}) = \{t \mid t \in \Omega, (\Sigma t_\lambda U_\lambda) \in \Gamma_c^j(\mathcal{U}; \mathcal{N})\}. \quad (4.13)$$

Because the various closed sets Γ_c^j are mutually disjoint, the corresponding compact sets Ω_c^j also have this property. It is therefore possible to construct open neighborhoods $\omega_j \subset \Omega$ of the sets Ω_c^j that have disjoint closures $\bar{\omega}_j$. Moreover, because of condition (b) of Definition 5, the neighborhoods ω_j can be constructed so that the polar cones

$$C^+(\bar{\omega}_j) = \{\delta \mid \delta \in \mathbf{R}^{3n-4}, (\delta', \delta) > 0 \text{ for all } \delta' \in \bar{\omega}_j\} \quad (4.14)$$

are nonempty. Finally, let

$$\mathfrak{L}_j^+(\mathcal{N}) = \{K \mid K \in \mathcal{N}, \Gamma_c^j(\mathcal{U}; \mathcal{N}) \cap \mathcal{C}_c(K) \text{ is nonempty}\}. \quad (4.15)$$

The structure of $T_c(K)$ at Type I points is then given by the following theorem, which is proved in Appendix D.

Theorem 4: Let $\bar{K} \in \mathfrak{L}^+$ be a Type I point. Let $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ be any simple coordinate system with local coordinates $z_\lambda = U_\lambda \cdot K$. Let $\mathcal{U} = \{U_1, \cdots, U_{3n-4}\}$, and let $\mathcal{N}$ be some product neighborhood of $\bar{K}$ that satisfies the conditions of Definition 5. Let ω_j be the neighborhoods of the $\Omega_c^j(\mathcal{U}; \mathcal{N})$ defined in the preceding paragraph. Finally, let the SAC condition be valid. Then there exists a neighborhood $\mathcal{N}_1$ of $\bar{K}$, $\mathcal{N}_1 \subset (\mathcal{N} \cap \mathcal{W})$, such that the restriction of $T_c[\psi]$ to $\mathcal{B}(\mathcal{N}_1)$ has the representation

$$T_c[\psi] = \int dK \psi(K) T_c^0(K) + \sum_j \lim_{\substack{|\delta| \to 0 \\ \delta \in C^+(\bar{\omega}_j)}} \int dK \psi(K) T_c^j(K'(K, \delta)). \quad (4.16)$$

The summation runs over the indices that label the $\Gamma_c^j(\mathcal{U}; \mathcal{N})$, and the quantity $K'(K, \delta)$ is defined by

$$K'(K, \delta) = \Pi_{\bar{K}}(\Pi_{\bar{K}}^{-1}(K) + i\delta). \quad (4.17)$$

The function $T_c^0(K)$ is infinitely differentiable on $\mathcal{N}_1$, and the functions $T_c^j(K)$ are holomorphic on the sets

$$\mathcal{E}_j = \{K \mid K \in \mathcal{W}_c \cap \Delta_c(\bar{K}), \operatorname{Im} \Pi_{\bar{K}}^{-1}(K) \in C^+(\bar{\omega}_j)\}. \quad (4.18)$$

Moreover, each limit function

$$T_c^j(K) = \lim_{\substack{|\delta| \to 0 \\ \delta \in C^+(\bar{\omega}_j)}} T_c^j(K'(K, \delta)) \quad (4.19)$$

exists and is infinitely differentiable on $\mathcal{N}_1 - \mathfrak{L}_j^+(\mathcal{N}_1)$.

Thus, aside from an infinitely differentiable background term, the amplitude $T_c(K)$ can be represented at Type I points as the sum of a finite number of terms, each with its own $i\epsilon$ prescription.[17]

5. CAUSAL DISPLACEMENTS AS GRADIENTS TO LANDAU SURFACES

In order to apply Theorem 4 at a point $\bar{K} \in \mathfrak{L}^+$, one must establish that $\bar{K}$ is of Type I. This is done by exploiting the very close connection between the causal displacement vectors U at $\bar{K}$ and the normal vectors to the various Landau surfaces that pass through $\bar{K}$. For example, when $\bar{K}$ belongs to only one positive-α Landau surface $\mathfrak{L}^+[\mathfrak{D}]$, there is essentially only one causal displacement U at $\bar{K}$, and this displacement can be identified with the normal to $\mathfrak{L}^+[\mathfrak{D}]$. The continuity of the normal then implies that $\bar{K}$ is of Type I. This result, and a number of related ones, are contained in the theorems that follow.

First we note that all points of $\mathcal{M}_0$ are Type II points. [Recall that $\mathcal{M}_0$ is the set of all points $K = (k_1, \cdots, k_n)$ of $\mathcal{M}$ at which two initial- or two final-particle momenta k_i are collinear.] This result is seen as follows. Let $\bar{K} \in \mathcal{M}_0$, and let $\bar{k}_1$ and $\bar{k}_2$ be collinear initial-particle momenta. (Similar arguments hold for collinear final-particle momenta.) Then, for every product wavefunction ψ that does not vanish at $\bar{K}$, every $\epsilon > 0$, and every U of the form $U = (u, 0, \cdots, 0)$,

[17] This property is referred to as the "general $i\epsilon$ rule." See H. P. Stapp, J. Math. Phys. 9, 1548 (1968).

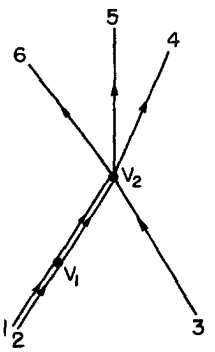


FIG. 5. A causal diagram $\mathfrak{D}$ with two initial-particle momenta collinear.

the various displaced velocity cones $V_\epsilon(\psi_i, u_i)$ always intersect in a way that allows the conditions of Definition 4 to be satisfied with a diagram $\mathfrak{D}$ of the type illustrated in Fig. 5. Thus, for any $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ that defines a simple coordinate system at $\bar{K}$, and for any product neighborhood $\mathcal{N}$ of $\bar{K}$, it is always possible to find a connected path in $\Gamma_c(\mathfrak{U}; \mathcal{N})$ that connects $\bar{U} = (\bar{u}, 0, \cdots, 0) \in \Gamma_c(\mathfrak{U}; \mathcal{N})$ with $-\bar{U} \in \Gamma_c(\mathfrak{U}; \mathcal{N})$. For this reason condition (b) of Definition 5 cannot be satisfied.

To classify points $\bar{K} \in \mathfrak{W}$ that do not lie in $\mathcal{M}_0$, some additional notation is introduced. The symbol $\bar{\mathfrak{D}}$ represents a fixed causal space–time diagram. The symbol $V(\bar{\mathfrak{D}}) = (v_1(\bar{\mathfrak{D}}), \cdots, v_m(\bar{\mathfrak{D}}))$ represents the set of space–time vectors that give the positions of the vertices of $\bar{\mathfrak{D}}$. The symbol $K(\bar{\mathfrak{D}}) = (k_1(\bar{\mathfrak{D}}), \cdots, k_n(\bar{\mathfrak{D}}))$ represents the set of mathematical momenta associated with the external lines of $\bar{\mathfrak{D}}$.

Definition 6: A diagram $\mathfrak{D}$ is *similar* to a diagram $\bar{\mathfrak{D}}$ if and only if its lines and vertices can be labeled so that $\mathfrak{D}$ and $\bar{\mathfrak{D}}$ have the same matrix ϵ of structure constants ϵ_{jr}, and the same types of particles associated with corresponding lines. The set of causal diagrams similar to $\bar{\mathfrak{D}}$ is denoted by $[\bar{\mathfrak{D}}]$.

This definition of $\mathfrak{D} \in [\bar{\mathfrak{D}}]$ does not require $V(\mathfrak{D})$ to coincide with $V(\bar{\mathfrak{D}})$, nor $K(\mathfrak{D})$ to coincide with $K(\bar{\mathfrak{D}})$. It does require each line of $\bar{\mathfrak{D}}$ to have a positive time-like image in any diagram $\mathfrak{D} \in [\bar{\mathfrak{D}}]$. Moreover, any line $\bar{L}_i$ of $\bar{\mathfrak{D}}$ and its image L_i in $\mathfrak{D} \in [\bar{\mathfrak{D}}]$ must be associated with the same type of particle.

Definition 7: A *contraction* $\mathfrak{D}'$ of $\bar{\mathfrak{D}}$ is a nontrivial diagram lying on the boundary of $[\bar{\mathfrak{D}}]$ that is formed by shrinking to points some, but not all, of the internal lines of $\bar{\mathfrak{D}}$. The notation $\mathfrak{D}' \subset \bar{\mathfrak{D}}$ means that $\mathfrak{D}'$ is a contraction of $\bar{\mathfrak{D}}$.

Definition 8: The positive-α Landau surface $\mathfrak{L}^+[\bar{\mathfrak{D}}]$ is the set of points K such that $K = K(\bar{\mathfrak{D}})$ for some $\mathfrak{D} \in [\bar{\mathfrak{D}}]$:

$$\mathfrak{L}^+[\bar{\mathfrak{D}}] = \{K \mid K = K(\mathfrak{D}), \mathfrak{D} \in [\bar{\mathfrak{D}}]\}. \tag{5.1}$$

The *restricted* positive-α Landau surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is the set of points K of $\mathfrak{L}^+[\bar{\mathfrak{D}}]$ that do not lie on $\mathfrak{L}^+[\mathfrak{D}']$ for any contraction $\mathfrak{D}'$ of $\bar{\mathfrak{D}}$:

$$\mathfrak{L}_0^+[\bar{\mathfrak{D}}] = \mathfrak{L}^+[\bar{\mathfrak{D}}] - \bigcup_{\mathfrak{D}' \subset \bar{\mathfrak{D}}} \mathfrak{L}^+[\mathfrak{D}']. \tag{5.2}$$

It is clear from Definition 8 that the set $\mathfrak{L}^+$ is the union of the restricted positive-α surfaces $\mathfrak{L}_0^+$.

The restricted surfaces $\mathfrak{L}_0^+$ are of interest because of the following result:

Theorem 5: If $\bar{\mathfrak{D}}$ is any fixed nontrivial connected causal diagram, and if $\bar{K} \in (\mathfrak{L}_0^+[\bar{\mathfrak{D}}] - \mathcal{M}_0)$, then there exists a neighborhood $\mathcal{N} \subset (\mathfrak{W} - \mathcal{M}_0)$ of $\bar{K}$ in which $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is an analytic submanifold of codimension 1.[10]

This theorem, which is proved in Appendix E, means, in particular, that nonmanifold points such as acnodes and cusps[18] cannot lie on $\mathfrak{L}^+ - \mathcal{M}_0$. The set $\mathfrak{L}^+ - \mathcal{M}_0$ is the union of manifolds of codimension 1 in $\mathfrak{W}$.

By virtue of Theorem 5, the normal vector to a surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is well defined (to within a scale factor) at each point $K \notin \mathcal{M}_0$ of that surface. The content of (b) of the next theorem is that this normal vector (appropriately scaled) is the n-particle displacement $U = (u_1, \cdots, u_n)$ that generates (by displacing lines originally passing through some common origin) the positions of the external lines of any diagram $\mathfrak{D}$ that satisfies $K(\mathfrak{D}) = K$. [Henceforth, the phrase "U generates $\mathfrak{D}$" will mean that $U = (u_1, \cdots, u_n)$ generates, by displacements $u_1, \cdots, u_n$ of lines originally passing through the origin, the positions of the external lines of $\mathfrak{D}$.]

Theorem 6: Let $\bar{\mathfrak{D}}$ be any fixed nontrivial connected causal space–time diagram, and let $\bar{K} \in \mathfrak{W}$ be a point of $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$. Then there is a full $4n$-dimensional neighborhood $\mathcal{N}(\bar{K})$ of $\bar{K}$ and a real analytic function $\Lambda(K)$, holomorphic in K over $\mathcal{N}(\bar{K})$, such that:

(a) The gradient $\nabla\Lambda(K)$ is nonzero at each point of $\mathcal{N}(\bar{K})$ and

$$\mathfrak{L}_0^+[\bar{\mathfrak{D}}] \cap \mathcal{N}(\bar{K}) = \{K \mid K \in \mathfrak{W} \cap \mathcal{N}(\bar{K}), \Lambda(K) = 0\}; \tag{5.3}$$

(b) if $K(\mathfrak{D}) \in \mathfrak{L}_0^+[\bar{\mathfrak{D}}] \cap \mathcal{N}(\bar{K})$ for some $\mathfrak{D} \in [\bar{\mathfrak{D}}]$, and if $U = (u_1, \cdots, u_n)$ is a set of n displacements that

[18] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, London, 1966), p. 104.

generates the diagram $\mathfrak{D}$, then U must have the form

$$U = \lambda\nabla\Lambda(K(\mathfrak{D})) + U_0(K(\mathfrak{D})), \tag{5.4a}$$

where $\lambda > 0$ and $U_0(K(\mathfrak{D}))$ is of the form (4.8). In other words,

$$u_i^\nu = \lambda \frac{\partial\Lambda}{\partial k_{i\nu}} + a^\nu + t_i k_i^\nu, \tag{5.4b}$$

where λ, a^ν, and t_i are real constants that depend only on the indicated indices and λ is strictly positive.

If two surfaces $\mathfrak{L}_0^+[\mathfrak{D}_1]$ and $\mathfrak{L}_0^+[\mathfrak{D}_2]$ coincide in some neighborhood of $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$, the two surfaces cannot be oriented in opposite ways. This follows from Theorem 7.

Theorem 7: Let $\mathfrak{D}_1$ and $\mathfrak{D}_2$ be two fixed nontrivial connected causal space–time diagrams, and let $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$ belong to both $\mathfrak{L}_0^+[\mathfrak{D}_1]$ and $\mathfrak{L}_0^+[\mathfrak{D}_2]$. Let the corresponding real analytic functions from Theorem 6(a) be $\Lambda_1(K)$ and $\Lambda_2(K)$. If $\mathfrak{L}_0^+[\mathfrak{D}_1]$ and $\mathfrak{L}_0^+[\mathfrak{D}_2]$ coincide in some neighborhood $\mathcal{N} \subset \mathcal{W}$ of $\bar{K}$, then $\nabla\Lambda_1(\bar{K}) = \lambda\nabla\Lambda_2(\bar{K}) + U_0(\bar{K})$, where $\lambda > 0$ and $U_0(\bar{K})$ is of the form (4.8).

A proof of Theorem 7 is given in Appendix F.

At points $\bar{K} \in \mathcal{W}$ not in $\mathcal{M}_0$, displacements $U_0(\bar{K})$ of the form (4.8) produce no essential changes in a diagram $\mathfrak{D}$. Their only effects are a common translation of all external lines of $\mathfrak{D}$ and displacements of these lines along themselves. The parameter λ fixes the scale of the diagram. Thus, Theorem 6(b) says that if $K = K(\mathfrak{D})$, where $\mathfrak{D} \in [\bar{\mathfrak{D}}]$ and $K \in \mathfrak{L}_0^+[\bar{\mathfrak{D}}]$, then the positions of the external lines of $\mathfrak{D}$ are obtained (essentially uniquely) by regarding the various components of $\nabla\Lambda(K)$ as the displacements of the corresponding external lines of $\mathfrak{D}$. Theorem 7 says that the sense of the causal direction along $\nabla\Lambda(K)$ is an *intrinsic feature* of the surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$; this sense does not depend on the particular class of similar diagrams $[\bar{\mathfrak{D}}]$ that might be used to define the given surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$.

To classify a point $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$ it is necessary to determine the complete set of displacements U that generate diagrams $\mathfrak{D}$ that satisfy $K(\mathfrak{D}) = \bar{K}$. The following two theorems give the structure of these sets. The first theorem is special; the second is general.

Theorem 8: Let $\bar{\mathfrak{D}}$ be a fixed nontrivial connected causal space–time diagram, and let $K(\bar{\mathfrak{D}}) = \bar{K}$ be a point of $\mathcal{W}$. If $\bar{K} = K(\mathfrak{D})$, where $\mathfrak{D}$ belongs either to $[\bar{\mathfrak{D}}]$ or to $[\mathfrak{D}']$ for some contraction $\mathfrak{D}'$ of $\bar{\mathfrak{D}}$, then any displacement U that generates $\mathfrak{D}$ is of the form

$$U = \sum_g \lambda_g \nabla\Lambda_g(\bar{K}) + U_0(\bar{K}), \tag{5.5}$$

where $\lambda_g \geq 0$ for all g, and U_0 is of the form (4.8). The (finite) sum in (5.5) runs over the indices g that label diagrams $\mathfrak{D}_g \subset \bar{\mathfrak{D}}$ or $\mathfrak{D}_g = \bar{\mathfrak{D}}$ for which $\bar{K} \in \mathfrak{L}_0^+[\mathfrak{D}_g]$.

This result is proved in Appendix E.

Theorem 9: Let $\bar{K}$ belong to $\mathfrak{L}^+ - \mathcal{M}_0$. Let I be a minimal set of indices g such that any restricted surface $\mathfrak{L}_0^+[\mathfrak{D}]$ that contains $\bar{K}$ coincides near $\bar{K}$ with one of the surfaces $\mathfrak{L}_0^+[\mathfrak{D}_g]$ for some $g \in I$. (The set I is known to be finite.[19]) If $K = K(\bar{\mathfrak{D}})$ for some connected causal space–time diagram $\bar{\mathfrak{D}}$, then any displacement U that generates $\bar{\mathfrak{D}}$ is of the form

$$U = \sum_{g \in I} \lambda_g \nabla\Lambda_g(\bar{K}) + U_0(\bar{K}), \tag{5.6}$$

where $\lambda_g \geq 0$ for all g and $U_0(\bar{K})$ is of the form (4.8).

Theorem 9 is a trivial consequence of Theorems 7 and 8. The characterization (5.6) of the displacements that generate diagrams $\mathfrak{D}$ for which $\bar{K} = K(\mathfrak{D})$ will be used to show that almost all points of $\mathfrak{L}^+ - \mathcal{M}_0$ are of Type I.

To show that a point $\bar{K}$ of $\mathfrak{L}^+ - \mathcal{M}_0$ is of Type I, it is not necessary to consider the sets $\Gamma_c(\mathfrak{U}; \mathcal{N})$ for all sets $\mathfrak{U}$ that define simple local coordinate systems at $\bar{K}$ or for product neighborhoods $\mathcal{N}$ of $\bar{K}$. It is sufficient to consider instead the sets

$$\Gamma_c(\mathfrak{U}; K) = \{U \mid U \in \mathcal{C}_c(K) \cap \Gamma(\mathfrak{U})\} \tag{5.7}$$

for any one (fixed) set $\mathfrak{U}$.

Theorem 10: Let $\bar{\mathfrak{U}} = \{\bar{U}_1, \cdots, \bar{U}_{3n-4}\}$ define a simple coordinate system at $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$. Then the point $\bar{K}$ is of Type I if and only if $\Gamma_c(\bar{\mathfrak{U}}; \bar{K})$ can be covered by a finite number of disjoint closed subsets $\Gamma_c^j(\bar{\mathfrak{U}}; \bar{K})$ of $\Gamma(\bar{\mathfrak{U}})$, each of which can be contained in a corresponding set of the form (4.11). Theorem 4 remains true if the ω_j are taken to be open neighborhoods (with disjoint closures) of the corresponding sets $\Omega_c^j(\bar{\mathfrak{U}}; \bar{K})$.

This theorem is proved in Appendix F.

Theorem 10 shows that the structure of $\mathcal{C}_c(\bar{K})$ determines whether a point $\bar{K} \in \mathfrak{L}^+ - \mathcal{M}_0$ is of Type I. To determine the structure of $\mathcal{C}_c(\bar{K})$ at these points we use the following theorem, which is proved in Appendix F.

[19] H. P. Stapp, J. Math. Phys. **8**, 1606 (1967).

Theorem 11: If $\bar{K}$ belongs to $\mathfrak{L}^+ - \mathcal{M}_0$, then the set $\mathcal{C}_c(\bar{K})$ consists of all displacements U that generate connected causal diagrams $\mathfrak{D}$ that satisfy $\bar{K} = K(\mathfrak{D})$.

Combining Theorems 9, 10, and 11, we obtain the following theorem.

Theorem 12: Let $\bar{K}$ be a point of $\mathfrak{L}^+ - \mathcal{M}_0$. Let I be a minimal set of indices g such that any restricted surface $\mathfrak{L}_0^+[\mathfrak{D}]$ that contains $\bar{K}$ coincides near $\bar{K}$ with one of the restricted surfaces $\mathfrak{L}_0^+[\mathfrak{D}_g]$ for some $g \in I$. If the vectors $\nabla\Lambda_g(\bar{K})$ and the ($4n$-dimensional) vectors F_ρ ($1 \leq \rho \leq n+4$) defined by

$$(F_\rho)_r^\mu = \delta_{\rho r}\bar{k}_r^\mu, \quad 1 \leq \rho \leq n \tag{5.8a}$$

and

$$(F_\rho)_r^\mu = \delta_\rho^{\mu+n+1}, \quad n+1 \leq \rho \leq n+4, \tag{5.8b}$$

are linearly independent, then the point $\bar{K}$ is of type I. Furthermore, the representation of $T_c(K)$ in Theorem 4 has only one boundary-value term at $\bar{K}$.

The proof is trivial. The vectors F_ρ form a basis for $\mathcal{C}_0(\bar{K})$. If the vectors $\nabla\Lambda_g(\bar{K})$ and F_ρ are linearly independent, there exists a set $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ that contains all the $\nabla\Lambda_g(\bar{K})$, $g \in I$, and defines a simple coordinate system at $\bar{K}$ (Appendix B). The set $\mathcal{C}_c(K) \cap \Gamma(\mathfrak{U})$ is then trivially contained in a *single* set $\Gamma^+(\mathfrak{U}, e)$ of the form (4.11). This implies that $\bar{K}$ is of Type I and that only one boundary-value term is required in the representation (4.16) of $T_c(K)$.[20]

Theorem 12 is applicable, in particular, to the case where $\bar{K}$ belongs to only one surface $\mathfrak{L}_0^+[\mathfrak{D}]$:

Corollary: If only one surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ passes through $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$, the point $\bar{K}$ is of Type I. Moreover, only one boundary-value term is needed in the representation (4.16) of $T_c(K)$.

In the situation described in the corollary only one boundary-value term $T_c^1(K)$ is needed in Theorem 4. By taking the neighborhood $\mathcal{N}$ of Theorem 4 small enough, the region of holomorphy of $T_c^1(K)$ can be expanded to include any given point in $\Delta_c(\bar{K}) \cap \mathcal{W}_c$ in the upper half-plane of $\sigma(K; \bar{K}) = \nabla\Lambda(\bar{K}) \cdot K$. The argument is similar to that in Theorem 1A and will not be repeated.

The corollary includes, of course, the special case in which $\mathfrak{L}^+$ consists near $\bar{K}$ of a single restricted surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$.[21] It also includes more complicated cases. For example, a point $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$ that lies on the edge of the surface $\mathfrak{L}_0^+[\mathfrak{D}_\tau]$ of the triangle diagram $\mathfrak{D}_\tau$ does not lie on $\mathfrak{L}_0^+[\mathfrak{D}_\tau]$. It lies on the surface $\mathfrak{L}_0^+[\mathfrak{D}]$ of a contraction $\mathfrak{D}$ of $\mathfrak{D}_\tau$. If these two surfaces are the only parts of $\mathfrak{L}^+$ that penetrate some neighborhood of $\bar{K}$, then the corollary applies.

The hypothesis of the corollary is satisfied at almost all points $\bar{K}$ of $\mathfrak{L}^+ - \mathcal{M}_0$. This is a consequence of the fact that only a finite number of distinct surfaces $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ intersect any bounded neighborhood $\mathcal{N} \subset \mathfrak{W}$ of $\bar{K}$.[19] The union of their intersections is therefore of zero measure in $\mathcal{N} \cap \mathfrak{L}^+$, and the complement of that union contains almost all points of $\mathcal{N} \cap \mathfrak{L}^+$. That is, in any bounded open set $\mathcal{N}$ of $\mathfrak{W}$, the set of points $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$ which lie on only one surface $\mathfrak{L}_0^+$ contains almost all points of $\mathcal{N} \cap (\mathfrak{L}^+ - \mathcal{M}_0)$.

A second consequence of Theorems 9, 10, and 11 is that if all the surfaces $\mathfrak{L}_0^+[\mathfrak{D}]$ that pass through $\bar{K}$ come from diagrams $\mathfrak{D}$ that are contractions of the same fixed diagram $\bar{\mathfrak{D}}$, then $\bar{K}$ is of Type I.[22]

Theorem 13: A point $\bar{K} \in (\mathfrak{L}^+ - \mathcal{M}_0)$ is of Type I if there is a nontrivial connected causal space–time diagram $\mathfrak{D}$ such that the diagrams $\mathfrak{D}_g$ of Theorem 9 are all contractions of $\mathfrak{D}$. In such a circumstance only one boundary-value term is needed in the representation of Theorem 4 of $T_c(K)$ at $\bar{K}$.

The proof is given in Appendix E.

It is not known if all points of $\mathfrak{L}^+ - \mathcal{M}_0$ are of Type I. Any counterexample would have to lie on at least four different surfaces $\mathfrak{L}_0^+[\mathfrak{D}_g]$. Two of these $\mathfrak{D}_g$ would have to be contractions of some diagram $\mathfrak{D}_1$ and two would have to be contractions of some other diagram $\mathfrak{D}_2$. But all four $\mathfrak{D}_g$ could not be contractions of any single diagram. We have not succeeded in finding such a case. In any event such points would be rare, and in a sense accidental, because their existence requires the intersection of surfaces $\mathfrak{L}_0^+[\mathfrak{D}_g]$ corresponding to contractions of one diagram $\mathfrak{D}_1$ to intersect the intersection of surfaces $\mathfrak{L}_0^+[\mathfrak{D}_g]$ corresponding to contractions of another "unrelated" diagram $\mathfrak{D}_2$. Two unrelated diagrams are diagrams that are not both contractions of any single diagram. It seems probable that singularities associated with unrelated diagrams will be additive and hence independent. A proof should emerge from the study of discontinuity formulas. That, however, is a subject in itself.[23,24]

[20] Points described by Theorem 12 are almost-simple points in the terminology of Ref. 17.

[21] Such points are called simple points in Ref. 17.

[22] It is a semisimple point in the terminology of Ref. 17.

[23] J. Coster and H. P. Stapp, J. Math. Phys. **10**, 371 (1969).

[24] J. Coster and H. P. Stapp, "Physical-Region Discontinuity Equations, Lawrence Radiation Laboratory Report UCRL-18512 (submitted to Phys. Rev.).

A concrete example of the analysis of this section is given in a nonmathematical version of the present work prepared by the first-named author and published elsewhere.[25]

6. SUMMARY

In this work the relationship between continuity properties of scattering functions in the physical region and macroscopic space–time phenomena has been examined. It was shown how singularities on Landau surfaces can be regarded as caused by processes in which the transfer of energy–momentum is carried by physical particles.

The algebraic equivalence of the Landau equations and corresponding space–time diagrams was emphasized earlier by Norton and Coleman.[16] The present work extends the algebraic result of Norton and Coleman by showing (in the course of proving Theorem 1) that if the scattering functions are infinitely differentiable except on the Landau surfaces, then the transition amplitude drops off faster than any inverse power of a scale parameter, unless the space–time collision regions are situated so that the momentum–energy can be carried from the initial particles to the final particles by means of physical particles. We also obtain the more difficult converse: If transition amplitudes fall off faster than any inverse power of the scaling parameter when the space–time collision regions are not causally connected via physical particles, then the scattering functions must be infinitely differentiable except on the Landau surfaces. Moreover, apart from infinitely differentiable singularities, the $i\epsilon$ prescriptions associated with the Landau surfaces coincide with those of perturbation theory.

Note Added in Proof: The infinitely differentiable background term is removed in a recent work by Iagolnitzer and Stapp.

APPENDIX A

The proof that the weak asymptotic causality condition is valid in nonrelativistic quantum mechanics is based on an inequality of Brenig and Haag.[26] Let ϕ_t^0 be the state at time t that would develop from an asymptotic initial-particle state ϕ if there were no interactions between the particles, and let ϕ_t be the corresponding state if there are interactions. Similarly, let ψ_t^0 be the state at time t that would develop into the asymptotic final particle state ψ if there were no interactions, and let ψ_t be the corresponding state if there are interactions. Then the transition amplitude $\langle\psi|\, T\,|\phi\rangle$ can be written

$$\langle\psi|\, T\,|\phi\rangle = \langle\psi_t|\,\phi_t\rangle - \langle\psi_t^0|\,\phi_t^0\rangle, \tag{A1}$$

where t is any arbitrary time. From (A1) an inequality follows:

$$|\langle\psi|\, T\,|\phi\rangle| \leq \|\psi_t - \psi_t^0\|\,\|\phi_t - \phi_t^0\| + \|\psi_t - \psi_t^0\| + \|\phi_t - \phi_t^0\|. \tag{A2}$$

The norm $\|\cdot\|$ in (A2) is defined for all functions $f(\mathbf{x}_1, \cdots, \mathbf{x}_m, t)$ by

$$\|f_t\| = \langle f_t\,|\,f_t\rangle^{\frac{1}{2}} \equiv \left\{\int d\mathbf{x}_1 \cdots d\mathbf{x}_m\, |f(\mathbf{x}_1, \cdots, \mathbf{x}_m, t)|^2\right\}^{\frac{1}{2}}, \tag{A3}$$

and it is assumed that ϕ_t^0 and ψ_t^0 have unit norms. The quantities $\|\psi_t - \psi_t^0\|$ and $\|\phi_t - \phi_t^0\|$ are bounded by the inequalities[26]

$$\|\phi_t - \phi_t^0\| \leq \int_{-\infty}^{t} dt'\, \|V\phi_{t'}^0\| \tag{A4a}$$

and

$$\|\psi_t - \psi_t^0\| \leq \int_{t}^{\infty} dt'\, \|V\psi_{t'}^0\|, \tag{A4b}$$

where V is the interaction Hamiltonian.

Let the asymptotic initial and final particles now be displaced by amounts $u_i\tau$ and let the displaced initial and final particles be represented by $\phi^{U\tau}$ and $\psi^{U\tau}$. For these displaced particles, the inequality (A2) leads to the following inequality:

$$|T[\phi\psi; U\tau]| \leq F[\phi; U, \tau]G[\psi; U, \tau] + F[\phi; U, \tau] + G[\psi; U, \tau], \tag{A5}$$

where

$$F[\phi; U, \tau] = \int_{-\infty}^{(\hat{t}+\epsilon)\tau} dt'\, \|V\phi_{t'}^{U\tau,0}\| \tag{A6a}$$

and

$$G[\psi; U, \tau] = \int_{(\hat{t}-\epsilon)\tau}^{\infty} dt'\, \|V\psi_{t'}^{U\tau,0}\|. \tag{A6b}$$

Here $\hat{t}$ is any arbitrary time, the number ϵ is positive, and the scale parameter τ is greater than 1.

If the potential V has a finite range R, the integrals that define $\|V\phi_t^{U\tau,0}\|$ and $\|V\psi_t^{U\tau,0}\|$ are restricted to the domain

$$A(R) = \{(\mathbf{x}_1, \cdots, \mathbf{x}_m)\,|\,|\mathbf{x}_i - \mathbf{x}_j| \leq R \text{ for all } i \text{ and } j\}. \tag{A7}$$

(Here it is assumed that there are m particles in all stages of the reaction; no creation or annihilation of particles is allowed.) Thus, the quantity $\|V\phi_t^{U\tau,0}\|$ has

[25] C. Chandler, Phys. Rev. 174, 1749 (1968).
[26] W. Brenig and R. Haag, Fortschr. Physik 7, 183 (1959).

the form

$$\|V\phi_{t'\tau}^{U\tau,0}\| = \left\{\tau^{3m}\int_{A(R\tau^{-1})} d\hat{\mathbf{x}}_1 \cdots d\hat{\mathbf{x}}_m \times \left|V(\hat{\mathbf{x}}_1\tau, \cdots, \hat{\mathbf{x}}_m\tau)\prod_{\text{initial}} \tilde{\psi}_i([\hat{\mathbf{x}}_i - \mathbf{u}_i]\tau, [\hat{t}' - u_{i0}]\tau)\right|^2\right\}^{\frac{1}{2}}. \tag{A8}$$

Now the wavefunctions $\tilde{\psi}_i([\hat{x} - u_i]\tau)$, considered as function of $\hat{x}$, collapse uniformly into the cones $\hat{V}_\epsilon(\psi_i; u_i)$ as τ becomes infinite. Consequently, if U belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi\psi)$ so that the initial particle cones $\hat{V}_\epsilon(\psi_i; u_i)$ are well separated before $(\hat{t} + \epsilon)$, then for some sufficiently large $\hat{\tau}$ the product wavefunction in (A8) is of rapid decrease in τ (and $|\hat{t}'|$) uniformly in $(\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_m, \hat{t}')$ for $(\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_m)$ in $A(R\hat{\tau}^{-1})$ and $\hat{t}' \leq (\hat{t} + \epsilon)$. Thus if V is bounded (or even merely integrable), the function $F[\phi; U, \tau]$ (and by similar arguments $G[\psi; U, \tau]$) is of rapid decrease when U belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi\psi)$. Then the inequality (A5) implies that the weak asymptotic causality condition is satisfied for any given U in $\mathcal{A}(\hat{t}, \epsilon, \phi\psi)$.

To extend the analysis to compact sets Γ of $\mathcal{A}(\hat{t}, \epsilon, \phi\psi)$ it is only necessary to observe that the velocity cones $\hat{V}_\epsilon(\psi_i, u_i)$ never come closer in the appropriate regions $\hat{x}_0 \leq (\hat{t} + \epsilon)$ and $\hat{x}_0 \geq (\hat{t} - \epsilon)$ than some distance $\delta(\Gamma)$. The number $\hat{\tau}$ is chosen so that $R \ll \delta(\Gamma)\hat{\tau}$, and the analysis proceeds as before. This insures that the WAC condition is satisfied uniformly on compact subsets of $\mathcal{A}(\hat{t}, \epsilon, \phi\psi)$. These arguments can be extended also to the case of potentials that have decreasing exponential bounds at large r.

The same ideas can be formulated in a classical theory by considering a statistical ensemble of classical experiments in which the momentum–space probability functions $P_i(\mathbf{k})$ of the initial and final particles have small compact support and in which the spatial distributions $P_i(\mathbf{x}, t)$ at time $t = 0$ fall off faster than any power of $|\mathbf{x}|^{-1}$.

Let $V_i(x)$ be the velocity cone that corresponds to the support of $P_i(\mathbf{k})$ and that has its tip at $x = (x_0, \mathbf{x})$:

$$V_i(x) = \{x' \mid x' - x = \lambda(\omega_i(\mathbf{k}), \mathbf{k}), \quad \lambda \in \mathbf{R}, \mathbf{k} \in \text{supp } P_i\}. \tag{A9}$$

Here $\omega_i(\mathbf{k})$ is $(\mathbf{k}^2 + m_i^2)^{\frac{1}{2}}$. Furthermore, let

$$V_i(x; r) = \bigcup_{|\mathbf{x}'-\mathbf{x}|\leq r} V_i(x_0, \mathbf{x}'). \tag{A10}$$

Now, if the trajectory of the ith freely moving particle passes through a point $x' = (x_0, \mathbf{x}')$ for which $|\mathbf{x}' - \mathbf{x}| < r$, then the trajectory must lie entirely in $V_i(x; r)$. This means that the fraction of the trajectories in the statistical ensemble for which particle i remains always inside $V_i(x; r)$ is just

$$\bar{P}_i(x; r) = \int_{|\mathbf{x}'-\mathbf{x}|\leq r} dx' P_i(\mathbf{x}', x_0). \tag{A11}$$

The rapid fall off of $P_i(\mathbf{x}, 0)$ for large $|\mathbf{x}|$ implies that

$$D_i(r\tau) \equiv 1 - \bar{P}_i(0; r\tau) \tag{A12}$$

goes rapidly to zero as τ becomes infinite: $D_i(r\tau) \Rightarrow 0$.

The stipulation in the weak asymptotic causality condition is (essentially) that the displaced velocity cones of the initial particles do not intersect for $t \leq \epsilon\tau$ and that the displaced velocity cones of the final particles do not intersect for $t \geq -\epsilon\tau$. The condition that the displaced cones do not intersect in these regions means that when $\tau = 1$ the minimum (Euclidean) distance between the cones in the regions $\hat{t} \leq \epsilon$ and $\hat{t} \geq -\epsilon$ is nonzero. If D_0 is this minimum distance, the minimum distance when τ is arbitrary is $D_0\tau$, which becomes infinite as τ becomes infinite.

Since the displaced cones $V_i(u_i\tau)$ have a minimum spatial separation $D_0\tau$ in the appropriate regions $\pm t \leq \epsilon\tau$, they can be replaced by slightly larger regions $V_i(u_i\tau; r\tau)$ that have a minimum spatial separation $d_0\tau > 0$.

Let the initial and final particles of the classical treatment be subjected to the displacements $u_i\tau$. The corresponding displaced spatial distributions $P_i^{U\tau}(\mathbf{x}, t)$ are given by

$$P_i^{U\tau}(\mathbf{x}, t) = P_i(\mathbf{x} - \mathbf{u}_i\tau, t - u_{i0}\tau). \tag{A13}$$

Thus, the probability that the freely moving particle i remains always inside $V_i(u_i\tau; r\tau)$ is

$$\bar{P}_i^{U\tau}(u\tau; r\tau) = \int_{|\mathbf{x}'-\mathbf{u}_i\tau|<r\tau} dx' P_i^{U\tau}(\mathbf{x}', u_{i0}\tau) = \bar{P}_i(0; r\tau). \tag{A14}$$

The probability that every particle i remains inside its displaced region $V_i(u\tau, r\tau)$ is $\prod_i \bar{P}_i(0; r\tau)$. This number rapidly approaches unity as τ becomes infinite.

Let us suppose that the interaction between the particles has a finite range R, in the sense that a set of particles do not interact unless the distance between some pair of them becomes less than R. But for sufficiently large τ the distance $d_0\tau$ of closest approach of the regions $V_i(u_i\tau; r\tau)$ is greater than R. Thus for this value of τ there will be no interaction between initial particles in the region $t \leq \epsilon\tau$ for those members of the ensemble for which each initial particle is in its region $V_i(u_i\tau, r\tau)$. The fraction of the members for which these conditions are realized (simultaneously for all particles) rapidly approaches unity. Consequently, the probability that the initial particles interact in $t \leq \epsilon\tau$ rapidly approaches zero as τ becomes

infinite. Similarly, the probability that the final particles interact in $t \geq -\epsilon\tau$ rapidly approaches zero as τ becomes infinite.

The fact that the fraction of members of the ensemble that have reactions in $t < 0$ decreases rapidly as τ becomes infinite means that the difference between the classical joint probability function

$$P_{\text{in}}^{U\tau}(\mathbf{X}, \mathbf{V}, t) = P_{\text{in}}^{U\tau}(\mathbf{x}_1, \cdots, \mathbf{x}_m; \mathbf{v}_1, \cdots, \mathbf{v}_m; t) \tag{A15}$$

and its unperturbed value

$$P_{\text{in}}^{U\tau,0}(\mathbf{X}, \mathbf{V}, t) = \prod_{\text{initial}} P_i^{U\tau,0}(\mathbf{x}_i, \mathbf{v}_i; t) \tag{A16}$$

must, when integrated, become small as τ becomes large:

$$\int d\mathbf{X}\, d\mathbf{V}\, |P_{\text{in}}^{U\tau}(\mathbf{X}, \mathbf{V}, 0) - P_{\text{in}}^{U\tau,0}(\mathbf{X}, \mathbf{V}, 0)| \Rightarrow 0. \tag{A17a}$$

Similarly, we must have

$$\int d\mathbf{X}\, d\mathbf{V}\, |P_{\text{out}}^{U\tau}(\mathbf{X}, \mathbf{V}, 0) - P_{\text{out}}^{U\tau,0}(\mathbf{X}, \mathbf{V}, 0)| \Rightarrow 0. \tag{A17b}$$

The classical expression for the overlap probability is

$$T = \int d\mathbf{X}\, d\mathbf{V} \min \{P_{\text{in}}^{U\tau}(\mathbf{X}, \mathbf{V}, 0), P_{\text{out}}^{U\tau}(\mathbf{X}, \mathbf{V}, 0)\}. \tag{A18}$$

This gives the fraction of the members of the "in" ensemble that can occur as members of the "out" ensemble, or conversely. (If in a certain "bin" the in ensemble has n_1 members and the out ensemble has n_2 members, the minimum of n_1 and n_2 is the maximum number of members common to both ensembles.) It follows from (A17) that T differs from its unperturbed value

$$T^0 = \int d\mathbf{X}\, d\mathbf{V} \min \{P_{\text{in}}^{U\tau,0}(\mathbf{X}, \mathbf{V}, 0), P_{\text{out}}^{U\tau,0}(\mathbf{X}, \mathbf{V}, 0)\}, \tag{A19}$$

by a term that goes rapidly to zero as τ becomes infinite.

Thus, for a fixed U in $\mathcal{A}(0, \epsilon, \psi)$, the weak asymptotic causality condition is valid in a classical model with finite range interactions. The analysis is extended to compact sets Γ of $\mathcal{A}(0, \epsilon, \psi)$ in the same way as in the quantum mechanical case.

APPENDIX B

By way of establishing notation, we give a constructive proof of the following well-known proposition: the restricted complex mass-shell $\mathcal{W}_c$ is a $(3n-4)$-dimensional analytic submanifold of $\mathbf{C}^{4n}$.

Proof: Let the n-tuples $K = (k_1, \cdots, k_n)$ of complex momentum vectors be associated with points $z = (z_1, \cdots, z_{4n})$ of $\mathbf{C}^{4n}$ through the equations

$$z_{4i+\mu-3} = k_{i\mu}, \quad 1 \leq i \leq n,\ 0 \leq \mu \leq 3. \tag{B1}$$

Then the set $\mathcal{M}_c$ can be written as

$$\mathcal{M}_c = \{z \mid z \in \mathbf{C}^{4n}, f_1(z) = \cdots = f_{n+4}(z) = 0\}, \tag{B2}$$

where the functions $f_i(z)$ are defined by

$$f_i(z) = \tfrac{1}{2} \sum_{\mu=0}^{3} g^{\mu\mu}(z_{4i+\mu-3})^2 - \tfrac{1}{2}m_i^2, \quad 1 \leq i \leq n, \tag{B3a}$$

and by

$$f_i(z) = g^{i-n-1,i-n-1} \sum_{j=1}^{n} z_{4j+i-n-4}, \quad n+1 \leq i \leq n+4. \tag{B3b}$$

(The metric is $g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$.)

Consider the Jacobian matrix

$$J_{ij}(z) = \frac{\partial f_i}{\partial z_j}(z). \tag{B4}$$

Explicit computation shows that the set of points of $\mathcal{M}_c$ where rank J is less than $n+4$ is exactly $\mathcal{M}_c - \mathcal{W}_c$. Therefore, at every point $\bar{K}$ (or $\bar{z}$) of $\mathcal{W}_c$ a (nonsingular) set of coordinates for $\mathbf{C}^{4n}$ can be defined by

$$F_i(z) = f_i(z), \quad 1 \leq i \leq n+4, \tag{B5a}$$

and

$$F_i(z) = \sum_{j=1}^{4n} E_{ij} z_j, \quad n+5 \leq i \leq 4n. \tag{B5b}$$

The $(3n-4)$ fixed real vectors

$$E_i = (E_{i1}, \cdots, E_{i4n}) \tag{B6}$$

appearing in (B5b) are any vectors which, together with the $n+4$ vectors

$$E_i(\bar{z}) = \left(\frac{\partial f_i}{\partial z_1}(\bar{z}), \cdots, \frac{\partial f_i}{\partial z_{4n}}(\bar{z})\right), \quad 1 \leq i \leq n+4, \tag{B7}$$

form a set of linearly independent vectors. The functions F_i define a coordinate system in a sufficiently small neighborhood $\Delta_c(\bar{K}) \subset \mathbf{C}^{4n}$ of $\bar{K} = K(\bar{z})$. It follows[27] from (B2) that the set $\Delta_c(\bar{K}) \cap \mathcal{W}_c$ is a $(3n-4)$-dimensional analytic submanifold of $\mathbf{C}^{4n}$. Since this construction can be made for any point $\bar{K} \in \mathcal{W}_c$, the proposition is proved.

Remark: The mapping $F: \mathbf{C}^{4n} \to \mathbf{C}^{4n}$ defined by (B5), and hence also its inverse, is real analytic. It

[27] Reference 10, p. 18. Following other authors we use the terminology analytic submanifold, instead of complex submanifold.

follows that the mapping $\Pi_{\bar{K}}$ introduced in Sec. 2B is also real analytic.

Remark: The vectors E_k for $n+5 \leq k \leq 4n$ can be associated with n-particle displacement vectors $U_\lambda = (u_{\lambda 1}, \cdots, u_{\lambda n})$, $1 \leq \lambda \leq 3n-4$, in the following way:

$$u^\mu_{\lambda j} = E_{\lambda+n+4, 4j+\mu-3}, \quad 1 \leq j \leq n, \quad 0 \leq \mu \leq 3. \tag{B8}$$

The local coordinates (B5b) of $\mathfrak{W}_c$ then become

$$F_\lambda = U_\lambda \cdot K = \sum_{j,\mu} u^\mu_{\lambda j} k_{j\mu}, \quad 1 \leq \lambda \leq 3n-4, \tag{B9}$$

where the bar indicates the relabeling of indices. Thus, the local coordinate system constructed in the proof of the proposition is a "simple" coordinate system. [See Eq. (2.13).]

Remark: For any point $\bar{K} \in \mathfrak{W}$ the set of $3n-4$ linearly independent vectors U_λ defined above provides a unique decomposition of any displacement vector U into the sum

$$U = \sum_{\lambda=1}^{3n-4} t_\lambda U_\lambda + U_0(\bar{K}), \tag{B10a}$$

where

$$U^\mu_{0j}(\bar{K}) = \bar{t}_j \bar{k}^\mu_j + a^\mu, \quad 1 \leq j \leq n, \tag{B10b}$$

is a causal displacement for any ϕ such that supp ϕ contains $\bar{K}$. The displacement $U_0(\bar{K})$ displaces each particle of the set specified by $\bar{K}$ along its own trajectory, and gives a single over-all displacement to all particles. Thus $U_0(\bar{K})$ is a member of the causal set $\mathfrak{C}_0(\bar{K})$ defined below Eq. (4.8).

The z_j defined above are simply components of the vectors k_i. In the rest of the paper the z's denote the $(3n-4)$ variables of a real local coordinate system.

Notice that $U_0(\bar{K})$ belongs to the null space of the matrix $\partial K/\partial z$. That is,

$$U_0[K(\bar{z})] \cdot \frac{\partial K(\bar{z})}{\partial z_\lambda} \equiv \sum_{j\mu} U^\mu_{0j} \frac{\partial \bar{k}_{j\mu}}{\partial z_\lambda} = 0, \quad 1 \leq \lambda \leq 3n-4. \tag{B11}$$

This follows from the restrictions on K imposed by (B2). Moreover, at any point of $\mathfrak{W}$ all vectors in the null space of $\partial K(\bar{z})/\partial z$ are of the form $U_0[K(\bar{z})]$, since this null space has dimension $n+4$.

APPENDIX C

A. Proof of Theorem 1

Let $\mathcal{A}(\hat{t}, \epsilon, \phi)$ be the set described in assumption (c). This set contains $-\nabla\Lambda(\bar{K})$, and is thus nonempty. It is, in fact, open in the topology induced by the Euclidean norm

$$\|U - U'\| = \left\{\sum_{i,\nu} |u_{i\nu} - u'_{i\nu}|^2\right\}^{\frac{1}{2}}. \tag{C1}$$

To see this, define for any neighborhood N of any displacement $\bar{U}$ in $\mathcal{A}(\hat{t}, \epsilon, \phi)$ the set

$$\hat{V}_\epsilon(\phi_i, N) = \bigcup_{U \in N} \hat{V}_\epsilon(\phi_i; u_i). \tag{C2}$$

Every two initial-particle cones $\hat{V}_\epsilon(\phi_i; \bar{u}_i)$ and $\hat{V}_\epsilon(\phi_j; \bar{u}_j)$ are separated by some finite (Euclidean) distance d_0 in $\hat{D}^-(\hat{t}, \epsilon)$. Therefore the sets $\hat{V}_\epsilon(\phi_i, N)$ and $\hat{V}_\epsilon(\phi_j, N)$ are separated in $\hat{D}^-(t, \epsilon)$ by a distance $d'_0 \geq (d_0 - 2\Delta)$, where Δ is the diameter of N. If Δ is chosen small enough, then the distance $d_0 - 2\Delta$ is positive, and the sets $\hat{V}_\epsilon(\phi_i, N)$ and $\hat{V}_\epsilon(\phi_j, N)$ are disjoint in $\hat{D}^-(\hat{t}, \epsilon)$. Similar arguments hold for each pair of initial particles and each pair of final particles. Thus every U in some neighborhood of $\bar{U}$ belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi)$. Since $\bar{U}$ is an arbitrary point of $\mathcal{A}(\hat{t}, \epsilon, \phi)$, this set is open.

According to hypothesis, the displacement $V = -\nabla\Lambda(\bar{K})$ belongs to $\mathcal{A}(\hat{t}, \epsilon, \phi)$. Since $\mathcal{A}(\hat{t}, \epsilon, \phi)$ is open, there exists a neighborhood N of V with compact closure $\bar{N}$ contained in $\mathcal{A}(\hat{t}, \epsilon, \phi)$. The WAC condition then requires that $T[\phi; U\tau] \Rightarrow 0$ uniformly on $\bar{N}$. The symbols N and $\bar{N}$ hereafter designate these two sets.

If the relation

$$\text{supp } \psi \subset \text{supp } \phi \tag{C3}$$

is true, the relation

$$\mathcal{A}(\hat{t}, \epsilon, \phi) \subset \mathcal{A}(\hat{t}, \epsilon, \psi) \tag{C4}$$

is also true. Thus it follows from WAC that the rapid decrease $T[\psi; U\tau] \Rightarrow 0$ is obtained uniformly on $\bar{N}$ for any fixed product wavefunction ψ in $\mathfrak{B}(\mathfrak{W})$ with support satisfying (C3).

Let $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ be any set of n-particle displacements that define a simple coordinate system $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ with local coordinates $F_\lambda = U_\lambda \cdot K$, and let

$$\Gamma(\mathfrak{U}) = \{U \mid U = \Sigma t_\lambda U_\lambda, \; t = (t_1, \cdots, t_{3n-4}) \in \Omega\}, \tag{C5}$$

where Ω is the unit sphere in $\mathbf{R}^{3n-4}$. Then, the set $\Gamma(\mathfrak{U})$ is contained in the union of the finite number of open sets $\Gamma^\pm_\rho(\mathfrak{U})$ constructed as follows. It is shown in Appendix B that the displacement $\nabla\Lambda(\bar{K})$ can be written as

$$\nabla\Lambda(\bar{K}) = \nabla_1\Lambda(\bar{K}) + U_0(\bar{K}), \tag{C6}$$

where $U_0(\bar{K})$ is a causal displacement belonging to

$\mathcal{C}_0(\bar{K})$, and

$$\nabla_1\Lambda(\bar{K}) = \Sigma\gamma_\lambda U_\lambda . \tag{C7}$$

The vector $\gamma = (\gamma_1, \cdots, \gamma_{3n-4})$ must be nonzero, since otherwise $-\nabla\Lambda(\bar{K})$ would not belong to $\mathcal{A}(\hat{t}, \epsilon, \phi)$. Let the normalization of $\Lambda(\bar{K})$ be such that $\gamma \equiv e_1$ is a vector of Ω, and let $e_2, \cdots, e_{3n-4}$ be any $3n-5$ other vectors in Ω which, together with e_1, form an orthonormal basis for $\mathbf{R}^{3n-4}$. For any α, $0 < \alpha < 1$, a finite open covering of Ω is given by the sets

$$\Omega_1^\pm = \{t \mid t \in \Omega, \pm(t, e_1) > (1-\alpha^2)^{\frac{1}{2}}\} \tag{C8a}$$

and

$$\Omega_\rho^\pm = \{t \mid t \in \Omega, \pm(t, e_\rho) > \beta(r-1)^{-\frac{1}{2}}\}, \quad 2 \leq \rho \leq 3n-4 = r, \tag{C8b}$$

where $\alpha > \beta > 0$, and (t, t') is the usual inner product $(t, t') = \Sigma t_\lambda t'_\lambda$ of $\mathbf{R}^{3n-4}$. The set $\Gamma(\mathcal{U})$ is thus covered by the open sets

$$\Gamma_\rho^\pm(\mathcal{U}) = \{U \mid U = \Sigma t_\lambda U_\lambda, t \in \Omega_\rho^\pm\}. \tag{C9}$$

The crucial step is to show that for any $0 < \alpha < 1$ there is some (real) neighborhood

$$\mathcal{N}_0 \subset (\mathcal{N} \cap \Delta_c'(\bar{K}) \cap \Delta_c(\bar{K}) \cap \text{supp } \phi)$$

of $\bar{K}$, such that for any fixed product wavefunction ψ in $\mathcal{B}(\mathcal{N}_0)$ the transition amplitude $T[\psi; U\tau]$ is of rapid decrease ($T[\psi; U\tau] \Rightarrow 0$) uniformly on

$$\Gamma_0(\mathcal{U}) = \Gamma(\mathcal{U}) - \Gamma_1^+(\mathcal{U}). \tag{C10}$$

Since

$$\Gamma_0(\mathcal{U}) \subset \left[\bigcup_{\rho\geq 2}\Gamma_\rho^+(\mathcal{U})\right] \cup \left[\bigcup_{\rho\geq 1}\Gamma_\rho^-(\mathcal{U})\right], \tag{C11}$$

it is sufficient to prove the uniform rapid decrease on the closed sets $\Gamma_1^-(\mathcal{U})$ and $\Gamma_\rho^\pm(\mathcal{U})$, $(\rho \geq 2)$.

For any fixed product $\psi \in \mathcal{B}(\mathcal{U})$ satisfying (C3) the uniform rapid decrease of $T[\psi; U\tau]$ on $\Gamma_1^-(\mathcal{U})$ is a consequence of the WAC condition, provided α is small enough so that $\Gamma_1^-(\mathcal{U}) \subset [N \cap \Gamma(\mathcal{U})]$. (If the original α is not small enough, then a smaller one can be used.) To use this fact let $\mathcal{O} \subset \text{supp } \phi$ be an open set with the property

$$\mathcal{O} \cap \mathcal{U} \subset (\mathcal{N} \cap \Delta_c'(\bar{K}) \cap \Delta_c(\bar{K})),$$

and let $\mathcal{O}' \subset \mathcal{O}$ be an open neighborhood of $\bar{K}$ with the property $\mathcal{O}' \subset \text{supp } \psi' \subset \mathcal{O}$, where ψ' is a product wavefunction. Let $\mathcal{N}_1$ be the intersection $\mathcal{N}_1 = \mathcal{O}' \cap \mathcal{U}$ of $\mathcal{O}'$ and $\mathcal{U}$. Finally, let $\chi \in \mathcal{B}(\mathcal{U})$ be a product wavefunction that is unity in supp ψ' and zero outside supp ϕ. Then for any product ψ in $\mathcal{B}(\mathcal{N}_1)$, the wavefunction $\bar{\psi} \equiv \psi\chi$ satisfies $T[\psi - \bar{\psi}; U\tau] \equiv 0$. Since $\bar{\psi}$ is a product wavefunction in $\mathcal{B}(\mathcal{U})$ that satisfies (C3), $T[\bar{\psi}; U\tau]$ is of rapid decrease uniformly on $\Gamma_1^-(\mathcal{U})$. Thus $T[\psi; U\tau]$ also has this property.

The uniform rapid decrease on the other sets $\Gamma_\rho^\pm$, $\rho \geq 2$ is a consequence of the smoothness requirement on $T(K)$. Let $z = (z_1, \cdots, z_{3n-4})$ be the local coordinates for which $T(z)$ is smooth in the variables $(z_2, \cdots, z_{3n-4})$. Let U be some displacement in $\Gamma_\sigma^\pm$ and let $h_\sigma(U)$ be the coordinate transformation defined by

$$\zeta_1 = z_1, \tag{C12a}$$

$$\zeta_\sigma = U \cdot K(z), \tag{C12b}$$

$$\zeta_\rho = \sum_\lambda e_{\rho\lambda} U_\lambda \cdot K(z), \quad 2 \leq \rho \leq 3n-4, \quad \rho \neq \sigma, \tag{C12c}$$

where the vectors $e_\rho = (e_{\rho 1}, \cdots, e_{\rho,3n-4})$ are the orthonormal basis vectors used in (C8), and the U_λ are as in (C5). Define

$$\phi_\rho(z) = \sum_\lambda e_{\rho\lambda} U_\lambda \cdot K(z) = V_\rho \cdot K(z), \quad 1 \leq \rho \leq 3n-4, \tag{C13}$$

and let $Q(z)$ be the determinant of the square matrix $Q_{\rho\lambda} = \partial\phi_\rho/\partial z_\lambda$. Finally, for any K in $\mathcal{N}_1$ (with $\mathcal{N}_1$ taken sufficiently small) write

$$\nabla\Lambda(K) = \Sigma g_\rho(K) V_\rho + U_0(K), \tag{C14}$$

where $U_0(K)$ belongs to $\mathcal{C}_0(K)$. The functions $g_\rho(K)$ are continuous, and $g_\rho(\bar{K}) = \delta_{\rho 1}$. Using the readily verified relation $U_0 \cdot \partial K/\partial z = 0$, one finds by explicit calculation that the Jacobian $H_\sigma(z, U)$ of the transformation $h_\sigma(U)$ is

$$H_\sigma(z, U) = Q(z)(g_1(z)X_\sigma - g_\sigma(z)X_1), \tag{C15}$$

where $U = \Sigma X_\rho V_\rho$. Thus if U belongs to $\Gamma_\sigma^\pm$, then the Jacobian does not vanish on the set

$$D_\sigma(\bar{K}) = \{z \mid K(z) \in \mathcal{N}_1 \equiv \mathcal{N}_1(\bar{K}), |Q(z)| > \epsilon, |g_\sigma(z)g_1^{-1}(z)| < \beta(r-1)^{-\frac{1}{2}}\}. \tag{C16}$$

[The open sets $D_\sigma(\bar{K})$ always contain $\bar{z}$, and hence are nonempty, for all $\sigma \geq 2$.] Therefore, if z belongs to D_σ and U belongs to $\Gamma_\sigma^\pm$ then the holomorphic transformation $h_\sigma(U)$ can be inverted, giving the $z_\lambda = z_\lambda(\zeta, X)$ as holomorphic functions of ζ and X. Then because $z_1(\zeta, X)$ is simply ζ_1, the smoothness of $T(z)$ in the variables $(z_2, \cdots, z_{3n-4})$ implies the smoothness of $T'(\zeta, X) = T(z(\zeta, X))$ in the variables $(\zeta_2, \cdots, \zeta_{3n-4}, X)$ when ζ belongs to $h_\sigma(U)D_\sigma$ and U belongs to $\Gamma_\sigma^\pm$. The proof of this is deferred to the end (Lemma 1).

Let $\mathcal{N}_\sigma = \Pi'_{\bar{K}}(D_\sigma)$. Then for U in $\Gamma_\sigma^\pm$ and ψ in $\mathcal{B}(\mathcal{N}_\sigma)$, the amplitude $T[\psi; U\tau]$ can be written

$$T[\psi; U\tau] = \int d\zeta_\sigma e^{-i\zeta_\sigma\tau} f(\zeta_\sigma, X), \tag{C17}$$

where

$$f(\zeta_\sigma, X) = \int d\zeta_1 \cdots d\zeta_{\sigma-1}\, d\zeta_{\sigma+1} \cdots d\zeta_{3n-4} H'_\sigma(\zeta, X) \times J'(\zeta, X)\bar{\psi}'(\zeta, X)T'(\zeta, X) \quad \text{(C18)}$$

is a distribution in ζ_σ that depends on X. The function J is the holomorphic Jacobian associated with the local coordinate system $(\Delta'_c(\bar{K}), \Pi'_{\bar{K}}, D'_c(\bar{K}))$. The holomorphy of $H'_\sigma(\zeta, X) = H_\sigma(z(\zeta, X))$ and $J'(\zeta, X) = J(z(\zeta, X))$, and the smoothness of T and $\bar{\psi} = \psi \circ \Pi'_{\bar{K}}$, imply the infinite differentiability of f in ζ_σ and in X for all $U \in \Gamma^{\pm}_\sigma$. (See Lemma 1.) The function f must also have compact support since the function

$$\bar{\psi}'(\zeta, X) = \bar{\psi}(z(\zeta, X))$$

does. It follows therefore, for all U in $\Gamma^{\pm}_\sigma$, that all derivatives $\partial^n f/\partial\zeta^n_\sigma$ are absolutely summable and hence that the integrals

$$I_n(X) = \int d\zeta_\sigma \left| \frac{\partial^n f}{\partial \zeta^n_\sigma}(\zeta_\sigma, X) \right| \quad \text{(C19)}$$

are bounded for U in $\Gamma^{\pm}_\sigma$. Equation (C17) then implies[28] that $T[\psi; U\tau] \Rightarrow 0$ uniformly on $\Gamma^{\pm}_\sigma$. Since the index σ was arbitrary, the amplitude $T[\psi; U\tau] \Rightarrow 0$, uniformly on the set $\Gamma_0(\mathfrak{V})$ defined in (C10) for all product wavefunctions ψ in $\mathcal{B}(\mathcal{N}_0)$, where $\mathcal{N}_0 \equiv \cap\mathcal{N}_\sigma$ is open in $\mathfrak{V}$ and contains $\bar{K}$.

To complete the proof let $\mathcal{N}' \subset \otimes\mathcal{M}_i$ be a neighborhood of $\bar{K}$, and let $\bar{\mathcal{N}}' \cap \mathfrak{V}$ be a subset of $\mathcal{N}_0$. Let χ be a product wavefunction in $\mathcal{B}(\mathcal{N}_0)$ with unit value on $\mathcal{N}'$. Then, for any ψ in $\mathcal{B}(\mathcal{N}' \cap \mathfrak{V})$, one has

$$T[\psi] = T[\psi\chi]. \quad \text{(C20)}$$

If the notation $\tilde{T}(t) \equiv T[\chi; \Sigma t_\lambda U_\lambda]$ is introduced, the amplitude $T[\psi]$ can be written in the form of the convolution[29]

$$T[\psi] = \int dt \tilde{\psi}(-t)\tilde{T}(t), \quad \text{(C21)}$$

where

$$\tilde{\psi}(t) = (2\pi)^{-(3n-4)} \int dz e^{-i(z,t)}(\psi \circ \Pi_K)(z). \quad \text{(C22)}$$

The z_λ in (C22) are the local coordinates $U_\lambda \cdot K$. Define

$$\theta(\Omega^+_1; t) = \begin{cases} 1, & \text{if } t = 0 \text{ or } t\,|t|^{-1} \equiv \hat{t} \in \bar{\Omega}^+_1, \\ 0, & \text{otherwise.} \end{cases} \quad \text{(C23)}$$

Equation (C21) can then be rewritten

$$T[\psi] = \int dt \tilde{\psi}(-t)[\tilde{T}^0(t) + \tilde{T}^1(t)], \quad \text{(C24)}$$

where

$$\tilde{T}^0(t) = [1 - \theta(\Omega^+_1, t)]\tilde{T}(t) \quad \text{(C25)}$$

and

$$\tilde{T}^1(t) = \theta(\Omega^+_1, t)\tilde{T}(t). \quad \text{(C26)}$$

The results of the preceding paragraph show that $\tilde{T}^0(t) \Rightarrow 0$ uniformly in $t\,|t|^{-1}$ as $|t| \to \infty$. Therefore, the function $\tilde{T}^0(t)$ has an infinitely differentiable Fourier transform $T^0(z)$, and[30]

$$\int dt \tilde{\psi}(-t)\tilde{T}^0(t) = \int dz(\psi \circ \Pi_{\bar{K}})(z)T^0(z). \quad \text{(C27)}$$

Let $J(z)$ be the Jacobian appropriate to the local coordinates $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$. Define

$$T^0_J(z) \equiv J^{-1}(z)T^0(z)$$

and let $T^0(K) \equiv T^0_J(\Pi^{-1}_{\bar{K}}(K))$. Then (C27) becomes

$$\int dt \tilde{\psi}(-t)\tilde{T}^0(t) = \int dK \psi(K)T^0(K), \quad \text{(C28)}$$

where $T^0(K)$ is infinitely differentiable on $\mathcal{N}' \cap \mathfrak{V}$.

The function $\tilde{T}^1(t)$ is not necessarily of rapid decrease when $|t| \to \infty$, but it has at most polynomial growth.[31] Hence, the function $\exp[-(\delta, t)]\tilde{T}^1(t)$ is of rapid decrease when δ belongs to

$$C^+ = \{\delta \mid (\delta, \hat{t}) > 0 \text{ for all } \hat{t} \in \bar{\Omega}^+_1\}, \quad \text{(C29)}$$

and it has a Fourier transform $T^1(z)$ that is holomorphic for $\operatorname{Im} z$ in C^+.[32] If $T^1_J(z) \equiv J^{-1}(z)T^1(z)$ is introduced for z in

$$E^+ = \{z \mid z \in \Delta_c(\bar{K}), \operatorname{Im} z \in C^+\}, \quad \text{(C30)}$$

the second term in (C24) becomes[30] after simple manipulation

$$\int dt \tilde{\psi}(-t)\tilde{T}^1(t) = \lim_{\substack{|\delta| \to 0 \\ \delta \in C^+}} \int dz(\psi \circ \Pi_{\bar{K}})(z)J(z)T^1_J(z + i\delta). \quad \text{(C31)}$$

From this it follows that

$$\int dt \tilde{\psi}(-t)\tilde{T}^1(t) = \lim_{\substack{|\delta| \to 0 \\ \delta \in C^+}} \int dK \psi(K)T^1(K'(K, \delta)), \quad \text{(C32)}$$

where $T^1(K) = T^1_J(\Pi^{-1}_K(K))$ and

$$K'(K, \delta) = \Pi_{\bar{K}}(\Pi^{-1}_{\bar{K}}(K) + i\delta). \quad \text{(C33)}$$

This completes the proof.

Lemma 1: Suppose $T[\psi]$ has the representation

$$T[\psi] = \int_S dz \left[\frac{d^m}{dz^m_1} J\psi(z)\right] F(z), \quad \text{(C34)}$$

where S is some domain, and $F(z)$ is a function that is continuous in z_1 and has continuous partial derivatives of all orders in the variables $(z_2, \cdots, z_n)$. Let

[28] H. Bremermann, Ref. 7, p. 85.
[29] See Lemma 2 of Appendix C.
[30] H. Bremermann, Ref. 7, p. 159.
[31] Reference 7, p. 164.
[32] R. F. Streater and A. S. Wightman, Ref. 7, p. 53.

$h: S \to S'$ be some nonsingular holomorphic mapping from $S \subset \mathbf{R}^n$ onto $S' \subset \mathbf{R}^n$ such that $z_1' \equiv h_1(z) = z_1$. Then there is a function $G(z')$, $z' = h(z)$, which is continuous in z_1' and has continuous partial derivatives of all orders in $(z_2', \cdots, z_{3n-4}')$, such that

$$T[\psi] = \int_{S'} dz' \left[\frac{d^m}{dz_1'^m} J\psi(h^{-1}(z'))\right] G(z'). \quad \text{(C35)}$$

Proof: Under the mapping h, the operator d^m/dz_1^m transforms into a differential operator

$$D = \sum_{p'} \sum_{p=1}^{m} h_{pp'}(z') D_p^{p'} \frac{d^p}{dz_1'^p}, \quad \text{(C36)}$$

where the $h_{pp'}$ are holomorphic functions and the $D_p^{p'}$ are derivative monomials in the variables $(z_2', \cdots, z_{3n-4}')$. The quantity $T[\psi]$ then has the form

$$T[\psi] = \int_{S'} dz' H'(z')[DJ\psi(h^{-1}(z'))]F(h^{-1}(z')). \quad \text{(C37)}$$

The function $F'(z') \equiv F(h^{-1}(z'))$ also has the property that it is continuous in the first variable z_1' and C^∞ in the other variables $(z_2', \cdots, z_{3n-4}')$. The function $H'(z')$ is the holomorphic Jacobian for the transformation h. For each p and p' the derivatives $D_p^{p'}$ can be transferred (through partial integrations) to the functions $H'h_{pp'}F'$. This transforms (C37) into the form

$$T[\psi] = \int_{S'} dz' \sum_{p=1}^{m} G_p(z') \frac{d^p}{dz_1'^p} J\psi(h^{-1}(z')). \quad \text{(C38)}$$

The functions $G_p(z')$ also have the property that they are continuous in the first variable z_1' and C^∞ in the others. Through further partial integrations the derivatives $d^p/dz_1'^p$ can all be transformed into derivatives $d^m/dz_1'^m$, yielding

$$T[\psi] = \int_{S'} dz' \left[\frac{d^m}{dz_1'^m} J\psi(h^{-1}(z'))\right] \sum_p G_p'(z'). \quad \text{(C39)}$$

The function $G = \sum_p G_p'$ is the function required by the Lemma.

Lemma 2: Equation (C21) is valid:

$$T[\psi] = T[\chi\psi] = \int dt \tilde{\psi}(-t)\tilde{T}(t). \quad \text{(C40)}$$

Proof: Since χ belongs to $\mathfrak{D}$, the functional $T[\chi\psi] = F[\psi]$ is a continuous linear functional on the space $\mathcal{E}$ of functions ψ that possess continuous partial derivatives of all orders. (Note that the support of ψ is not restricted here.) This is because $\chi\psi$ belongs to $\mathfrak{D}$ for every ψ in $\mathcal{E}$. The functional F then belongs to $\mathcal{E}'$, and the result[33] of Bremermann is directly applicable, yielding (C40) for all $\psi \in \mathfrak{D}$ that satisfy (C20).

[33] H. Bremermann, Ref. 7, p. 166. The lemma is Problem 6 on p. 166. The proof is a straightforward application of the technique of Riemann sums found on p. 49.

B. Proof of Theorem 1A

The proof consists of two parts. The first is a demonstration that the number α, the simple coordinate system of Theorem 1, the number ϵ, and the set $\mathcal{N}_\epsilon$ can be chosen so that $[\mathcal{N} \cap C_\epsilon^+(\bar{K})] \subset \mathcal{E}_\alpha$. The second consists of the necessary generalization of the way the limit (3.7) is taken.

Choose a simple coordinate system $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ in which $z_1 = \sigma(K; \bar{K}) = \nabla\Lambda(\bar{K}) \cdot K$. Such a choice is, of course, possible only if $\nabla\Lambda(\bar{K})$ does not belong to $\mathcal{C}_0(\bar{K})$. [See (B11).] But if $\nabla\Lambda(\bar{K})$ belonged to $\mathcal{C}_0(\bar{K})$ the various displaced velocity cones $\hat{V}(\phi_i, -\partial_i\Lambda(\bar{K}))$ would have a common point and assumption (c) of Theorem 1 could not be satisfied. Thus, coordinates with $z_1 = \sigma$ can be chosen.

Let $\mathcal{N}'' \subset [\Delta_c(\bar{K}) \cap \mathcal{W}_i]$ be a complex neighborhood of $\bar{K}$ such that its closure $\bar{\mathcal{N}}''$ is also contained in $\Delta_c(\bar{K}) \cap \mathcal{W}_c$ and the set $\Pi_{\bar{K}}^{-1}(\bar{\mathcal{N}}'')$ is convex. Then for $K \in [\mathcal{N}'' \cap C_\epsilon^+(\bar{K})]$ one has

$$\text{Im } z_1(K) \geq \sup_{U \in R_\epsilon} \text{Im } (U \cdot K) = \epsilon \, \|\text{Im } K\|. \quad \text{(C41)}$$

Since $\Pi_{\bar{K}}^{-1}(\bar{\mathcal{N}}'')$ is convex, there is some $A > 0$ such that, for $K \in \mathcal{N}''$,

$$\|\text{Im } K\| \geq |\text{Im } z| \, A, \quad \text{(C42)}$$

where $z = \Pi_{\bar{K}}^{-1}(K)$. For since the mapping $\Pi_{\bar{K}}$ is holomorphic, the functions $f_{i\nu}(x, y) \equiv \text{Im } k_{i\nu}(x + iy)$ have derivatives of all orders for $(x + iy)$ in $\bar{\mathcal{N}}''$ and can therefore be expanded about $y = 0$ by using the Taylor formula[34]:

$$f_{i\nu}(x, y) = f_{i\nu}(x, 0) + \sum_\lambda y_\lambda \frac{\partial f_{i\nu}}{\partial y_\lambda}(x, ty), \quad \text{(C43)}$$

where t, $0 \leq t \leq 1$, is some number that depends in general on y. Since $k_{i\nu}(z)$ is real when z is real, $f_{i\nu}(x, 0) = 0$. We can therefore write

$$\|\text{Im } K\| = |y| \, \Delta(x, y), \quad \text{(C44)}$$

where

$$\Delta(x, y) = |y|^{-1} \left\{\sum_{i\nu} \left| \sum_\lambda y_\lambda \frac{\partial f_i}{\partial y_\lambda}(x, ty) \right|^2\right\}^{\frac{1}{2}}. \quad \text{(C45)}$$

Consider now

$$A = \inf_{\substack{(x+iy) \in \Pi \bar{K}^{-1}(\bar{\mathcal{N}}'') \\ y \neq 0}} \Delta(x, y). \quad \text{(C46)}$$

If $A = 0$, there must be a sequence of points (x_n, y_n), with $y_n \neq 0$, such that

$$\lim_{n \to \infty} \Delta(x_n, y_n) = 0. \quad \text{(C47)}$$

[34] T. M. Apostol, *Mathematical Analysis* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1957), p. 124.

Moreover, because $\bar{\mathcal{N}}''$ is closed, the sequence (x_n, y_n) approaches a limit $(\bar{x}, \bar{y})$ with $(\bar{x} + i\bar{y}) \in \Pi_{\bar{K}}^{-1}(\bar{\mathcal{N}}'')$. If $\bar{y} \neq 0$, then $\|\text{Im } K(\bar{x} + i\bar{y})\| = 0$. But for a simple coordinate system the vanishing of the imaginary part of $K(\bar{x} + i\bar{y})$ implies $\bar{y} = 0$. This precludes the case $\bar{y} \neq 0$. To discuss the case where $\bar{y} = 0$, we first define $w_n = y_n |y_n|^{-1}$. The sequence w_n, suitably restricted to a subsequence, is convergent to some $\bar{w}$ with unit norm. The continuity of the derivatives $\partial f_{iv}/\partial y_\lambda$ further implies that

$$A = \lim_{n\to\infty} \Delta(x_n, y_n) = \left\{ \sum_{iv} \left| \sum_\lambda \bar{w}_\lambda \frac{\partial f_{iv}}{\partial y_\lambda}(\bar{x}, 0) \right|^2 \right\}^{\frac{1}{2}}. \quad \text{(C48)}$$

If $A = 0$, the equations

$$\sum_\lambda \bar{w}_\lambda \frac{\partial f_{iv}}{\partial y_\lambda}(\bar{x}, 0) = 0 \quad \text{(C49)}$$

must be satisfied for all i and v. Because the real analyticity of the $k_{iv}(z)$ implies

$$\frac{\partial k_{iv}}{\partial z_\lambda}(\bar{x}) = \frac{\partial f_{iv}}{\partial y_\lambda}(\bar{x}, 0), \quad \text{(C50)}$$

the equations (C49) state that the vectors $V_\lambda = (v_{i\lambda}, \cdots, v_{n\lambda})$, with $v_{i\lambda,v} = \partial k_{iv}/\partial z_\lambda$, are linearly dependent. These vectors V_λ form the rows of the Jacobian matrix of the mapping $\Pi_{\bar{K}}(z)$. Since this Jacobian has maximal rank in $D_c(\bar{K})$, the rows cannot be linearly dependent. This contradiction implies that A cannot be zero. Consequently A is greater than zero.

For any $\epsilon > 0$ one can find an $0 < \alpha < 1$ such that $\alpha < \epsilon A$. Then (C41) and (C42) imply

$$\text{Im } z_1 > |\text{Im } z|\, \alpha. \quad \text{(C51)}$$

This implies that z is in $\Pi_{\bar{K}}^{-1}(\mathcal{E}_\alpha)$. Thus,

$$[\mathcal{N}'' \cap C_\epsilon^+(\bar{K})] \subset \mathcal{E}_\alpha$$

and the first part of the theorem is proved.

It is clear that $\mathcal{N}''$ can be chosen so that

$$[\mathcal{N}'' \cap \mathfrak{W}] \subset \mathcal{N}'$$

where $\mathcal{N}'$ is the neighborhood of Theorem 1. Theorem 1 therefore implies Theorem 1A, provided the manner of taking the limit (3.7) can be converted to that of (3.14). Let $z_\lambda(x, s)$, $1 \leq \lambda \leq 3n - 4$, be any uniformly continuous functions of $x \in D(\bar{K}) \equiv D_c(\bar{K}) \cap \mathbf{R}^{3n-4}$ and s, $0 \leq s \leq 1$, which have the following three properties: (a) partial derivatives (with respect to x) of all orders exist and are continuous in both x and s; (b) $z(x, 0) = x$ for all x; and (c) $z(x, s)$ belongs to $\Pi_{\bar{K}}^{-1}(\mathcal{E}_\alpha)$ for all x and $s > 0$. We want to show that

$$T[\psi] = \lim_{s\to 0} \int dK \psi(K)[T^0(K) + T^1(K'(K, s))], \quad \text{(C52)}$$

where $T^0(K)$ and $T^1(K)$ are the functions of Theorem 1 and

$$K'(K, s) = \Pi_{\bar{K}}(z(\Pi_{\bar{K}}^{-1}(K), s)) \quad \text{(C53)}$$

Since $[\mathcal{N}'' \cap C_\epsilon^+(\bar{K})] \subset \mathcal{E}_\alpha$, all paths $K'(K, s)$ of the type allowed by the theorem are of the type (C53). Thus a proof of (C52) proves also Theorem 1A.

The relevant term in (C52) is the one involving $T^1(K'(K, s))$. In terms of local coordinates it can be written

$$I(s) \equiv \int dK \psi(K) T^1(K'(K, \delta)) = \int dx J(x) \psi(x) T_J^1(z(x, s)). \quad \text{(C54)}$$

The function $T_J^1(z)$ was defined in the proof of Theorem 1 as $J^{-1}(z)T^1(z)$, where $T^1(z)$ is defined for $\text{Im } z \in C^+(\alpha)$ and $z \in D_c(\bar{K})$ as

$$T^1(z) = (2\pi)^{-(3n-4)} \int dt e^{+i(z,t)} \tilde{T}^1(t). \quad \text{(C55)}$$

For $s > 0$ and $x \in D(\bar{K})$, the quantity $\text{Im } (z(x, s), t)$ is bounded from below by $\eta(s)\,|t| > 0$, for all (real) $t \neq 0$ in the support of $\tilde{T}^1(t)$. This lower bound is a consequence of the continuity of $z(x, s)$ and the assumption that $z(x, s) \in \Pi_{\bar{K}}^{-1}(\mathcal{E}_\alpha)$ for $s > 0$. The integral (C55) therefore converges uniformly in z and the integrations in (C54) can be interchanged[35]:

$$I(s) = \int dt \tilde{T}^1(t) \Phi(-t, s), \quad \text{(C56)}$$

where

$$\Phi(t, s) = (2\pi)^{-(3n-4)} \int dx J(x) J^{-1}(z(x, s)) \psi(x) e^{-i[z(x,s),t]}. \quad \text{(C57a)}$$

The next step is to show that the integral (C56) converges uniformly in s in some strip $0 \leq s \leq s_0$, where $0 < s_0 \leq 1$. The function $\tilde{T}^1(t)$ is continuous and of at most polynomial growth, so there is some integer N such that $\tilde{T}^1(t)(1 + |t|^{2N})^{-1}$ is absolutely summable. On the other hand, the function

$$(1 + |t|^{2N})\Phi(-t, s)$$

is bounded in both t and s for t in the support of $\tilde{T}^1(t)$ and s in some strip $0 \leq s \leq s_0$. To see this, consider the functions $z_\lambda(x, s)$ as a mapping ζ from $D(\bar{K})$ into $D_c(\bar{K})$ for each s. Let $W(x, s)$ be the Jacobian of ζ. Assumption (a) about $z(x, s)$ implies that $W(x, s)$ is continuous in both x and s and assumption (b) implies that $W(x, 0) = 1$ for all x. It follows that there exists s_0, $0 < s_0 \leq 1$, such that $W(x, s)$ does not vanish on any product set of the form

[35] Reference 35, p. 445.

$P \times I$, where $I = \{s \mid 0 \leq s \leq s_0\}$ and P is any compact subset of $D(\bar{K})$. For any $s \in I$, therefore, the mapping ζ can be inverted on P. Since supp ψ is a compact subset of $D(\bar{K})$, this result can be applied to (C57), yielding

$$\Phi(t, s) = (2\pi)^{-(3n-4)} \times \int_{\Gamma(s)} dz J(x(z)) J^{-1}(z) W^{-1}(z) \psi(x(z)) e^{-i(z,t)}. \tag{C57b}$$

The contours $\Gamma(s)$ in (C57b) are the images under ζ of $D(\bar{K}) \cap$ supp ψ for various values of s. The sets $\Gamma(s)$ are compact for all $s \in I$. Consider now the function $(1 + |t|^{2N})\Phi(-t, s)$:

$$(1 + |t|^{2N})\Phi(-t, s) = (2\pi)^{-(3n-4)} \int_{\Gamma(s)} dz J(x(z)) J^{-1}(z) W^{-1}(z) \psi(x(z)) \times \left[1 + (-1)^N \left(\sum_\lambda \frac{\partial^2}{\partial z_\lambda^2}\right)^N\right] e^{i(z,t)}. \tag{C58}$$

Partial integrations of (C58) yield

$$(1 + |t|^{2N})\Phi(-t, s) = (2\pi)^{-(3n-4)} \times \int_{\Gamma(s)} dz W^{-1}(z) F_N(z) e^{i(z,t)}, \tag{C59}$$

where

$$F_N(z) = W(z)\left[1 + (-1)^N \left(\sum_\lambda \frac{\partial^2}{\partial z_\lambda^2}\right)^N\right] \times J(x(z)) J^{-1}(z) W^{-1}(z) \psi(x(z)). \tag{C60}$$

Equation (C59) can be rewritten:

$$(1 + |t|^{2N})\Phi(-t, s) = (2\pi)^{-3n-4} \int dx F_N(z(x, s)) e^{i(z,t)}. \tag{C61}$$

The continuity of the mapping ζ in both x and s, and the continuity of the functions J, W, and ψ ensure the boundedness of $F_N(z(x, s))$ on $D(\bar{K}) \times I$. The boundedness of $e^{i(z,t)}$ for all t in the support of $\tilde{T}^1(t)$ is ensured by the fact that z is either real ($s = 0$) or in $\Pi_{\bar{K}}^{-1}(\mathcal{E}_a)$. Thus, the function $(1 + |t|^{2N})\Phi(-t, s)$ is bounded in both t and s, with t in the support of $\tilde{T}^1(t)$ and $0 \leq s \leq s_0$, and the integral (C56) converges uniformly.[36]

The order of the limit $s \to 0$ and the integration over t can therefore be interchanged:

$$\lim_{s \to 0} I(s) = \int dt T^1(t) \Phi(-t, 0). \tag{C62}$$

Because $\Phi(-t, 0)$ is just $\tilde{\psi}(-t)$, equation (C62) is

$$\lim_{s \to 0} I(s) = \int dt \tilde{T}^1(t) \tilde{\psi}(-t). \tag{C63}$$

[36] Reference 35, p. 438.

Since it was shown in the proof of Theorem 1 that

$$\lim_{\substack{|\delta| \to 0 \\ \delta \in C^+(x)}} \int dK \psi(K) T^1[K'(K, \delta)] = \int dt \tilde{T}^1(t) \tilde{\psi}(-t), \tag{C64}$$

the proof is complete.

APPENDIX D

A. Proof of Theorem 3

Consider an arbitrary point $\bar{K}$ of $\mathfrak{W} - \mathfrak{L}^+$, and let $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ be a set of linearly independent displacements that define a simple coordinate system $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ at $\bar{K}$. Because $\mathfrak{W} - \mathfrak{L}^+$ is open,[19] there exists a (product) neighborhood $\mathcal{N}_1' \subset \otimes \mathcal{M}_i$, $(\bar{\mathcal{N}}_1' \cap \mathcal{M}) \subset (\Delta_c(\bar{K}) \cap [\mathfrak{W} - \mathfrak{L}^+])$, of $\bar{K}$ such that $\bar{\mathcal{N}}_1'$ is the support of some product wavefunction χ'. Because $\mathfrak{U}$ defines a set of local coordinates at $\bar{K}$, the set $\mathcal{N}_1'$ can be chosen small enough so that the set $\Gamma(\mathfrak{U})$ defined in (4.9) has an empty intersection with $\mathcal{C}_0(\chi')$. [See Eq. (B10).]

Consider next a product χ satisfying supp $\chi \equiv \bar{\mathcal{N}}_1 \subset \mathcal{O} \subset \bar{\mathcal{N}}_1'$, where $\mathcal{O}$ is open. Then

$$\mathcal{C}_c(\chi) \subset \mathcal{C}_0(\chi'). \tag{D1}$$

To prove (D1) assume the converse: Suppose there is a U in $\mathcal{C}_c(\chi)$ that is not in $\mathcal{C}_0(\chi')$. Because the points of $\bar{\mathcal{N}}' \cap \mathcal{M}$ lie in $\mathfrak{W} - \mathfrak{L}^+$ and hence in $\mathcal{M} - \mathcal{M}_0$, we can assume that no two initial k_i are collinear in $\bar{\mathcal{N}}'$ and no two final k_i are collinear in $\bar{\mathcal{N}}'$. Then, because U is not in $\mathcal{C}_0(\chi')$, one can find some $\epsilon > 0$ such that the sets $\bar{V}_\epsilon(\chi_i, u_i)$ of Def. 4 have no common point. But then the diagram $\mathfrak{D}_\epsilon$ required by Def. 4, and by the fact that U is in $\mathcal{C}_c(\chi)$, must be a nontrivial diagram. This diagram $\mathfrak{D}_\epsilon$ belongs to $\mathcal{C}_c(K) - \mathcal{C}_0(K)$ for some K in $\bar{\mathcal{N}}' \cap \mathcal{M}$. But then this K lies on $\mathfrak{L}^+$, contrary to the definition of $\bar{\mathcal{N}}'$. This contradiction proves (D1).

Because $\Gamma(\mathfrak{U})$ does not intersect $\mathcal{C}_0(\chi)$, it does not intersect $\mathcal{C}_c(\chi)$, and is therefore a (compact) subset of $\mathcal{A}_c(\chi)$.

Since $\Gamma(\mathfrak{U})$ is a compact subset of $\mathcal{A}_c(\chi)$, the SAC condition implies that $\tilde{T}_c(t) = T_c[\chi; \sum t_\lambda U_\lambda] \Rightarrow 0$ uniformly in $t\,|t|^{-1}$ as $|t| \to \infty$.

Let the product wavefunction χ have unit value on the closure $\bar{\mathcal{N}}_2$ of some neighborhood $\mathcal{N}_2 \subset \mathfrak{W}$ of $\bar{K}$, where $\bar{\mathcal{N}}_2$ is a subset of $\mathcal{N}_1 \cap \mathfrak{W}$. If ψ belongs to $\mathcal{B}(\mathcal{N}_2)$, then $T_c[\psi] = T_c[\psi\chi]$. This relation can be rewritten[29]

$$T_c[\psi] = \int dt \tilde{\psi}(-t) \tilde{T}_c(t), \tag{D2}$$

where

$$\tilde{\psi}(t) = (2\pi)^{-(3n-4)} \int dz e^{-i(z,t)} (\psi \circ \pi_{\bar{K}})(z) \tag{D3}$$

is defined just as in Appendix C. Since $\tilde{T}_c(t)$ is of rapid decrease uniformly in $t\,|t|^{-1}$ when $|t|$ becomes infinite, it has an infinitely differentiable Fourier transform $T_c(z)$.[37] Moreover, the convolution theorem[30] can be used to convert (D2) to

$$T_c[\psi] = \int dz(\psi \circ \pi_{\bar{K}})(z)T_c(z). \tag{D4}$$

Let $J(z)$ be the weight function (Jacobian) appropriate to the mapping $\Pi_{\bar{K}}$, and let $T_{cJ}(z) = J^{-1}(z)T_c(z)$. Finally, let $T_c(K) = T_{cJ}(\Pi_{\bar{K}}^{-1}(K))$. This function is infinitely differentiable on $\mathcal{N}_2$, and

$$T_c[\psi] = \int dK\psi(K)T_c(K) \tag{D5}$$

for every wavefunction in $\mathcal{B}(\mathcal{N}_2)$. The distribution $T_c(K)$ is therefore infinitely differentiable on $\mathcal{N}_2$ and hence at $\bar{K}$.

B. Proof of Theorem 4

Let $\mathcal{N} \subset [\mathcal{W} \cap \Delta_c(\bar{K})]$ be a neighborhood that satisfies the conditions of Def. 5. Then, any neighborhood $\mathcal{N}_1$ of $\bar{K}$ fulfills the conditions of the theorem if its closure $\bar{\mathcal{N}}_1$ is contained in $\mathcal{N}$.

To prove this, let χ be a product wavefunction in $\mathcal{B}(\mathcal{N})$ with unit value on $\bar{\mathcal{N}}_1$, and let $T_c[\chi;\, U(t)] \equiv \tilde{T}_c(t)$ for any displacement U of the form $U(t) = \sum t_\lambda U_\lambda$. Being the Fourier transform of a distribution with compact support, $\tilde{T}_c(t)$ is infinitely differentiable.[31] If ψ is any wavefunction in $\mathcal{B}(\mathcal{N}_1)$, the transition amplitude $T_c[\psi]$ can be written[29]

$$T_c[\chi\psi] = T_c[\psi] = \int dt\tilde{\psi}(-t)\tilde{T}_c(t), \tag{D6}$$

where $\tilde{\psi}$ is defined in (D3). The domain of integration is broken up in the following way. Let

$$\omega_0 = \Omega - \bigcup_{i\geq 1} \bar{\omega}_i, \tag{D7}$$

and for all $i \geq 0$ let

$$C(\omega_i) = \{t \mid t \neq 0,\ t\,|t|^{-1} \in \omega_i\}. \tag{D8}$$

Define the step functions

$$\theta_i(t) = \begin{cases} 1, & \text{if } t \in C(\omega_i), \\ 0, & \text{if } t \notin \bar{C}(\omega_i), \end{cases} \tag{D9}$$

and adjust the (finite) values on the boundaries of the $C(\omega_i)$ so that

$$\sum_{i\geq 0} \theta_i(t) = 1 \tag{D10}$$

for all t. Equation (D6) then becomes

$$T[\psi] = \sum_{i\geq 0} \int dt\tilde{\psi}(-t)\tilde{T}^i(t), \tag{D11}$$

where subscripts c are now dropped and

$$\tilde{T}^i(t) = \theta_i(t)\tilde{T}(t). \tag{D12}$$

Consider first the term $\tilde{T}^0(t)$. Because the set $\bar{\omega}_0$ corresponds to a closed subset of $\Gamma(\mathfrak{U}) - \Gamma_c(\mathfrak{U};\, \mathcal{N})$, the SAC condition implies that, as $|t|$ increases, the function $\tilde{T}^0(t)$ is of rapid decrease, $\tilde{T}^0(t) \Rightarrow 0$, uniformly in $t\,|t|^{-1}$ for t in $\bar{C}(\omega_0)$. Since $\tilde{T}^0(t)$ vanishes for $t \notin \bar{C}(\omega_0)$, the restriction that t belong to $\bar{C}(\omega_0)$ can be removed: $\tilde{T}^0(t) \Rightarrow 0$ as $|t| \to \infty$. This means that $\tilde{T}^0(t)$ has an infinitely differentiable Fourier transform $T^0(z)$, and the first term in (D11) can be written[30]

$$\int dt\tilde{\psi}(-t)\tilde{T}^0(t) = \int dz(\psi \circ \pi_{\bar{K}})(z)T^0(z). \tag{D13}$$

Let $J(z)$ be the weight function appropriate to the mapping $\Pi_{\bar{K}}$, and let $T_J^0(z) = J^{-1}(z)T^0(z)$. The definition $T^0(K) = T_J^0(\Pi_{\bar{K}}^{-1}(K))$ allows (D13) to be written in the desired form

$$\int dt\tilde{\psi}(-t)\tilde{T}^0(t) = \int dK\psi(K)T^0(K), \tag{D14}$$

where $T^0(K)$ is infinitely differentiable on $\mathcal{N}_1$.

The functions $\tilde{T}^i(t)$, $i \geq 1$, which vanish for $t \notin \bar{C}(\omega_i)$, are of at most polynomial growth as $|t|$ becomes infinite. Any exponential damps this polynomial growth, and hence the functions

$$\exp\{-(\delta, t)\}\tilde{T}^i(t)$$

are of exponential decrease as $|t| \to \infty$ uniformly in $t\,|t|^{-1}$ for any $\delta \in C^+(\bar{\omega}_i)$. The function

$$\exp\{-(\delta, t)\}\tilde{T}^i(t)$$

has a Fourier transform $T^i(x + i\delta)$ that is holomorphic for $\delta \in C^+(\bar{\omega}_i)$.[32] It is evident that

$$\int dt\tilde{\psi}(-t)\tilde{T}^i(t) = \lim_{\substack{|\delta|\to 0 \\ \delta\in C^+(\bar{\omega}_i)}} \int dt\tilde{\psi}(-t)e^{-(\delta,t)}\tilde{T}^i(t). \tag{D15}$$

The convolution theorem[30] can be used to write (D15) as

$$\int dt\tilde{\psi}(-t)\tilde{T}^i(t) = \lim_{\substack{|\delta|\to 0 \\ \delta\in C^+(\bar{\omega}_i)}} \int dx(\psi \circ \pi_K)(x)T^i(x + i\delta). \tag{D16}$$

Define $T_J^i(z) = J^{-1}(z)T^i(z)$ for z in the set

$$E_i = \{z \mid z \in D_c(\bar{K}), \operatorname{Im} z \in C^+(\bar{\omega}_i)\}, \tag{D17}$$

and define $T^i(K) = T_J^i(\Pi_{\bar{K}}^{-1}(K))$. Then, (D16) takes

[37] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, 1937), p. 174. The case here is actually the (trivial) extension to several dimensions of the result of Titchmarsh.

the form

$$\int dt\tilde{\psi}(-t)\tilde{T}^i(t) = \lim_{\substack{|\delta|\to 0 \\ \delta\in C^+(\bar{\omega}_i)}} \int dK\psi(K)T^i(K'(K,\delta)), \tag{D18}$$

where

$$K'(K,\delta) = \Pi_R(\Pi_R^{-1}(K) + i\delta). \tag{D19}$$

The function $T^i(K)$ is holomorphic on $\mathcal{E}_i = \Pi_R(E_i)$. Equations (D14) and (D18) combine to yield the desired representation (4.16).

It remains to show that if $\bar{K}$ is any point in $\mathcal{N}_1 - \mathfrak{L}_i^+$, then the limit function (4.19) exists and is infinitely differentiable at $\bar{K}$. By virtue of (D18) the function $T^i(K)$ exists as a distribution. It is only necessary to show that it is infinitely differentiable. Let $\mathcal{N}_2$ be a neighborhood of $\bar{K}$ with closure contained in $\mathcal{N}_1 - \mathfrak{L}_i^+$, and let ψ be a product wavefunction in $\mathcal{B}(\mathcal{N}_2)$. Then it follows from the results just derived that $T[\psi; U]$ has the formal representation

$$T[\psi; U(t)] \equiv \tilde{T}(\psi, t) = \int dzJ(z)\psi(z)e^{-i(z,t)}\left\{\sum_j T_J^j(z)\right\} \tag{D20a}$$

$$= \int dt'\tilde{\psi}(t-t')\sum \tilde{T}^i(t'), \tag{D20b}$$

where $U(t) = \sum t_\lambda U_\lambda$, and the various $T_J^j(z)$ are distributions. Because $\mathcal{N}_2$ contains no points of $\mathfrak{L}_i^+$ the displacements in $\Gamma_c^i(\mathfrak{U}; \mathcal{N}_1)$ belong to $\mathcal{A}_c(\psi)$. The image of $\Gamma_c^i(\mathfrak{U}; \mathcal{N}_1)$ in Ω is in $\bar{\omega}_i$. Since the sets $\bar{\omega}_j$ $(j > 0)$ are disjoint, there is a neighborhood ω_i' of $\bar{\omega}_i$ with closure $\bar{\omega}_i'$ that does not intersect any of the sets $\bar{\omega}_j$ with $j \neq 0, i$. The set $\bar{\omega}_i'$ is therefore the image in Ω of a set $\Gamma_c'^i(\mathfrak{U}; \mathcal{N}_1) \subset \Gamma(\mathfrak{U})$ that is a subset of $\mathcal{A}_c(\psi)$. The SAC condition then requires that $\tilde{T}(\psi, \hat{t}\tau)$ be of rapid decrease, $\tilde{T}(\psi, \hat{t}\tau) \Rightarrow 0$, as $\tau \to \infty$, uniformly in $\hat{t} = t\,|t|^{-1}$ for $\hat{t}$ in $\bar{\omega}_i'$. This requirement is also satisfied by the first $(j = 0)$ contribution to (D20) since $T^0(z)$ is infinitely differentiable. [See (D13).] For $j \neq 0$, i the set $\bar{\omega}_i'$ is a subset of $\Omega - \bar{\omega}_j$. According to Lemma 3, proved below, the contributions $j \neq 0$, i to (D20) must, therefore, also be of rapid decrease uniformly on $\bar{\omega}_i'$. Therefore, the ith term of (D20) is also of rapid decrease uniformly on $\bar{\omega}_i'$. But by virtue of Lemma 3 the ith term must be of rapid decrease uniformly also on the complement of ω_i'. Thus for all t we have

$$\lim_{|t|\to\infty} |t|^N \int dzJ(z)\psi(z)T_J^i(z)e^{-i(z,t)} = 0 \tag{D21}$$

for all integers N. This implies that $T_J^i(z)$ is infinitely differentiable in the interior of the support of ψ.[37] Since ψ can be chosen to be nonzero at $\bar{z}$, the function $T_J^i(z)$ must be infinitely differentiable at $\bar{z}$. Thus $T^i(K)$ is, by definition, infinitely differentiable at $\bar{K}$.

This completes the proof.

Lemma 3: Let ω be an open subset of Ω, and let $\bar{\omega}'$ be a closed subset of $\Omega - \bar{\omega}$. Define

$$G(t) = \int_{\bar{C}(\omega)} dt'\tilde{\psi}(t-t')\tilde{T}(\chi, t'), \tag{D22}$$

where $\tilde{T}(\chi, t)$ and $\tilde{\psi}(t)$ are defined in the proof of Theorem 4. Then, for every integer N, the limit

$$\lim_{\tau\to\infty} \tau^N G(\hat{t}\tau) = 0 \tag{D23}$$

is obtained uniformly in $\hat{t} = t\,|t|^{-1}$ on $\bar{\omega}'$.

Proof: The function $\tilde{T}(\chi; t)$ is of at most polynomial growth as $|t| \to \infty$.[31] There is, therefore, an integer p for which $(1 + |t|^p)^{-1}\tilde{T}(\chi; t)$ is bounded. Let A be that bound. Then

$$|G(\hat{t}\tau)| \leq A\int_{\bar{C}(\omega)} dt'(1 + |t'|^p)\,|\tilde{\psi}(\hat{t}\tau - t')|, \tag{D24}$$

$$\leq A\int_{\bar{C}(\omega)} dt'(1 + |t'|^p)C_q\,|\hat{t}\tau - t'|^{-q}. \tag{D25}$$

The τ dependence of the right-hand side of (D25) can be explicitly exhibited:

$$|G(\hat{t}\tau)| \leq AC_q\tau^{3n-4-q}\{A_0(\hat{t}) + \tau^p A_p(\hat{t})\}, \tag{D26}$$

where

$$A_r(\hat{t}) = \int_{\bar{C}(\omega)} dt'\,|t'|^r\,|\hat{t} - t'|^{-q}, \quad r = 0, p. \tag{D27}$$

Since $\bar{\omega}' \subset \Omega - \bar{\omega}$, the magnitude $|t - t'|$ is bounded from below by a positive number when $\hat{t}$ is restricted to $\bar{\omega}'$. It follows that the integrals $A_r(\hat{t})$ are bounded on $\bar{\omega}'$ if q is chosen large enough. In fact, if N is any positive integer, the number q can be chosen large enough that τ^{N+1} times the right-hand side of (D26) is (uniformly) bounded on $\bar{\omega}'$. It follows that $G(\hat{t}\tau)$ satisfies (D23) uniformly on $\bar{\omega}'$.

APPENDIX E

A. Proof of Theorem 5

Let $\bar{\mathfrak{D}}$ be a fixed nontrivial connected causal space-time diagram, and let $\bar{K}$ belong to $(\mathfrak{L}_0^+[\bar{\mathfrak{D}}] - \mathcal{M}_0)$. According to Theorem 6, the set $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is locally the set of zeros (on $\mathfrak{W}$) of a real analytic function $\Lambda(K)$ that has nonzero gradient near $\bar{K}$. This would immediately imply that $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is an analytic submanifold of $\mathfrak{W}$ of codimension 1 at $\bar{K}$ were it not for the possibility that the gradient of Λ with respect to local coordinates

might vanish, even though $\nabla\Lambda(\bar{K})$ does not. To rule out this possibility, let $\bar{\Lambda}(z) = \Lambda(K(z))$, where z is a set of local coordinates at $\bar{K}$ for $\mathfrak{W}$. Then,

$$\frac{\partial\bar{\Lambda}}{\partial z_\lambda} = \frac{\partial\Lambda}{\partial K}\frac{\partial K}{\partial z_\lambda} = \sum_{i,\mu}\frac{\partial\Lambda}{\partial k_{i\mu}}\frac{\partial k_{i\mu}}{\partial z_\lambda}, \quad 1 \leq \lambda \leq 3n-4. \tag{E1}$$

Now, any vector $\nabla\Lambda \equiv \partial\Lambda/\partial K$ that causes (E1) to vanish for all λ is of the form $U_0(K)$. [See (B11).] But if $\nabla\Lambda$ were of this form, the displacements U which generate diagrams $\mathfrak{D} \in [\bar{\mathfrak{D}}]$, $\bar{K} = K(\mathfrak{D})$, would also be of this form. Hence, because $\bar{K}$ is not in $\mathcal{M}_0$, the diagrams $\mathfrak{D} \in [\bar{\mathfrak{D}}]$, $\bar{K} = K(\mathfrak{D})$, would be trivial. This is contrary to hypothesis. Thus, $\partial\bar{\Lambda}/\partial z_\lambda$ is nonzero and the surface $\mathfrak{L}_0^+[\bar{\mathfrak{D}}]$ is an analytic submanifold of $\mathfrak{W}$ of codimension 1 at $\bar{K}$.

B. Proof of Theorem 6

Let $\bar{\mathfrak{D}}$ be a fixed nontrivial connected causal space–time diagram with n external lines and m vertices, and let $\mathfrak{D} \in [\bar{\mathfrak{D}}]$. Let

$$Q_r[K(\mathfrak{D})] = \sum_j |\epsilon_{jr}|\, k_j(\mathfrak{D}) \tag{E2}$$

be the sum of the mathematical energy–momentum vectors k_j of the external lines attached to vertex r of $\mathfrak{D} \in [\bar{\mathfrak{D}}]$. Energy–momentum conservation at vertex r then gives

$$Q_r[K(\mathfrak{D})] = F_r[V(\mathfrak{D})], \tag{E3}$$

where

$$F_r(V) = \sum_j{}' \mu_j\epsilon_{jr}\,\|\Delta_j(V)\|^{-1}\Delta_j(V). \tag{E4}$$

The primed sum extends only over internal lines. The vectors $\Delta_j(V)$ are defined by

$$\Delta_j(V) = \sum_r \epsilon_{jr}v_r, \tag{E5}$$

and the quantity

$$\|\Delta_j(V)\| = [\Delta_j(V)\cdot\Delta_j(V)]^{\frac{1}{2}} \tag{E6}$$

is a Lorentz length. [The $\|\Delta_j[V(\mathfrak{D})]\|$ are all strictly positive for $\mathfrak{D} \in [\bar{\mathfrak{D}}]$, by definition.] The μ_j and ϵ_{jr} are the masses and structure constants of $\bar{\mathfrak{D}}$. Equation (E3) is obtained by first expressing $Q_r[K(\mathfrak{D})]$ in terms of the momentum–energies associated with the internal lines incident upon vertex r, and then using the identity $\|\Delta_j\| \equiv \|\alpha_j p_j\| = \alpha_j\mu_j$ to eliminate α_j.

In terms of the quantities just defined, the positive-α Landau surface $\mathfrak{L}^+[\bar{\mathfrak{D}}]$ is the intersection of the mass shell $\mathcal{M}$ with the set

$$S[\bar{\mathfrak{D}}] = \{K = (k_1, \cdots, k_n) \mid Q_r(K) = F_r[V; \bar{\mathfrak{D}}],\ V \in \Omega^+\}, \tag{E7}$$

where the argument $\bar{\mathfrak{D}}$ in $F_r[V; \bar{\mathfrak{D}}] = F_r(V)$ emphasizes the dependence of the ϵ_{jr} and μ_j in $F_r(V)$ upon $\bar{\mathfrak{D}}$, and

$$\Omega^+ \equiv \{V \mid \Delta_j(V) \text{ are positive timelike}\}. \tag{E8}$$

If $\bar{V} \in \Omega^+$ and $\bar{K} \in \mathfrak{L}^+[\bar{\mathfrak{D}}]$ satisfy $Q_r(\bar{K}) = F_r(\bar{V})$, then the set

$$\Omega^+(\bar{K}) = \{V \mid V \in \Omega^+, Q_r(\bar{K}) = F_r(V)\}, \tag{E9}$$

consists precisely of those points V which satisfy

$$\Delta_j(V) = \lambda_j\Delta_j(\bar{V}), \text{ all internal lines } j, \tag{E10}$$

where the λ_j are strictly positive scalars. For if V satisfies (E10), it clearly belongs to $\Omega^+(\bar{K})$. [See (E4).] Conversely, if V belongs to $\Omega^+(\bar{K})$, the vectors

$$D_r(V) = F_r(V) - F_r(\bar{V}) \tag{E11}$$

must vanish. This gives

$$\sum_r v_r\cdot D_r(V) = \sum_j{}' \mu_j\{\|\Delta_j(V)\| - \|\Delta_j(\bar{V})\|^{-1}\Delta_j(V)\cdot\Delta_j(\bar{V})\} = 0. \tag{E12}$$

Each term in the braces is nonpositive, hence each must vanish. This implies (E10).

Condition (E10) is essentially the condition that V belong to the null space of the Jacobian matrix $H(\bar{V})$ defined by

$$H_{r\mu,sv}(\bar{V}) = \frac{\partial F_{r\mu}}{\partial v_s^v}(\bar{V}) \tag{E13a}$$

$$= \sum_{kj}{}' \epsilon_{jr}\epsilon_{js}\mu_j\,\|\Delta_j(\bar{V})\|^{-3} \times \{\|\Delta_j(\bar{V})\|^2\, g_{\mu v} - \Delta_{j\mu}(\bar{V})\Delta_{jv}(\bar{V})\}. \tag{E13b}$$

The null space of $H(\bar{V})$ consists of all m-tuples $W = (w_1, \cdots, w_m)$ of four-vectors for which the equations

$$\sum_{r\mu} w_r^\mu H_{r\mu,sv}(\bar{V}) = 0 \tag{E14}$$

are satisfied for all s and v. It is evident from (E13b) that all vectors that satisfy (E10) belong to this null space. Hence, the set $\Omega^+(\bar{K})$ is contained in the null space of $H(\bar{V})$. Conversely, any vector V in the null space of $H(\bar{V})$ must satisfy (E10), without the restriction to positive λ_j. For if (E14) is true, the equation

$$\sum_{r\mu,sv} w_r^\mu w_s^v H_{r\mu,sv}(\bar{V}) = 0 = \sum_j{}' \mu_j\,\|\Delta_j(\bar{V})\|^{-3}\{\|\Delta_j(\bar{V})\|^2\cdot\|\Delta_j(W)\|^2 - [\Delta_j(\bar{V})\cdot\Delta_j(W)]^2\} \tag{E15}$$

is also true. Since each term in the braces is nonpositive, each must vanish. This implies (E10), without the restriction to positive λ_j.

Explicit computation shows that any linear combination of the vectors $\bar{V}$ and E_ρ $(0 \leq \rho \leq 3)$, where E_ρ is a $4m$-dimensional vector with components

$$(E_\rho)^\mu_r = \delta^\mu_\rho, \tag{E16}$$

belongs to the null space of $H(\bar{V})$. Since $\Delta_j(E_\rho) = 0$ for all ρ and j, the vectors $\bar{V} \in \Omega^+(\bar{K})$ and E_ρ must be linearly independent. The dimension $N(\bar{V})$ of the null space of $H(\bar{V})$ must therefore be at least five: $N(\bar{V}) \geq 5$.

On the other hand, $N(\bar{V})$ cannot be greater than 5. For suppose it were. Then there would exist some $\overline{W}$, linearly independent of the vectors $\bar{V}$ and E_ρ, such that $\overline{W}$, $\bar{V}$, and the E_ρ, and hence also any linear combination of them, belong to the null space of $H(\bar{V})$. Consider the identity

$$\Delta_j(\overline{W} + \alpha\bar{V}) \equiv \lambda_j(\alpha)\Delta_j(\bar{V}) = (\bar{\lambda}_j + \nabla)\Delta_j(\bar{V}), \tag{E17}$$

where the $\bar{\lambda}_j$ are defined by

$$\Delta_j(\overline{W}) = \bar{\lambda}_j\Delta_j(\bar{V}). \tag{E18}$$

The number α can obviously be chosen so that $\lambda_j(\alpha) \geq 0$ for all j and $\lambda_j(\alpha) = 0$ for some j. Let $\{\alpha_\lambda\}$ be a sequence $\alpha_\lambda \to \alpha$, with $\alpha_\lambda > \alpha$, and introduce

$$W_\lambda = \overline{W} + \alpha_\lambda\bar{V}. \tag{E19}$$

The vectors W_λ belong to $\Omega^+(\bar{K})$ and they converge to a limit, $W' = \overline{W} + \alpha\bar{V}$, which is not zero since $\overline{W}$ is linearly independent of $\bar{V}$. The set of four-vectors W' defines a diagram $\mathfrak{D}'$ which is a contraction of $\bar{\mathfrak{D}}$. The diagram $\mathfrak{D}'$ cannot be a trivial diagram because the trivial diagrams are generated only by linear combinations of the E_ρ, and W' cannot be one of these because of the linear independence of the $\overline{W}$, $\bar{V}$, and E_ρ. The function $F'_s(V)$ corresponding to the vertex s of $\mathfrak{D}'$ is simply the sum of the functions $F_r(V)$ corresponding to those vertices r of $\bar{\mathfrak{D}}$ that unite to form vertex s in the contraction of $\bar{\mathfrak{D}}$ that gives $\mathfrak{D}'$. The function $Q'_s(K)$ corresponding to the vertex s of $\mathfrak{D}'$ is formed in the same way from the $Q_r(K)$ of $\bar{\mathfrak{D}}$. Thus we obtain

$$Q'_s(\bar{K}) = F'_s(W_\lambda) \tag{E20}$$

for each value of λ. Since the F'_s do not depend on those internal lines of $\bar{\mathfrak{D}}$ that are contracted in forming $\mathfrak{D}'$, the limit can be taken: $Q'_s(\bar{K}) = F'_s(W')$. But then $\bar{K}$ belongs to $\mathfrak{L}^+[\mathfrak{D}']$. This contradicts the assumption of the theorem. Thus the quantity $N(\bar{V})$ cannot be greater than 5. But then $N(\bar{V})$ is exactly 5, and the matrix $H(\bar{V})$ has rank

$$R(\bar{V}) \equiv 4m - N(\bar{V}) = 4m - 5 \equiv R. \tag{E21}$$

The knowledge that $N(\bar{V}) = 5$ is itself useful. It says that all vectors W in the null space of $H(\bar{V})$ are of the form

$$W = \lambda\bar{V} + \sum_\rho a^\rho E_\rho. \tag{E22}$$

Thus all $V \in \Omega^+(K)$ are of this form. Variations of the scalars a_ρ simply translate the entire diagram, and variations of λ merely change the scaling of the diagram. Thus (E22) tells us that there is essentially only one diagram $\mathfrak{D}$ from the set $[\bar{\mathfrak{D}}]$ that satisfies $K(\mathfrak{D}) = \bar{K}$.

The vectors Q_r satisfy the four conditions $\sum_r Q^\mu_r = 0$, $0 \leq \mu \leq 3$. Thus we may consider the reduced space in which one of the four-vectors Q_r is eliminated. Similarly one of the four-vectors v_r is eliminated by requiring $\sum_r v^\nu_r = 0$. Since the eliminated rows and columns are linear combinations of the remaining ones, the reduced $4(m-1)$-dimensional matrix H still has rank $R = 4m - 5$.

Following the procedure of Goursat[38] one can now construct a function $\Phi(Q)$ of the remaining $(m-1)$ Q's which is real analytic at $Q = \bar{Q} \equiv Q(\bar{K})$, which has a nonvanishing gradient $\nabla\Phi(Q)$ at $Q = \bar{Q}$, and which vanishes on the set

$$\mathcal{R}(\Omega') = \{Q = (Q_1, \cdots, Q_m) \mid Q_r = F_r(V),\ V \in \Omega'\}, \tag{E23}$$

for some neighborhood $\Omega' \subset \Omega^+$ of $\bar{V}$. The construction of $\Phi(Q)$ goes as follows. Since the rank of the reduced H, which we will call $\tilde{H}$, is just one less than the maximum possible rank $4(m-1)$, one may, by virtue of the implicit function theorem, arrange the Q^μ_r and the v^μ_r so that the first $R = 4m - 5$ of the v^μ_r (called x_i's) can be expressed as real analytic functions $\bar{x}_i(X_1, \cdots, X_R, t) = \bar{x}_i(X, t)$ of the first R of the Q_r's (called X_j's) and the final v^μ_r (called t). These expressions $\bar{x}_i(X, t)$ for the v^μ_r's are then inserted into the expression for the final Q^μ_r (called T). This gives

$$T = \tilde{T}(\bar{x}_1(X, t), \cdots, \bar{x}_R(X, t), t) = \bar{T}[X, t]. \tag{E24}$$

Differentiation of (E24) gives

$$\frac{\partial\bar{T}}{\partial t} = \sum_{i=1}^{R} \frac{\partial\tilde{T}}{\partial\bar{x}_i}\frac{\partial\bar{x}_i}{\partial t} + \frac{\partial\tilde{T}}{\partial t}. \tag{E25}$$

Similarly, one has

$$\bar{X}_j(\bar{x}_1(X, t), \cdots, \bar{x}_R(X, t), t) = \bar{X}_j[X, t] \equiv X_j, \tag{E26}$$

[38] E. Goursat, *A Course in Mathematical Analysis* (Ginn and Co., Boston, Mass., 1904), Vol. 1, p. 56. The arguments are a slight extension of those of Goursat.

which upon differentiation gives

$$\frac{\partial \bar{X}_j}{\partial t} = 0 = \sum_{i=1}^{R} \frac{\partial \tilde{X}_j}{\partial \tilde{x}_i} \frac{\partial \tilde{x}_i}{\partial t} + \frac{\partial \tilde{X}_j}{\partial t}, \quad 1 \leq j \leq R. \tag{E27}$$

Equations (E25) and (E27) can be combined and simplified by writing $T = \bar{X}_0$ and $t = \tilde{x}_0$ and by recognizing that the matrix $\partial \tilde{X}_j / \partial x_i$ is just $\tilde{H}_{ji}$:

$$\frac{\partial \bar{X}_j}{\partial t} = \sum_{i=0}^{R} \tilde{H}_{ji} \frac{\partial \tilde{x}_i}{\partial t}, \quad 0 \leq j \leq R. \tag{E28}$$

Multiplication by the matrix C of cofactors of $\tilde{H}$ yields

$$\sum_j C_{ij} \frac{\partial \bar{X}_j}{\partial t} = (\det \tilde{H}) \frac{\partial \tilde{x}_i}{\partial t}. \tag{E29}$$

This equation, when combined with (E27) and the fact that $\det \tilde{H} = 0$, yields

$$C_{00} \frac{\partial \bar{T}}{\partial t} = 0. \tag{E30}$$

But C_{00} is the cofactor (minor) of $\tilde{H}$ that was chosen to be nonzero. There is, therefore, a full neighborhood of the image $(\bar{X}, \bar{t})$ of $\bar{V}$ in which

$$\frac{\partial \bar{T}}{\partial t} = 0. \tag{E31}$$

This implies that $\bar{T}$ is independent of t:

$$T = \bar{T}[X]. \tag{E32}$$

Since the X_j and T are just the Q_r^μ, Eq. (E32) can be rewritten

$$T - \bar{T}[X] \equiv \Phi(Q) = 0. \tag{E33}$$

This defines the real analytic function $\Phi(Q)$. It is evident from (E33) that $\nabla\Phi(Q)$ is nonzero at $\bar{Q}$. The neighborhood Ω' of $\bar{V}$ is chosen small enough that C_{00} is nonzero and $\bar{T}[X]$ is single-valued and holomorphic on the image $\mathcal{R}(\Omega')$ of Ω'.

We now show that there exists a $4m$-dimensional neighborhood $\mathcal{N}(\bar{Q})$ of $\bar{Q}$ such that

$$\mathcal{R}(\Omega^+) \cap \mathcal{N}(\bar{Q}) = \{Q \mid Q \in \mathcal{N}(\bar{Q}), \Phi(Q) = 0\} \cap \mathcal{E}, \tag{E34}$$

where

$$\mathcal{E} = \{Q \mid Q = (Q_1, \cdots, Q_m), \textstyle\sum Q_j = 0\}. \tag{E35}$$

The fact that $\mathcal{R}(\Omega^+)$ is confined to $\mathcal{E}$ follows immediately from (E3) and (E4) by explicit computation. The nontrivial content of (E34) is that, subject to this restriction, the zeros of Φ exactly coincide with $\mathcal{R}(\Omega^+)$ in some neighborhood of $\bar{Q}$.

The construction of the function Φ ensures that it vanishes on $\mathcal{R}(\Omega')$:

$$\mathcal{R}(\Omega') \subset \{Q \mid \Phi(Q) = 0\} \cap \mathcal{E}. \tag{E36}$$

To show (E34) we first show that a neighborhood $\mathcal{N}'(\bar{Q})$ of $\bar{Q}$ can be chosen so that

$$\mathcal{R}(\Omega^+) \cap \mathcal{N}'(\bar{Q}) \subset \mathcal{R}(\Omega'). \tag{E37}$$

Suppose this were not true. Then one could find a sequence of points $Q(\lambda) \to \bar{Q}$ such that, for each value of λ, $Q(\lambda)$ is in $\mathcal{R}(\Omega^+)$ but not in $\mathcal{R}(\Omega')$. Each of these points $Q(\lambda)$ is generated by a corresponding point $V(\lambda) \in \Omega^+$, which can be required to satisfy $\sum v_r(\lambda) = 0$ and $\sum' \|\Delta_j(V(\lambda))\| = 1$. [The value of the mapping F of (E4) is insensitive to such restrictions.] The points $V(\lambda)$ are then confined to a bounded region of V space. For if this were not true, the Euclidean norms of the difference vectors $\Delta_j(V(\lambda))$ would have to be unbounded for some j. This cannot be reconciled with the required boundedness of both their Lorentz norms and the energy components of all the $Q(\lambda)$.

Since the $V(\lambda)$ are confined to a bounded region, the infinite sequence of $V(\lambda)$ must have a subsequence that has a limiting point $V(\infty)$. If this limit point were in Ω^+, then the continuity of $F(V)$ would ensure that the image (under F) of $V(\infty)$ would be $\bar{Q}$. This would require that $V(\infty)$ have the form (E22). The normalization and translation conditions would then ensure that $V(\infty) \equiv \bar{V}$. This is not possible since the $V(\lambda)$ must all lie outside the neighborhood Ω' of $\bar{V}$. Thus $V(\infty)$ cannot be an element of Ω^+.

The only other possibility is that some of the $\|\Delta_j(V(\infty))\|$ are zero. The corresponding vectors $\Delta_j(V(\infty))$ must then also be zero. For if

$$\|\Delta_j(V(\lambda))\| \to 0$$

but $\Delta_j(V(\lambda)) \not\to 0$, then the energy parts of some of the Q_r are forced to become infinite, which contradicts the requirement $Q_r(\lambda) \to \bar{Q}_r$. Thus certain of the vectors $\Delta_j(V(\infty))$ must be zero. Not all can be zero because of the condition $\sum' \|\Delta_j(V(\lambda))\| = 1$. Thus, after appropriate scaling and over-all translation and specification of the individual external momenta incident on each vertex, the diagram corresponding to $V(\infty)$ is a contraction $\mathfrak{D}'$ of $\bar{\mathfrak{D}}$. Equation (E20) again yields a violation of our original hypothesis that $K \in \mathfrak{L}_0^+[\bar{\mathfrak{D}}]$. Thus none of the $\Delta_j(V(\infty))$ can vanish.

All alternatives having been ruled out, Eq. (E37) is established. It then follows from (E36) that there exists a $4m$-dimensional neighborhood $\mathcal{N}''(\bar{Q})$ of $\bar{Q}$ such that

$$\mathcal{R}(\Omega^+) \cap \mathcal{N}''(\bar{Q}) \subset \{Q \mid Q \in \mathcal{N}''(\bar{Q}), \Phi(Q) = 0\} \cap \mathcal{E}. \tag{E38}$$

This result is half of (E34).

To complete the proof of (E34) we construct a $4m$-dimensional neighborhood $\mathcal{N}''(\bar{Q})$ of $\bar{Q}$ such that

$$\{Q \mid Q \in \mathcal{N}''(\bar{Q}),\ \Phi(Q) = 0\} \cap \mathcal{E} \subset \mathcal{R}(\Omega^+) \cap \mathcal{N}''(\bar{Q}). \quad \text{(E39)}$$

Then (E34) is satisfied with $\mathcal{N}(\bar{Q}) = \mathcal{N}'(\bar{Q}) \cap \mathcal{N}''(\bar{Q})$.

To prove (E39), consider the equations

$$x_i = \bar{x}_i(X, t), \quad \text{(E40)}$$

where the functions on the right are those appearing in (E24). Combining (E40) with the condition $\sum v_r = 0$, one obtains a system of equations

$$v_r^\mu = \bar{v}_r^\mu(X, t) \quad \text{(E41)}$$

that gives all the v_r^μ as functions of the X_j $(1 \leq j \leq R)$ and t, where t is just one of the v_r^μ's. Let $\bar{X}$ be the projection of $\bar{Q}$ onto X-space, and let $\bar{t}$ be the value of t such that $\bar{V} \equiv V(\bar{X}, \bar{t})$ is the point of Ω' that satisfies $\bar{Q} = F(\bar{V})$ and $\sum' \Delta_j(\bar{V}) = 1$. Because of the non-singular nature of the mapping (E41) there are neighborhoods $\mathcal{N}_X$ and $\mathcal{N}_t$ of $\bar{X}$ and $\bar{t}$ such that the image [under (E41)] of $\mathcal{N}_X \otimes \mathcal{N}_t$ is contained in Ω'. Moreover, (E26) and (E33) imply that if the projection $X(Q)$ of Q onto X-space belongs to $\mathcal{N}_X$, if $Q \in \mathcal{E}$, and if $\Phi(Q) = 0$, then $Q = F[\bar{V}(X(Q), t)]$ for any $t \in \mathcal{N}_t$. Thus, every point of $\{\Phi(Q) = 0\}$ and $\mathcal{E}$ that satisfies $X(Q) \in \mathcal{N}_X$ is generated by some point V in Ω'. Taking $\mathcal{N}''(\bar{Q})$ to be the set $\{Q \mid X(Q) \in \mathcal{N}_X\}$, we have (E39). Thus (E34) is proved.

The proof of Theorem 6 is completed by transforming the preceding results from Q-space into K-space. Thus one defines

$$\Lambda(K) = \Phi(Q(K)) \quad \text{(E42)}$$

and lets $\mathcal{N}(\bar{K})$ be any K-space neighborhood of $\bar{K}$ with image [under (E2)] in Q-space contained in $\mathcal{N}(\bar{Q})$. Since $\mathcal{S}[\bar{\mathfrak{D}}]$ is the K-space image of $\mathcal{R}(\Omega^+)$, (E34) becomes (5.3). [All points of $\mathfrak{L}^+ \cap \mathcal{N}$ belong to $\mathcal{M}$, and hence also to $\mathcal{W}$ if $\mathcal{N}(\bar{K}) = \mathcal{N}$ is a small enough neighborhood of $\bar{K} \in \mathcal{W}$.]

If K is a point of $\mathfrak{L}^+[\bar{\mathfrak{D}}] \cap \mathcal{N}(\bar{K})$ then the point $Q(K)$ lies in $\mathcal{R}(\Omega^+) \cap \mathcal{N}(\bar{Q})$. Hence, by virtue of (E37), $Q(K)$ lies in $\mathcal{R}(\Omega')$. Thus there is a $V(K)$ in Ω' such that $F[V(K)] = Q(K)$. For all points in Ω' we have $C_{00} \neq 0$. Thus the rank, $R[V(K)]$ of $H(V(K))$ is $4m - 5$. The arguments that gave (E22) show also that any vector in the null space of $H(V(K))$ is of the form

$$W = \lambda V(K) + \sum_\rho a^\rho E_\rho. \quad \text{(E43)}$$

However, the gradient $\nabla\Phi(Q)$ at $Q = Q(K)$ belongs to the null space of $H(V(K))$, as is seen from

$$\frac{\partial\Phi[F(V)]}{\partial v_s^\nu} = \sum_{r\mu} \frac{\partial\Phi(F_{r\mu})}{\partial F_{r\mu}} \frac{\partial F_{r\mu}(V)}{\partial v_s^\nu}, \quad \text{(E44a)}$$

$$= \sum_{r\mu} \frac{\partial\Phi(Q_{r\mu})}{\partial Q_{r\mu}} H_{r\mu,s\nu}(V) \quad \text{(E44b)}$$

$$= \sum_{r\mu} (\nabla\Phi)_r^\mu H_{r\mu,s\nu}(V) \equiv 0. \quad \text{(E44c)}$$

Since $\nabla\Phi$ is nonzero we may rewrite (E43) (using new λ and a^ρ) as $v_r^\mu(K) = \lambda\nabla\Phi[Q(K)]_r^\mu + \sum_\rho a^\rho(E_\rho)_r^\mu$ or, more briefly, as

$$V(K) = \lambda\nabla\Phi + \sum a^\rho E_\rho, \quad \text{(E45)}$$

where the sign of Φ is chosen so that λ is positive.

The positions of the vertices $v_r(K)$ determine the positions of the lines of the corresponding diagram. In particular the position of the external line L_i is generated by the displacement

$$u_i = \lambda(\nabla\Phi)_{r(i)} + \sum a^\rho(E_\rho)_{r(i)}, \quad \text{(E46)}$$

where $r(i)$ labels the vertex to which L_i is connected. The general displacement that generates this position of L_i is obtained by adding an arbitrary translation of this line along itself:

$$u_i^\mu = \lambda(\nabla\Phi)_{r(i)}^\mu + \sum a^\rho(E_\rho)_{r(i)}^\mu + t_i k_i^\mu. \quad \text{(E47)}$$

The E_ρ is independent of r [See (E16).] and can be considered a set of vectors over i, rather than r. Since $Q_{r(i)}$ is a sum of terms containing k_i, we can write

$$\frac{\partial\Phi}{\partial k_{i\mu}} = \sum \frac{\partial\Phi}{\partial Q_{r\nu}} \frac{\partial Q_{r\nu}}{\partial k_{i\mu}} = \frac{\partial\Phi}{\partial Q_{r(i)\mu}} \equiv (\nabla\Phi)_{r(i)}^\mu. \quad \text{(E48)}$$

Substitution of (E48) into (E47) then gives

$$u_i^\mu = \lambda \frac{\partial\Phi[Q(K)]}{\partial k_{i\mu}} + \sum a^\rho \delta_\rho^\mu + t_i k_i^\mu, \quad \text{(E49)}$$

which is just (5.4).

C. Proof of Theorem 8

Let $\bar{V} = V(\bar{\mathfrak{D}})$. It was shown below (E14) that the null space of $H(\bar{V})$ contains $\bar{\Omega}^+(\bar{K})$, the closure of the set $\Omega^+(\bar{K})$. The set $\bar{\Omega}^+(\bar{K})$ contains the vectors $V = V(\mathfrak{D})$ for all diagrams $\mathfrak{D}$ that satisfy $\bar{K} = K(\mathfrak{D})$ and belong either to $[\bar{\mathfrak{D}}]$ or to $[\mathfrak{D}']$ for some $\mathfrak{D}' \subset \bar{\mathfrak{D}}$. Hence, the null space of $H(\bar{V})$ contains all points V that correspond to the diagrams $\mathfrak{D}$ of the theorem.

Let the vectors E_ρ defined by (E16) together with the vectors of the set $\{V_1, \cdots, V_p\}$, where $p = N(\bar{V}) - 4$, be a basis for the null space of $H(\bar{V})$. Thus, any vector W of this null space has a unique representation

$$W = \sum_{i=1}^{p} \lambda_i V_i + \sum_{\rho=0}^{3} a^\rho E_\rho. \quad \text{(E50)}$$

Because of (E10), the vector W must satisfy the equations

$$\Delta_j(W) = \alpha_j(W)\Delta_j(\bar{V}) = \left(\sum_i \lambda_i X_{ij}\right)\Delta_j(\bar{V}), \quad \text{(E51)}$$

where the X_{ij} are defined by

$$\Delta_j(V_i) = X_{ij}\Delta_j(\bar{V}). \quad \text{(E52)}$$

Because $\bar{\mathfrak{D}}$ is connected, the condition $\Delta_j(W) = 0$ for all j implies $\lambda_i = 0$ for all i. This in turn implies the linear independence of the vectors $\mathbf{R}_i = (X_{i1}, X_{i2}, \cdots)$. These vectors $\mathbf{R}_i$ form a basis for the space of vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots)$ appearing in (E51). Through (E51) the vector $\boldsymbol{\alpha}(W)$ specifies W up to an over-all translation $\sum a^\rho E_\rho$.

In terms of $\boldsymbol{\alpha}$ vectors, the set $\bar{\Omega}^+(\bar{K})$ has the following description. For any W in $\bar{\Omega}^+(\bar{K})$ the vector $\boldsymbol{\alpha}(W)$ is a linear combination, $\boldsymbol{\alpha}(W) = \sum \lambda_i(W)\mathbf{R}_i$, of the $\mathbf{R}_i$. The vector W is in $\bar{\Omega}^+(\bar{K})$ if and only if the vectors $\boldsymbol{\lambda}(W) = (\lambda_1(W), \cdots, \lambda_p(W))$ and $\mathbf{C}_j = (X_{1j}, \cdots, X_{pj})$ satisfy $\boldsymbol{\lambda}(W) \cdot \mathbf{C}_j \geq 0$ for all j. (The index j labels the internal lines of $\bar{\mathfrak{D}}$.) From this description it is clear that $\bar{\Omega}^+(\bar{K})$ is convex and starlike [$W \in \bar{\Omega}^+(\bar{K})$ implies $\lambda W \in \bar{\Omega}^+(\bar{K})$ for all $\lambda \geq 0$].

Consider a nonzero vector $\boldsymbol{\alpha}(W)$ corresponding to a point W of $\bar{\Omega}^+(\bar{K})$. If all other points W' of $\bar{\Omega}^+(\bar{K})$ give an $\boldsymbol{\alpha}(W')$ proportional to $\boldsymbol{\alpha}(W)$ then $p = 1$ and the dimension of the null space of $H(\bar{V})$ is five. In this case no contraction $\mathfrak{D}' \subset \mathfrak{D}$ can give point $K(\mathfrak{D}') = \bar{K}$ and Theorem 6 gives the desired result. Alternatively, if there is a $\boldsymbol{\alpha}(W')$ not proportional to $\boldsymbol{\alpha}(W)$, then let P be the plane through the origin that contains both $\boldsymbol{\alpha}(W)$ and $\boldsymbol{\alpha}(W')$. The intersection of P with the image A^+ in $\boldsymbol{\alpha}$-space of $\bar{\Omega}^+(\bar{K})$ is two-dimensional, convex, and starlike. The boundaries of $P \cap A^+$ are therefore two half-lines originating at the origin which, because A^+ contains no vector $\boldsymbol{\alpha}$ with any negative components, must intersect in an angle less than π. Let $\boldsymbol{\alpha}(W_1)$ and $\boldsymbol{\alpha}(W_2)$ be vectors in A^+ that define these two boundary rays. In terms of these vectors, the original vector $\boldsymbol{\alpha}(W)$ has the representation

$$\boldsymbol{\alpha}(W) = y_1\boldsymbol{\alpha}(W_1) + y_2\boldsymbol{\alpha}(W_2), \quad \text{(E53)}$$

where y_1 and y_2 are strictly positive.

Because $\boldsymbol{\alpha}(W_1)$ and $\boldsymbol{\alpha}(W_2)$ lie in the boundary of $P \cap A^+$, the corresponding vectors $\boldsymbol{\lambda}(W_1)$ and $\boldsymbol{\lambda}(W_2)$ must be orthogonal to some of the vectors $\mathbf{C}_i$. The vector $\boldsymbol{\lambda}(W_1)$ is orthogonal to $\mathbf{C}_i$, $i \in I_1$, and the vector $\boldsymbol{\lambda}(W_2)$ is orthogonal to $\mathbf{C}_i$, $i \in I_2$.

There are two alternatives for the vector $\boldsymbol{\alpha}(W_1)$. The first is that $\boldsymbol{\alpha}(W_1)$ and its positive multiples are the only vectors in A^+ for which the corresponding vectors $\boldsymbol{\lambda}$ are orthogonal to the vectors $\mathbf{C}_i$, $i \in I_1$. The second is that there is some second linearly independent vector $\boldsymbol{\alpha}(W_1')$ in A^+ with $\boldsymbol{\lambda}(W_1')$ orthogonal to the vectors $\mathbf{C}_i$, $i \in I_1$.

In the first case the vector W_1 satisfies $W_1 = V(\mathfrak{D}_1)$, where $\bar{K} = K(\mathfrak{D}_1)$ and the diagram $\mathfrak{D}_1$ cannot be further contracted at $\bar{K}$. Thus, the point $\bar{K}$ belongs to $\Gamma_0^+[\mathfrak{D}_1]$, and W_1 must have the form (E45). Equation (E53) allows one to write

$$W = \lambda_1\nabla\Phi + y_2W_2 + \sum a^\rho E_\rho, \quad \text{(E54)}$$

where λ_1 is positive.

In the second case the analysis just performed on $\boldsymbol{\alpha}(W)$ is applied to $\boldsymbol{\alpha}(W_1)$. The plane P_1 corresponding to P contains $\boldsymbol{\alpha}(W_1')$ and $\boldsymbol{\alpha}(W_1)$. The intersection $P_1 \cap A^+$ has boundary rays defined by $\boldsymbol{\alpha}(W_{11})$ and $\boldsymbol{\alpha}(W_{12})$ such that the corresponding vectors $\boldsymbol{\lambda}(W_{11})$ and $\boldsymbol{\lambda}(W_{12})$ are both orthogonal not only to the vectors $\mathbf{C}_i$, $i \in I_1$, but to some additional $\mathbf{C}_i$ as well. In terms of these new vectors $\boldsymbol{\alpha}(W_{11})$ and $\boldsymbol{\alpha}(W_{12})$, the vector $\boldsymbol{\alpha}(W_1)$ can be written

$$\boldsymbol{\alpha}(W_1) = y_{11}\boldsymbol{\alpha}(W_{11}) + y_{12}\boldsymbol{\alpha}(W_{12}), \quad \text{(E55)}$$

where y_{11} and y_{12} are strictly positive.

The entire analysis is then repeated with $\boldsymbol{\alpha}(W_{11})$, $\boldsymbol{\alpha}(W_{12})$, and $\boldsymbol{\alpha}(W_2)$, etc. At each stage at least one new $\mathbf{C}_i$ is added to the previous set of $\mathbf{C}_i$'s. Since the number of $\mathbf{C}_i$'s is finite, the procedure must terminate. At that stage all the vectors into which W is decomposed are associated with diagrams that have no further contractions. Thus we obtain

$$W = \sum \lambda_g\nabla\Phi_g + \sum a^\rho E_\rho, \quad \text{(E56)}$$

where $\lambda_g \geq 0$ and the sum runs over those diagrams $\mathfrak{D}_g \subset \bar{\mathfrak{D}}$ or $\mathfrak{D}_g = \bar{\mathfrak{D}}$ that satisfy $\bar{K} = K(\mathfrak{D}_g)$, but which have no contractions that do.

The arguments following (E42) complete the proof.

D. Proof of Theorem 13

Because of Theorems 10 and 11, it is sufficient to show that the set (5.5) is convex, apart from vectors of the form $U_0(\bar{K})$. In particular, we wish to show that the simultaneous equations

$$\bar{U} = \sum \lambda_g\nabla\Lambda_g(\bar{K}) + U_0(\bar{K}), \quad \lambda_g \geq 0, \quad \text{(E57a)}$$

and

$$-\bar{U} = \sum \lambda_g'\nabla\Lambda_g(\bar{K}) + U_0'(\bar{K}), \quad \lambda_g' \geq 0, \quad \text{(E57b)}$$

imply that $\bar{U} = U_0(\bar{K}) = -U_0'(\bar{K})$. Adding (E57a) to (E57b) we obtain

$$W = \sum_g (\lambda_g + \lambda_g')\nabla\Lambda_g(\bar{K}), \quad \text{(E58)}$$

where $W = -U_0(\bar{K}) - U_0'(\bar{K})$. Define [see (E42)]

$$V = \sum_g (\lambda_g + \lambda_g')\nabla\Phi_g(\bar{Q}). \quad \text{(E59)}$$

This V gives the positions of all the vertices of a diagram with external lines specified by W. Because all the $\mathfrak{D}_g$ are contractions of $\mathfrak{D}$, we have

$$\Delta_j(V) = \sum_g (\lambda_g + \lambda_g')\Delta_j(\nabla\Phi_g) = \sum_g (\lambda_g + \lambda_g')X_{gj}\Delta_j(V(\bar{\mathfrak{D}})). \quad \text{(E60)}$$

Because the λ_g, λ_g', X_{gj}, and $\Delta_j(V(\bar{\mathfrak{D}}))$ are all nonnegative, so are the $\Delta_j(V)$. But the positions of the external lines of V are given by $W = -U_0(\bar{K}) - U_0'(\bar{K})$. Therefore V must be a trivial diagram, since for $\bar{K} \notin \mathcal{M}_0$ no nontrivial connected causal diagram can have its external lines coincident with those of a trivial diagram. But if V is trivial then (E60) implies that $\lambda_g + \lambda_g'$ is zero, for all g. Thus λ_g and λ_g' vanish separately and $U = U_0(\bar{K}) = -U_0'(\bar{K})$.

To complete the proof the $\mathfrak{U} = (U_1, \cdots, U_{3n-4})$ of Theorem 10 is chosen to contain a subset S of the set of vectors $\nabla\Lambda_g(\bar{K})$ such that S together with the $n + 4$ vectors of $U_0(\bar{K})$ are a set of linearly independent vectors that span a space that contains all of the vectors $\nabla\Lambda_g(\bar{K})$. The set of vectors of the form $\sum' \lambda_g\nabla\Lambda_g(\bar{K})$ with $\lambda_g \geq 0$ and $\sum \lambda_g \neq 0$, where $\sum'$ is over S, is a convex set by the argument given above. Then Theorem 11 insures that $\Gamma_c(U; \bar{K})$ is contained in a single set of the form (4.11), and Theorem 10 completes the proof.

APPENDIX F

A. Proof of Theorem 7

The first step of the proof is to show that $\nabla\Lambda_1(\bar{K}) = \lambda\nabla\Lambda_2(\bar{K}) + U_0(\bar{K})$, where λ is real and $U_0(\bar{K})$ is of the form (4.8). The more difficult second step is to show that the number λ must be positive.

Since $\mathfrak{L}_0^+[\mathfrak{D}_2]$ is a submanifold of codimension 1 in $\mathfrak{V}$ at $\bar{K}$, there exists a local coordinate system $(\Delta_c(\bar{K}), \Pi_{\bar{K}}, D_c(\bar{K}))$ at $\bar{K}$ such that $z_1 = \Lambda_2(K)$. [See (E1).] The fact that $\mathfrak{L}_0^+[\mathfrak{D}_1]$ and $\mathfrak{L}_0^+[\mathfrak{D}_2]$ coincide in some neighborhood $\mathcal{N}$ of $\bar{K}$ means that in some neighborhood $N \subset \Pi_{\bar{K}}^{-1}(\mathcal{N} \cap \Delta_c(\bar{K}))$ of $\bar{z} = \Pi_{\bar{K}}^{-1}(\bar{K})$, the function $\bar{\Lambda}_1(z) \equiv \Lambda_1(K(z))$ vanishes whenever $z_1 = 0$. That is, in some neighborhood $N' \subset N$ of $\bar{z}$, the real analytic function $\bar{\Lambda}_1(z)$ has a power series expansion

$$\bar{\Lambda}_1(z) = \sum_{m=1}^{\infty} a_m(z_2, \cdots, z_{3n-4})z_1^m, \quad \text{(F1)}$$

where the a_m are real analytic functions. Explicit computation then shows that

$$\frac{\partial\bar{\Lambda}_1}{\partial z_j}(\bar{z}) = \lambda\frac{\partial\bar{\Lambda}_2}{\partial z_j}(\bar{z}), \quad \text{(F2)}$$

where λ is some real number. Since $\Lambda_1(K)$ and $\Lambda_2(K)$ are real analytic functions of K, we have

$$\frac{\partial\Lambda_g}{\partial z_j} = (\nabla\Lambda_g)\frac{\partial K}{\partial z_j}, \quad (g = 1, 2). \quad \text{(F3)}$$

[See (E1).] Equations (F2) and (F3) combine to yield

$$[\nabla\Lambda_1(\bar{K}) - \lambda\nabla\Lambda_2(\bar{K})]\frac{\partial K}{\partial z_j}(\bar{z}) = 0 \quad \text{(F4)}$$

for all j. Since the only vectors that are annihilated by the matrix $\partial K/\partial z$ have the form (4.8) [see (B11)], Eq. (F4) implies that $\nabla\Lambda_1(\bar{K}) = \lambda\nabla\Lambda_2(\bar{K}) + U_0(\bar{K})$, where λ is real and $U_0(\bar{K})$ has the form (4.8).

We first examine the case where λ is strictly positive; the other case ($\lambda < 0$) will then be easy to rule out.

For each value of g ($g = 1, 2$), Eq. (E42) gives

$$\frac{\partial\bar{\Lambda}_g(z)}{\partial z_\lambda} = \sum_{r\nu i\mu}\frac{\partial\Phi_g(Q[K(z)])}{\partial Q_{r\nu}}\frac{\partial Q_{r\nu}}{\partial k_{i\mu}}\frac{\partial k_{i\mu}}{\partial z_\lambda} \quad \text{(F5a)}$$

$$= \sum_{i\mu}\frac{\partial\Phi_g}{\partial Q_{i\mu}}\frac{\partial k_{i\mu}}{\partial z_\lambda}, \quad \text{(F5b)}$$

where Q_i is the vertex momentum Q_r that depends on k_i. According to (B11) the left side of this equation determines $\partial\Phi_g/\partial Q_{i\mu}$, apart from vectors of the form $U_0[K(z)]$. Then, in view of (E45), $\nabla\bar{\Lambda}_g$ determines the positions of the external vertices of the diagrams $\mathfrak{D}_g$ apart from scalings, over-all translations, and translations of the position of the vertex v_i that is connected to L_i along L_i. The L_i are here considered to be complete lines, not just line segments.

It is useful to introduce diagrams $\mathfrak{D}_1(z)$ and $\mathfrak{D}_2(z)$ that differ from the original diagrams $\mathfrak{D}_1$ and $\mathfrak{D}_2$ by scaling and choice of origin. The fact that $\mathfrak{D}_1$ is a nontrivial connected causal diagram ensures that there is a pair of vertices, v_I and v_F, such that v_I is connected to two initial lines, v_F is connected to two final lines, and v_F is in the positive light cone of v_I. Let the position and scale of $\mathfrak{D}_1(z)$ be fixed by placing v_I at the origin and requiring that $|v_I - v_F| = 1$. According to the results established above, the external lines whose intersection defines v_I in $\mathfrak{D}_1$ must also cross in $\mathfrak{D}_2$, and similarly for v_F. Thus the position and scale of $\mathfrak{D}_2(z)$ can also be fixed by placing v_I at the origin and normalizing so that $|v_I - v_F| = 1$. (In effect, λ is normalized to unity.)

Diagrams constructed according to the rule just described are here called *adjusted* diagrams. The result (E45) is also applicable to them. In particular, Eq. (E45) implies that for a sufficiently small neighborhood $\mathcal{N}(\bar{z})$ of $\bar{z}$ that does not intersect the Landau surfaces for any contractions of $\mathfrak{D}_g$, each point z of $\mathcal{N}(\bar{z}) \subset \{\bar{\Lambda}_g(z) = 0\}$ corresponds to a unique adjusted

diagram $\mathfrak{D}_g(z) \in [\mathfrak{D}_g(\bar{z})]$. This is because all ambiguities of translation and scaling have been removed.

The vertex of $\mathfrak{D}_g(z)$ that is connected to the external line L_i is called $v_{gi}(z)$, and the line parallel to k_i passing through $v_{gi}(z)$ is called $L_{gi}(z)$.

The arguments given above show that for each i, $L_{1i}(z)$ coincides with $L_{2i}(z)$, but they do not show that $v_{1i}(z)$ coincides with $v_{2i}(z)$; these two points could be different points of $L_i(z)$. The main part of the proof consists in showing that the vertices $v_{1i}(z)$ and $v_{2i}(z)$ do in fact coincide if either is connected to two different initial lines (including L_i) or two different final lines (including L_i).

Let L_i be an initial-particle line. If both v_{1i} and v_{2i} are connected to the same additional initial external line L_j $(i \neq j)$, then $v_{1i}(\bar{z})$ and $v_{2i}(\bar{z})$ must coincide. For since $\bar{K}$ does not lie on $\mathcal{M}_0$, the two different initial lines intersect in at most one point. More generally, suppose that v_{1i} is connected to the two initial lines L_i and L_j $(i \neq j)$, and that v_{2i} is connected to the two initial lines L_i and L_k $(i \neq k, j \neq k)$. Then again $v_{1i}(\bar{z})$ and $v_{2i}(\bar{z})$ must coincide. For a small rotation of the two intersecting lines L_i and L_j about the axis $k_i(\bar{z}) + k_j(\bar{z})$ through $v_{1i}(\bar{z})$ gives a nearby point z' of $\{\Lambda_1(z) = 0\}$. This is because the sum $k_i + k_j$ is not changed. The new point z' must belong also to $\{\Lambda_2(z) = 0\}$. Thus there must be a point $v_{2i}(z') = v_{2k}(z')$. But then $L_i(z')$ must intersect the line $L_k(z') = L_k(\bar{z})$. This can be true for several z' near $\bar{z}$ only if $L_k(\bar{z})$ passes through the point $v_{1i}(\bar{z})$. This implies that the point $v_{2i}(\bar{z})$ must coincide with $v_{1i}(\bar{z})$.

We now show that this result $(v_{1i}(\bar{z}) = v_{2i}(\bar{z}))$ also holds provided only that the vertex v_{1i} is contained in two initial lines L_i and L_j $(i \neq j)$. For every z in some neighborhood of $\bar{z}$ the line $L_i(z)$ contains both $v_{1i}(z)$ and $v_{2i}(z)$. The point $v_{1i}(z)$ may or may not be a vertex of $\mathfrak{D}_2(z)$. In either case one can construct a causal diagram $\mathfrak{D}_3(z)$ containing $v_{1i}(z)$ as an external vertex and with external lines coinciding with those of $\mathfrak{D}_2(z)$. One simply regards the part of $L_i(z)$ lying between $v_{1i}(z)$ and $v_{2i}(z)$ as an internal line of $\mathfrak{D}_3(z)$, and similarly for all lines L_k that in diagram $\mathfrak{D}_1$ are connected to v_{1i}. All the conditions for a causal diagram are satisfied by these diagrams $\mathfrak{D}_3(z) \in [\mathfrak{D}_3(\bar{z})]$. Since the external lines $L_i(z)$ of $\mathfrak{D}_3(z)$ are the same as those of $\mathfrak{D}_2(z)$, we see that in some neighborhood N of $\bar{z}$ the surface $\mathfrak{L}^+[\mathfrak{D}_3(\bar{z})]$ contains the surface $\{z \mid \overline{\Lambda}_2(z) = 0\}$.

We now show that $\bar{K}$ belongs to $\mathfrak{L}_0^+[\mathfrak{D}_3(\bar{z})]$. Suppose that this is not true, and that $\bar{K}$ belongs to $\mathfrak{L}^+[\mathfrak{D}_3'(\bar{z})]$, where $\mathfrak{D}_3'(\bar{z}) \subset \mathfrak{D}_3(\bar{z})$. According to the arguments of Appendix E [see (E10)] the internal lines of $\mathfrak{D}_3'(\bar{z})$ must be parallel to the corresponding internal lines of $\mathfrak{D}_3(\bar{z})$. But then there would be a diagram $\mathfrak{D}_2'(\bar{z})$ contained in $\mathfrak{D}_3'(\bar{z})$ that would have the same external lines as $\mathfrak{D}_2(\bar{z})$. This diagram $\mathfrak{D}_2'(\bar{z})$ would be either an element of $[\mathfrak{D}_2(\bar{z})]$ that has the same external lines as $\mathfrak{D}_2(\bar{z})$, or a contraction of such a diagram. The conditions of the theorem ensure that no contraction of $\mathfrak{D}_2(\bar{z})$ has the same external lines as $\mathfrak{D}_2(\bar{z})$. And (E45) shows that the only element of $[\mathfrak{D}_2(\bar{z})]$ that has the same external lines as $\mathfrak{D}_2(\bar{z})$ is $\mathfrak{D}_2(\bar{z})$ itself. This would make $\mathfrak{D}_2'(\bar{z})$ identical to $\mathfrak{D}_2(\bar{z})$. But then the contraction $\mathfrak{D}_3'(\bar{z})$ of $\mathfrak{D}_3(\bar{z})$ would be identical to $\mathfrak{D}_3(\bar{z})$, which is not possible. It follows that $\bar{K}$ belongs to $\mathfrak{L}_0^+[\mathfrak{D}_3(\bar{z})]$.

The surface $\mathfrak{L}_0^+[\mathfrak{D}_3(\bar{z})]$ is a submanifold of $\mathfrak{W}$ of codimension 1 in a neighborhood of $\bar{K}$. In the space of local coordinates z, let $\mathfrak{L}_0^+[\mathfrak{D}_3(\bar{z})]$ be represented by $\{\overline{\Lambda}_3(z) = 0\}$. Since $\mathfrak{L}_0^+[\mathfrak{D}_3(\bar{z})]$ contains $\mathfrak{L}_0^+[\mathfrak{D}_2(\bar{z})]$ in some neighborhood of $\bar{K}$, and since both are submanifolds of $\mathfrak{W}$ of codimension 1, it follows from the arguments leading to (F1) that the two surfaces are identical in some neighborhood of $\bar{K}$.

A rotation of the lines $L_i(z)$ and $L_j(z)$ which intersect at $v_{1i}(z)$ about the axis $k_i(z) + k_j(z)$ takes one to a nearby point on $\{\overline{\Lambda}_1(z) = 0\}$, and hence on $\{\overline{\Lambda}_3(z) = 0\}$. The vertices of the unique corresponding diagram $\mathfrak{D}_3(z)$ must be the same as those of $\mathfrak{D}_3(\bar{z})$, since the positions of the vertices depend only on the $Q[K(z)]$, by virtue of (E45), and these remain unaltered. However, the vertex of $\mathfrak{D}_3(z)$ at $v_{2i}(z)$ will not coincide with the vertex of $\mathfrak{D}_3(\bar{z})$ at $v_{2i}(\bar{z})$ for arbitrary rotations unless $v_{2i}(\bar{z}) = v_{1i}(\bar{z})$. This is the desired result.

Since $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are interchangeable, the above result shows that $v_{1i}(\bar{z})$ and $v_{2i}(\bar{z})$ must coincide if either is connected to two different initial lines.

Similar arguments hold for vertices connected to final lines.

The preceding result is useful in the following way. For any $v_I'(\bar{z})$ that is connected to two initial lines there is a v_F' connected to two final lines that lies in the positive light cone of v_I' or coincides with v_I'. Thus the original v_I and v_F can be chosen to satisfy the additional conditions that there is no v_F' positive timelike relative to v_F, and there is no v_I' negative timelike relative to v_I. Then the total external momentum Q_F at v_F must be positive timelike, and Q_I at v_I must be negative timelike. This is because the internal lines connected to v_F must all terminate at v_F, and the internal lines connected to v_I must all originate at v_I.

The above discussion refers to the case in which the signs of $\nabla\overline{\Lambda}_1$ and $\nabla\overline{\Lambda}_2$ are the same. If these signs were opposite, the external lines and vertices of $\mathfrak{D}_2(\bar{z})$ would be obtained from those of $\mathfrak{D}_1(\bar{z})$ by reflection through the origin. But this clearly cannot give a causal diagram, for the vertex $\hat{v}_F$ $(\equiv -v_F)$ of $\mathfrak{D}_2(\bar{z})$ would have no vertices $\hat{v}_I'$ $(\equiv -v_I')$ of $\mathfrak{D}_2(\bar{z})$ lying

negative timelike to it. No internal lines could terminate on it and Q_F could not be positive timelike, contrary to fact. Thus the gradients $\nabla\Lambda_1$ and $\nabla\Lambda_2$ must have the same sign.

B. Proof of Theorem 10

The first step is to show that if the conditions of Definition 5 are met for any particular set $\bar{\mathfrak{U}} = \{\bar{U}_1, \cdots, \bar{U}_{3n-4}\}$ that defines a simple coordinate system at $\bar{K}$, they are met for all such sets. To see this consider an n-particle displacement U. According to (B10), it has the unique representation

$$U = \sum_{\lambda=1}^{3n-4} t_\lambda \bar{U}_\lambda + U_0(\bar{K}), \tag{F6}$$

where $U_0(\bar{K})$ is linearly independent of the vectors in the set $\bar{\mathfrak{U}}$. Provided $t = (t_1, \cdots, t_{3n-4})$ is not zero, the projection of U onto $\Gamma(\bar{\mathfrak{U}})$ is $\bar{V} = \sum \hat{t}_\lambda \bar{U}_\lambda$, where $\hat{t}_\lambda = \sum t_\lambda |t|^{-1}$. Since (F6) is valid for any displacement U, it is valid in particular for the members of any set $\mathfrak{U} = \{U_1, \cdots, U_{3n-4}\}$ that define a simple coordinate system at $\bar{K}$. For these U_γ, Eq. (F6) becomes

$$U_\gamma = \sum_\lambda t_{\gamma\lambda} \bar{U}_\lambda + U_{0,\gamma}(\bar{K}). \tag{F7}$$

Finally, since any displacement U has a unique representation of the type (F6) with the $\bar{U}_\lambda$ replaced by the U_γ, we have

$$U = \sum_\gamma \beta_\gamma U_\gamma + U_0'(\bar{K}) = \sum_\lambda \left(\sum_\gamma \beta_\gamma t_{\gamma\lambda}\right) \bar{U}_\lambda + U_0(\bar{K}). \tag{F8}$$

That is,

$$t_\lambda = \sum_\gamma \beta_\gamma t_{\gamma\lambda}. \tag{F9}$$

Because both $\mathfrak{U}$ and $\bar{\mathfrak{U}}$ define simple coordinate systems at $\bar{K}$, the matrix M of coefficients $t_{\gamma\lambda}$ is nonsingular. Therefore, the vector t, which defines the projection of U onto $\Gamma(\bar{\mathfrak{U}})$, also uniquely defines the projection of U onto $\Gamma(\mathfrak{U})$. Thus, the sets $\Gamma_c^j(\bar{\mathfrak{U}}; \mathcal{N})$ of Definition 5 are isomorphic to the corresponding sets $\Gamma_c^j(\mathfrak{U}; \mathcal{N})$ for any other choice of the set $\mathfrak{U}$. Moreover, if e is some vector in $\mathbf{R}^{3n-4}$, the Eq. (F9) yields $(t, e) = (\beta, Me)$. Hence, if the projection of U onto $\Gamma(\bar{\mathfrak{U}})$ is in $\Gamma^+(\bar{\mathfrak{U}}, e)$, the projection of U onto $\Gamma(\mathfrak{U})$ is in $\Gamma^+(\mathfrak{U}, Me)$. This proves the statement that if the conditions of Definition 5 are met for any particular set $\bar{\mathfrak{U}} = \{\bar{U}_1, \cdots, \bar{U}_{3n-4}\}$ that defines a simple coordinate system at $\bar{K}$, they are met for all such sets. Next we prove the following lemma.

Lemma 1: For any $\delta > 0$ one can find a product neighborhood $\mathcal{N}$ of any $\bar{K} \in \mathcal{M} - \mathcal{M}_0$ such that

$$\Gamma_c(\mathfrak{U}; \mathcal{N}) \subset \Gamma_c(\mathfrak{U}; \bar{K}, \delta), \tag{F10}$$

where

$$\Gamma_c(\mathfrak{U}; \bar{K}, \delta) = \{U \mid (U + \Delta) \in \mathcal{C}_c(\bar{K}) \cap \Gamma(\mathfrak{U}); \quad \Delta = \sum d_\lambda U_\lambda, [\sum d_\lambda^2]^{\frac{1}{2}} \leq \delta\}. \tag{F11}$$

Proof: To prove this we first express $\Gamma_c(\mathfrak{U}; \mathcal{N})$ in a different way. Let $\hat{V}_\epsilon(k_i'; u_i)$ be the set obtained from $\hat{V}_\epsilon(\psi_i; u_i)$ by replacing supp ψ_i by k_i'. Let $C(K)$ be the set of connected causal diagrams $\mathfrak{D}$ that satisfy $K = K(\mathfrak{D})$, $\sum v_r(\mathfrak{D}) = 0$, and $\sum' \|\Delta_j(V(\mathfrak{D}))\| = 1$. Define

$$\Gamma_c'(\mathfrak{U}; K, K', \epsilon) = \{U' \mid U' = \sum t_\lambda' U_\lambda, \quad v_{r(i)}(\mathfrak{D}) \in \hat{V}_\epsilon(k_i'; u_i' + a), \mathfrak{D} \in C(K)\}, \tag{F12}$$

where $v_{r(i)}(\mathfrak{D})$ is the vertex of $\mathfrak{D}$ that is connected to the external line L_i, and a is a real vector giving an over-all translation. Define

$$\Gamma_c(\mathfrak{U}; K, K', \epsilon) = \{U \mid U \in \Gamma(\mathfrak{U}), \quad \beta U = U' \in \Gamma_c'(\mathfrak{U}; K, K', \epsilon), \beta > 0\}. \tag{F13}$$

Finally, define

$$\Gamma_c(\mathfrak{U}; \mathcal{N}, \epsilon) = \{U \mid U \in \Gamma_c(\mathfrak{U}; K, K', \epsilon), \quad K \text{ and } K' \text{ in } \mathcal{N}\}. \tag{F14}$$

Then for sufficiently small product $\mathcal{N}$ of $\bar{K} \in \mathcal{M} - \mathcal{M}_0$ we have

$$\Gamma_c(\mathfrak{U}; \mathcal{N}) \subset \bigcap_{\epsilon > 0} \Gamma_c(\mathfrak{U}; \mathcal{N}, \epsilon). \tag{F15}$$

To prove (F15), assume it is false. Then for some $\bar{\epsilon} > 0$ there is some U in $\Gamma_c(\mathfrak{U}; \mathcal{N})$ that is not in $\Gamma_c(\mathfrak{U}; \mathcal{N}, \bar{\epsilon})$. Since this U is in $\Gamma_c(\mathfrak{U}; \mathcal{N})$ one can, for each $\epsilon > 0$, find a $\mathfrak{D}_\epsilon$ that satisfies the conditions of Definition 4, with this U and with supp $\psi = \bar{\mathcal{N}}$. Thus a sequence of $\mathfrak{D}_\epsilon$ can be constructed for any sequence $\epsilon_i \to 0$. The norms $N(\mathfrak{D}_{\epsilon_i}) = \sum' \|\Delta_j(V(\mathfrak{D}_{\epsilon_i}))\|$ either approach zero or they do not. If they do not, then the normalized $\bar{\epsilon}_i \equiv \epsilon_i / N(\mathfrak{D}_{\epsilon_i})$ must reach values less than $\bar{\epsilon}$. But then $\Gamma_i(\mathfrak{U}; \mathcal{N}, \bar{\epsilon})$ would contain U, contrary to assumption. Thus the norms $N(\mathfrak{D}_{\epsilon_i})$ must approach zero. This means the diagrams $\mathfrak{D}_{\epsilon_i}$ approach trivial diagrams. But for sufficiently small $\bar{\mathcal{N}} \equiv$ supp ψ about $\bar{K} \in \mathcal{M} - \mathcal{M}_0$ the $U \in \Gamma(\mathfrak{U})$ cannot satisfy the conditions of Definition 4 with any trivial (or nearly trivial) $\mathfrak{D}_\epsilon$ for ϵ smaller than some $\epsilon_1 > 0$, because the various $V_\epsilon(\psi_i; u_i)$ can have no common point (or nearly common point) in these circumstances. This rules out the possibility that the norms $N(\mathfrak{D}_{\epsilon_i})$ approach zero, and hence proves (F15).

Because of (F15) it is sufficient for the completion of the proof of Lemma 1 to prove the following lemma.

Lemma 2: For any $\delta > 0$ one can find a product neighborhood $\mathcal{N}$ of any $\bar{K} \in \mathcal{M} - \mathcal{M}_0$ such that

$$\bigcap_{\epsilon > 0} \Gamma_c(\mathfrak{U}; \mathcal{N}, \epsilon) \subset \Gamma_c(\mathfrak{U}; \bar{K}, \delta). \tag{F10}$$

Proof: To prove Lemma 2, assume it is false. Then there must be some $\delta > 0$ such for any product neighborhood $\mathcal{N}$ of $\bar{K}$ there is some $U(\mathcal{N})$ that belongs to $\Gamma_c(\mathfrak{U}; \mathcal{N}, \epsilon)$ for all $\epsilon > 0$, but does not belong to $\Gamma_c(\mathfrak{U}; \bar{K}, \delta)$. Thus, for any sequence $\{\epsilon_s, \mathcal{N}_s\}$, $s = 1, 2, \cdots$, with $\epsilon_s \to 0$ and $\mathcal{N}_s \to \bar{K}$, there is a sequence of U_s such that

$$U_s \in \Gamma_c(\mathfrak{U}; \mathcal{N}_s, \epsilon_s) \tag{F16a}$$

and

$$U_s \notin \Gamma_c(\mathfrak{U}; \bar{K}, \delta). \tag{F16b}$$

Each U_s satisfying (F16a) corresponds to a $U_s' = U/\beta$ that generates a diagram $\mathfrak{D}_s$, satisfying

$$v_{r(i)}(\mathfrak{D}_s) \in \hat{V}_{\epsilon_s}(k_{is}'; u_{is} + a_s) \tag{F17a}$$

and

$$\mathfrak{D}_s \in C(K_s), \tag{F17b}$$

where $K_s \to \bar{K}$ and $K_s' \to \bar{K}$.

It has been shown elsewhere[19] that the number of different positive-α Landau surfaces that pass through any bounded region is finite. The infinite sequence $\mathfrak{D}_s$ must therefore be divided between a finite number of classes $[\mathfrak{D}]$, at least one of which must have an infinite number of the diagrams $\mathfrak{D}_s$. Let this class be denoted by $[\mathfrak{D}_1]$, and let the $\mathfrak{D}_s$ not in $[\mathfrak{D}_1]$ be disregarded. The sets $V(\mathfrak{D}_s)$ are confined to a bounded region and must have at least one accumulation point $\bar{V} = V(\bar{\mathfrak{D}})$. The argument for this was given in Appendix E below (E37). The arguments of Appendix E [see (E20)] also show that $\bar{K} = K(\bar{\mathfrak{D}})$.

If we can show that the sequence $\{U_s\}$ has an accumulation point $\bar{U}$ in $\mathcal{C}_c(\bar{K}) \cap \Gamma(\mathfrak{U})$, we shall have established a contradiction with (F16b), and shall therefore have proved Lemma 2.

Each displacement U_s corresponds to a unique displacement U_s' in $\Gamma_c'(\mathfrak{U}; K_s, K_s', \epsilon_s)$. The points U_s' have a unique limit point $\bar{U}'$ defined by the condition

$$\bar{U}' \in \{U' \mid U' = \sum t_\lambda' U_\lambda, v_{r(i)} \in \hat{V}_0(\bar{k}_i; u_i' + a)\}. \tag{F18}$$

The fact that the $\bar{U}'$ defined by (F18) is unique follows from (B10), since the various U' that satisfy the second condition in (F18) differ by vectors of the form $U_0(\bar{K})$. The U_s' of (F16a) satisfy, according to (F14) and (F12), the condition

$$U_s' \in \{U' \mid U' = \sum t_\lambda' U_\lambda, \quad v_{r(i)}^s \equiv v_{r(i)}(\mathfrak{D}_s) \in \hat{V}_\epsilon(k_i'; u_i + a), k' \in \bar{\mathcal{N}}_s\}. \tag{F19}$$

The continuity properties of the set on the right of (F19) ensure that the U_s' approach the $\bar{U}'$ of (F18).

If the $\bar{U}'$ is nonzero, then the $U_s = U_s'/\beta_s$ must approach the limit $\bar{U} = \bar{U}'/\bar{\beta}$ where $\bar{\beta}^2 = \sum t_\lambda^2(\bar{U}')$ is nonzero. This $\bar{U}$ would lie in $\mathcal{C}_c(\bar{K}) \cap \Gamma(U)$, thus contradicting (F16b). Thus the proof will be completed by showing that $\bar{U}'$ is nonzero.

To see that the vector $\bar{U}'$ is different from zero notice first that, because $\bar{\mathfrak{D}}$ is nontrivial, the earliest vertex $\bar{v}_I$ must be definitely earlier than the latest vertex $\bar{v}_F$. By virtue of the stability requirement, the initial vertex $\bar{v}_I$ must be connected to at least two initial lines and the final vertex $\bar{v}_F$ must be connected to two final lines. Because $\bar{K}$ does not belong to $\mathcal{M}_0$, the initial lines connected to $\bar{v}_I$ meet only at $\bar{v}_I$ and the final lines connected to $\bar{v}_F$ meet only at $\bar{v}_F$. Such a configuration does not allow the $\bar{U}'$ of (F18) to be zero, since $\bar{U}' = 0$ means that all the external lines pass through a common point [see (B10)]. This completes the proof of Lemma 2 and, by virtue of (F15), the proof of Lemma 1.

Given Lemma 1 and the result proved just before it, the proof of Theorem 10 is trivial.

C. Proof of Theorem 11

Let $\tilde{\mathcal{C}}_c(\bar{K})$ be the set of U that generate $\mathfrak{D}$ that satisfy $K(\mathfrak{D}) = \bar{K}$. What must be shown is that for each $\bar{K} \in \mathcal{M} - \mathcal{M}_0$, we have

$$\mathcal{C}_c(\bar{K}) = \tilde{\mathcal{C}}_c(\bar{K}). \tag{F20}$$

It is obvious that $\tilde{\mathcal{C}}_c(\bar{K}) \subset \mathcal{C}_c(\bar{K})$ and that $\mathcal{C}_0(\bar{K}) \subset \tilde{\mathcal{C}}_c(\bar{K})$. What must be shown is that for each $\bar{K} \in \mathcal{M} - \mathcal{M}_0$, we have

$$\mathcal{C}_c(\bar{K}) - \mathcal{C}_0(\bar{K}) \subset \tilde{\mathcal{C}}_c(\bar{K}). \tag{F21}$$

To prove this, first define

$$\mathcal{C}_c'(\bar{K}, \epsilon) = \{U' \mid V_{r(i)}(\mathfrak{D}) \in \bar{\hat{V}}_\epsilon(\bar{R}_i; u_i' + a), \mathfrak{D} \in C(K)\},$$

where $C(K)$ is defined above (F12), and define

$$\mathcal{C}_c(\bar{K}, \epsilon) = \{U \mid \beta U = U' \in \mathcal{C}_c'(\bar{K}, \epsilon), \beta > 0\}. \tag{F22}$$

Then, for $\bar{K} \in \mathcal{M} - \mathcal{M}_0$, we have

$$\mathcal{C}_c(\bar{K}) - \mathcal{C}_c(\bar{K}) \subset \bigcap_{\epsilon > 0} \mathcal{C}_c(\bar{K}, \epsilon). \tag{F23}$$

The proof of (F23) is the same as the proof of (F15), except for the obvious substitutions. It remains only to show that

$$\bigcap_{\epsilon > 0} \mathcal{C}_c(\bar{K}, \epsilon) \subset \tilde{\mathcal{C}}_c(\bar{K}). \tag{F24}$$

The proof of this is similar to the proof of Lemma 2. If (F24) were not true, then there would be some $U \notin \tilde{\mathcal{C}}_c(\bar{K})$ that belongs to each $\mathcal{C}_c(\bar{K}, \epsilon)$ on the left. Thus for each $\epsilon > 0$ there would be a $\mathfrak{D}_\epsilon \in C(\bar{K})$ such that the conditions of Def. 4 can be satisfied with this U, and with supp ψ replaced by $\bar{K}$. A sequence $\epsilon_s \to 0$ gives then a corresponding sequence $\mathfrak{D}_s \in C(\bar{K})$. As in Lemma 1, the $V(\mathfrak{D}_s)$ must accumulate at a $\bar{V}$ that corresponds to a $\bar{\mathfrak{D}}$ that satisfies $K(\bar{\mathfrak{D}}) = \bar{K}$. But then U would belong to $\tilde{\mathcal{C}}_c(\bar{K})$. This contradiction proves (F24), and hence also the theorem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Application of the Kirillov Theory to the Representations of $O(2, 1)$*

S. A. DUNNE
Physics Department, Imperial College, London

(Received 12 October 1966)

The Kirillov construction is applied to the semisimple Lie group $O(2, 1)$. All the unitary irreducible representations (except the supplementary series) are found, provided that complex subalgebras and complex points on orbits are admitted. The characters of the representations are calculated and the relation of their Fourier transforms to the orbits is examined.

INTRODUCTION

It was shown by Kirillov[1] that a simple construction gives all the unitary continuous irreducible representations of any nilpotent Lie group. This construction, which is described in Sec. I, correlates these representations with the orbits of the coadjoint representation of the group, specifying a certain one-dimensional representation of a particular subgroup from which each irreducible representation is induced.

It is of interest to know whether Kirillov's construction gives all the irreducible unitary representations of other groups. Bernat[2] has shown that it does for exponential solvable Lie groups. For other classes of groups, including semisimple Lie groups, it seems to give all the representations, provided that the method is extended by complexifying the groups suitably. Of course, one then begins to lose the neat geometrical interpretation given by Kirillov.

The complexified Kirillov construction was applied by Streater[3] to the harmonic-oscillator group (a nonexponential solvable Lie group), yielding all the irreducible unitary representations. In this paper we apply it to the semisimple groups $O(3)$ and $O(2, 1)$, the latter being of interest because it is the smallest semisimple, noncompact Lie group. The problem of complexification has been studied by Kostant.[4]

Kirillov also showed that, for nilpotent groups, the characters of the irreducible representations are Fourier transforms of delta functions on the corresponding orbits. We investigate the relationship for $O(2, 1)$ in Sec. V in order to see how it is modified for semisimple groups.

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research OAR through the European Office Aerospace Research U.S. Air Force.

[1] A. A. Kirillov, Usp. Mat. Nauk. **106**, 57 (1962) [Russ. Math. Surv. **17**, 53 (1962)].

[2] P. Bernat, Doctoral thesis, Ann. Sci. Ecole Normale Superieure, 1965, Fasc 1, p. 37.

[3] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).

[4] B. Kostant (private communication).

I. KIRILLOV THEORY

Here we outline the Kirillov theory for any Lie group.

Let $\mathbf{G}$ be the real Lie algebra of the Lie group G. Then $\mathbf{G}$ is an n-dimensional real vector space. We define the dual space $\mathbf{G}'$ as the space of real linear functionals over $\mathbf{G}$; it is also an n-dimensional vector space. We write the functional as $\mathbf{p} \cdot \mathbf{x}$, where $\mathbf{x} \in \mathbf{G}$, $\mathbf{p} \in \mathbf{G}'$.

We consider the adjoint representation of the group G in the space $\mathbf{G}$:

$$g \to \rho(g).$$

The co-adjoint representation $\rho'(g)$ of G in $\mathbf{G}'$ is then defined by

$$[\rho'(g)\mathbf{p}] \cdot [\rho(g)\mathbf{x}] = \mathbf{p} \cdot \mathbf{x}.$$

Then, the orbit containing the point $\mathbf{p} \in \mathbf{G}'$ is the set of points $\{\rho'(g)\mathbf{p}\}$, where g runs through the Lie group G. The orbits are disjoint subsets which fill $\mathbf{G}'$.

Given a point $\mathbf{p} \in \mathbf{G}'$, a Lie subalgebra $\mathbf{H} \in \mathbf{G}$ is said to be subordinate to $\mathbf{p}$ if

$$\mathbf{p} \cdot [\mathbf{h}_1, \mathbf{h}_2] = 0 \quad \text{for all} \quad \mathbf{h}_1, \mathbf{h}_2 \in \mathbf{H}.$$

Then $(i\mathbf{p} \cdot \mathbf{h})$ and $\exp i\mathbf{p} \cdot \mathbf{h}$ are one-dimensional representations of $\mathbf{H}$ and $\exp \mathbf{H} \subset G$, respectively. $\mathbf{H}$ is said to be maximally subordinate to $\mathbf{p}$ if there is no other subalgebra of larger dimension subordinate to $\mathbf{p}$. The Kirillov construction takes one point $\mathbf{p}$ on each orbit, and for each such point takes the one-dimensional representation $U(\exp \mathbf{h}) = \exp i\mathbf{p} \cdot \mathbf{h}$ of $\exp \mathbf{H}$, where $\mathbf{H}$ is a subalgebra maximally subordinate to $\mathbf{p}$. The corresponding representation of G is taken to be the representation induced by $U(\exp \mathbf{h})$. Thus each orbit yields one representation of G.

We note here that the representation T of a group G, induced by the representation $U(g)$ of a subgroup H, is defined on functions (in general, vector functions) on the group G which satisfy

$$F(hg) = U(h)F(g) \quad \text{for} \quad h \in H, g \in G.$$

Then the induced representation is

$$[T(g_2)F](g_1) = F(g_1 g_2).$$

An induced representation can be put in the form of a multiplier representation on the space of functions on the coset space G/H:

$$[T(g)f](z) = U[g(z)] \cdot f(zg),$$

where $g(z) \in H$.

Then, if $d\mu(z)$ is an invariant measure on G/H and $U(h)$ is unitary, T is unitary with the inner product

$$\int_{G/H} \overline{F_1}(g)F_2(g)\, d\mu(z), \quad z = Hg.$$

If G is a nilpotent Lie group, the above construction yields each inequivalent irreducible unitary representations once. Moreover, if $\mathbf{p}_1$ and $\mathbf{p}_2$ are points on the same orbit and $\mathbf{H}_1$, $\mathbf{H}_2$ are corresponding maximally subordinate subalgebras, the corresponding induced representations are equivalent. Also, let $C(e^x)$ be the character of the induced representation T (calculated, if necessary, as a distribution on the group). Then,

$$\int C(e^x)e^{i\mathbf{p}\cdot\mathbf{x}}\, d\mathbf{x} = \delta[\psi(\mathbf{p})],$$

where $\psi(\mathbf{p}) = 0$ is the equation of the corresponding orbit, $d\mathbf{x}$ is the Euclidean measure on $\mathbf{G}'$ (invariant for nilpotent groups), and δ is the Dirac delta function.

II. UNITARY IRREDUCIBLE REPRESENTATIONS OF $O(3)$

In this section we illustrate the Kirillov theory by applying it to $O(3)$, obtaining the usual $(2j + 1)$-dimensional representations (j = integer). This will be helpful for comparison with $O(2, 1)$.

The elements of $O(3)$ and of its covering group $SU(2)$ can be written as $g \equiv \exp(\mathbf{x} \cdot \mathbf{J})$, where the anti-Hermitian infinitesimal generators $\mathbf{J} \equiv (J_1, J_2, J_3)$ satisfy the relations

$$[J_2, J_3] = J_1, \quad [J_3, J_1] = J_2, \quad [J_1, J_2] = J_3. \quad (1)$$

The group multiplication law is most easily written down by using the fundamental representation of $SU(2)$, in which $\mathbf{J} = -\frac{1}{2}i\boldsymbol{\sigma}$ and

$$g \equiv \exp(\mathbf{x} \cdot \mathbf{J}) = \cos(\tfrac{1}{2}r) \cdot I - i \sin(\tfrac{1}{2}r)\mathbf{x} \cdot \boldsymbol{\sigma}/r.$$

Here $r^2 = x^2 + y^2 + z^2$, where

$$0 \leq r \leq \pi, \quad \text{for } O(3), \quad g \equiv -g,$$
$$0 \leq r \leq 2\pi, \quad \text{for } SU(2).$$

Thus we can write

$$g \equiv (\alpha, \alpha^*, \beta, \beta^*), \quad \alpha\alpha^* + \beta\beta^* = 1, \quad (2)$$

with the group composition law

$$g_1g_2 = (\beta_1^*\alpha_2 + \alpha_1\beta_2, \beta_1\alpha_2^* + \alpha_1^*\beta_2^* - \alpha_1^*\alpha_2 + \beta_1\beta_2, -\alpha_1\alpha_2^* + \beta_1^*\beta_2^*). \quad (3)$$

These parameters are related to the Euclidean coordinates in the Lie algebra space by

$$\exp(-\tfrac{1}{2}i\mathbf{x} \cdot \boldsymbol{\sigma}) = \begin{pmatrix} \beta^* & -i\alpha \\ -i\alpha^* & \beta \end{pmatrix},$$

i.e.,

$$\alpha = [(x - iy)/r] \sin(\tfrac{1}{2}r),$$
$$\beta = \cos(\tfrac{1}{2}r) + i(z/r) \sin(\tfrac{1}{2}r),$$
$$\alpha^* = [(x + iy)/r] \sin(\tfrac{1}{2}r),$$
$$\beta^* = \cos(\tfrac{1}{2}r) - i(z/r) \sin(\tfrac{1}{2}r).$$

We shall need these formulas for complex values of x, y, z also. The asterisk is equivalent to complex conjugation for real x, y, z only, and in that case we write $g \equiv (\alpha, \beta)$. For $O(3)$, we identify the elements $(\alpha, \alpha^*, \beta, \beta^*)$ and $(-\alpha, -\alpha^*, -\beta, -\beta^*)$.

The adjoint and co-adjoint representations are realized as rotations of three-dimensional Euclidean spaces, and thus the orbits are spheres $p_1^2 + p_2^2 + p_3^2 =$ const in the dual space.

We first carry out the Kirillov construction on the real group $O(3)$ without complexifying. We take $(0, 0, a)$, $a \geq 0$, as a representative point on the orbit $p_1^2 + p_2^2 + p_3^2 = a^2$. The only real subordinate subalgebras are one-dimensional; we choose the one generated by J_3.

Thus we have to find the representation of $O(3)$ induced by the representation

$$U(h) \equiv U(e^{zJ_3}) = e^{iaz}$$

of the subgroup $O(2)$.

We note immediately that only orbits of integer radius will yield one-valued representations of $O(3)$ [half-integer for $SU(2)$].

To find the representation we put

$$F(h_1g_2) = U(h_1)F(g_2),$$

for

$$h_1 = (0, e^{\frac{1}{2}az}), \quad g_2 = (\alpha_2, \beta_2).$$

Therefore,

$$F(e^{-\frac{1}{2}iz}\alpha_2, e^{\frac{1}{2}iz}\beta_2) = e^{iza}F(\alpha_2, \beta_2).$$

Take

$$e^{\frac{1}{2}iz} = \alpha_2/|\alpha_2|.$$

Then

$$F(\alpha_2, \beta_2) = (\alpha_2^*/|\alpha_2|)^{2a}F(|\alpha_2|, \alpha_2\beta_2/|\alpha_2|)$$

and, therefore,

$$F(\alpha, \beta) = (\alpha^*/|\alpha|)^{2a}f(-\beta/\alpha^*),$$

using (2). The action of the group $[T_a(g_2)F](g_1) = F(g_1g_2)$ reduces to

$$[T_a(\alpha, \beta)f](w) = \left(\frac{\beta^* - \alpha^*w}{|\beta^* - \alpha^*w|}\right)^{2a} f\left(\frac{\alpha + \beta w}{\beta^* - \alpha^*w}\right), \quad (4)$$

where f is a function on the complex plane, since $w = -\beta_1/\alpha_1^*$.

Make the transformation

$$\tilde{f}(w) = Uf(w) = (1 + |w|^2)^a f(w), \tag{5}$$

$$\tilde{T} = UTU^{-1}. \tag{6}$$

Then,

$$(\tilde{T}_a\tilde{f})(w) = (\beta^* - \alpha^* w)^{2a}\tilde{f}(w') \tag{7}$$

where

$$w' = (\alpha + \beta w)/(\beta^* - \alpha^* w)$$

The space of analytic functions of w is obviously an invariant subspace of $\tilde{T}$, but *a priori* there is no reason for $\tilde{f}(w)$ to be analytic. Therefore, the induced representation is reducible.

In order to obtain the irreducible representations by the Kirillov method, we must complexify the parameters x, y, z, and require either (i) that the representation in the form $\tilde{T}$ be analytic in these parameters or (ii) that we use functions analytic in x, y, z from the start. In the latter case, complex subalgebras are allowed. Method (ii) is more consistent and yields the same result as (i), as we now show.

When x, y, z are complexified, α, α^*, β, β^* are complex variables subject only to Eq. (2). There is in this case a two-dimensional (complex) subalgebra, e.g., $\{J_3, J_1 + iJ_2\}$, subordinate to $(0, 0, a)$.

We seek functions $F(\alpha, \alpha^*, \beta, \beta^*)$, analytic in x, y, z, with $\alpha\alpha^* + \beta\beta^* = 1$, satisfying

$$F(h_1g_2) = U(h_1)F(g_2)$$

for

$$h_1 = (\alpha_1, 0, \beta_1, \beta_1^{-1}), \quad g_2 = (\alpha_2, \alpha_2^*, \beta_2, \beta_2^*),$$

and $U(h_1) = e^{iaz}$, where $e^{\frac{1}{2}iz} = \beta_1$. Therefore,

$$F(\beta_1^{-1}\alpha_2 + \alpha_1\beta_2, \beta_1\alpha_2^*, \beta_1\beta_2, -\alpha_1\alpha_2^* + \beta_1^{-1}\beta_2) = (\beta_1)^{2a}F(\alpha_2, \alpha_2^*, \beta_2, \beta_2^*).$$

Take $\alpha_1 = \beta_2^*$, $\beta_1^{-1} = \alpha_2^*$. Therefore,

$$F(\alpha_2, \alpha_2^*, \beta_2, \beta_2^*) = (\alpha_2^*)^{2a}F(1, 1, \beta_2/\alpha_2^*, 0) = (\alpha_2^*)^{2a}\tilde{f}(-\beta_2/\alpha_2^*), \quad \text{say.}$$

The action of the group reduces directly to

$$(\tilde{T}_a\tilde{f})(w) = (\beta^* - \alpha^* w)^{2a}\tilde{f}(w'). \tag{7}$$

$f(w)$ is analytic in the complex plane, since F, α^*, and $-\beta/\alpha^*$ are analytic functions of x, y, z. Since

$$\frac{d^2w'}{d^2w} = \left|\frac{dw'}{dw}\right|^2 = \frac{1}{|\beta^* - \alpha^* w|^4} = \frac{(1 + |w'|^2)^2}{(1 + |w|^2)^2},$$

the invariant measure on the complex plane is

$$\frac{d^2w}{(1 + |w|^2)^2}, \quad \text{for } x, y, z \text{ real},$$

$$d^2w = du\,dv, \quad \text{where} \quad w = u + iv.$$

Therefore, the unitary norm is

$$\|f\|^2 = \iint \frac{|f(w)|^2\,d^2w}{(1 + |w|^2)^2}.$$

Therefore,

$$\|\tilde{f}\|^2 = \iint \frac{|\tilde{f}(w)|^2\,d^2w}{(1 + |w|^2)^{2a+2}}. \tag{8}$$

The only analytic functions, for which this norm is finite, are polynomials of (at most) degree $2a$. $\tilde{T}$ is irreducible on these functions, since it transforms w^n into $(\beta^* - \alpha^* w)^{2a-n}(\alpha + \beta w)^n$, $0 \le n \le 2a$, which is a linear combination of all such functions w^n.

Thus, by using analytic functions of the complexified parameters and demanding unitary representations of the real group, we have obtained irreducible representations, which are, of course, the usual $(2a + 1)$-dimensional ones D^a (a = half-odd-integer gives the double-valued representations).

The generators, eigenfunctions, and matrix elements can easily be written down from (7) and (8):

$$J_1 = -aw + \tfrac{1}{2}(1 + w^2)\,d/dw,$$

$$J_2 = -iaw - \tfrac{1}{2}i(1 - w^2)\,d/dw,$$

$$J_3 = -ia + iwd/dw.$$

Hence, $\mathbf{J}^2 = -a(a + 1)$. (N.B. These are the anti-Hermitian generators, $-i$ times the usual angular-momentum generators.)

The eigenfunctions of J_3 are

$$f_{a,m}(w) = N_{a,m}w^{a-m}, \quad -a \le m \le a, \tag{9}$$

where

$$N_{a,m}^{-2} = \iint \frac{|w|^{2a-2m}d^2w}{(1 + |w|^2)^{2a+2}} = \frac{\pi}{2a + 1}\binom{2a}{a - m}^{-1}. \tag{10}$$

The corresponding eigenvalues are $-im$.

In this basis, the matrix element of a general rotation (α, β) is

$$(f_n, Tf_m) = N_{a,n} \cdot N_{a,m} \times \iint w^{*a-n}\frac{(\beta^* - \alpha^* w)^{a+m}(\alpha + \beta w)^{a-m}d^2w}{(1 + |w|^2)^{2a+2}} = \frac{N_{a,m}}{N_{a,n}} \times \text{coefficient of } w^{a-n} \text{ in } (\beta^* - \alpha^* w)^{a+m}(\alpha + \beta w)^{a-m}.$$

We may note that the action of the transformation on the complex plane is the same as the representation of rotations of a sphere by stereographic projection onto a touching plane.

The fixed points w_1, w_2 of

$$w \to w' = (\alpha + \beta w)/(\beta^* - \alpha^* w)$$

correspond to the projections of the poles of the axis of rotation. Using these to characterize a rotation, we can write down the (unnormalized) eigenfunctions of a general rotation as

$$(w - w_1)^{a+m}(w - w_2)^{a-m}, \quad -a \leq m \leq a.$$

We now consider the characters of the group and their Fourier transforms. It has been remarked by Kirillov that, unlike nilpotent groups, the Euclidean measure on the Lie algebra is not the invariant measure on the group for semisimple groups. In the case of $O(3)$, the invariant group measure is

$$dg = \left(\frac{\sin \frac{1}{2}r}{\frac{1}{2}r}\right) dx\, dy\, dz, \quad r^2 = x^2 + y^2 + z^2.$$

Therefore, in order to obtain a covariant function on the dual space, we must first multiply the character by

$$\frac{\sin \frac{1}{2}r}{\frac{1}{2}r}$$

before taking the Fourier transform with respect to the Euclidean measure.

Now $C_a(g) = \sin (a + \frac{1}{2})r/\sin \frac{1}{2}r$. Therefore we need to evaluate

$$\iiint \frac{\sin (a + \frac{1}{2})r}{\sin \frac{1}{2}r} \cdot \frac{\sin \frac{1}{2}r}{\frac{1}{2}r} \cdot e^{-ip\cdot x} d^3\mathbf{x}.$$

This gives

$$\tilde{C}_a(g) = 8\pi^2 \delta[p_1^2 + p_2^2 + p_3^2 - (a + \tfrac{1}{2})^2]. \quad (11)$$

Therefore the (modified) Fourier transforms of the characters are delta functions on the orbits. However, we do not arrive back at the orbit we started from, but at one of radius increased by $\frac{1}{2}$. The latter orbit leads to $D^{a+\frac{1}{2}}$ [double-valued for $O(3)$, single-valued for $SU(2)$]. We shall find something similar for $O(2, 1)$.

III. IRREDUCIBLE UNITARY REPRESENTATIONS OF $O(2, 1)$

In this section we derive the unitary irreducible representations of $O(2, 1)$ by a modification of the Kirillov method, and we obtain the representations which were first derived by Bargmann.[5,6]

The group elements can be written as

$$g = e^{x\cdot J} = (\alpha, \alpha^*, \beta, \beta^*), \quad \beta\beta^* - \alpha\alpha^* = 1, \quad (12)$$

where

$$[J_1, J_2] = -J_3, \quad [J_2, J_3] = J_1, \quad [J_3, J_1] = J_2, \quad (13)$$

$$\alpha = [(x - iy)/\rho] \sin (\tfrac{1}{2}\rho),$$

$$\beta = \cos (\tfrac{1}{2}\rho) + i(z/\rho) \sin (\tfrac{1}{2}\rho),$$

$$\alpha^* = [(x + iy)/\rho] \sin (\tfrac{1}{2}\rho),$$

$$\beta^* = \cos (\tfrac{1}{2}\rho) - (iz/\rho) \sin (\tfrac{1}{2}\rho),$$

$$\rho^2 = z^2 - x^2 - y^2, \quad \rho \text{ real or imaginary},$$

$$(\alpha, \alpha^*, \beta, \beta^*) \equiv (-\alpha, -\alpha^*, -\beta, -\beta^*), \quad \text{for } O(2, 1).$$

(α, β) gives the spinor group, with fundamental (non-unitary) representation

$$\begin{pmatrix} \beta^* & \alpha \\ \alpha^* & \beta \end{pmatrix}, \quad |\beta|^2 - |\alpha|^2 = 1.$$

Hence, the multiplication law is

$$\alpha_{12} = \beta_1^* \alpha_2 + \alpha_1 \beta_2, \quad \beta_{12} = \alpha_1^* \alpha_2 + \beta_1 \beta_2, \quad \text{etc.} \quad (14)$$

$O(2, 1)$ can be obtained by analytic continuation from $O(3)$, taking

$$x \to -ix, \quad y \to -iy, \quad z \to z,$$

$$\alpha \to -i\alpha, \quad \alpha^* \to -i\alpha^*, \quad \beta \to \beta, \quad \beta^* \to \beta^*,$$

$$J_1 \to iJ_1, \quad J_2 \to iJ_2, \quad J_3 \to J_3.$$

The real group splits into the sectors

$$0 \leq \rho \leq \pi, \quad \text{denoted I},$$

$$0 \leq i\rho < \infty, \quad \text{denoted II}.$$

For x, y, z real, the asterisk denotes complex conjugation.

The adjoint and co-adjoint representations are transformations of a pseudo-Euclidean $(2 + 1)$ space, and therefore the orbits are hyperboloids in the dual space:

$$p_1^2 + p_2^2 - p_3^2 = \text{const.}$$

These are of two kinds:

hyperboloids of one sheet,

$$p_1^2 + p_2^2 - p_3^2 = a^2 \geq 0;$$

hyperboloids of two sheets,

$$p_1^2 + p_2^2 - p_3^2 = -a^2 \leq 0$$

(each sheet of the latter is a distinct orbit).

Consider first orbits of the second kind. We complexify the algebra for the same reasons as for $O(3)$ in the last section and use functions analytic in x, y, z. The deduction of the representation runs parallel to that for $O(3)$.

We induce from the representation $U(h_1) = e^{iaz}$ of the subgroup generated by $\{J_1 - iJ_2, J_3\}$, a maximally

[5] V. Bargmann, Ann. Math. **48**, 568 (1947).

[6] A. Barut and C. Fronsdal, Proc. Roy. Soc. (London) **A287**, 532 (1965).

subordinate complex subalgebra to $(0, 0, a)$ on the orbit $p_1^2 + p_2^2 - p_3^2 = -a^2$.

The required functions satisfy

$$F(\beta_1^*\alpha_2, \beta_1\alpha_2^* + \alpha_1^*\beta_2^*, \alpha_1^*\alpha_2 + \beta_1\beta_2, \beta_1^*\beta_2^*) = (\beta_1)^{2a}F(\alpha_2, \beta_2), \quad \alpha_1 = 0.$$

Take $\beta_1 = \beta_2^*$, $\alpha_1^* = -\alpha_2^*$; then,

$$F(\alpha_2, \beta_2) = (\beta_2^*)^{-2a}F(\alpha_2/\beta_2^*, 0, 1, 1) = (\beta_2^*)^{-2a}f(\alpha_2/\beta_2^*), \quad \text{say.}$$

Note that $|\alpha_2/\beta_2^*| < 1$ for real z, y, z.

The action of the group reduces to

$$(T_a f)(w) = (\beta^* + \alpha^* w)^{-2a} f(w'), \tag{15}$$

$$w' = (\alpha + \beta w)/(\beta^* + \alpha^* w), \tag{16}$$

where f is analytic in $|w| < 1$. This could have been obtained by analytic continuation from (7) for $O(3)$ (with $a \to -a$ and $w \to -iw$, $w^* \to -iw^*$ also).

Unlike $O(3)$, $(0, 0, -a)$ yields an inequivalent representation; it is not on the same orbit.

For $a < 0$, make the transformation from (15):

$$g(w) = Uf(w) = (1 - |w|^2)^{2a} f^*(w), \tag{17}$$

$$\tilde{T} = UTU^{-1}. \tag{18}$$

Then,

$$(\tilde{T}_a g)(w) = (\beta^* + \alpha^* w)^{2a} g(w'). \tag{19}$$

This is of the form (15) for $-a$, but g is the conjugate of an analytic function. This conjugate representation is not equivalent to the original one. [For $O(3)$ we can prove equivalence by transforming $w \to w^{*-1}$; here this is not possible, since we require that $|w| < 1$.]

Again, a is an integer for one-valuedness. We have derived the discrete series, and its conjugate series, of irreducible unitary representations of $O(2, 1)$. We need only consider the properties of one of these series, say $a > 0$.

The unitary norm for T is

$$\|f\|^2 = \iint_{|w|<1} |f(w)|^2 (1 - |w|^2)^{2a-2} d^2w. \tag{20}$$

The function space, therefore, consists of all functions analytic in $|w| < 1$; these can be expressed as power series in w. T_a transforms w^n into

$$(\beta^* + \alpha^* w)^{-2a-n}(\alpha + \beta w)^n, \quad n \geq 0,$$

and therefore the representation is irreducible.

The infinitesimal generators are

$$J_1 = -aw + \tfrac{1}{2}(1 - w^2)\, d/dw,$$
$$J_2 = -iaw - \tfrac{1}{2}i(1 + w^2)\, d/dw,$$
$$J_3 = ia + iw\, d/dw.$$

Hence,

$$Q = J_3^2 - J_1^2 - J_2^2 = a(1 - a).$$

The eigenfunctions of J_3 are

$$K_{a,m} w^{m-a}, \quad m \geq a, \tag{21}$$

where

$$K_{a,m}^{-2} = \iint_{|w|<1} |w|^{2m-2a} (1 - |w|^2)^{2a-2} d^2w = \frac{\pi}{2a - 1} \cdot \binom{a + m - 1}{m - a}^{-1}. \tag{22}$$

The corresponding eigenvalues are im.

In this basis, the matrix element of a finite transformation is

$$(f_n, Tf_m) = K_{a,m}K_{a,n} \iint_{|w|<1} w^{*n-a}(\beta^* + \alpha^* w)^{-a-m} \times (\alpha + \beta w)^{m-a}(1 - |w|^2)^{2a-2} d^2w, \tag{23}$$

which may be written

$$(f_n, Tf_m) = (K_{a,m}/K_{a,n}) \times \text{coefficient of } w^{n-a} \text{ in the expansion of } (\beta^* + \alpha^* w)^{-a-m} \times (\alpha + \beta w)^{-a+m}. \tag{24}$$

We note that all these properties of the discrete series of unitary irreducible representations of $O(2, 1)$ can be derived by suitable analytic continuation from those of $O(3)$.

Consider now the orbit $(p_1^2 + p_2^2 - p_3^2) = M^2$. We choose on it the point $(M, 0, 0)$ and find a maximal subordinate subalgebra $H = \{J_1, J_2 + J_3\}$. This is a real subalgebra and we do not have to complexify the parameters.

For $g_1 \in e^H$,

$$\alpha_1 + \beta_1 = \alpha_1^* + \beta_1^* = e^{\frac{1}{2}x}.$$

Therefore,

$$U(g_1) = e^{iMx} = (\alpha_1 + \beta_1)^{2iM}.$$

M is not restricted to integral values since the subgroup is noncompact.

The required functions satisfy

$$F(\alpha_1\beta_2 + \beta_1^*\alpha_2, \beta_1\beta_2 + \alpha_1^*\alpha_2) = (\alpha_1 + \beta_1)^{2iM}F(\alpha_2, \beta_2), \quad \alpha_1 + \beta_1 = \alpha_1^* + \beta_1^*.$$

Take

$$\alpha_1 = \alpha_2 e^{i\vartheta}, \quad \beta_1 = \beta_2^* e^{-i\vartheta},$$

where

$$e^{2i\vartheta} = (\alpha_2^* + \beta_2^*)/(\alpha_2 + \beta_2).$$

Then,

$$\alpha_1 + \beta_1 = \alpha_1^* + \beta_1^* = |\alpha_2 + \beta_2|^{-1}$$

and

$$F(\alpha_2, \beta_2) = |\alpha_2 + \beta_2|^{2iM} F(0, -e^{-i\vartheta})$$
$$= |\alpha_2 + \beta_2|^{2iM} f\left(\frac{\alpha_2 + \beta_2}{\alpha_2^* + \beta_2^*}\right), \quad \text{say.}$$

The action of the group reduces to

$$(Tf)(e^{i\phi}) = |\beta^* + \alpha^* e^{i\phi}|^{2iM} f\left(\frac{\alpha + \beta e^{i\phi}}{\beta^* + \alpha^* e^{i\phi}}\right). \quad (25)$$

The function space consists of (well-behaved) functions on the unit circle.

There is no invariant measure on the unit circle, but we can find a unitary norm if we take M complex, with Im $(M) = \frac{1}{2}$. Then,

$$\|f\|^2 = \oint |f(e^{i\phi})|^2 \, d\phi \quad (26)$$

is invariant. Another possibility is to take Re $(M) = 0$; then

$$\|f\|^2 = \oint\oint \bar{f}(e^{i\phi_1}) f(e^{i\phi_2}) \, |e^{i\phi_1} - e^{i\phi_2}|^{-2-2iM} \, d\phi_1 \, d\phi_2 \quad (27)$$

is invariant.

These two types of representations are the principal and supplementary series, respectively, both labeled by a continuous parameter. Some of their properties are given below. Although it is not possible to derive them by the Kirillov construction from real points on orbits, there is some correspondence between them and the orbits, at least for the principal series, as we show in the next section.

The principal series of unitary irreducible representations is realized on functions on the unit circle:

$$(T_s f)(e^{i\phi}) = |\beta^* + \alpha^* e^{i\phi}|^{-1+2is} f(e^{i\phi'}), \qquad M = s + i/2, \quad (28)$$

where

$$e^{i\phi'} = \frac{\alpha + \beta e^{i\phi}}{\beta^* + \alpha^* e^{i\phi}}, \quad (29)$$

$$\|f\|^2 = \oint |f(e^{i\phi})|^2 \, d\phi < \infty. \quad (26)$$

$f(e^{i\phi})$ can be expressed as a Fourier series:

$$\sum_{-\infty}^{\infty} C_m e^{im\phi}.$$

The functions

$$f_m = (2\pi)^{-\frac{1}{2}} e^{im\phi}, \quad -\infty < m < \infty, \quad (30)$$

are eigenfunctions of J_3 with eigenvalues im. The Casimir operator Q is equal to $s^2 + \frac{1}{4}$.

All the inequivalent representations of the principal series are given by $s \geq 0$.

The supplementary series has a transformation law of the same form as (28):

$$(T_\sigma f)(e^{i\phi}) = |\beta^* + \alpha^* e^{i\phi}|^{-1-2\sigma} f(e^{i\phi'}), \qquad M = i/2 + i\sigma, \quad (31)$$

with the norm

$$\|f\|^2 = \oint\oint \bar{f}(e^{i\phi_1}) f(e^{i\phi_2}) \, |e^{i\phi_1} - e^{i\phi_2}|^{-1+2\sigma} \, d\phi_1 \, d\phi_2. \quad (32)$$

Again, f can be expressed as a Fourier series. We require $0 < \sigma < \frac{1}{2}$ in order that the norm be finite and positive.

The eigenfunctions of J_3 are

$$f_m = N_{\sigma,m} e^{im\phi}, \quad -\infty < m < \infty,$$

where

$$N_{\sigma,m}^{-2} = \oint\oint e^{-im(\phi_1 - \phi_2)} \, |e^{i\phi_1} - e^{i\phi_2}|^{-1+2\sigma} \, d\phi_1 \, d\phi_2$$
$$= 2\pi \oint e^{-im\phi} \, |1 - e^{i\phi}|^{-1+2\sigma} \, d\phi$$
$$= 2^{1+2\sigma} \cdot \pi^{\frac{3}{2}} \cdot \frac{\Gamma(\sigma)\Gamma(\frac{1}{2} - \sigma + |m|)}{\Gamma(\frac{1}{2}\sigma)\Gamma(\frac{1}{2} + \sigma + |m|)}.$$

The Casimir operator Q is equal to $-\sigma^2 + \frac{1}{4}$.

IV. CHARACTERS OF $O(2, 1)$

In this section we find the characters of the irreducible unitary representations of $O(2, 1)$ and their modified Fourier transforms.

The characters of a representation are class functions, i.e., they are the same for group elements belonging to the same conjugation class. The conjugation classes of $O(2, 1)$ are completely specified by the value of ρ and the sign of z, and, where necessary, we can choose the group elements as below:

(i) conjugation classes of type I, $0 < \rho \leq \pi$:

group element $\alpha = 0, \quad \beta = e^{\frac{1}{2}iz}$,

$$z = \epsilon(z) \cdot \rho, \; x = y = 0;$$

(ii) conjugation classes of type II, $0 < i\rho$:

group element $\alpha = \cosh \frac{1}{2}|\rho|, \quad \beta = \sinh \frac{1}{2}|\rho|$,

$$x = |\rho|, \; y = z = 0.$$

There are two exceptional conjugation classes, namely

(a) $\rho = 0, \quad \mathbf{x} \neq 0$,

(b) $x = y = z = 0$.

A. Discrete Series

Take the eigenfunctions of J_3 as a basis. Then the character is

$$C_a(\rho) = \sum_{m=a}^{\infty} (f_m, Tf_m). \quad .$$

For type I,

$$C_a^{\mathrm{I}}(\rho) = \frac{2a+1}{\pi}\sum_{m=a}^{\infty}\binom{a+m-1}{m-a} \times \iint_{|w|<1} w^{*m-a}e^{iaz}(e^{iz}w)^{m-a}(1-|w|^2)^{2a-2}d^2w, \qquad \alpha = 0.$$

Using (21), (22), and (23), we obtain

$$C_a^{\mathrm{I}}(\rho) = \frac{2a-1}{\pi}\oint\oint_{|w|<1}(1-e^{iz}|w|^2)^{-2a} \times e^{iaz}(1-|w|^2)^{2a-2}d^2w.$$

Therefore,

$$C_a^{\mathrm{I}}(\rho) = \frac{ie^{i(a-\frac{1}{2})z}}{2\sin(\frac{1}{2}z)}, \quad \text{with} \quad z = \epsilon(z)\cdot\rho. \tag{33}$$

For type II, a more careful procedure is required to evaluate the character. Take

$$y = z = 0, \quad t = \tanh(\tfrac{1}{2}|\rho|), \quad c = \cosh(\tfrac{1}{2}|\rho|).$$

Now

$$(f_m, Tf_m) = \text{coeff of } w^{m-a} \text{ in } c^{-2a}(1+tw)^{-a-m}(t+w)^{-a+m}$$

$$= \text{coeff of } w^0 \text{ in } c^{-2a}(1+tw)^{-a} \times (1+tw^{-1})^{-a}\left(\frac{1+tw^{-1}}{1+tw}\right)^m, \quad m \geq a,$$

from (24). We can expand this as a power series in w in the annulus $t < |w| < t^{-1}$, since $0 < t < 1$. If we replace w by w^{-1}, we obtain

$$(f_m, Tf_m) = \text{coeff of } w^0 \text{ in } C^{-2a}(1+tw)^{-a} \times (1+tw^{-1})^{-a}\left(\frac{1+tw^{-1}}{1+tw}\right)^{-m},$$

which we can also expand in this annulus. Therefore,

$$C_a^{\mathrm{II}}(\rho) = \frac{c^{-2a}}{2}\left(\sum_{-\infty}^{\infty} - \sum_{-a+1}^{a-1}\right)\frac{1}{2\pi i} \times \oint\frac{dw}{w}(1+tw)^{-a}(1+tw^{-1})^{-a}\left(\frac{1+tw^{-1}}{1+tw}\right)^m,$$

where the contour of integration lies in the annulus $t < |w| < t^{-1}$. To evaluate the infinite sum, take the contour $|w| = 1$ and sum inside the integral:

$$\sum_{-\infty}^{\infty}\left(\frac{1+tw^{-1}}{1+tw}\right)^m = 2\pi\delta(\phi) = \frac{\pi}{t}[\delta(\vartheta)/(1+t) + \delta(\vartheta-\pi)/(1-t)],$$

where

$$w = e^{i\vartheta}, \quad \frac{1+tw^{-1}}{1+tw} = e^{i\phi}.$$

Then the integral yields for this part

$$\frac{\cosh(a-\frac{1}{2})|\rho|}{2\sinh\frac{1}{2}|\rho|}. \tag{34}$$

The finite sum becomes

$$\frac{-c^{-2a}}{2}\cdot\frac{1}{2\pi i}\oint dw\,\frac{(1+tw^{-1})^{-2a+1} - (1+tw)^{-2a+1}}{w^2-1}.$$

Take the contour in the annulus $1 < |w| < t^{-1}$ (i.e., outside $w = \pm 1$).

The integral is the difference of two parts. The first is zero (change the variable to w^{-1}—the contour is now inside ± 1). The second part yields

$$\sinh[(a-\tfrac{1}{2})|\rho|]/2\sinh(\tfrac{1}{2}|\rho|). \tag{35}$$

Therefore, for elements belonging to conjugation classes of type II,

$$C_a^{\mathrm{II}}(\rho) = e^{-(a-\frac{1}{2})|\rho|}/2\sinh(\tfrac{1}{2}|\rho|). \tag{36}$$

By analytic continuation from $O(3)$, the invariant group measure is

$$\left|\frac{\sin\frac{1}{2}\rho}{\frac{1}{2}\rho}\right|^2 d^3\mathbf{x},$$

which can also be verified directly.

Therefore, the Fourier transform of the character that we must use is

$$\iiint C_a(\rho)\cdot\frac{\sin\frac{1}{2}\rho}{\frac{1}{2}\rho}\cdot e^{-i\mathbf{p}\cdot\mathbf{x}}d^3\mathbf{x} = \iiint\left[\frac{ie^{(a-\frac{1}{2})\epsilon(z)\rho}}{\epsilon(z)\rho}\vartheta(\rho^2) + \frac{e^{-(a-\frac{1}{2})|\rho|}}{|\rho|}\vartheta(-\rho^2)\right]e^{-i\mathbf{p}\cdot\mathbf{x}}d^3\mathbf{x}.$$

This can be evaluated using Fourier transforms of Bessel functions; it yields

$$8\pi^2\delta(p_1^2+p_2^2-p_3^2+(a-\tfrac{1}{2})^2)\vartheta(p_3). \tag{37}$$

This is the delta function on the orbit through $(0, 0, a-\frac{1}{2})$. As in the case for $O(3)$, we do not get back to the original orbit, but to one displaced by $\frac{1}{2}$.

B. Principal Series

From (26), (28), and (30) we find the following expression for the character:

$$C_s(\rho) = (2\pi)^{-1}\sum_{-\infty}^{\infty}\oint e^{-im\phi}|\beta^* + \alpha^*e^{i\phi}|^{-1+2is}e^{im\phi'}d\phi = \oint|\beta^* + \alpha^*e^{i\phi}|^{-1+2is}\delta(\phi-\phi')\,d\phi,$$

where

$$e^{i\phi'} = \frac{\alpha + \beta e^{i\phi}}{\beta^* + \alpha^* e^{i\phi}}.$$

Now, $\phi - \phi' = 0$ only has roots for elements of type II. Therefore,

$$C_s^{\mathrm{I}}(\rho) = 0, \quad \text{for} \quad \rho > 0. \tag{38}$$

Taking $y = z = 0$, we obtain

$$C_s^{\mathrm{II}}(\rho) = \oint c^{-1+2is}|1 + te^{i\phi}|^{-1+2is}\delta(\phi - \phi')\, d\phi.$$

Now

$$\delta(\phi - \phi') = (\pi/t)[\delta(\phi)/(1 + t) + \delta(\phi - \pi)/(1 - t)].$$

Therefore,

$$C_s^{\mathrm{II}}(\rho) = \frac{\cos(s|\rho|)}{\sinh(\frac{1}{2}|\rho|)}, \quad \text{for} \quad i\rho > 0. \tag{39}$$

Evaluating the Fourier transform as for the discrete series, we obtain

$$\iiint C_s(\rho) \cdot \left(\frac{\sin\frac{1}{2}\rho}{\frac{1}{2}\rho}\right) \cdot e^{-i\mathbf{p}\cdot\mathbf{x}} d^3\mathbf{x} = 8\pi^2\delta(p_1^2 + p_2^2 - p_3^2 - s^2). \tag{40}$$

This is the delta function on the orbit through $(s, 0, 0)$; the principal series representations were induced from a subalgebra subordinate to the point $(s + \frac{1}{2}i, 0, 0)$.

C. Supplementary Series

If we replace *is* by σ in the principal series, we obtain

$$C_\sigma(\rho) = \cosh(\sigma|\rho|)/\sinh\tfrac{1}{2}|\rho|\, \vartheta(-\rho^2). \tag{41}$$

This result can also be evaluated directly by a similar method to that given in Sec. IV.B. This character has no finite Fourier transform. Below we list the characters of the various representations and their Fourier transforms:

$$\rho^2 = z^2 - x^2 - y^2, \quad \begin{cases} \rho^2 > 0, & \text{for type I,} \\ \rho^2 < 0, & \text{for type II.} \end{cases}$$

Discrete Series: $a > 0$:

$$C_a^{\mathrm{I}}(\rho) = \frac{ie^{i(a-\frac{1}{2})\epsilon(z)\rho}}{2\sin[\frac{1}{2}\epsilon(z)\rho]}, \tag{33}$$

$$C_a^{\mathrm{II}}(\rho) = \frac{e^{-(a-\frac{1}{2})|\rho|}}{2\sinh(\frac{1}{2}|\rho|)}. \tag{36}$$

Fourier transform:

$$\tilde{C}_a(\mathbf{p}) = 8\pi^2\delta(p_1^2 + p_2^2 - p_3^2 + (a - \tfrac{1}{2})^2)\vartheta(p_3). \tag{37}$$

(For the conjugate series, take complex conjugates.)

Principal Series: $s \geq 0$:

$$C_s^{\mathrm{I}}(\rho) = 0, \tag{38}$$

$$C_s^{\mathrm{II}}(\rho) = \frac{\cos(s|\rho|)}{\sinh(\frac{1}{2}|\rho|)}. \tag{39}$$

Fourier transform:

$$\tilde{C}_s(\mathbf{p}) = 8\pi^2\delta(p_1^2 + p_2^2 - p_3^2 - s^2). \tag{40}$$

Supplementary Series: $0 < \sigma < \frac{1}{2}$:

$$C_\sigma^{\mathrm{I}}(\rho) = 0, \tag{41a}$$

$$C_\sigma^{\mathrm{II}}(\rho) = \frac{\cosh(\sigma|\rho|)}{\sinh(\frac{1}{2}|\rho|)}. \tag{41b}$$

There is no Fourier transform.

These characters have the following completeness and orthogonality relations:

$$\sum_{a=1}^{\infty} \bar{C}_a^{\mathrm{I}}(\rho)C_a^{\mathrm{I}}(\rho') + \sum_{a=1}^{\infty} \bar{C}_{-a}^{\mathrm{I}}(\rho)C_{-a}^{\mathrm{I}}(\rho') = \frac{2\pi\delta(\rho - \rho')}{4\sin^2(\frac{1}{2}\rho)}\vartheta[\epsilon(z)\cdot\epsilon(z')], \tag{42}$$

$$\frac{1}{2\pi}\left[\int_0^\pi \vartheta(z) + \int_0^\pi \vartheta(-z)\right] \times 4\sin^2(\tfrac{1}{2}\rho)_a(\rho)C_{a'}(\rho)\, d\rho = \delta_{a,a'}. \tag{43}$$

$-a$ denotes the representation conjugate to a.

$$\int_0^\infty ds\bar{C}_s^{\mathrm{II}}(\rho)C_s^{\mathrm{II}}(\rho') = \frac{2\pi\delta(|\rho| - |\rho'|)}{4\sinh^2(\frac{1}{2}|\rho|)}, \tag{44}$$

$$\frac{1}{2\pi}\int_0^\infty 4\sinh^2(\tfrac{1}{2}|\rho|)\bar{C}_s(\rho)C_{s'}(\rho)\, d|\rho| = \delta(s - s'). \tag{45}$$

Therefore, the characters of the principal series and of the two discrete series for conjugation classes of type I form a complete and orthogonal basis for the class functions of $O(2, 1)$.

V. CONCLUSION

The orbits $p_1^2 + p_2^2 - p_3^2 = -a^2 < 0$ of $O(2, 1)$ for integer values of a are found to be in one-to-one correspondence with the representations (unitary, irreducible) of the two discrete series. These are given by the Kirillov construction, provided that the group is complexified and analytic functions specified. We find that the (suitably defined) Fourier transforms of the characters of these representations are delta functions—not on the corresponding orbits, but on the orbits with a replaced by $a - \frac{1}{2}$.

The orbits $p_1^2 + p_2^2 - p_3^2 = M^2 > 0$ of $O(2, 1)$ are found to be in one-to-one correspondence with the representations of the principal series. However, since

these cannot be induced from unitary representations, M has to be complexified in order that the Kirillov construction yield the unitary representations. The Fourier transforms of the characters are delta functions on these orbits.

The remaining irreducible unitary representations of $O(2, 1)$, the supplementary series, are obtained by analytic continuation from part of the principal series; they have no direct relation with the Kirillov theory.

Therefore we find that the Kirillov theory can be applied to some extent to $O(2, 1)$; this is probably true for other noncompact semisimple Lie groups. There remain many questions to be answered, but it is hoped that the Kirillov method will be useful for finding representations of more complicated noncompact Lie groups which are currently of interest to physicists for the classification of elementary particles.

ACKNOWLEDGMENT

The author is grateful to Dr. R. F. Streater for suggesting the problem and for his encouragement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Solution of the Dimer Problem by the S-Matrix Method

CHUNGPENG FAN
Institute for Theoretical Physics, State University of New York, Stony Brook, New York

(Received 14 July 1968)

The S-matrix approach of Hurst is applied to rederive the partition function for the dimer problem. By using every two lattice sites as a unit cell and a set of six creation and annihilation operators at each cell, the expression for the partition is reduced to a form equivalent to the vacuum expectation value of an S matrix with a quadratic interaction Lagrangian.

INTRODUCTION

Recently Hurst[1] and Gibberd[2] have shown that the Ising problem can be solved easily in terms of creation and annihilation operators in a way that the partition function is expressed in a similar form as the vacuum expectation value of the S matrix in quantum field theory. The solubility of the Ising model was related to the quadratic nature of the corresponding Lagrangian in the S matrix. This new method, referred to as the S-matrix method was also proved elsewhere[3] to be useful in solving many other two-dimensional models in statistical mechanics, such as the modified KDP model of ferroelectricity.[4] In this note we apply it to the solution of the dimer problem which has been solved by Kasteleyn, Temperley, and Fisher[5] by using the Pfaffian method and reformulated recently in terms of transfer matrix by Lieb.[6] Our purpose in doing so is to bring further attention to the usefulness of a method in handling statistical mechanical problems which has already been proved powerful in the theory of elementary particles.

[1] C. A. Hurst, J. Math. Phys. **7**, 305 (1966).
[2] R. W. Gibberd and C. A. Hurst, J. Math. Phys. **8**, 1427 (1967).
[3] C. Fan and F. Y. Wu, Phys. Rev. **179**, 560 (1969).
[4] F. Y. Wu, Phys. Rev. **168**, 539 (1968).
[5] M. E. Fisher, Phys. Rev. **124**, 1664 (1961); H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); P. W. Kasteleyn, Physica **27**, 1209 (1962); J. Math. Phys. **4**, 287 (1963).
[6] E. H. Lieb, J. Math. Phys. **8**, 2339 (1968).

I. THE PARTITION FUNCTION

We consider a rectangular planar lattice of n rows and $2m$ columns closely packed with dimers. We group every two neighbors on a horizontal row into a unit cell; thus we have nm cells. This is shown in Fig. 1. It can easily be seen that for every unit cell there are ten allowed configurations. The unit cells are ordered helically from 1 to N (nm), from the left to the right and from the bottom to the top. To each cell j we assign three bonds: the bond $(h)_j$ that connects this cell horizontally to the cell to its right and the bonds $(r)_j$ and $(l)_j$ that connect this cell vertically the cell above it. $(r)_j$ stands to the right of $(l)_j$. To each of these bonds we assign a pair of creation and annihilation operators $a_j^{(h)}$, $a_j^{(h)\dagger}$, etc. These operators form an anticommutative set. For convenience we define the following set of six operators:

$$\begin{aligned} A^{(1)}(j) &= a_{j-1}^{(h)}, \\ A^{(2)}(j) &= a_{j-m}^{(l)}, \\ A^{(3)}(j) &= a_{j-m}^{(r)}, \\ A^{(4)}(j) &= a_j^{(h)\dagger}, \\ A^{(5)}(j) &= a_j^{(r)\dagger}, \\ A^{(6)}(j) &= a_j^{(l)\dagger}, \end{aligned} \tag{1}$$

which represent the six connections in each unit cell.

these cannot be induced from unitary representations, M has to be complexified in order that the Kirillov construction yield the unitary representations. The Fourier transforms of the characters are delta functions on these orbits.

The remaining irreducible unitary representations of $O(2, 1)$, the supplementary series, are obtained by analytic continuation from part of the principal series; they have no direct relation with the Kirillov theory.

Therefore we find that the Kirillov theory can be applied to some extent to $O(2, 1)$; this is probably true for other noncompact semisimple Lie groups. There remain many questions to be answered, but it is hoped that the Kirillov method will be useful for finding representations of more complicated noncompact Lie groups which are currently of interest to physicists for the classification of elementary particles.

ACKNOWLEDGMENT

The author is grateful to Dr. R. F. Streater for suggesting the problem and for his encouragement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Solution of the Dimer Problem by the S-Matrix Method

CHUNGPENG FAN
Institute for Theoretical Physics, State University of New York, Stony Brook, New York

(Received 14 July 1968)

The S-matrix approach of Hurst is applied to rederive the partition function for the dimer problem. By using every two lattice sites as a unit cell and a set of six creation and annihilation operators at each cell, the expression for the partition is reduced to a form equivalent to the vacuum expectation value of an S matrix with a quadratic interaction Lagrangian.

INTRODUCTION

Recently Hurst[1] and Gibberd[2] have shown that the Ising problem can be solved easily in terms of creation and annihilation operators in a way that the partition function is expressed in a similar form as the vacuum expectation value of the S matrix in quantum field theory. The solubility of the Ising model was related to the quadratic nature of the corresponding Lagrangian in the S matrix. This new method, referred to as the S-matrix method was also proved elsewhere[3] to be useful in solving many other two-dimensional models in statistical mechanics, such as the modified KDP model of ferroelectricity.[4] In this note we apply it to the solution of the dimer problem which has been solved by Kasteleyn, Temperley, and Fisher[5] by using the Pfaffian method and reformulated recently in terms of transfer matrix by Lieb.[6] Our purpose in doing so is to bring further attention to the usefulness of a method in handling statistical mechanical problems which has already been proved powerful in the theory of elementary particles.

[1] C. A. Hurst, J. Math. Phys. **7**, 305 (1966).
[2] R. W. Gibberd and C. A. Hurst, J. Math. Phys. **8**, 1427 (1967).
[3] C. Fan and F. Y. Wu, Phys. Rev. **179**, 560 (1969).
[4] F. Y. Wu, Phys. Rev. **168**, 539 (1968).
[5] M. E. Fisher, Phys. Rev. **124**, 1664 (1961); H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); P. W. Kasteleyn, Physica **27**, 1209 (1962); J. Math. Phys. **4**, 287 (1963).
[6] E. H. Lieb, J. Math. Phys. **8**, 2339 (1968).

I. THE PARTITION FUNCTION

We consider a rectangular planar lattice of n rows and $2m$ columns closely packed with dimers. We group every two neighbors on a horizontal row into a unit cell; thus we have nm cells. This is shown in Fig. 1. It can easily be seen that for every unit cell there are ten allowed configurations. The unit cells are ordered helically from 1 to N (nm), from the left to the right and from the bottom to the top. To each cell j we assign three bonds: the bond $(h)_j$ that connects this cell horizontally to the cell to its right and the bonds $(r)_j$ and $(l)_j$ that connect this cell vertically the cell above it. $(r)_j$ stands to the right of $(l)_j$. To each of these bonds we assign a pair of creation and annihilation operators $a_j^{(h)}$, $a_j^{(h)\dagger}$, etc. These operators form an anticommutative set. For convenience we define the following set of six operators:

$$\begin{aligned} A^{(1)}(j) &= a_{j-1}^{(h)}, \\ A^{(2)}(j) &= a_{j-m}^{(l)}, \\ A^{(3)}(j) &= a_{j-m}^{(r)}, \\ A^{(4)}(j) &= a_j^{(h)\dagger}, \\ A^{(5)}(j) &= a_j^{(r)\dagger}, \\ A^{(6)}(j) &= a_j^{(l)\dagger}, \end{aligned} \tag{1}$$

which represent the six connections in each unit cell.

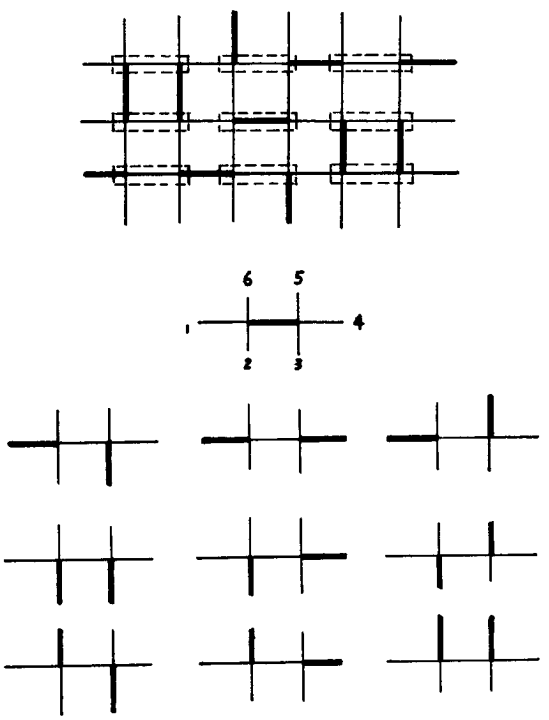


FIG. 1. Configurations of dimers and cells and the ordering of the connecting bonds.

As shown in Fig. 1, they are arranged in a counterclockwise order starting from the left.

We give a weight of x to every horizontal bond and y to every vertical bond. The weights are always allotted to the corresponding creation operators. Thus the partition function is

$$Z = \langle 0| \prod_{j=1}^{N} \{x + (A^{(3)}(j) + xA^{(4)}(j) + yA^{(5)}(j)) \times (A^{(1)}(j) + A^{(2)}(j) - yA^{(6)}(j))\} |0\rangle, \quad (2)$$

where the first term in the bracket represents the configuration without external bonds while the nine terms in the operator product are those each with two external bonds. More specifically, we have

$$Z = \langle 0| \prod_{j=1}^{N} \{x + a_{j-m}^{(r)} a_{j-1}^{(h)} + x a_j^{(h)\dagger} a_{j-1}^{(h)} + y a_j^{(r)\dagger} a_{j-1}^{(h)} + a_{j-m}^{(r)} a_{j-m}^{(l)} + x a_j^{(h)\dagger} a_{j-m}^{(l)} + y a_j^{(r)\dagger} a_{j-m}^{(l)} + y a_j^{(l)\dagger} a_{j-m}^{(r)} + x y a_j^{(l)\dagger} a_j^{(h)\dagger} + y^2 a_j^{(l)\dagger} a_j^{(r)\dagger}\} |0\rangle, \quad (3)$$

where $\langle 0|$ is the vacuum state as defined in quantum field theory.

Using the anticommutative property of those operators, it can easily be rewritten as an exponential function of a quadratic form:

$$Z = x^N \langle 0| T\left\{\exp \tfrac{1}{2} \sum_{j=1}^{N} \sum_{p,q=1}^{6} k_{pq} A^p(j) A^q(j)\right\} |0\rangle, \quad (4)$$

where the k matrix is

$$[k] = \begin{bmatrix} 0 & 0 & -1/x & -1 & -y/x & 0 \\ 0 & 0 & -1/x & -1 & -y/x & 0 \\ 1/x & 1/x & 0 & 0 & 0 & -y/x \\ 1 & 1 & 0 & 0 & 0 & -y \\ y/x & y/x & 0 & 0 & 0 & -y^2/x \\ 0 & 0 & y/x & y & y^2/x & 0 \end{bmatrix}, \quad (5)$$

where T denotes an ordering operator which requires the terms obtained on expanding the exponential to be written from right to left in order of increasing j.

By means of Wick's theorem we sum up the series expansion of the partition function and finally obtain the expression for its logarithm:

$$\log Z = \tfrac{1}{2} \sum_{r=1}^{N} \log \det [I - kA] + N \log x, \quad (6)$$

where I is the identity matrix and A is the matrix of Fourier transform of the cell-ordered contractions:

$$[A] = \begin{bmatrix} 0 & 0 & 0 & w^{-r} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w^{-mr} \\ 0 & 0 & 0 & 0 & w^{-mr} & 0 \\ -w^{r} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -w^{mr} & 0 & 0 & 0 \\ 0 & -w^{mr} & 0 & 0 & 0 & 0 \end{bmatrix} \quad (7)$$

and $w = \exp(2\pi i/N)$.

In the limit of infinite lattice size the summation is reduced to a double integral

$$\log Z = \frac{2mn}{2 \times 8\pi^2} \int_0^{2\pi} d\theta \int_0^{2\pi} d\varphi \times \log 4(x^2 \sin^2 \tfrac{1}{2}\theta + y^2 \sin^2 \varphi) = \frac{2nm}{(2\pi)^2} \int_0^{\pi} d\theta \int_0^{\pi} d\varphi \times \log 2[(x^2 + y^2) - x^2 \cos^2 \theta - y^2 \cos^2 \varphi] \quad (8)$$

which is the previous result for the dimer problem.[7]

Finally, we have to point out two facts which we have postponed until now for discussion. First, it requires a proof to verify that in the expansion of (3) all the graphs associated with the operator products have the same parity. The parity of a graph, even or odd, is the permutation needed to bring all the creation operators and their corresponding annihilation operators into pairs. Secondly, the edge effect should be clarified.

[7] E. W. Montroll, in *Applied Combinatorial Mathematics*, E. F. Beckenbach, Ed. (John Wiley & Sons, New York, 1964), Chap. 4.

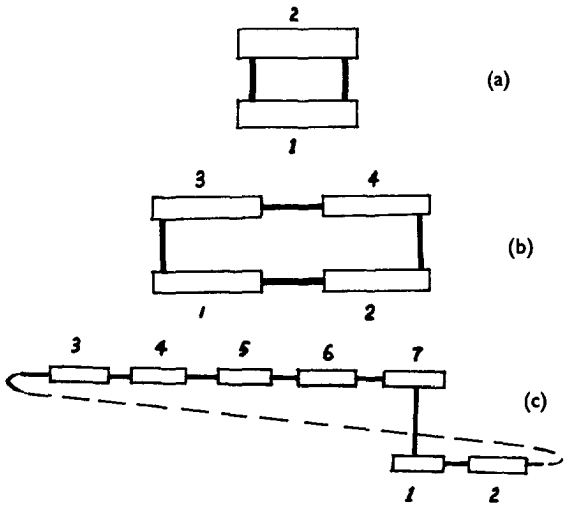


FIG. 2. Types of connected graph: (a) dumbbell graphs; (b) closed polygons; (c) closed loops involving boundary cells.

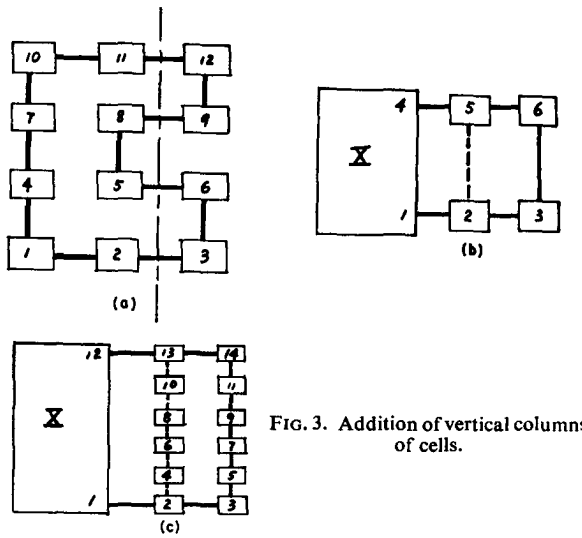


FIG. 3. Addition of vertical columns of cells.

II. PARITY OF GRAPHS

Because there is no term involving four bonds from the same cell, all the graphs are thus nonintersecting union of the following three types of connected graphs: (a) dumbbell graphs, where two vertically neighboring cells are connected with both right and left vertical bonds; (b) closed polygons, where the cells are connected into a closed loop without crossing or intersecting; (c) closed loops involving cells at the boundaries. They are shown in Fig. 2.

For the dumbbells we have the parity equal to 1,

$$p(a_1^{(r)}a_1^{(l)}a_1^{(l)\dagger}a_1^{(r)\dagger}) = p(a_1^{(r)}a_1^{(r)\dagger}) = 1; \qquad (9)$$

with two more cells added, we have, for Fig. 2(b), the smallest polygon

$$p(a_2^{(r)}a_3^{(h)}a_3^{(h)\dagger}a_1^{(l)}a_2^{(r)\dagger}a_1^{(h)}a_1^{(l)\dagger}a_1^{(h)\dagger}) = p(a_2^{(r)}a_1^{(l)}a_2^{(r)\dagger}a_1^{(h)}a_1^{(l)\dagger}a_1^{(h)\dagger}) = p(a_2^{(r)}a_2^{(r)\dagger}a_1^{(h)}a_1^{(l)}a_1^{(l)\dagger}a_1^{(h)\dagger}) = 1. \qquad (10)$$

For a general graph as in Fig. 3(a), since every cell contributes a pair of operators, therefore, if we fix the order of the two operators at each cell, we can permute the pairs without changing the parity. Thus, the pair 3 can permute with 4 and 5, 7 and 8 can permute with 3 and 6, while 10 and 11 can permute with 3, 6, and 9; we have then

$$\underset{12}{(a_9^{(r)}a_{11}^{(h)})}\underset{9}{(a_9^{(r)\dagger}a_8^{(h)})}\underset{6}{(a_3^{(r)}a_5^{(h)})}\underset{3}{(a_3^{(r)\dagger}a_2^{(h)})}\underset{11}{(a_{11}^{(h)\dagger}a_{10}^{(h)})}\underset{10}{(a_{10}^{(h)\dagger}a_7^{(l)})}\underset{8}{(a_8^{(h)\dagger}a_5^{(l)})}\underset{7}{(a_7^{(l)\dagger}a_4^{(r)})}\underset{5}{(a_5^{(l)\dagger}a_5^{(h)\dagger})}\underset{4}{(a_4^{(r)\dagger}a_1^{(l)})}\underset{2}{(a_2^{(h)\dagger}a_1^{(h)})}\underset{1}{(a_1^{(l)\dagger}a_1^{(h)\dagger})} \qquad (11)$$

which in simpler notation is

$$(12, 9, 6, 3)(11, 10, 8, 7, 5, 4, 2, 1). \qquad (12)$$

In other words, we can order the pairs in the two groups across the partition in a way as if these two groups are independent of each other. We use this result to complete our proof by means of induction method.

We proceed to consider a graph consisting of a group of operators called X with terminal cells 1 and 4 connected to cells 2 and 5 as shown in Fig. 3(b). We wish to compare the parities of the graphs before and after the addition of a vertical column of cells (3, 6). Before the addition

$$p[a_2^{(r)}a_4^{(h)}a_2^{(r)\dagger}a_1^{(h)}(X)] = p[a_1^{(h)}a_4^{(h)}(X)]; \qquad (13)$$

after the addition of (3, 6) and removal of the vertical bond between 2 and 5, we have

$$p[a_3^{(r)}a_5^{(h)}a_5^{(h)\dagger}a_4^{(h)}a_3^{(r)\dagger}a_2^{(h)}a_2^{(h)\dagger}a_1^{(h)}(X)] = p[a_3^{(r)}a_4^{(h)}a_3^{(r)\dagger}a_2^{(h)}a_2^{(h)\dagger}a_1^{(h)}(X)] = p[a_1^{(h)}a_4^{(h)}(X)]. \qquad (14)$$

For a more general case of adding a column of six cells as shown in Fig. 3(c), we have before the addition

$$p[a_{10}^{(r)}a_{12}^{(h)}a_{10}^{(r)\dagger}a_8^{(l)}a_8^{(l)\dagger}a_6^{(r)}a_6^{(r)\dagger}a_4^{(l)}a_4^{(l)\dagger}a_1^{(h)}(X)] = p[a_{10}^{(r)}a_{12}^{(h)}a_{10}^{(r)\dagger}a_1^{(h)}(X)] = p[a_1^{(h)}a_{12}^{(h)}(X)] \qquad (15)$$

and after addition

$$p[a_{11}^{(r)}a_{13}^{(h)}a_{13}^{(h)\dagger}a_{12}^{(h)}a_{11}^{(r)\dagger}a_9^{(l)}a_9^{(l)\dagger}a_7^{(r)}a_7^{(r)\dagger} \times a_5^{(l)}a_5^{(l)\dagger}a_3^{(r)}a_3^{(r)\dagger}a_2^{(h)}a_2^{(h)\dagger}a_1^{(h)}(X)] = p[a_{11}^{(r)}a_{12}^{(h)}a_{11}^{(r)\dagger}a_1^{(h)}(X)] = p[a_1^{(h)}a_{12}^{(h)}(X)]. \qquad (16)$$

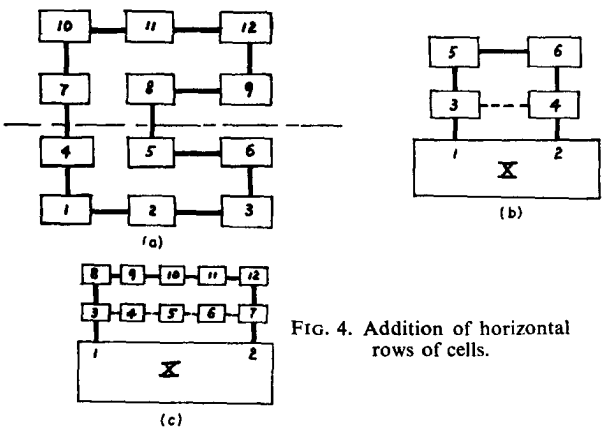


FIG. 4. Addition of horizontal rows of cells.

So, indeed, the parity does not change according to our rules of ordering.

The whole procedure can be similarly applied to the case of horizontal connections. The corresponding graphs are shown in Fig. 4.

Since every closed polygon can be constructed from the dumbbell graph by successively adding necessary rows and columns of cells and removing the redundant bonds, we therefore have in fact completed the proof for the correct parity of the closed polygons.

III. THE EDGE EFFECT

As for the third type of graphs involving the boundary cells as in Fig. 2(c), the parity is found unfortunately to be negative:

$$p(a_1^{(l)} \cdots a_3^{(h)} a_3^{(h)\dagger} a_2^{(h)} a_2^{(h)\dagger} a_1^{(h)} a_1^{(l)\dagger} a_1^{(h)\dagger}) = p(a_1^{(l)} a_1^{(h)} a_1^{(l)\dagger} a_1^{(h)\dagger}) = -1. \quad (17)$$

However, in addition to this trouble, we also have complication due to the edge effect. For instance, along the first row at the bottom the annihilation operators are not defined because $j - m$ and $j - 1$ are negative; there are just no bonds beyond to be annihilated. Likewise along the top row there are no bonds to be created. Here we have to define

$$a_j^{(l)} = -a_{j+N}^{(l)\dagger}, \quad a_j^{(r)\dagger} = -a_{j+N}^{(r)\dagger}, \quad \text{for } 1 - m \leq j \leq 0, \quad (18)$$

which means we impose toroidal boundary condition for the vertical columns. In any case the problem pertaining to edge effect is troublesome only as a matter of mathematical rigor, since physically when the lattice size approaches infinity in the thermodynamic limit, it will become asymptotically insignificant because the boundary to area ratio is then almost zero. This has been discussed by Hurst and Green.[8]

It is found difficult to formulate the monomer–dimer problem in this S-matrix method, as the graphs cannot be proved to have the same parity with both linear and quadratic terms of operators being involved.

ACKNOWLEDGMENT

The author wishes to express his thanks and gratitude to Professor C. N. Yang for drawing his attention to this problem and for his encouragement and guidance.

[8] H. S. Green and C. A. Hurst, *Order–Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964).

Lie-like Approach to the Theory of Representations of Finite Groups

A. GAMBA
Istituto di Fisica dell'Università, Genova, Italy

(Received 18 March 1968)

Using characters, one can set up the theory of representations for both finite and continuous groups. For continuous groups, another approach—the Lie theory—is also possible. It is shown that a similar theory, based on commutators, can be developed also in the case of finite groups. Essentially, the group algebra of a finite group is converted into a Lie algebra by replacing the usual associative product by the product $x \circ y = xy - yx$. Then the resulting Lie algebra is a direct sum of special unitary Lie algebras.

1. INTRODUCTION

The theory of representations of continuous groups can be developed along two different lines:

(A) The first approach is an almost straightforward generalization of the theory of characters of finite groups.[1]

(B) The second approach—the Lie approach—is based on the algebra of the generators of the infinitesimal elements.[2]

The literature is usually rather partial towards either one or the other method. In books following line A, it is useless to search for words like "root diagram," "maximum weight," etc.; often the very name of Lie is missing. Similarly, in books following line B, even the definition of such a basic concept as the character of a representation is often missing.

Yet both approaches are useful and complementary. There are problems (for example, the splitting of a degenerate level in a crystal) that can be attacked only with approach A, whereas in other cases the approach B—through the use of commutators—gives a better insight into the quantum-mechanical nature of a problem.

With finite groups, there seems to be only one approach, namely A. In this paper we would like to show briefly that an approach of type B can be developed also in the case of finite groups. Of course, in a finite group there are no infinitesimal elements. There are, however, commutators on which a Lie-like theory can be developed. It is hoped that the new approach will give a better insight into the theory of groups of finite order and show more clearly relations between discrete and continuous groups.

[1] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, N.J., 1946); F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1938); H. Boerner, *Darstellungen von Gruppen* (Springer-Verlag, Berlin, 1967).

[2] E. Cartan, "Sur la structure des groupes de transformations finis et continus," Thèse, Paris, 1894; G. Racah, *Group Theory and Spectroscopy* (Institute of Advanced Studies, Princeton, N.J., 1951). See, for other references, R. E. Behrends *et al.*, Rev. Mod. Phys. **34**, 1 (1962).

2. THE BASIC THEOREM

The easiest way to arrive at a Lie-like theory for finite groups is to consider the regular representation. It is well known that the regular representation is reducible and contains every irreducible representation of dimensionality n_i exactly n_i times, this fact being at the root of the formula

$$n_1^2 + n_2^2 + \cdots + n_k^2 = N \tag{1}$$

that relates the dimensionalities of the k irreducible representations $n_1, n_2, \cdots, n_k$ to the order N of the group. Therefore, the matrix of the regular representation corresponding to an arbitrary element of the group can always be written as a sum of smaller matrices with a total maximum number of different matrix elements equal to the order of the group.

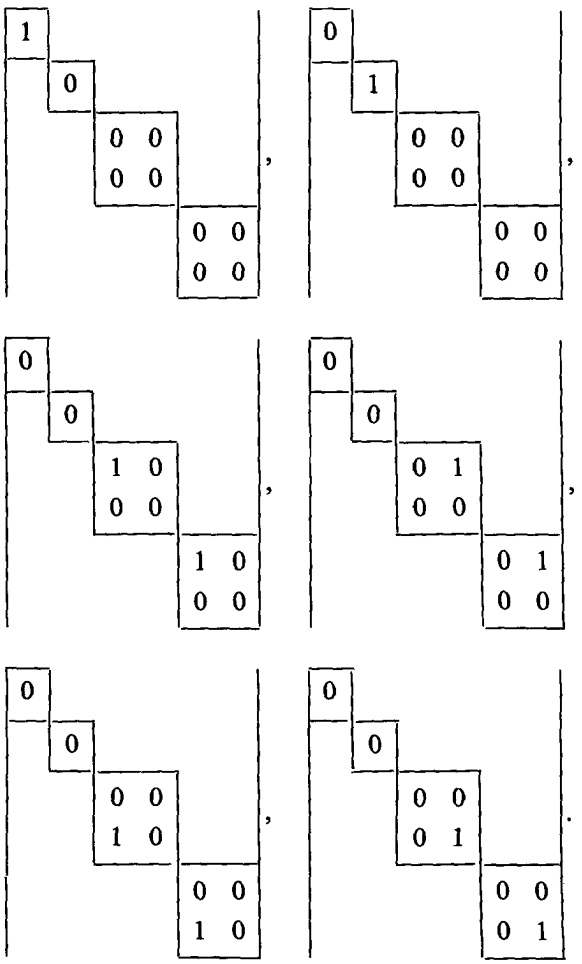
For example, take the group π_3, the symmetric permutation group on three variables, of order 6. In the regular representation, every element $A, B, \cdots$ can be written in the form

$$A = \begin{array}{|c|c|cc|cc|} \hline a & & & & & \\ \hline & b & & & & \\ \hline & & c & d & & \\ & & e & f & & \\ \hline & & & & c & d \\ & & & & e & f \\ \hline \end{array} \,,$$

$$B = \begin{array}{|c|c|cc|cc|} \hline a' & & & & & \\ \hline & b' & & & & \\ \hline & & c' & d' & & \\ & & e' & f' & & \\ \hline & & & & c' & d' \\ & & & & e' & f' \\ \hline \end{array} \,, \text{ etc.,} \tag{2}$$

with a maximum of 6 different matrix elements (a, b, c, d, e, f, or a', b', c', d', e', f', etc.).

It is, therefore, always possible to substitute the original N group elements, $A, B, \cdots$, with other N linear combinations of group elements, having all the $a, b, c, \cdots$ equal to zero, except one. In the above example the new elements will be of the form



(3)

By inspection we arrive, therefore, at the following:

Theorem: It is always possible to rearrange linearly the elements of a finite group into sets of "operators," in such a way that operators belonging to different sets commute, whereas operators of the same set have the commutation relations of the generators of the infinitesimal elements of some special unitary group $SU(n)$.

3. AN EXAMPLE

Consider the group π_4, the symmetric permutation group on four variables. It is a group of order 24, sufficiently complex to make the results not trivial and yet sufficiently simple to allow a complete solution within a reasonable amount of space. The multiplication table of this group is identical with the multiplication table of the octahedral group given in Lomont's book,[3] whose notation for the elements will be used.[4]

The results of Sec. 2 amount to saying that the 24 elements can be rearranged into 24 "operators" separated into 8 sets. Operators belonging to different sets commute. Operators belonging to the same set have the indicated commutation relations. The eight sets are the following:

$$\text{(I)}\quad C_1 = \mathbf{I}, \tag{4}$$

$$\text{(II)}\quad C_2 = \mathbf{5} + \mathbf{10} + \mathbf{13} + \mathbf{16} + \mathbf{18} + \mathbf{21}, \tag{5}$$

$$\text{(III)}\quad C_3 = \mathbf{2} + \mathbf{4} + \mathbf{9} + \mathbf{11} + \mathbf{12} + \mathbf{14} + \mathbf{20} + \mathbf{22}, \tag{6}$$

$$\text{(IV)}\quad C_4 = \mathbf{1} + \mathbf{3} + \mathbf{7} + \mathbf{8} + \mathbf{15} + \mathbf{23}, \tag{7}$$

$$\text{(V)}\quad C_5 = \mathbf{6} + \mathbf{17} + \mathbf{19}. \tag{8}$$

The above sets, each containing only one operator, are the well-known class operators that commute with every element of the group.

(VI) A set of three operators, with the commutation relation of the three generators of the infinitesimal rotations of $SU(2)$:

$$[L_0, L_+] = L_+, \quad [L_0, L_-] = -L_-, \quad [L_+, L_-] = 2L_0. \tag{9}$$

They are the following ones:

$$L_0 = (1 + 2\omega)^{-1}\{\mathbf{2} + \mathbf{4} + \mathbf{20} + \mathbf{22} - \mathbf{9} - \mathbf{11} - \mathbf{12} - \mathbf{14}\}, \quad \omega = e^{i(2\pi/3)}, \tag{10}$$

$$L_+ = \tfrac{1}{12}\{\mathbf{1} + \mathbf{7} + \mathbf{16} + \mathbf{18} + \omega(\mathbf{3} + \mathbf{5} + \mathbf{21} + \mathbf{23}) + \omega^2(\mathbf{8} + \mathbf{10} + \mathbf{13} + \mathbf{15})\}, \tag{11}$$

$$L_- = \tfrac{1}{12}\{\mathbf{1} + \mathbf{7} + \mathbf{16} + \mathbf{18} + \omega^2(\mathbf{3} + \mathbf{5} + \mathbf{21} + \mathbf{23}) + \omega(\mathbf{8} + \mathbf{10} + \mathbf{13} + \mathbf{15})\}. \tag{12}$$

(VII) A set of eight operators, with the commutation relations of the generators of the infinitesimal rotations of $SU(3)$, as given in standard form in Behrends *et al.* [Ref. 2, formula (II.12)]. They are the

[3] J. S. Lomont, *Applications of Finite Groups* (Academic Press Inc., New York, 1959), p. 33.

[4] Boldface type indicates a group element in Lomont's notation, lightface type denotes ordinary numbers.

following ones:

$$E_1 = [1/8(6^{\frac{1}{2}})]\{12 + 13 + 14 + 15 - 8 - 9 - 10 - 11\}, \quad (13)$$

$$E_{-1} = [1/8(6^{\frac{1}{2}})]\{2 + 8 + 13 + 20 - 4 - 10 - 15 - 22\}, \quad (14)$$

$$E_2 = [1/8(6^{\frac{1}{2}})]\{2 + 3 + 4 + 5 - 20 - 21 - 22 - 23\}, \quad (15)$$

$$E_{-2} = [1/8(6^{\frac{1}{2}})]\{5 + 11 + 12 + 23 - 3 - 9 - 14 - 21\}, \quad (16)$$

$$E_3 = [1/8(6^{\frac{1}{2}})]\{1 + 9 + 12 + 16 - 7 - 11 - 14 - 18\}, \quad (17)$$

$$E_{-3} = [1/8(6^{\frac{1}{2}})]\{2 + 7 + 16 + 22 - 1 - 4 - 18 - 20\}, \quad (18)$$

$$H_1 = [1/16(3^{\frac{1}{2}})]\{2(6 - 17) + 1 + 5 + 7 + 21 - 3 - 16 - 18 - 23\}, \quad (19)$$

$$H_2 = \tfrac{1}{48}\{-4(19) + 2(6 + 10 + 13 + 17 - 8 - 15) + 1 + 3 + 7 + 23 - 5 - 16 - 18 - 21\}. \quad (20)$$

(VIII) Another set of eight operators with an $SU(3)$ algebra, as follows:

$$E'_1 = [1/8(6^{\frac{1}{2}})] \times \{8 + 10 + 12 + 14 - 9 - 11 - 13 - 15\}, \quad (21)$$

$$E'_{-1} = [1/8(6^{\frac{1}{2}})] \times \{2 + 10 + 15 + 20 - 4 - 8 - 13 - 22\}, \quad (22)$$

$$E'_2 = [1/8(6^{\frac{1}{2}})] \times \{2 + 4 + 21 + 23 - 3 - 5 - 20 - 22\}, \quad (23)$$

$$E'_{-2} = [1/8(6^{\frac{1}{2}})] \times \{3 + 11 + 12 + 21 - 5 - 9 - 14 - 23\}, \quad (24)$$

$$E'_3 = [1/8(6^{\frac{1}{2}})] \times \{7 + 9 + 12 + 18 - 1 - 11 - 14 - 16\}, \quad (25)$$

$$E'_{-3} = [1/8(6^{\frac{1}{2}})] \times \{1 + 2 + 18 + 22 - 4 - 7 - 16 - 20\}, \quad (26)$$

$$H'_1 = [1/16(3^{\frac{1}{2}})]\{2(6 - 17) + 3 + 16 + 18 + 23 - 1 - 5 - 7 - 21\}, \quad (27)$$

$$H'_2 = \tfrac{1}{48}\{-4(19) + 2(6 + 8 + 15 + 17 - 10 - 13) + 5 + 16 + 18 + 21 - 1 - 3 - 7 - 23\}. \quad (28)$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Analytical Solutions of the Neutron Transport Equation in Arbitrary Convex Geometry*

ALAN G. GIBBS
Theoretical Physics Section, Reactor Physics Department, Battelle Memorial Institute Pacific Northwest Laboratory, Richland, Washington

(Received 1 July 1968)

The integral equation describing the transport of monoenergetic, isotropically scattered neutrons in a one-, two-, or three-dimensional body of arbitrary convex shape, containing distributed sources, is considered. An exact representation of the neutron density $\rho(\mathbf{r})$ is obtained, involving a superposition of functions belonging to the null space of a simple differential operator. In general, when a countable basis is chosen to span the null space, the coefficients in the expansion of $\rho(\mathbf{r})$ satisfy a coupled system of singular integral equations which is reducible to a system of Fredholm equations. If no sources are present, an exact criticality condition is also obtained. Some techniques for evaluating the expansion coefficients are given and several examples are considered.

1. INTRODUCTION

Analytical solutions of the equations of neutron transport and radiative transfer have been obtained only for a small class of highly idealized problems. Perhaps the simplest such problem concerns the transport of monoenergetic neutrons in a plane, homogeneous, isotropically scattering medium. An expression for the neutron density $\rho(x)$ due to a point source in such a medium was first obtained by the Wiener–Hopf technique,[1] although this problem can also be solved by the method of singular eigenfunctions introduced by Case in 1960.[2] In principle, Case's method can be applied to any plane geometry problem, and it leads to a precise representation of the angular density $\psi(x, \mu)$ as a superposition of elementary functions. However, the expansion coefficients are defined as the solution of a singular integral equation and can be found explicitly only for a few special problems. Thus, strictly speaking, Case's method does not provide a complete solution of the problem, but rather a mapping of one problem (the integro-differential transport equation) into another (the singular integral equation for the coefficients). Nonetheless, Case's representation gives considerable insight into the mathematical structure of the solution and, for many practical problems, excellent approximations exist for the expansion coefficients.

Since 1960, a great many papers have appeared extending Case's work to include anisotropic scattering, time and energy dependence, and media with piecewise-variable properties. Most of this work has been restricted to plane geometry, although a notable exception occurs in the work of Mitsis,[3] who considered the criticality problems for the sphere and infinite cylinder. By using techniques analogous to Case's, he showed that these problems can also be reduced to the solution of singular integral equations. Case and Zweifel[4] have observed that the same equations can be obtained by exploiting a certain replication property of the kernel of the integral transport equation.

In contrast to the number of one-dimensional transport problems which have been treated successfully either by the Wiener–Hopf technique or by Case's method and its extensions, little progress has been made in solving transport problems in two and three dimensions. While Bareiss and Abu-Shumays[5] have exhibited a vast collection of separable solutions of the three-dimensional transport equation, it is not yet clear how a minimal number of these solutions is to be selected to treat any given problem or how the corresponding expansion coefficients are to be determined. In an alternative approach, Williams[6] and Kaper[7] have attempted to reduce the same equation to an "equivalent" one-dimensional equation by postulating a simple form for the neutron density in two of the dimensions—e.g., $\exp[i(B_x x + B_y y)]$—and solving exactly for the functional dependence in the remaining dimension. While both of the preceding approaches have produced significant insights, neither

* Work supported by Battelle Memorial Institute.

[1] K. M. Case, F. de Hoffman, and G. Placzek, *Introduction to the Theory of Neutron Diffusion, Vol. I* (Los Alamos Scientific Laboratory, Los Alamos, New Mexico, 1953).

[2] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).

[3] G. J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problems," Report No. ANL-6787, Argonne National Laboratory, Argonne, Illinois, 1963.

[4] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Reading, Mass., 1967), Appendix 1.

[5] E. H. Bareiss and I. K. Abu-Shumays, "On the Structure of the Isotropic Transport Operator in Three Independent Space Variables," Report No. ANL-7328, Argonne National Laboratory, Argonne, Illinois, 1967.

[6] M. M. R. Williams, Nukleonik **9**, 305 (1967).

[7] H. G. Kaper, J. Math. Phys. **10**, 286 (1969).

has yet led to an exact solution of a realistic three-dimensional transport problem.

Recently, Williams[8] has exploited the replication properties of the kernel of the integral transport equation to obtain a precise representation of the neutron density in some simple two-dimensional problems. As in Case's method, the expansion coefficients must be determined by solving a singular integral equation.

In the present paper, the replication method is extended to obtain a representation of the neutron density valid in an arbitrary convex body. The representation will be seen to involve a superposition of functions belonging to the null space of a simple differential operator. Many equivalent representations are possible, depending on the basis chosen to span the null space. If a countable basis is chosen, the coefficients in the expansion of the neutron density must generally be obtained as the solution of a coupled system of singular integral equations, although in special cases the equations may be uncoupled.

The basic formalism is developed in Secs. 2 through 6. In Secs. 7 and 8, some simple examples are considered as illustrations. The infinite-medium case is discussed in Secs. 9 and 10, and techniques for obtaining approximate values of the expansion coefficients and approximate criticality conditions are discussed in Secs. 11 and 12.

2. BASIC EQUATIONS

Since the present analysis applies to problems in either one, two, or three dimensions, it is convenient to use a common notation to denote certain quantities, regardless of their dimensionality. Thus, in what follows, $\mathbf{r}$ and ∇^2 denote, respectively, the position coordinate(s) and the Laplacian operator in the appropriate coordinate system; similarly, $\int d\mathbf{r}$ denotes integration over the (one, two, or three) position coordinate(s).

We consider an arbitrary, convex, finite region D, with surface $\dot{D}$, in one-, two-, or three-dimensional space. The origin of the coordinate system will be taken to be inside D or on $\dot{D}$. We assume that neutrons scatter isotropically, and let c denote the mean number of secondaries per collision. If length is measured in units of the mean free path, the monoenergetic neutron density $\rho(\mathbf{r})$ satisfies the equation[1]

$$\rho(\mathbf{r}) = c\int_D \rho(\mathbf{r}')K(|\mathbf{r}-\mathbf{r}'|)\,d\mathbf{r}' + \rho_0(\mathbf{r}), \quad \mathbf{r}\in D. \tag{2.1}$$

Here $\rho_0(\mathbf{r})$ is the uncollided density, due to sources inside D or to neutrons entering D from outside. In the three-dimensional case, the kernel $K(|\mathbf{r}|)$ is given by

$$K(|\mathbf{r}|) = \frac{e^{-r}}{4\pi r^2} = \int_0^1 \frac{e^{-r/\mu}}{4\pi r}\frac{d\mu}{\mu^2}. \tag{2.2a}$$

The kernels for one- and two-dimensional problems are obtained by integrating Eq. (2.2a) over the superfluous coordinates; thus

$$K(|\mathbf{r}|) = \frac{1}{2\pi}\int_0^1 K_0\left(\frac{r}{\mu}\right)\frac{d\mu}{\mu^2} \tag{2.2b}$$

in two dimensions and

$$K(|x|) = \tfrac{1}{2}E_1(|x|) = \int_0^1 \frac{\mu}{2}\,e^{-|x|/\mu}\frac{d\mu}{\mu^2} \tag{2.2c}$$

in one dimension.

In what follows, we assume that $\rho_0(\mathbf{r})$ is produced by a distribution $S(\mathbf{r})$ of *isotropic* sources in D. For this case,

$$\rho_0(\mathbf{r}) = \int_D S(\mathbf{r}')K(|\mathbf{r}-\mathbf{r}'|)\,d\mathbf{r}', \tag{2.3}$$

and Eq. (2.1) becomes

$$\rho(\mathbf{r}) = \int_D [c\rho(\mathbf{r}') + S(\mathbf{r}')]K(|\mathbf{r}-\mathbf{r}'|)\,d\mathbf{r}', \quad \mathbf{r}\in D. \tag{2.4}$$

More general problems, involving anisotropic sources or incident neutrons, can be treated by writing the density as

$$\rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \rho_1(\mathbf{r}). \tag{2.5}$$

Substitution of this form into Eq. (2.1) shows that $\rho_1(\mathbf{r})$ satisfies Eq. (2.4) with $S(\mathbf{r}) = c\rho_0(\mathbf{r})$. It is thus sufficient to study Eq. (2.4).

We note that the right-hand sides of Eqs. (2.2) give an integral representation of $K(|\mathbf{r}|)$ in terms of the infinite-medium Green's function of the Helmholtz equation, i.e.,

$$K(|\mathbf{r}|) = \int_0^1 G(\mathbf{r},\mu)\frac{d\mu}{\mu^2}, \tag{2.6}$$

where

$$(-\nabla^2 + \mu^{-2})G(\mathbf{r}-\mathbf{r}',\mu) = \delta(\mathbf{r}-\mathbf{r}'). \tag{2.7}$$

In exploiting this fact, we generalize the approach used by Mitsis to solve the criticality problem for the infinite cylinder. By inserting the representation, Eq. (2.6), into Eq. (2.4) and interchanging the order of integration, we obtain

$$\rho(\mathbf{r}) = \int_0^1 F(\mathbf{r},\mu)\frac{d\mu}{\mu^2}, \quad \mathbf{r}\in D, \tag{2.8}$$

where

$$F(\mathbf{r},\mu) = \int_D [c\rho(\mathbf{r}') + S(\mathbf{r}')]G(\mathbf{r}-\mathbf{r}',\mu)\,d\mathbf{r}'. \tag{2.9}$$

[8] M. M. R. Williams, J. Math. Phys. **9**, 1873, 1885 (1968).

Equations (2.8) and (2.9) may be thought of as a transform pair, in the sense that if the actual $\rho(\mathbf{r})$ satisfying Eq. (2.4) is used in Eq. (2.9), then the resulting $F(\mathbf{r}, \mu)$, when substituted into Eq. (2.8), reproduces that $\rho(\mathbf{r})$. In deriving these coupled equations for $\rho(\mathbf{r})$ and $F(\mathbf{r}, \mu)$, we began with an equation for $\rho(\mathbf{r})$ alone. This process can be continued to obtain an equation for $F(\mathbf{r}, \mu)$ alone; by applying the operator $(-\nabla^2 + \mu^{-2})$ to Eq. (2.9) and using Eqs. (2.7) and (2.8), we obtain

$$(-\nabla^2 + \mu^{-2})F(\mathbf{r}, \mu) - c\int_0^1 F(\mathbf{r}, \mu)\frac{d\mu}{\mu^2} = S(\mathbf{r}). \quad (2.10)$$

The $F(\mathbf{r}, \mu)$ corresponding to the actual solution of Eq. (2.4) must be a solution of Eq. (2.10), although this equation may have many other solutions as well. We will construct the general solution of Eq. (2.10), which contains certain arbitrary coefficients, and then select the coefficients so that Eqs. (2.8) and (2.9) are mutually consistent. The density $\rho(\mathbf{r})$ will then be given by Eq. (2.8).

3. SOLUTION OF THE HOMOGENEOUS EQUATION

To obtain the general solution of Eq. (2.10) we need both a particular solution $F_p(\mathbf{r}, \mu)$ and the general solution of the homogeneous equation

$$(-\nabla^2 + \mu^{-2})F_H(\mathbf{r}, \mu) - c\int_0^1 F_H(\mathbf{r}, \mu)\frac{d\mu}{\mu^2} = 0. \quad (3.1)$$

We defer construction of $F_p(\mathbf{r}, \mu)$ until later. Equation (3.1) can be solved by separation of variables: Let

$$F_H(\mathbf{r}, \mu) = f(\nu, \mu)R(\nu, \mathbf{r}), \quad (3.2)$$

where ν is a separation parameter. Substitution of this ansatz into Eq. (3.1) leads to

$$\frac{\nabla^2 R(\nu, \mathbf{r})}{R(\nu, \mathbf{r})} = \frac{1}{\mu^2} - \frac{c}{f(\nu, \mu)}\int_0^1 f(\nu, \mu)\frac{d\mu}{\mu^2}. \quad (3.3)$$

By the usual argument, each side of Eq. (3.3) is a constant, which we denote by $1/\nu^2$. Thus,

$$\left(\frac{1}{\mu^2} - \frac{1}{\nu^2}\right)f(\nu, \mu) = c\int_0^1 f(\nu, \mu)\frac{d\mu}{\mu^2} \quad (3.4)$$

and

$$\left(-\nabla^2 + \frac{1}{\nu^2}\right)R(\nu, \mathbf{r}) = 0. \quad (3.5)$$

Equation (3.4) is independent of the domain D and is thus the same equation as the one obtained for the cylindrical criticality problem by Mitsis, whose results apply here. In particular, with the (arbitrary) normalization chosen to be

$$\int_0^1 f(\nu, \mu)\frac{d\mu}{\mu^2} = 1, \quad (3.6)$$

Mitsis found that Eq. (3.4) has one discrete solution

$$f(\nu_0, \mu) = \frac{c\nu_0^2\mu^2}{\nu_0^2 - \mu^2} \quad (3.7)$$

plus a continuum of singular solutions

$$f(\nu, \mu) = \frac{c\nu^2\mu^2}{\nu^2 - \mu^2} + \mu^2\lambda(\mu)\delta(\mu - \nu) \quad (3.8)$$

for $\nu \in (0, 1)$. Here ν_0 is the root of the equation

$$c\nu_0 \tanh^{-1}\nu_0^{-1} = 1, \quad (3.9)$$

and is tabulated in Ref. 1. The function $\lambda(\mu)$ is given by

$$\lambda(\mu) = 1 - c\mu \tanh^{-1}\mu, \quad (3.10)$$

and is the same function as introduced by Case.[2] As observed by Mitsis,

$$f(\nu, \mu) = \mu^2[\phi_\nu(\mu) + \phi_{-\nu}(\mu)], \quad (3.11)$$

where the $\phi_\nu(\mu)$ are Case's singular eigenfunctions. However, the functions $\phi_\nu(\mu)$ are convenient only for treating problems in plane geometry; the functions $f(\nu, \mu)$ arise more naturally in arbitrary geometry and should be considered fundamental.

From Eq. (3.11) and the completeness theorem proven by Case for the $\phi_\nu(\mu)$, it follows that the functions $f(\nu, \mu)$ form a complete set in the space of all functions $h(\mu)$, defined on $0 \leq \mu \leq 1$, which satisfy the requirements:

(a) $h(\mu)/\mu^2$ satisfies the Hölder condition[9] at each point of the open interval $0 < \mu < 1$;

(b) near the points $\mu = 0$ and $\mu = 1$, $h(\mu)/\mu^2$ has, at worst, weak singularities.

Specifically, this means that for each such function $h(\mu)$ there exist expansion coefficients $A(\nu_0)$ and $A(\nu)$ such that

$$h(\mu) = A(\nu_0)f(\nu_0, \mu) + \int_0^1 A(\nu)f(\nu, \mu)\,d\nu, \quad (3.12)$$

where the integral is to be interpreted as a Cauchy principal value. The expansion converges to the even extension of $h(\mu)$ for $-1 \leq \mu \leq 0$.

Some additional properties of the functions $f(\nu, \mu)$, analogous to properties of Case's functions, are easily obtained. From Eqs. (3.4) and (3.6) it follows that

$$\int_0^1 f(\nu, \mu)f(\nu', \mu)\frac{d\mu}{\mu^2} = 0, \quad \nu \neq \nu', \quad (3.13)$$

[9] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Groningen, The Netherlands, 1953).

and by applying the Poincaré-Bertrand theorem[9] we find that

$$\int_0^1 f(\nu,\mu)\left[\int_0^1 A(\nu')f(\nu',\mu)\,d\nu'\right]\frac{d\mu}{\mu^2} = \frac{\nu^2}{g(c,\nu)}A(\nu) \tag{3.14}$$

for $\nu \in (0, 1)$, where $A(\nu)$ is an arbitrary function satisfying the Hölder condition. Here,

$$g(c,\nu) = [\lambda^2(\nu) + (\tfrac{1}{2}\pi c\nu)^2]^{-1} \tag{3.15}$$

is the function tabulated in Ref. 1. Finally, we find by direct computation that

$$\int_0^1 f^2(\nu_0,\mu)\frac{d\mu}{\mu^2} = \frac{c}{2}\nu_0^4\left\{\frac{c}{\nu_0^2-1} - \frac{1}{\nu_0^2}\right\}. \tag{3.16}$$

It is convenient to denote this last expression by $\nu_0^2/g(c,\nu_0)$; i.e., we extend the definition of $g(c,\nu)$ to the point $\nu = \nu_0$ by setting

$$g(c,\nu_0) = \frac{2}{c\nu_0^2}\left\{\frac{c}{\nu_0^2-1} - \frac{1}{\nu_0^2}\right\}^{-1}. \tag{3.17}$$

The above results can be used to obtain the expansion coefficients in Eq. (3.12). In particular, it is of considerable utility to represent the function $h(\mu) = 1$ as a superposition of the $f(\nu,\mu)$. Since μ^{-2} has a second-order pole at $\mu = 0$, Case's completeness theorem does not guarantee that this expansion is possible; however, formal application of the above properties of $f(\nu,\mu)$ yields

$$1 = \frac{g(c,\nu_0)}{\nu_0^2}f(\nu_0,\mu) + \int_0^1 g(c,\nu)f(\nu,\mu)\frac{d\nu}{\nu^2}. \tag{3.18}$$

It is shown in Appendix A that this representation is, in fact, rigorous.

In what follows, it is often necessary to exhibit expressions involving both discrete and continuum terms, and to condense notation we write

$$h(\nu_0) + \int_0^1 h(\nu)\,d\nu \equiv \int_\nu h(\nu)\,d\nu, \tag{3.19}$$

whatever the function $h(\nu)$ may be. In particular, Eq. (3.18) becomes

$$1 = \int_\nu g(c,\nu)f(\nu,\mu)\frac{d\nu}{\nu^2}. \tag{3.20}$$

This concludes the discussion of the properties of the functions $f(\nu,\mu)$.

The functions $R(\nu,\mathbf{r})$, which also arise in solving Eq. (3.1), satisfy Eq. (3.5), i.e., they belong to the null space of the operator $(-\nabla^2 + \nu^{-2})$. If a basis is chosen for the null space, any solution $R(\nu,\mathbf{r})$ can be expressed as a linear combination of the basis elements. We note that problems of practical interest do not require consideration of the most general null space of $(-\nabla^2 + \nu^{-2})$ but only of some *restricted null space*. For example, it follows from Eq. (2.9) that for a finite domain D we need consider only functions which are bounded in D. In addition, the solutions must be invariant under an integral number of rotations about any axis. Thus, some possible choices of bases are

$$\{e^{\pm x/\nu}\},\quad \{I_k(r/\nu)e^{ik\theta},\ k = 0, \pm1, \cdots\},$$

and

$$\{i_l(r/\nu)Y_l^m(\theta,\phi);\ l = 0, 1, \cdots;\ m = 0, \pm1, \cdots, \pm l\}$$

for one-, two-, or three-dimensional problems, respectively. Here I_k and i_l denote the cylindrical and spherical modified Bessel functions, and Y_l^m the spherical harmonics. The above choices will be referred to as the *standard bases*, although we are by no means constrained to use only these. For example, the set $\{I_k(r/\nu)e^{ik\theta}\}$ should be most convenient when D is a circular cylinder (although it is not limited to this case), while, if D is an elliptic cylinder, elliptic coordinates and a basis involving Mathieu functions may be more convenient. This point will be further illustrated in Sec. 8.

How many elements are required to form an acceptible basis? In plane geometry, any basis of the restricted null space consists of at most two linearly independent combinations of $e^{\pm x/\nu}$. [If the problem is symmetric in x, only the even combination cosh (x/ν) is required.] In two- or three-dimensional problems, an infinite number of elements is usually required unless a high degree of symmetry is present. As long as only one position coordinate is nonperiodic (e.g., r), while the other(s) are periodic (e.g., θ or θ, ϕ), the basis is countable. However, if Cartesian coordinates are used, this simplification is lost and a continuum basis is required. For example, it can be shown that the set $\{\exp(\alpha x/\nu)\cosh(1-\alpha^2)^{\frac{1}{2}}(y/\nu),\ -1 \le \alpha \le 1\}$ is admissible for two-dimensional problems symmetric in y. The analysis following applies to problems involving either countable or continuum bases, although the former case is somewhat more transparent. Thus, in the present paper, we assume that one of the standard bases, or some other countable basis, has been chosen. To a large extent, the use of a continuum basis requires only the replacement of certain sums by integrals; however, this topic will be treated in detail in a later paper.

Let $\{R_n(\nu,\mathbf{r})\}$ denote the chosen basis, where n may denote, for example, the single index k or the pair of indices (l, m) of the standard bases. In terms

of this basis, the most general admissible solution of Eq. (15) is

$$R(\nu, \mathbf{r}) = \sum_n A_n(\nu) R_n(\nu, \mathbf{r}), \tag{3.21}$$

where the $A_n(\nu)$ are arbitrary coefficients. The most general admissible solution of the homogeneous equation (3.1) is thus

$$F_H(\mathbf{r}, \mu) = \int_\nu f(\nu, \mu)\left(\sum_n A_n(\nu) R_n(\nu, \mathbf{r})\right) d\nu = \sum_n \int_\nu A_n(\nu) f(\nu, \mu) R_n(\nu, \mathbf{r})\, d\nu. \tag{3.22}$$

4. THE GENERAL SOLUTION

To complete the solution of Eq. (2.10), it is necessary to construct a particular solution $F_p(\mathbf{r}, \mu)$. This will be accomplished through application of the following theorem.

Theorem: Let D_1 denote some region in space, with surface $\dot{D}_1$, and let $\varphi(\mathbf{r}, \mu)$ be a solution of the equation

$$(-\nabla^2 + \mu^{-2})\varphi(\mathbf{r}, \mu) = S_1(\mathbf{r}), \quad \mathbf{r} \in D_1. \tag{4.1}$$

Let $\varphi(\mathbf{r}, \mu)$ satisfy the boundary condition

$$\varphi(\mathbf{r}, \mu) = T(\mathbf{r}), \quad \mathbf{r} \in \dot{D}_1. \tag{4.2}$$

Here $S_1(\mathbf{r})$ and $T(\mathbf{r})$ are given functions, *independent of* μ. Then the function

$$F_p(\mathbf{r}, \mu) \equiv \int_\nu g(c, \nu) f(\nu, \mu) \varphi(\mathbf{r}, \nu) \frac{d\nu}{\nu^2} \tag{4.3}$$

satisfies the equation

$$(-\nabla^2 + \mu^{-2})F_p(\mathbf{r}, \mu) - c\int_0^1 F_p(\mathbf{r}, \mu) \frac{d\mu}{\mu^2} = S_1(\mathbf{r}), \quad \mathbf{r} \in D_1, \tag{4.4}$$

and $F_p(\mathbf{r}, \mu)$ satisfies the same boundary condition as $\varphi(\mathbf{r}, \mu)$ on $\dot{D}_1$.

Proof: Write

$$(-\nabla^2 + \mu^{-2}) = (-\nabla^2 + \nu^{-2}) + (\mu^{-2} - \nu^{-2}) \tag{4.5}$$

and substitute Eq. (4.3) into the left-hand side of Eq. (4.4). By using Eqs. (4.1), (3.4), and (3.20), we obtain

$$(-\nabla^2 + \mu^{-2})F_p(\mathbf{r}, \mu) - c\int_0^1 F_p(\mathbf{r}, \mu) \frac{d\mu}{\mu^2} = \int_\nu g(c, \nu) f(\nu, \mu) \frac{d\nu}{\nu^2} \cdot S_1(\mathbf{r}) + \int_\nu g(c, \nu)\left[\left(\frac{1}{\mu^2} - \frac{1}{\nu^2}\right) f(\nu, \mu) - c\int_0^1 f(\nu, \mu) \frac{d\mu}{\mu^2}\right] \times \varphi(\mathbf{r}, \nu) \frac{d\nu}{\nu^2} = S_1(\mathbf{r}), \tag{4.6}$$

as required. For $\mathbf{r}$ on $\dot{D}_1$, from Eqs. (4.3) and (4.2) we obtain

$$F_p(\mathbf{r}, \mu) = \int_\nu g(c, \nu) f(\nu, \mu) \frac{d\nu}{\nu^2} \cdot T(\mathbf{r}) = T(\mathbf{r}), \tag{4.7}$$

as required. This completes the proof.

The theorem will be used to construct an $F_p(\mathbf{r}, \mu)$ which vanishes as $r \to \infty$. To do this, let D_1 be the infinite medium, let $T(\mathbf{r}) = 0$, and let $S_1(\mathbf{r})$ be zero outside D and equal to $S(\mathbf{r})$ inside D An appropriate solution of Eq. (4.1) is

$$\varphi(\mathbf{r}, \mu) = \int_D G(\mathbf{r} - \mathbf{r}', \mu) S(\mathbf{r}')\, d\mathbf{r}'. \tag{4.8}$$

Thus, the function

$$F_p(\mathbf{r}, \mu) = \int_\nu g(c, \nu) f(\nu, \mu)\left[\int_D G(\mathbf{r} - \mathbf{r}', \nu) S(\mathbf{r}')\, d\mathbf{r}'\right]\frac{d\nu}{\nu^2} \tag{4.9}$$

is a particular solution of Eq. (2.10) which vanishes as $r \to \infty$.

The most general admissible solution of Eq. (2.10) is obtained by combining Eqs. (4.9) and (3.22):

$$F(\mathbf{r}, \mu) = \int_\nu g(c, \nu) f(\nu, \mu)\left(\int_D G(\mathbf{r} - \mathbf{r}', \nu) S(\mathbf{r}')\, d\mathbf{r}'\right)\frac{d\nu}{\nu^2} + \sum_n \int_\nu A_n(\nu) f(\nu, \mu) R_n(\nu, \mathbf{r})\, d\nu. \tag{4.10}$$

By substituting this expression into Eq. (2.8) and using the normalization of $f(\nu, \mu)$, we obtain

$$\rho(\mathbf{r}) = \int_\nu g(c, \nu)\left(\int_D G(\mathbf{r} - \mathbf{r}'', \nu) S(\mathbf{r}'')\, d\mathbf{r}''\right)\frac{d\nu}{\nu^2} + \sum_n \int_\nu A_n(\nu) R_n(\nu, \mathbf{r})\, d\nu. \tag{4.11}$$

This representation of $\rho(\mathbf{r})$ in terms of the Green's function and solutions of the Helmholtz equation is the main result of this paper, but before discussing its implications we proceed to show how the coefficients $A_n(\nu)$ are evaluated.

5. EVALUATION OF THE COEFFICIENTS

For the one-dimensional criticality problems, Mitsis was able to deduce a boundary condition, satisfied by $F(\mathbf{r}, \mu)$ on $\dot{D}$, from which he could evaluate his (single) $A(\nu)$. Unfortunately, this approach does not apply to an arbitrary domain D; instead, the expression for $\rho(\mathbf{r})$ must be substituted into Eq. (2.9) and the $A_n(\nu)$ chosen so that the original $F(\mathbf{r}, \mu)$,

Eq. (4.10), is reproduced. This operation requires evaluation of the integrals

$$I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu) = \int_D G(\mathbf{r} - \mathbf{r}', \mu) G(\mathbf{r}' - \mathbf{r}'', \nu)\, d\mathbf{r}' \quad (5.1)$$

and

$$I_2(\mathbf{r}, n, \mu, \nu) = \int_D G(\mathbf{r} - \mathbf{r}', \mu) R_n(\nu, \mathbf{r}')\, d\mathbf{r}'. \quad (5.2)$$

It is easily shown that the above integrals can be expressed in terms of $G(\mathbf{r}, \mu)$ and the elements of the basis $\{R_n(\nu, \mathbf{r})\}$. The procedure is classic. To evaluate $I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu)$ we write the equations satisfied by $G(\mathbf{r}' - \mathbf{r}'', \nu)$ and $G(\mathbf{r}, -\mathbf{r}', \mu)$, considered as functions of $\mathbf{r}'$:

$$(-\nabla'^2 + \mu^{-2}) G(\mathbf{r} - \mathbf{r}', \mu) = \delta(\mathbf{r} - \mathbf{r}'), \quad (5.3)$$

$$(-\nabla'^2 + \nu^{-2}) G(\mathbf{r}' - \mathbf{r}'', \nu) = \delta(\mathbf{r}' - \mathbf{r}''). \quad (5.4)$$

Multiplying the first equation by $G(\mathbf{r}' - \mathbf{r}'', \nu)$, the second by $G(\mathbf{r} - \mathbf{r}', \mu)$, subtracting, integrating over all $\mathbf{r}' \in D$, and applying Green's theorem leads to

$$(\mu^{-2} - \nu^{-2}) I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu) = G(\mathbf{r} - \mathbf{r}'', \nu) - G(\mathbf{r} - \mathbf{r}'', \mu) + W(\mathbf{r}, \mathbf{r}'', \nu, \mu), \quad (5.5)$$

where

$$W(\mathbf{r}, \mathbf{r}'', \nu, \mu) = \int_{\dot{D}} \left[G(\mathbf{r}' - \mathbf{r}'', \nu) \frac{\partial}{\partial n'} G(\mathbf{r} - \mathbf{r}', \mu) - G(\mathbf{r} - \mathbf{r}', \mu) \frac{\partial}{\partial n'} G(\mathbf{r}' - \mathbf{r}'', \nu) \right] dS'. \quad (5.6)$$

Since $\mathbf{r}$ and $\mathbf{r}''$ are interior points of D, while the integration in Eq. (5.6) is over the surface, we note that

$$(-\nabla^2 + \mu^{-2}) W(\mathbf{r}, \mathbf{r}'', \nu, \mu) = 0 \quad (5.7)$$

and

$$(-\nabla''^2 + \nu^{-2}) W(\mathbf{r}, \mathbf{r}'', \nu, \mu) = 0. \quad (5.8)$$

Furthermore, $G(\mathbf{r} - \mathbf{r}', \mu)$ has an integrable singularity, so that $W(\mathbf{r}, \mathbf{r}'', \nu, \mu)$ is finite. Since $G(\mathbf{r}, \mu)$ is clearly unchanged by an integral number of rotations of the coordinate axes, it follows that $W(\mathbf{r}, \mathbf{r}'', \nu, \mu)$ belongs to the *restricted* null spaces of the operators in Eqs. (5.7) and (5.8). Thus we can write

$$W(\mathbf{r}, \mathbf{r}'', \nu, \mu) = \sum_{n,m} w_{nm}(\nu, \mu) R_n^*(\nu, \mathbf{r}'') R_m(\mu, \mathbf{r}), \quad (5.9)$$

where the $w_{nm}(\nu, \mu)$ are an appropriate set of coefficients. [The use of the complex-conjugate basis element $R_n^*(\nu, \mathbf{r})$ rather than $R_n(\nu, \mathbf{r})$ itself is for later convenience; for the standard bases, $\{R_n^*(\nu, \mathbf{r})\}$ is at most a rearrangement of $\{R_n(\nu, \mathbf{r})\}$, and in the general case it is an alternative basis.] It follows from Eq. (5.6) that $W(\mathbf{r}, \mathbf{r}'', \nu, \mu) = -W(\mathbf{r}'', \mathbf{r}, \mu, \nu)$. This result, together with the fact that $W(\mathbf{r}, \mathbf{r}'', \nu, \mu)$ is real, implies that

$$w_{nm}^*(\nu, \mu) = -w_{mn}(\mu, \nu). \quad (5.10)$$

Since $I_1(\mathbf{r}, \mathbf{r}'', \nu, \mu)$ is bounded as $\mu \to \nu$, we deduce from Eq. (5.5) that $W(\mathbf{r}, \mathbf{r}'', \nu, \nu) = 0$, which implies that

$$w_{nm}(\nu, \nu) = 0. \quad (5.11)$$

These last two results are independent of the domain D and basis $\{R_n(\nu, \mathbf{r})\}$. Finally, by substituting Eq. (5.9) into Eq. (5.5), we obtain

$$I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu) = \frac{\nu^2\mu^2}{\nu^2 - \mu^2} \left\{ G(\mathbf{r} - \mathbf{r}'', \nu) - G(\mathbf{r} - \mathbf{r}'', \mu) + \sum_{nm} w_{nm}(\nu, \mu) R_n^*(\nu, \mathbf{r}'') R_m(\mu, \mathbf{r}) \right\}. \quad (5.12)$$

An analogous argument leads to a representation for $I_2(\mathbf{r}, n, \mu, \nu)$:

$$I_2(\mathbf{r}, n, \mu, \nu) = \frac{\nu^2\mu^2}{\nu^2 - \mu^2} \left\{ R_n(\nu, \mathbf{r}) - \sum_m q_{nm}(\nu, \mu) R_m(\mu, \mathbf{r}) \right\}. \quad (5.13)$$

Here the coefficients $q_{nm}(\nu, \mu)$ are defined by the expansion [cf. Eqs. (5.6) and (5.9)]:

$$\sum_m q_{nm}(\nu, \mu) R_m(\mu, \mathbf{r}) = \int_{\dot{D}} \left[G(\mathbf{r} - \mathbf{r}', \mu) \frac{\partial}{\partial n'} R_n(\nu, \mathbf{r}') - R_n(\nu, \mathbf{r}') \frac{\partial}{\partial n'} G(\mathbf{r} - \mathbf{r}', \mu) \right] dS'. \quad (5.14)$$

Since $I_2(\mathbf{r}, n, \nu, \nu) < \infty$, it follows from Eq. (5.13) that

$$q_{nm}(\nu, \nu) = \delta_{nm}, \quad (5.15)$$

independent of D and the choice of basis. Since the quantities in wavy brackets in Eqs. (5.12) and (5.13) vanish for $\nu = \mu$, the factor $\nu^2\mu^2/(\nu^2 - \mu^2)$ can be replaced, whenever convenient, by $c^{-1}f(\nu, \mu)$.

The above representations of $I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu)$ and $I_2(\mathbf{r}, n, \mu, \nu)$ exhibit the most general replication property of the kernel $G(\mathbf{r} - \mathbf{r}', \mu)$; namely, when $G(\mathbf{r}' - \mathbf{r}'', \nu)$, or any element of a basis $\{R_n(\nu, \mathbf{r}')\}$, is multiplied by $G(\mathbf{r} - \mathbf{r}', \mu)$ and the result integrated over all $\mathbf{r}' \in D$, only linear combinations of the same functions are produced. The one-dimensional replication properties noted by Case and Zweifel[4] and the recent results of Williams[8] are special cases of the above.

At this point it is possible to state a criterion for a basis $\{R_n(\nu, \mathbf{r})\}$ to be sufficiently complete: since the

integral $I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu)$ arises from the source term in Eq. (4.11), only $I_2(\mathbf{r}, n, \mu, \nu)$ occurs in a source-free problem. Thus, if $S(\mathbf{r}) \equiv 0$, a prospective basis is sufficiently complete if the integral $I_2(\mathbf{r}, n, \mu, \nu)$ can be represented as in Eq. (5.13) for each value of n. If sources are present, we may require in addition that the integral $I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu)$ be representable as in Eq. (5.12), although it is not difficult to show [cf. Eq. (5.18) below] that it is sufficient for Eq. (5.12) to hold after multiplication by $S(\mathbf{r}'')$ and integration over all $\mathbf{r}'' \in D$.

By substituting $\rho(\mathbf{r})$ from Eq. (4.11) into Eq. (2.9) and using the above results and the representation of unity given in Eq. (3.20), we obtain

$$\begin{aligned} F(\mathbf{r}, \mu) = &\int_\nu g(c, \nu) f(\nu, \mu) \left(\int_D S(\mathbf{r}') G(\mathbf{r} - \mathbf{r}', \nu)\, d\mathbf{r}' \right) \frac{d\nu}{\nu^2} \\ &+ \sum_n \int_\nu A_n(\nu) f(\nu, \mu) R_n(\nu, \mathbf{r})\, d\nu \\ &- \sum_m \left\{ \sum_n \int_\nu A_n(\nu) f(\nu, \mu) q_{nm}(\nu, \mu)\, d\nu \right. \\ &\left. - \sum_n \int_\nu g(c, \nu) f(\nu, \mu) w_{nm}(\nu, \mu) S_n(\nu) \frac{d\nu}{\nu^2} \right\} R_m(\mu, \mathbf{r}), \end{aligned} \tag{5.16}$$

where

$$S_n(\nu) = \int_D S(\mathbf{r}'') R_n^*(\nu, \mathbf{r}'')\, d\mathbf{r}''. \tag{5.17}$$

The first two terms on the right-hand side of Eq. (5.16) are identical to the original $F(\mathbf{r}, \mu)$, Eq. (4.10). Thus $F(\mathbf{r}, \mu)$ will reproduce itself if the $A_n(\nu)$ can be chosen to make the final term vanish. Since the $R_m(\mu, \mathbf{r})$ are linearly independent, this requires that

$$\begin{aligned} \sum_n \int_\nu A_n(\nu) f(\nu, \mu) q_{nm}(\nu, \mu)\, d\nu \\ = \sum_n \int_\nu g(c, \nu) f(\nu, \mu) w_{nm}(\nu, \mu) S_n(\nu) \frac{d\nu}{\nu^2}, \quad \text{all } m. \end{aligned} \tag{5.18}$$

The nature of this equation becomes more evident if the discrete and continuum terms are written out separately and the explicit form of $f(\nu, \mu)$ is substituted from Eq. (3.8). We denote the (known) right-hand side of Eq. (5.18) by $\mu^2 \tilde{S}_m(\mu)$, i.e., we put

$$\tilde{S}_m(\mu) = \frac{1}{\mu^2} \sum_n \int_\nu g(c, \nu) f(\nu, \mu) w_{nm}(\nu, \mu) S_n(\nu) \frac{d\nu}{\nu^2}. \tag{5.19}$$

With these devices, Eq. (5.18) becomes

$$\begin{aligned} \lambda(\mu) A_m(\mu) + c \sum_n \int_0^1 \frac{\nu^2 q_{nm}(\nu, \mu)}{\nu^2 - \mu^2} A_n(\nu)\, d\nu \\ = \tilde{S}_m(\mu) - \sum_n A_n(\nu_0) \left(\frac{c\nu_0^2 q_{nm}(\nu_0, \mu)}{\nu_0^2 - \mu^2} \right), \quad \text{all } m, \end{aligned} \tag{5.20a}$$

which is seen to be a coupled system of singular integral equations for the $A_n(\nu)$.

At a later stage in the analysis, it may be convenient to regard Eq. (5.20a) as an equation for a set of modified expansion coefficients

$$A_n'(\nu) = M(\nu) \cdot A_n(\nu), \tag{5.21}$$

where the modifying factor $M(\nu)$ is a (known) continuous function to be chosen judiciously later. By multiplying and dividing the terms of Eq. (5.20a) by $M(\nu)$, we obtain

$$\begin{aligned} \lambda(\mu) A_m'(\mu) + c \sum_n \int_0^1 \frac{\nu^2}{\nu^2 - \mu^2} \frac{M(\mu)}{M(\nu)} q_{nm}(\nu, \mu) A_n'(\nu)\, d\nu \\ = M(\mu) \left[\tilde{S}_m(\mu) - \sum_n \frac{A_n'(\nu_0)}{M(\nu_0)} \left(\frac{c\nu_0^2 q_{nm}(\nu_0, \mu)}{\nu_0^2 - \mu^2} \right) \right], \quad \text{all } m. \end{aligned} \tag{5.20b}$$

As shown by Vekua,[10] the general procedure for solving a system of singular integral equations such as Eq. (5.20b) consists of first writing the kernels as the sum of a dominant singular part and a Fredholm part, and then using complex variable methods to invert the dominant part. This leads to a system of coupled Fredholm equations for the unknown quantities. Several simplifications occur for the present problem. To isolate the dominant part of Eq. (5.20b), we expand the singular term in partial fractions

$$\frac{\nu^2}{\nu^2 - \mu^2} = \frac{\nu}{2} \left(\frac{1}{\nu - \mu} + \frac{1}{\nu + \mu} \right), \tag{5.22}$$

and introduce the functions

$$H_{nm}(\nu, \mu) \equiv \{[M(\mu)/M(\nu)] q_{nm}(\nu, \mu) - \delta_{nm}\}/(\nu - \mu). \tag{5.23}$$

Using these devices in Eq. (5.20b) and recalling that $q_{nm}(\nu, \nu) = \delta_{nm}$, we obtain

$$\lambda(\mu) A_m'(\mu) + \frac{c}{2} \int_0^1 \frac{\nu}{\nu - \mu} A_m'(\nu)\, d\nu = \Phi_m(\mu), \quad \text{all } m, \tag{5.24}$$

where

$$\begin{aligned} \Phi_m(\mu) = M(\mu) \left\{ \tilde{S}_m(\mu) - \sum_n \frac{A_n'(\nu_0)}{M(\nu_0)} \left(\frac{c\nu_0^2 q_{nm}(\nu_0, \mu)}{\nu_0^2 - \mu^2} \right) \right\} \\ - \frac{c}{2} \sum_n \int_0^1 \frac{\nu [2\nu H_{nm}(\nu, \mu) + \delta_{nm}]}{\nu + \mu} A_n'(\nu)\, d\nu. \end{aligned} \tag{5.25}$$

Thus the equations for $A_n'(\nu)$ are seen to be coupled only through the Fredholm parts in the $\Phi_m(\mu)$, while the dominant parts of the equations, as represented by

[10] N. P. Vekua, *Systems of Singular Integral Equations* (P. Noordhoff, Groningen, The Netherlands, 1967).

the left-hand side of Eq. (5.24), are uncoupled. Furthermore, the dominant part is the same for each m, and is independent of the domain D and the choice of basis $\{R_n(\nu, \mathbf{r})\}$.

In order to convert the singular equations to Fredholm equations, we temporarily consider the functions $\Phi_m(\mu)$ to be known and "solve" Eq. (5.24) by the standard methods of Muskhelishvili[9] (or equivalently, by the "half-range orthogonality" of Kuščer *et al.*[11]). Since the left-hand side of this equation is independent of m, D, and the choice of basis, it is the same as in the equations obtained by Mitsis for the one-dimensional criticality problems, and his solution can be applied here. (See Ref. 3 for details.) The solution of Eq. (5.24) involves the functions

$$X(z) = \exp\left[-\frac{c}{2}\int_0^1 g(c,\mu)\left(1+\frac{c\mu^2}{1-\mu^2}\right)\ln(\mu - z)\,d\mu\right] \tag{5.26}$$

and

$$\gamma(\nu) = \tfrac{1}{2}c\nu X^-(\nu)/(\lambda(\nu) - \tfrac{1}{2}\pi i c\nu), \tag{5.27}$$

where

$$X^{\pm}(\nu) = \lim_{\epsilon\to 0} X(\nu \pm i\epsilon). \tag{5.28}$$

In particular, a solution exists only if $\Phi_m(\mu)$ satisfies the condition

$$\int_0^1 \gamma(\mu)\Phi_m(\mu)\,d\mu = 0, \quad \text{all } m. \tag{5.29}$$

If this condition is satisfied, the "solution" is given by

$$A'_m(\mu) = \lambda(\mu)g(c,\mu)\Phi_m(\mu) - \frac{1}{X^-(\mu)[\lambda(\mu) + \tfrac{1}{2}\pi i c\mu]}\int_0^1 \frac{\gamma(\nu)}{\nu - \mu}\Phi_m(\nu)\,d\nu. \tag{5.30}$$

Since $\Phi_m(\mu)$ actually involves the unknown functions, Eq. (5.30) is a system of Fredholm equations for the $A'_m(\mu)$. Both this equation and the constraint, Eq. (5.29), can be put in more transparent form by substituting $\Phi_m(\mu)$ from Eq. (5.25). The results can be expressed compactly by introducing the integral operators [defined on an arbitrary function $h(\mu)$]

$$T_1[h(\mu)] \equiv \lambda(\mu)g(c,\mu)h(\mu) - \frac{1}{X^-(\mu)[\lambda(\mu) + \tfrac{1}{2}\pi i c\mu]}\int_0^1 \frac{\gamma(\nu')}{\nu' - \mu}h(\nu')\,d\nu' \tag{5.31}$$

and

$$T_2[h(\mu)] \equiv \int_0^1 \gamma(\mu)h(\mu)\,d\mu. \tag{5.32}$$

[11] I. Kuščer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N.Y.) **30**, 411 (1964).

With this notation, Eq. (5.30) becomes

$$A'_m(\mu) + \frac{c}{2}\sum_n \int_0^1 K_{mn}(\mu,\nu)A'_n(\nu)\,d\nu = S'_m(\mu) - \sum_n B_{mn}(\mu)A'_n(\nu_0), \quad \text{all } m, \tag{5.33}$$

where

$$K_{mn}(\mu,\nu) = \nu T_1\left[\frac{2\nu H_{nm}(\nu,\mu) + \delta_{nm}}{\nu + \mu}\right],$$
$$B_{mn}(\mu) = \frac{c\nu_0^2}{M(\nu_0)}T_1\left[\frac{M(\mu)q_{nm}(\nu_0,\mu)}{\nu_0^2 - \mu^2}\right], \tag{5.34}$$
$$S'_m(\mu) = T_1[M(\mu)\tilde{S}_m(\mu)].$$

Similarly, the constraint, Eq. (5.29), becomes

$$\frac{c}{2}\sum_n \int_0^1 k_{mn}(\nu)A'_n(\nu)\,d\nu + \sum_n b_{mn}A'_n(\nu_0) = s'_m, \quad \text{all } m, \tag{5.35}$$

where

$$k_{mn}(\nu) = \nu T_2\left[\frac{2\nu H_{nm}(\nu,\mu) + \delta_{nm}}{\nu + \mu}\right],$$
$$b_{mn} = \frac{c\nu_0^2}{M(\nu_0)}T_2\left[\frac{M(\mu)q_{nm}(\nu_0,\mu)}{\nu_0^2 - \mu^2}\right], \tag{5.36}$$
$$s'_m = T_2[M(\mu)\tilde{S}_m(\mu)].$$

Since $M(\mu)$ is a continuous function of μ, since $R_n(\mu, \mathbf{r})$ is a continuous function of μ^{-2}, and since $I_2(\mathbf{r}, n, \mu, \nu)$ is bounded for a finite domain D, it follows from Eqs. (5.13) and (5.23) that the $H_{nm}(\nu, \mu)$ are bounded as $\mu \to \nu$. It is shown in Appendix B that the operator T_2 is bounded, and thus each of the kernels $K_{mn}(\mu, \nu)$ is bounded (in the sense of the L_2 norm); thus Eq. (5.33) is a Fredholm equation, as asserted above.

Since many choices of basis $\{R_m(\mu, \mathbf{r})\}$ are possible [each leading to different sets of functions $q_{nm}(\nu, \mu)$ and $H_{nm}(\nu, \mu)$], and since $M(\mu)$ can be chosen arbitrarily, it appears that the construction of a purely mathematical proof of existence of a solution of Eqs. (5.33) and (5.35) would be very difficult. However, the existence of a solution can be made highly plausible on physical grounds: since Eq. (5.33) is a Fredholm equation, it is solvable unless the corresponding adjoint homogeneous equation has a nontrivial solution. But according to the Fredholm theorems, this is possible only if the homogeneous form of Eq. (5.33), i.e., ($S'_m = 0$) itself has a nontrivial solution. Since such a solution is physically possible only if the system is critical, Eq. (5.33) should be solvable for any subcritical system.

The coefficients $A'_n(\mu)$ must be determined by solving Eq. (5.33) subject to the constraint, (5.35). The constraint serves to determine the discrete coefficients

$A'_n(\nu_0)$, or in the source-free problem ($s'_m = 0$) it leads to an exact criticality condition. The procedure will be clarified through examples in later sections.

6. CALCULATION OF THE KERNELS

We note that finding closed-form expressions for the kernels from Eqs. (5.34) and (5.36) is in general a difficult task, since the application of T_1 and T_2 requires the evaluation of integrals of the form $\int_0^1 \gamma(\mu)h(\mu)\,d\mu$, where $h(\mu)$ is some complicated function of μ. While these integrals in general must be evaluated numerically, an important exception occurs if the required $h(\mu)$ satisfies the following conditions:

(a) With μ replaced by a complex variable z, $h(z)$ is a meromorphic function and none of its poles lie in the interval (0, 1].

(b) $h(z)$ may have a pole at $z = 0$, and for $z \to \infty$, $h(z) \to h(\infty) < \infty$.

It is shown in Appendix B that, for such functions $h(\mu)$, the required integrals can be evaluated by contour integration, yielding

$$T_1[h(\mu)] = (1 - c)(\nu_0^2 - \mu^2)g(c, \mu)X(-\mu) \times \left\{X(0)/\mu \cdot \operatorname{Res} h(0) + \sum_j \frac{X(z_j)}{\mu - z_j} \operatorname{Res} h(z_j)\right\} \quad (6.1)$$

and

$$T_2[h(\mu)] = \sum_j X(z_j) \operatorname{Res} h(z_j) + X(0) \operatorname{Res} h(0) - h(\infty). \quad (6.2)$$

Here the z_j are the poles of $h(z)$, and "Res" denotes the residue. Numerical values of $X(-\mu)$ have been tabulated by Kowalska.[12]

When the above formulas apply, analytical expressions can be obtained for the kernels of Eqs. (5.33) and (5.35). It is sometimes possible to arrange this fortunate circumstance by making a judicious choice of $M(\mu)$; in other cases it will be necessary to resort to numerical computation.

7. EXAMPLES: THE ONE-DIMENSIONAL CRITICAL PROBLEMS

As both a test of the above formalism and an illustration of how the various quantities appearing in Eqs. (5.33) and (5.35) are evaluated for specific problems, it will be shown that when these equations are specialized to the one-dimensional critical problems, they reduce to the equations obtained by Mitsis. Techniques for solving Eqs. (5.33) and (5.35) will be discussed in a later section.

[12] K. Kowalska, *Tables of the Functions $X(c, -\nu)$ and $X^{\pm}(c_1, c_2)$* (Report No. 630/IX-A/PR, Institute of Nuclear Research, Warsaw, 1965).

The Critical Slab

Consider a bare, source-free slab of half-thickness b. Since the neutron density must be an even function of x, the restricted null space for this problem is the set of all even solutions of

$$\left(-\frac{d^2}{dx^2} + \frac{1}{\nu^2}\right)R(\nu, x) = 0. \quad (7.1)$$

Clearly, this set is spanned by the single function

$$R_1(\nu, x) = \cosh x/\nu \quad (7.2)$$

and thus the neutron density has the representation [cf. Eq. (4.11)]:

$$\rho(x) = A(\nu_0)\cosh\frac{x}{\nu_0} + \int_0^1 A(\nu)\cosh\frac{x}{\nu}\,d\nu, \quad (7.3)$$

where the coefficients are to be determined. As a first step, we find $q_{11}(\nu, \mu)$ by evaluating the integral $I_2(x, 1, \mu, \nu)$ [Eq. (5.2)]:

$$I_2(x, 1, \mu, \nu) = \int_{-b}^{b} \frac{\mu}{2}\exp\left(-\frac{|x - x'|}{\mu}\right)\cosh\frac{x'}{\nu}\,dx' = \frac{\nu^2\mu^2}{\nu^2 - \mu^2}\left\{\cosh\frac{x}{\nu} - e^{-b/\mu}\left[\cosh\frac{b}{\nu} + \frac{\mu}{\nu}\sinh\frac{b}{\nu}\right]\cosh\frac{x}{\mu}\right\}. \quad (7.4)$$

This result is of the form [cf. Eq. (5.13)]:

$$I_2(x, 1, \mu, \nu) = \frac{\nu^2\mu^2}{\nu^2 - \mu^2}\{R_1(\nu, x) - q_{11}(\nu, \mu)R_1(\mu, x)\}, \quad (7.5)$$

where

$$q_{11}(\nu, \mu) = e^{-b/\mu}\left[\cosh\frac{b}{\nu} + \frac{\mu}{\nu}\sinh\frac{b}{\nu}\right]. \quad (7.6)$$

Equation (7.5) verifies that the single function $R_1(\nu, x)$ does constitute a sufficiently complete basis for this problem [cf. the remarks preceding Eq. (5.16)]. As in Eq. (5.23),

$$H_{11}(\nu, \mu) = \left[\frac{M(\mu)}{M(\nu)}q_{11}(\nu, \mu) - 1\right]\Big/(\nu - \mu), \quad (7.7)$$

where we are free to select $M(\mu)$ at will. It is convenient to choose $M(\mu)$ so that $M(z)q_{11}(\nu, z)$ is a meromorphic function of z; thus we choose

$$M(\mu) = e^{b/\mu} \quad (7.8)$$

to eliminate the essential singularity of $q_{11}(\nu, z)$ at $z = 0$. The expansion coefficients in Eq. (7.3) will then be given by [cf. Eq. (5.21)]:

$$A(\nu) = \frac{1}{M(\nu)}A'(\nu) = e^{-b/\nu}A'(\nu) \quad (7.9)$$

and $A'(\nu)$ will be determined from Eq. (5.33).

With the above choice of $M(\mu)$, we find

$$\frac{M(\mu)q_{11}(\nu_0,\mu)}{\nu_0^2-\mu^2}=\frac{\cosh b/\nu_0+(\mu/\nu_0)\sinh b/\nu_0}{\nu_0^2-\mu^2} \tag{7.10}$$

and

$$\frac{2\nu H_{11}(\nu,\mu)+1}{\nu+\mu}=\frac{e^{-2b/\nu}}{\nu+\mu}. \tag{7.11}$$

The former function has poles at $\mu=\pm\nu_0$, while the latter function has a pole at $\mu=-\nu$. Both functions vanish as $\mu\to\infty$. Thus Eqs. (6.1) and (6.2) apply, and we find from Eqs. (5.34) and (5.36) that

$$\begin{aligned}K_{11}(\nu,\mu)&=(\nu_0^2-\mu^2)(1-c)g(c,\mu)\\&\quad\times X(-\mu)X(-\nu)\nu[e^{-2b/\nu}/(\nu+\mu)],\\k_{11}(\nu)&=\nu X(-\nu)e^{-2b/\nu},\\B_{11}(\mu)&=(\nu_0^2-\mu^2)(1-c)g(c,\mu)X(-\mu)\\&\quad\times\left\{\frac{c\nu_0X(\nu_0)}{2(\nu_0-\mu)}+\frac{c\nu_0X(-\nu_0)}{2(\nu_0+\mu)}\,e^{-2b/\nu_0}\right\},\\b_{11}&=\tfrac{1}{2}c\nu_0[X(-\nu_0)e^{-2b/\nu_0}-X(\nu_0)].\end{aligned} \tag{7.12}$$

Finally, Eqs. (5.33) and (5.35) reduce to

$$A'(\mu)+\tfrac{1}{2}c\int_0^1K_{11}(\nu,\mu)A'(\nu)\,d\nu=-B_{11}(\mu)A'(\nu_0) \tag{7.13}$$

and

$$\tfrac{1}{2}c\int_0^1k_{11}(\nu)A'(\nu)\,d\nu=-b_{11}A'(\nu_0). \tag{7.14}$$

These equations are identical to those obtained by Mitsis who applied Case's singular eigenfunction method to the problem.

The Critical Sphere

Consider a bare, source-free sphere of radius R. The set of all bounded, spherically symmetric solutions of the equation

$$(-\nabla^2+\nu^{-2})R(\nu,\mathbf{r})=0 \tag{7.15}$$

is spanned by the single function

$$R_0(\nu,r)=i_0\left(\frac{r}{\nu}\right)=\frac{\sinh r/\nu}{r/\nu}. \tag{7.16}$$

Thus the neutron density (4.11) is given by

$$\rho(\mathbf{r})=A(\nu_0)i_0\left(\frac{r}{\nu_0}\right)+\int_0^1A(\nu)i_0\left(\frac{r}{\nu}\right)d\nu. \tag{7.17}$$

By use of the addition theorem[13]

$$\begin{aligned}&\frac{\exp(-|\mathbf{r}-\mathbf{r}'|/\mu)}{4\pi|\mathbf{r}-\mathbf{r}'|}\\&=\frac{1}{2\pi^2\mu}\sum_{l=0}^{\infty}(2l+1)i_l\left(\frac{r_<}{\mu}\right)k_l\left(\frac{r_>}{\mu}\right)P_l(\cos\theta_{rr'}),\end{aligned} \tag{7.18}$$

[13] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).

the integral $I_2(r,0,\mu,\nu)$ of Eq. (5.2) is easily evaluated, and we find that

$$I_2(\mathbf{r},0,\mu,\nu)=\frac{\nu^2\mu^2}{\nu^2-\mu^2}\left\{i_0\left(\frac{r}{\nu}\right)-q_{00}(\nu,\mu)i_0\left(\frac{r}{\mu}\right)\right\}, \tag{7.19}$$

where

$$q_{00}(\nu,\mu)=\frac{e^{-R/\mu}}{\mu}\left[\nu\sinh\frac{R}{\nu}+\mu\cosh\frac{R}{\nu}\right]. \tag{7.20}$$

To eliminate the singularity of $q_{00}(\nu,\mu)$ at $\mu=0$, we choose

$$M(\mu)=(\mu/2)e^{R/\mu} \tag{7.21}$$

(the factor of 2 is for convenience), so that the coefficients in Eq. (7.17) are given by

$$A(\nu)=(2/\nu)e^{-R/\nu}A'(\nu). \tag{7.22}$$

The equations for $A'(\nu)$ are then found just as for the slab problem. In particular,

$$\begin{aligned}K_{00}(\nu,\mu)&=-(\nu_0^2-\mu^2)(1-c)g(c,\nu)\\&\quad\times X(-\mu)X(-\nu)[\nu e^{-2R/\nu}/(\nu+\mu)],\\k_{00}(\nu)&=-\nu X(-\nu)e^{-2R/\nu},\\B_{00}(\mu)&=(\nu_0^2-\mu^2)(1-c)g(c,\mu)X(-\mu)\\&\quad\times\left\{\frac{c\nu_0X(\nu_0)}{2(\nu_0-\mu)}-\frac{c\nu_0X(-\nu_0)}{2(\nu_0+\mu)}\,e^{-2R/\nu_0}\right\},\\b_{00}&=-(c\nu_0/2)[X(\nu_0)+e^{-2R/\nu_0}X(-\nu_0)].\end{aligned} \tag{7.23}$$

Finally, Eqs. (5.33) and (5.35) reduce to

$$A'(\mu)+\frac{c}{2}\int_0^1K_{00}(\mu,\nu)A'(\nu)\,d\nu=-B_{00}(\mu)A'(\nu_0), \tag{7.24}$$

$$\frac{c}{2}\int_0^1k_{00}(\nu)A'(\nu)\,d\nu=-b_{00}A'(\nu_0). \tag{7.25}$$

These equations are identical to Eqs. (4.6-5) and (4.6-6) of Mitsis' paper.

The Critical Infinite Cylinder

For this problem, Mitsis did not carry the analysis as far as for the slab and sphere, but stopped with equations corresponding to Eqs. (5.24) and (5.25) above. To obtain his results, we first note that once again the set of all bounded, rotationally invariant solutions of Eq. (7.15) is spanned by a single function

$$R_0(\nu,r)=I_0(r/\nu). \tag{7.26}$$

Thus,

$$\rho(\mathbf{r})=A(\nu_0)I_0\left(\frac{r}{\nu}\right)+\int_0^1A(\nu)I_0\left(\frac{r}{\nu}\right)d\nu. \tag{7.27}$$

The integral $I_2(r,0,\mu,\nu)$ can again be evaluated by

using the appropriate addition theorem[14]

$$\frac{1}{2\pi} K_0\left(\frac{|\mathbf{r}-\mathbf{r}'|}{\mu}\right) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} I_n\left(\frac{r_<}{\mu}\right)K_n\left(\frac{r_>}{\mu}\right)e^{in(\theta-\theta')}, \quad (7.28)$$

and we find that

$$I_2(\mathbf{r}, 0, \mu, \nu) = \frac{\nu^2\mu^2}{\nu^2-\mu^2}\left\{I_0\left(\frac{r}{\nu}\right) - q_{00}(\nu,\mu)I_0\left(\frac{r}{\mu}\right)\right\}, \quad (7.29)$$

where

$$q_{00}(\nu,\mu) = \frac{R}{\nu}K_0\left(\frac{R}{\mu}\right)I_1\left(\frac{R}{\nu}\right) + \frac{R}{\mu}I_0\left(\frac{R}{\nu}\right)K_1\left(\frac{R}{\mu}\right). \quad (7.30)$$

If we choose $M(\mu) = 1$, so that $A(\nu) = A'(\nu)$, then

$$H_{00}(\nu,\mu) = [q_{00}(\nu,\mu) - 1]/(\nu-\mu), \quad (7.31)$$

and Eqs. (5.24) and (5.25) become

$$\lambda(\mu)A(\mu) + \frac{c}{2}\int_0^1 \frac{\nu}{\nu-\mu}A(\nu)\,d\nu = \Phi(\mu), \quad (7.32)$$

$$\Phi(\mu) = -\left(\frac{c\nu_0^2 q_{00}(\nu,\mu)}{\nu_0^2-\mu^2}\right)A(\nu_0) - \frac{c}{2}\int_0^1 \frac{\nu[2\nu H_{00}(\nu,\mu)+1]}{\nu+\mu}A(\nu)\,d\nu. \quad (7.33)$$

The functions $q_{00}(\nu,\mu)$ and $H_{00}(\nu,\mu)$ defined above are identical to the functions $q(\nu,\mu)$ and $H(\nu,\mu)$ introduced by Mitsis, and Eqs. (7.32) and (7.33) are identical to his Eqs. (5.6-7) and (5.6-7a).

8. A MORE GENERAL EXAMPLE

The three simple problems considered above are inherently one-dimensional, and in each case only a single function $R(\nu, \mathbf{r})$ is required to span the appropriate restricted null space. However, if we consider the same simple geometries with nonsymmetric sources or consider more complex geometries, the null space is less restricted and more basis elements $R_n(\nu, \mathbf{r})$ are required.

As an illustration, consider a two- or three-dimensional region D described in a coordinate system $\mathbf{r} = (r_1, \mathbf{x})$. Here r_1 denotes one coordinate, and $\mathbf{x}$ the remaining one or two coordinate(s). Suppose that the surface D has the simple form $r_1 = R =$ constant. (For example, D might be an elliptic cylinder or an ellipsoid.) In addition, assume that Eq. (3.5) has solutions of the form

$$R(\nu, \mathbf{r}) = y(\nu, r_1)W(\nu, \mathbf{x}). \quad (8.1)$$

As discussed by Morse and Feshbach,[15] it will generally be possible to find a set of solutions

$$y_n(\nu, r_1)W_n(\nu, \mathbf{x}),$$

where the $W_n(\nu, \mathbf{x})$ are a complete set of functions, orthogonal with respect to some weight function $\rho_W(\mathbf{x})$,

$$\int \rho_W(\mathbf{x})W_n(\nu,\mathbf{x})W_m^*(\nu,\mathbf{x})\,d\mathbf{x} = \delta_{nm}. \quad (8.2)$$

The $y_n(\nu, r_1)$ are solutions of appropriate Sturm–Liouville equations and we let $y_{1n}(\nu, r_1)$ denote the solutions regular at $r_1 = 0$ and $y_{2n}(\nu, r_1)$ denote the solutions vanishing at $r_1 = \infty$. Thus we can take

$$\{R_n(\nu, \mathbf{r})\} = \{y_{1n}(\nu, r_1)W_n(\nu, \mathbf{x})\}. \quad (8.3)$$

As shown by Morse and Feshbach, the Green's function $G(\mathbf{r}-\mathbf{r}', \mu)$ can be expanded as

$$G(\mathbf{r}-\mathbf{r}',\mu) = -\left(\frac{h_1'}{h_{\mathbf{x}}'}\right)\rho_W(\mathbf{x}')\sum_m W_m(\mu,\mathbf{x})W_m^*(\mu,\mathbf{x}') \times \frac{y_{1m}(\mu, r_{1<})y_{2m}(\mu, r_{1>})}{\Delta[y_{1m}(\mu, r_1'), y_{2m}(\mu, r_1')]}. \quad (8.4)$$

Here h_1' and $h_{\mathbf{x}}'$ denote the scale factors of the coordinate system, and Δ denotes the Wronskian.

For the present problem, the $q_{nm}(\nu, \mu)$ are most easily obtained from the expansion in Eq. (5.14) rather than by direct evaluation of $I_2(\mathbf{r}, n, \mu, \nu)$. In the present case, the surface integral corresponds to an integral over all $\mathbf{x}$ with $r_1 = R$. By substituting the above expressions for $G(\mathbf{r}-\mathbf{r}', \mu)$ and $R_n(\nu, \mathbf{r}')$ into Eq. (5.14) and equating coefficients of $R_m(\mu, \mathbf{r})$, we deduce that

$$q_{nm}(\nu,\mu) = \frac{\Delta[y_{2m}(\mu, R), y_{1n}(\nu, R)]}{\Delta[y_{2m}(\mu, R), y_{1m}(\mu, R)]} \cdot \int \rho_W(\mathbf{x})W_n(\nu,\mathbf{x})W_m^*(\mu,\mathbf{x})\,d\mathbf{x}. \quad (8.5)$$

A further simplification occurs if $W_n(\mu, \mathbf{x})$ is independent of μ (as occurs for the important cases of the sphere and infinite cylinder, when the corresponding standard bases are used). For these cases the integral in Eq. (8.5) reduces to δ_{nm}, so that

$$q_{nm}(\nu,\mu) = \frac{\Delta[y_{2n}(\mu, R), y_{1n}(\nu, R)]}{\Delta[y_{2n}(\mu, R), y_{1n}(\mu, R)]}\delta_{nm}. \quad (8.6)$$

By substituting the explicit form of $y_{1n}(\mu, r_1)$ and

[14] A. Erdelyi, Ed., *Higher Transcendental Functions* (McGraw-Hill Book Co., New York, 1953), Vol. 2.

[15] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., New York, 1953), Chap. 7.

$y_{2n}(\mu, r_1)$, we find that for an infinite cylinder

$$q_{nm}(\nu, \mu) = \left\{\frac{R}{\nu} K_n\left(\frac{R}{\mu}\right) I'_n\left(\frac{R}{\nu}\right) - \frac{R}{\mu} K'_n\left(\frac{R}{\mu}\right) I_n\left(\frac{R}{\nu}\right)\right\}\delta_{nm}, \quad (8.7)$$

and for a sphere

$$q_{lm,l'm'}(\nu, \mu) = \frac{2R}{\pi\mu}\left\{\frac{R}{\nu} k_l\left(\frac{R}{\mu}\right) i'_l\left(\frac{R}{\nu}\right) - \frac{R}{\mu} k'_l\left(\frac{R}{\mu}\right) i_l\left(\frac{R}{\nu}\right)\right\}\delta_{ll'}\delta_{mm'}. \quad (8.8)$$

It is easily verified that the results given in Eqs. (7.30) and (7.20) are special cases of the above. Similarly, by using Eqs. (5.6) and (5.9), it is easily shown that if $W_n(\mu, \mathbf{x})$ is independent of μ, the coefficients $w_{nm}(\nu, \mu)$ which occur in the source term of Eq. (5.33) are also diagonal. In particular,

$$w_{nm}(\nu, \mu) = \frac{1}{2\pi}\left\{\frac{R}{\mu} K_n\left(\frac{R}{\nu}\right) K'_n\left(\frac{R}{\mu}\right) - \frac{R}{\nu} K_n\left(\frac{R}{\mu}\right) K'_n\left(\frac{R}{\nu}\right)\right\}\delta_{nm}, \quad (8.9)$$

for an infinite cylinder, and

$$w_{lm,l'm'}(\nu, \mu) = \frac{4R}{\mu\nu}\left\{\frac{R}{\mu} k_l\left(\frac{R}{\nu}\right) k'_l\left(\frac{R}{\mu}\right) - \frac{R}{\nu} k_l\left(\frac{R}{\mu}\right) k'_l\left(\frac{R}{\nu}\right)\right\}\delta_{ll'}\delta_{mm'}, \quad (8.10)$$

for a sphere.

Since both $q_{nm}(\nu, \mu)$ and $w_{nm}(\nu, \mu)$ are diagonal for the sphere and infinite cylinder, the equations for $A'_n(\nu)$ uncouple, and only the nth mode of the source distribution contributes to $A'_n(\nu)$. Thus, the solution of problems involving nonsymmetric sources in these simple geometries is in principle no more difficult than the solution of symmetric source problems—one has only to solve a separate Fredholm integral equation for the coefficient $A'_n(\nu)$ of each mode present in the source distribution. However, we note that in other geometries $q_{nm}(\nu, \mu)$ and $w_{nm}(\nu, \mu)$ are, in general, not diagonal and the coefficients $A'_n(\nu)$ must then be determined by solving a *coupled set* of Fredholm equations.

9. THE INFINITE MEDIUM

It is convenient to discuss the case of an infinite medium before considering methods to solve Eq. (5.33) for the $A'_n(\nu)$. Since the functions $R_n(\nu, \mathbf{r})$ may be unbounded as $r \to \infty$, some care is required to extend the above analysis rigorously to include this case. Only an heuristic argument will be given here. We consider the effects of extending a finite domain D in all directions until it becomes infinite. Since both $G(\mathbf{r} - \mathbf{r}', \mu)$ and its normal derivative become exponentially small as $r' \to \infty$, we deduce from Eqs. (5.6) and (5.9) that

$$w_{nm}(\nu, \mu) \to 0. \quad (9.1)$$

An inspection of Eq. (5.14) shows that the situation is more complex for the $q_{nm}(\nu, \mu)$. As r' becomes large for fixed $\mathbf{r}$, the dominant terms in the integrand are proportional to $\exp(\nu^{-1} - \mu^{-1})r'$; thus

$$q_{nm}(\nu, \mu) \to \begin{cases} 0, & \mu < \nu, \\ \infty, & \mu > \nu. \end{cases} \quad (9.2)$$

We note in particular that

$$q_{nm}(\nu_0, \mu) \to 0. \quad (9.3)$$

What effect the above results will have on the kernels of Eq. (5.33) depends on the choice of $M(\mu)$, but the essential features of the final solution can be deduced for the case $M(\mu) = 1$. Then from Eqs. (9.1), (5.19), (5.34), and (5.36) we deduce that $S'_m(\mu)$ and s'_m both go to zero. From Eqs. (9.3), (5.34), and (5.36) it follows that $B_{mn}(\mu)$ and b_{mn} both go to zero. Thus, Eq. (5.35) reduces to

$$\frac{c}{2}\sum_n \int_0^1 k_{mn}(\nu) A_n(\nu)\, d\nu = 0. \quad (9.4)$$

But from Eqs. (9.2) and (5.36) we conclude that $k_{mn}(\nu) \to \infty$; thus Eq. (9.4) can be satisfied only if $A_n(\nu) \to 0$ (sufficiently fast) for $\nu \in (0, 1)$. This result might have been anticipated on physical grounds, since past experience with one-dimensional problems indicates that a physically meaningful solution $\rho(\mathbf{r})$ can grow no faster than e^{r/ν_0} for large r.

From the above discussion (which can be made rigorous) it is seen that, for an infinite medium, every term in Eqs. (5.33) and (5.35) vanishes, and that $A_n(\nu) \equiv 0$ for $\nu \in (0, 1)$. Since the coefficients $B_{nm}(\mu)$ and b_{nm} vanish, it is not necessary for the discrete coefficients $A_n(\nu_0)$ to be zero. In fact, they can be chosen arbitrarily. The expression for the neutron density, Eq. (4.11), thus reduces to

$$\rho(\mathbf{r}) = \int_\nu g(c, \nu)\left(\int_D S(\mathbf{r}') G(\mathbf{r} - \mathbf{r}', \nu)\, d\mathbf{r}'\right)\frac{d\nu}{\nu^2} + \sum_n A_n(\nu_0) R_n(\nu_0, \mathbf{r}). \quad (9.5)$$

The first term on the right represents neutrons produced by the source distribution $S(\mathbf{r})$, while the remaining terms can be thought of as the contribution due to large sources located at $r = \infty$. It is plausible

that the discrete coefficients should be arbitrary, since any (physically reasonable) distribution of sources at infinity may be specified. In particular, if there are no sources at infinity, we require that

$$\lim_{r\to\infty} \rho(\mathbf{r}) = 0.$$

Since this excludes the $R_n(\nu_0, \mathbf{r})$ terms, Eq. (9.5) reduces to

$$\rho(\mathbf{r}) = \int_\nu g(c, \nu)\left(\int_D S(\mathbf{r}')G(\mathbf{r} - \mathbf{r}', \nu)\, d\mathbf{r}'\right)\frac{d\nu}{\nu^2}. \quad (9.6)$$

If we consider the elementary source $S(\mathbf{r}) = \delta(\mathbf{r})$, then

$$\rho(\mathbf{r}) = \int_\nu g(c, \nu)G(\mathbf{r}, \nu)\frac{d\nu}{\nu^2}. \quad (9.7)$$

Specialization of Eq. (9.7) to one, two, or three dimensions leads to expressions for the density due to a plane, line, or point source, respectively, in an infinite medium. These results have of course been obtained by other methods and are well known, but the above analysis provides a concise and simultaneous derivation of all three.

10. THE PARTIALLY INFINITE MEDIUM

Partially infinite media problems are typified by the one-dimensional Milne problem. In two and three dimensions it is easy to conceive of a wealth of related problems in which the domain D extends to infinity in certain directions, but remains finite in others. It appears that a general theory of such problems will be highly complex, although a few tentative ground rules can be given if physically meaningful solutions are to be obtained. The continuum coefficients $A_n(\nu)$ of basis elements $R_n(\nu, \mathbf{r})$ which grow faster than e^{r/ν_0} in a direction in which D is infinite should be set equal to zero. The corresponding discrete coefficients will be arbitrary; they represent sources at infinity. The remaining coefficients will be determined by solving Eqs. (5.33) and (5.35). The general theory of Sec. 5 has been applied to the familiar one-dimensional Milne problem and application of the above ground rules leads directly to the usual solution.[16]

11. CALCULATION OF THE EXPANSION COEFFICIENTS

The expansion coefficients $A_n(\nu)$ which appear in the representation of $\rho(\mathbf{r})$ in Eq. (4.11) must be determined either by solving the singular equation (5.24) directly or by solving the equivalent Fredholm equations (5.33) and (5.35). It was shown in Sec. 9 that, for the infinite medium, the $A_n(\nu)$ vanish for $\nu \in (0, 1)$. Thus, if the domain D is large, the contribution of the continuum $A_n(\nu)$ terms to $\rho(\mathbf{r})$ should be small, and an approximate calculation will suffice.

For numerical calculations, the expansion of $\rho(\mathbf{r})$ must be truncated to a finite number (say N) of elements of the basis $\{R_n(\nu, \mathbf{r})\}$. In this case, it is convenient to rewrite Eqs. (5.33) and (5.35) in matrix notation as

$$A'(\mu) + (c/2)KA'(\mu) + BA'(\nu_0) = S'(\mu) \quad (11.1)$$

and

$$(c/2)kA'(\mu) + bA'(\nu_0) = s'. \quad (11.2)$$

Here, $A'(\mu)$ and $A'(\nu_0)$ are $(N \times 1)$ column vectors, $B(\mu)$ and b are $(N \times N)$ matrices, and K and k are $(N \times N)$ matrix integral operators.

If b is invertible, a formal solution of Eqs. (11.1) and (11.2) can be obtained by first solving Eq. (11.2) for $A'(\nu_0)$:

$$A'(\nu_0) = b^{-1}[s' - (c/2)kA'(\mu)] \quad (11.3)$$

and then substituting the result into Eq. (11.1) and solving for $A'(\mu)$:

$$A'(\mu) = [I - (c/2)(B(\mu)b^{-1}k - K)]^{-1} \times [S'(\mu) - B(\mu)b^{-1}s']. \quad (11.4)$$

If

$$\|B(\mu)b^{-1}k - K\| < 2/c, \quad (11.5)$$

the inverse is defined by its Neumann series expansion and Eq. (11.4) gives a valid solution for $A'(\mu)$. When specialized to the critical slab and sphere problems, this algorithm reduces to the iterative scheme used with considerable success by Mitsis. The approach will only be successful if Eq. (11.5) holds. As shown in Sec. 7 above, the choice of $M(\mu)$ most convenient for evaluating the kernels of Eq. (5.33) for the slab and sphere leads to kernels which decrease as $e^{-2R/\nu}$ as the size of the body increases. Thus, for these simple geometries, Eq. (11.5) is satisfied for sufficiently large bodies. Unfortunately, there is no obvious assurance that this will be the case for more complex geometries; while physical intuition makes the solvability of Eqs. (11.1) and (11.2) highly plausible, it does not imply that the solution is obtainable as a Neumann series. If Eq. (11.5) is not satisfied, it should still be possible to find solutions of Eqs. (11.1) and (11.2) by discrete ordinates methods.

As an alternative approach, it might be possible to obtain the $A_n(\mu)$ directly from the singular equation (5.24) by extending the method of Bareiss and Neuman.[17]

[16] S. I. Schreiner, Dept. of Nuclear Engineering, University of Washington, private communication, 1968.

[17] E. H. Bareiss and C. P. Neuman, "Singular Integrals and Singular Integral Equations with a Cauchy Kernel and the Method of Symmetric Paring," Report No. ANL-6988, Argonne National Laboratory, Argonne, Illinois, 1965.

Considerable work will be required to prove rigorously the convergence of the various approximation methods outlined above and to determine the most efficient way of evaluating the $A_n(\mu)$.

12. CRITICALITY PROBLEMS

When no sources are present in D, it is more convenient to solve Eq. (11.1) for $A'(\mu)$:

$$A'(\mu) = -[I + (c/2)K]^{-1}B(\mu)A'(\nu_0) \quad (12.1)$$

and then substitute the result into Eq. (11.2) to obtain the (purely algebraic) equation

$$\{b - (c/2)k[I + (c/2)K]^{-1}B(\mu)\}A'(\nu_0) = 0. \quad (12.2)$$

Clearly, a nontrivial solution is possible only if

$$\det \{b - (c/2)k[I + (c/2)K]^{-1}B(\mu)\} = 0. \quad (12.3)$$

This is the exact criticality condition, implicitly relating the dimensions of the system to c. If $\|K\| < 2/c$, the inverse in Eq. (12.3) can be expanded in its Neumann series; in particular, if $\|K\| \ll 2/c$, we put $[I + (c/2)K]^{-1} \simeq I$; then Eq. (12.3) reduces to

$$\det [b - (c/2)kB(\mu)] = 0. \quad (12.4)$$

For the special cases of the slab and sphere, this expression reduces to the "first order" results of Mitsis. Better approximations can be obtained by retaining more terms in the expansion of the inverse.

13. SUMMARY AND DISCUSSION OF RESULTS

The main result of this paper is the representation of the neutron density $\rho(\mathbf{r})$ given by Eq. (4.11). The technique for obtaining the expansion coefficients $A_n(\nu)$ in this representation has been illustrated above, but a brief summary will be given here. First, a coordinate system and a prospective basis must be selected. Considerable freedom may be exercised in making the selection, although it will be convenient, when possible, to choose the coordinates and basis so the boundary $\dot{D}$ and the source distribution $S(\mathbf{r})$ have simple representations. The $q_{nm}(\nu, \mu)$ and $w_{nm}(\nu, \mu)$ must then be obtained either (as in Sec. 7) by evaluating $I_1(\mathbf{r}, \mathbf{r}'', \mu, \nu)$ and $I_2(\mathbf{r}, n, \mu, \nu)$ and expressing the results in the form of Eqs. (5.12) and (5.13) or (as in Sec. 8) by evaluating the surface integrals in Eqs. (5.6) and (5.14). These representations are possible if the prospective basis is sufficiently complete. The $A_n(\nu)$ must then be evaluated either by solving the singular equation (5.24) numerically or by solving the equivalent Fredholm equations (5.33) and (5.35). If the latter approach is chosen, $M(\mu)$ should, if possible, be chosen so the arguments of the operators T_1 and T_2 in Eqs. (5.34) and (5.36) satisfy the criteria of Sec. 6. This will facilitate evaluation of the kernels in Eqs. (5.33) and (5.35). Methods for solving these equations have been discussed in Sec. 12. Finally, $\rho(\mathbf{r})$ is obtained by returning to Eq. (4.11) and substituting the evaluated coefficients.

Since, in general, the expansion coefficients can not be evaluated exactly, the formalism given above does not provide a complete solution of the problem. However, as in Case's method, the resulting representation of $\rho(\mathbf{r})$ affords considerable insight into the structure of the solution, and the equations for the coefficients $A_n(\nu)$ readily lend themselves to systematic approximations—a feature not exhibited by the original equation [Eq. (2.1)] for $\rho(\mathbf{r})$.

The representation of $\rho(\mathbf{r})$ has a number of striking features which deserve comment. First, the basic structure of the solution, a superposition of discrete modes ($\nu = \nu_0$) and continuum modes [$\nu \in (0, 1)$], involving solutions of the Helmholtz equation, is seen to be the same for all problems of interest in one, two, or three dimensions, regardless of the shape of D or the choice of coordinate system and basis. This wide range of possible representations, corresponding to different choices of $\{R_n(\nu, \mathbf{r})\}$, is reminiscent of the many possible representations of the wavefunction in quantum mechanics.

In the present paper, only representations involving countable bases have been considered. Since the standard bases can be used to treat any problem of interest, this restriction does not imply any loss of generality; however, it may sometimes be more convenient to represent $\rho(\mathbf{r})$ in terms of a continuum basis, such as the one given as an example in Sec. 3. All of the above analysis has been extended to admit this possibility and these results, as well as others, will be reported in a later paper. The primary modification necessary when n takes on a continuum of values is that the coupled set of Fredholm equations for the $A_n(\nu)$ becomes a single, multidimensional Fredholm equation in the variables ν and n.

In the representation of $\rho(\mathbf{r})$, Eq. (4.11), continuum modes occur both in the source term and in the term involving the $R_n(\nu, \mathbf{r})$. Since the $A_n(\nu)$ vanish for $\nu \in (0, 1)$ in an infinite medium, the latter can be thought of as surface modes. Continuum modes are thus generated both by sources and by surfaces, and far from either, the density is well represented by the discrete ($\nu = \nu_0$) part of Eq. (4.11). These observations show that the results obtained previously for one-dimensional problems remain valid in the general case.

Finally, it should be noted that, while the entire analysis has been carried out for the transport

equation, it is highly plausible that a similar treatment will apply to any integral equation whose kernel has an integral representation involving the Green's function of a simple differential operator.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the encouragement and helpful criticisms provided by Professor N. J. McCormick of the University of Washington Nuclear Engineering Department.

APPENDIX A

Here we verify a result which was derived formally in Sec. 3:

$$1 = \frac{g(c, \nu_0)}{\nu_0^2} f(\nu_0, \mu) + \int_0^1 g(c, \nu) f(\nu, \mu) \frac{d\nu}{\nu^2}. \quad \text{(A1)}$$

Let $\Lambda(z)$ denote the function introduced by Case[2]:

$$\Lambda(z) = 1 - \frac{cz}{2} \int_{-1}^{1} \frac{d\mu}{\mu - z}. \quad \text{(A2)}$$

According to the Plemelj formulas,[9]

$$\Lambda^{\pm}(\nu) = \lim_{\epsilon \to 0} \Lambda(\nu \pm i\epsilon) = \lambda(\nu) \pm \tfrac{1}{2}\pi i c \nu, \quad \nu \in (0, 1). \quad \text{(A3)}$$

We deduce from Eq. (3.15) that

$$\Lambda^{+}(\nu)\Lambda^{-}(\nu) = 1/g(c, \nu), \quad \nu \in (0, 1), \quad \text{(A4)}$$

and from Eqs. (3.9) and (3.17) that

$$\Lambda'(\pm\nu_0) = \pm 2/[c\nu_0 g(c, \nu_0)]. \quad \text{(A5)}$$

Furthermore, $\Lambda(\pm\nu_0) = 0$. Now consider the integral

$$I = \frac{-\mu^2}{2\pi i} \int_C \frac{dz}{z(z^2 - \mu^2)\Lambda(z)}, \quad \text{(A6)}$$

where $\mu \in (-1, 1)$ and $C = C_1 \cup C_2$ is the contour shown in Fig. 1. The integral can be evaluated by summing the residues at $z = \pm\nu_0$; thus,

$$I = -2\mu^2/[\nu_0(\nu_0^2 - \mu^2)\Lambda'(\nu_0)]. \quad \text{(A7)}$$

On the other hand, I can be evaluated by adding the contributions from various parts of C. As C_1 is expanded to infinity, its contribution vanishes. The contribution from C_2, apart from the six semicircles around the poles at $\mu = 0$ and $\mu = \pm 1$, is

$$I_2 = \frac{\mu^2}{2\pi i} P \int_{-1}^{1} \frac{d\nu}{\nu(\nu^2 - \mu^2)} \left(\frac{1}{\Lambda^{-}(\nu)} - \frac{1}{\Lambda^{+}(\nu)} \right)$$
$$= c\mu^2 P \int_0^1 \frac{d\nu}{(\nu^2 - \mu^2)\Lambda^{+}(\nu)\Lambda^{-}(\nu)}. \quad \text{(A8)}$$

The contributions from the six semicircles are easily evaluated, and their sum is

$$I_3 = \lambda(\nu)/[\Lambda^{+}(\nu)\Lambda^{-}(\nu)] - 1. \quad \text{(A9)}$$

Equating I [Eq. (A7)] to the sum of I_2 and I_3, and using Eq. (A4), Eq. (A5), and the definition of $f(\nu, \mu)$, leads directly to Eq. (A1).

APPENDIX B

Here we verify the identities given in Sec. 6 for the operators T_1 and T_2. The procedure is identical to that used in Appendix A.

Let $h(z)$ be a meromorphic function, having no poles in (0, 1], which approaches a finite limit $h(\infty)$ as $z \to \infty$. To evaluate $T_2[h(\mu)]$, consider the integral

$$I = \frac{1}{2\pi i} \int_C X(z) h(z)\, dz, \quad \text{(B1)}$$

where $C = C_0 \cup C_1 \cup C_2$ is the contour shown in Fig. 2. Since $X(z)$ is analytic in the cut plane, we have

$$I = \sum_j X(z_j) \cdot \text{Res}\, h(z_j), \quad \text{(B2)}$$

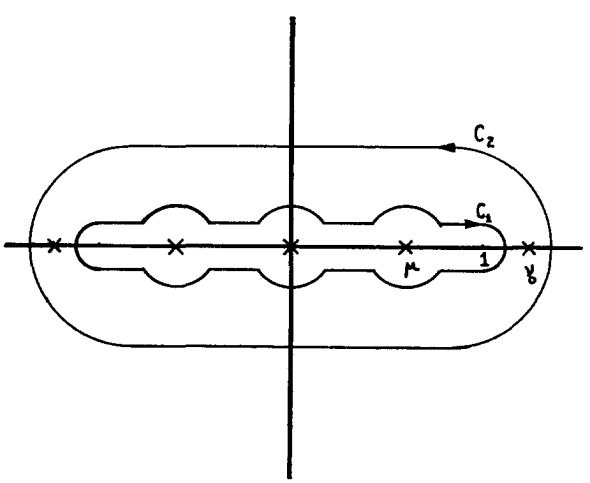


FIG. 1. Contour for evaluation of integral in Eq. (A6).

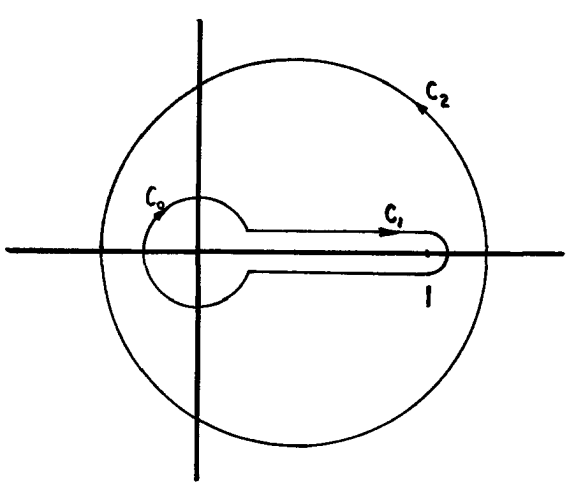


FIG. 2. Contour for evaluation of integral in Eq. (B1).

where the z_j are the poles of $h(z)$. On the other hand, I can be evaluated by adding the contributions from the parts of C. As the radius of C_0 goes to zero,

$$I_0 \to -X(0) \cdot \operatorname{Res} h(0). \tag{B3}$$

On C_1, the contribution is

$$I_1 = \frac{1}{2\pi i}\int_0^1 h(\mu)[X^+(\mu) - X^-(\mu)]\,d\mu = \int_0^1 \gamma(\mu)h(\mu)\,d\mu, \tag{B4}$$

where the last equation follows from the identity[2]

$$\frac{X^+(\mu)}{X^-(\mu)} = \frac{\Lambda^+(\mu)}{\Lambda^-(\mu)} \tag{B5}$$

and the definition of $\gamma(\mu)$. Let C_2 be expanded to infinity. Since $X(z) \sim 1/z$, clearly

$$I_2 \to h(\infty). \tag{B6}$$

Since $I = I_0 + I_1 + I_2$, the result given in Eq. (6.2) for $T_2[h(\mu)]$ follows.

To evaluate $T_1[h(\mu)]$, we need only write

$$P\frac{1}{\nu' - \mu} = \lim_{\epsilon \to 0} \frac{1}{2}\left(\frac{1}{\nu' - \mu + i\epsilon} + \frac{1}{\nu' - \mu - i\epsilon}\right) \tag{B7}$$

and apply the results already obtained for T_2. By defining

$$X(\mu) = \tfrac{1}{2}[X^+(\mu) + X^-(\mu)] \tag{B8}$$

and using Eq. (A4), Eq. (B5), and the identity[3]

$$1/[X^-(\mu)\Lambda^+(\mu)] = g(c,\mu)(\nu_0^2 - \mu^2)(1-c)X(-\mu), \tag{B9}$$

the result given in Eq. (6.1) is readily obtained.

To show that T_1 is a bounded operator, Eq. (B9) is used to rewrite Eq. (5.31) as

$$T_1[h(\mu)] = \lambda(\mu)g(c,\mu)h(\mu) - (1-c)(\nu_0^2 - \mu^2)X(-\mu)g(c,\mu) \times \int_0^1 \frac{\gamma(\nu)h(\nu)\,d\nu}{\nu - \mu}. \tag{B10}$$

According to the triangle inequality,

$$\|T_1[h(\mu)]\| = \left(\int_0^1 |T_1[h(\mu)]|^2\,d\mu\right)^{\frac{1}{2}} \le \|\lambda(\mu)g(c,\mu)h(\mu)\| + |1-c| \cdot \|g(c,\mu)(\nu_0^2 - \mu^2)X(-\mu)G(\mu)\|, \tag{B11}$$

where

$$G(\mu) = \int_0^1 \frac{\gamma(\nu)h(\nu)\,d\nu}{\nu - \mu}. \tag{B12}$$

But

$$\|g(c,\mu)(\nu_0^2 - \mu^2)X(-\mu)G(\mu)\| \le \max|g(c,\mu)(\nu_0^2 - \mu^2)X(-\mu)| \cdot \|G(\mu)\|, \tag{B13}$$

and $\|G(\mu)\|$ can be estimated by using the theorem given by Titchmarsh[18] concerning the norm of a Hilbert transform: If we put

$$\gamma_1(\nu) = \begin{cases} \pi\gamma(\nu), & \text{for } \nu \in (0,1), \\ 0, & \text{otherwise,} \end{cases} \tag{B14}$$

then

$$G(\mu) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\gamma_1(\nu)h(\nu)\,d\nu}{\nu - \mu}, \tag{B15}$$

and according to the theorem

$$\|G(\mu)\|^2 = \int_0^1 |G(\mu)|^2\,d\mu \le \int_{-\infty}^{\infty} |G(\mu)|^2\,d\mu = \int_{-\infty}^{\infty} |\gamma_1(\mu)h(\mu)|^2\,d\mu. \tag{B16}$$

Thus,

$$\|G(\mu)\| \le \pi \cdot \max|\gamma(\mu)| \cdot \|h(\mu)\|. \tag{B17}$$

To complete the estimate of $T_1[h(\mu)]$, consider the remaining term on the right-hand side of Eq. (B11), and use Eq. (3.15):

$$\|\lambda(\mu)g(c,\mu)h(\mu)\| = \left(\int_0^1 \frac{\lambda^2(\mu)}{\lambda^2(\mu) + (\frac{1}{2}\pi c\mu)^2} g(c,\mu)h^2(\mu)\,d\mu\right)^{\frac{1}{2}} \le \left(\int_0^1 g(c,\mu)h^2(\mu)\,d\mu\right)^{\frac{1}{2}} \le \max[g(c,\mu)]^{\frac{1}{2}} \cdot \|h(\mu)\|. \tag{B18}$$

Combining Eqs. (B11), (B13), (B17), and (B18) yields

$$\|T_1[h(\mu)]\| \le \{\max[g(c,\mu)]^{\frac{1}{2}} + \pi|1-c| \cdot \max[g(c,\mu) \cdot (\nu_0^2 - \mu^2)X(-\mu)] \cdot \max|\gamma(\mu)|\}\,\|h(\mu)\|. \tag{B19}$$

From the properties of $g(c,\mu)$ (Ref. 1), $X(-\mu)$ (Ref. 12), and the definition of $\gamma(\mu)$, it follows that the quantity in wavy brackets is finite and thus T_1 is a bounded operator.

[18] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, 1937), Theorem 90.

Wave Propagation in Sinusoidally Stratified Plasma Media*

K. F. CASEY AND J. R. MATTHES
Electrical Engineering Department, Air Force Institute of Technology, Dayton, Ohio
AND
C. YEH
Department of Engineering, University of California, Los Angeles, California

(Received 1 August 1968)

The problem of the propagation of electromagnetic waves in a sinusoidally stratified plasma media is treated analytically. The propagation characteristics of TE and TM waves are determined, respectively, from the characteristic equations of the resultant Mathieu and Hill equations. Detailed dispersion characteristics of TE and TM waves in an infinite stratified plasma medium and in waveguides filled longitudinally with this stratified dispersion material are given. It is found that, although the stop-band and pass-band structures exist for the ω–β diagrams of both TE and TM waves, detailed dispersion properties for TE and TM waves are quite different for most frequency ranges except when $[(\omega/\omega_{p0})^2 - 1]^2 \gg \delta^2$, where ω is the frequency of the propagating waves, ω_{p0} is the average plasma frequency of the inhomogeneous plasma medium, and δ is the amplitude of the sinusoidally varying term for the electron-density profile ($0 \leq \delta \leq 1$).

I. INTRODUCTION

The problem of electromagnetic wave propagation in a sinusoidally stratified medium is not only of interest from a theoretical point of view but also possesses many possible applications. For example, a section of waveguide filled with this type of inhomogeneous dielectric may be used as a band-pass filter in the millimeter-wave or optical region. The use of an ultrasonic standing wave as a modulating device for certain pressure-sensitive media, such as carbon disulfide, pentane, or nitric acid at optical frequencies to achieve a sinusoidally varying dielectric medium may be proposed.[1] Other applications of wave interactions in periodic media, such as the deflection of laser beams,[2] the measurement of acoustic properties in crystals,[3] and Čerenkov radiation in periodically stratified media,[4] can also be found. In recent years, the problem of wave propagation in sinusoidally stratified dielectric nondispersive media has been considered in detail by several authors.[5] However, the results are not applicable for dispersive media.

The purpose of this investigation is to consider the propagation characteristics of waves in a dispersive sinusoidally varying medium. Specifically, the dispersive medium is assumed to be an inhomogeneous cold plasma with a sinusoidally varying free electron density profile.

Two types of waves may exist: one with its electric vector transverse to the direction of the inhomogeneity, called a TE wave, and the other with its magnetic vector transverse to the direction of inhomogeneity, called a TM wave. Both types of waves will be treated. Detailed dispersion characteristics of TE and TM waves in an infinite stratified plasma medium and in waveguides filled longitudinally with this inhomogeneous plasma will be presented.

It is hoped that these results will be useful in the diagnostics of plasmas[6] and in the study of wave propagation in solids.[7]

II. FORMULATION OF THE PROBLEM

It is assumed that the inhomogeneous plasma medium under consideration fills the entire space and possesses a plasma frequency

$$\omega_p^2(z) = n(z)e^2/m\epsilon_0, \tag{1}$$

where z is the axial coordinate, $n(z)$ is the free electron density, e is the electron charge, m is the electron mass, and ϵ_0 is the free-space permittivity.

The source-free vector wave equations in this medium are

$$\nabla \times \nabla \times \mathbf{E} - k_0^2(\epsilon(z)/\epsilon_0)\mathbf{E} = 0, \tag{2}$$

$$\nabla \times \nabla \times \mathbf{H} - \frac{\nabla\epsilon(z)}{\epsilon(z)} \times \nabla \times \mathbf{H} - k_0^2\left(\frac{\epsilon(z)}{\epsilon_0}\right)\mathbf{H} = 0, \tag{3}$$

* Supported partly by the Office of Naval Research and partly by the National Science Foundation.

[1] C. Yeh and Z. A. Kaprielian, "On Inhomogeneously Filled Waveguides," USCEC Dept. 84-206, Electrical Engineering Department, University of Southern California, Los Angeles, 1963.

[2] M. G. Cohn and E. I. Gordon, Bell System Tech. J. **44**, 693 (1965).

[3] H. H. Parker, E. F. Kelly, and D. I. Bulef, Appl. Phys. Letters **5**, 7 (1964).

[4] K. F. Casey, C. Yeh, and Z. A. Kaprielian, Phys. Rev. **140**, B768 (1965).

[5] T. Tamir, H. C. Wang, and A. A. Oliner, IEEE Trans. Microwave Theory Tech., MT **12**, 323 (1964).

[6] M. A. Heald and C. B. Wharton, *Plasma Diagnostics with Microwaves* (John Wiley & Sons, Inc., New York, 1965).

[7] L. Brillouin, *Wave Propagation in Periodic Structures* (Dover Publications, Inc., New York, 1953).

where $\mathbf{E}$ and $\mathbf{H}$ are, respectively, the electric and magnetic field vectors, $k_0^2 = \omega^2\mu_0\epsilon_0$, and a time dependence $e^{-i\omega t}$ is assumed. μ_0 is the free-space permeability. The dielectric permittivity of the inhomogeneous cold plasma is related to the plasma frequency by the following relation:

$$\epsilon(z) = \epsilon_0[1 - \omega_p^2(z)/\omega^2]. \tag{4}$$

It can be shown that all field components in this medium can be obtained from the scalar quantities $\Phi(x, y, z)$ and $\Psi(x, y, z)$ as follows[1]:

$$\mathbf{E}^{(m)} = \nabla \times [\Phi(x, y, z)\mathbf{e}_z], \tag{5}$$

$$\mathbf{H}^{(m)} = (-i/\omega\mu_0)\nabla \times \nabla \times [\Phi(x, y, z)\mathbf{e}_z], \tag{6}$$

for transverse electric waves, and

$$\mathbf{H}^{(e)} = \nabla \times [\Psi(x, y, z)\mathbf{e}_z], \tag{7}$$

$$\mathbf{E}^{(e)} = (i/\omega\epsilon(z))\nabla \times \nabla \times [\Psi(x, y, z)\mathbf{e}_z], \tag{8}$$

for transverse magnetic waves. $\mathbf{e}_z$ is the unit vector in the z direction. Upon substituting (5) into (3) and (7) into (4), carrying out the vector operations, and separating variables in rectangular coordinates, one obtains

$$\Phi(x, y, z) = \begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(sx)\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(wy)U^{(1),(2)}(z), \tag{9}$$

$$\Psi(x, y, z) = \begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(px)\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(qy)V^{(1),(2)}(z), \tag{10}$$

where s, w, p, and q are separation constants. $U^{(1),(2)}(z)$ and $V^{(1),(2)}(z)$ satisfy, respectively, the differential equations

$$\left\{\frac{d^2}{dz^2} + \left[k_0^2\left(\frac{\epsilon(z)}{\epsilon_0}\right) - s^2 - w^2\right]\right\}U^{(1),(2)}(z) = 0 \tag{11}$$

and

$$\left\{\frac{d^2}{dz^2} - \left(\frac{d\epsilon(z)}{dz}\right)\frac{1}{\epsilon(z)}\frac{d}{dz} + \left[k_0^2\left(\frac{\epsilon(z)}{\epsilon_0}\right) - p^2 - q^2\right]\right\} \times V^{(1),(2)}(z) = 0. \tag{12}$$

Introducing the dimensionless variable $\xi = \pi z/d$, where d has the dimension of length and will be defined later, it can be shown that Eqs. (11) and (12) can be put into the following form:

$$\left[\frac{d^2}{d\xi^2} + \lambda(\xi)\right]W^{(1),(2)}(\xi) = 0, \tag{13}$$

where

$$\lambda(\xi) = \frac{d^2}{\pi^2}\left[k_0\left(\frac{\epsilon(\xi)}{\epsilon_0}\right) - s^2 - w^2\right] \tag{14}$$

if

$$W^{(1),(2)}(\xi) = U^{(1),(2)}(\xi), \tag{15}$$

and

$$\lambda(\xi) = \frac{1}{\epsilon^2(\xi)}\left\{[\epsilon^{\frac{1}{2}}(\xi)]'' + \frac{\epsilon'(\xi)}{\epsilon(\xi)}[\epsilon^{\frac{1}{2}}(\xi)]' + \frac{d^2}{\pi^2}\left[k_0\left(\frac{\epsilon(\xi)}{\epsilon_0}\right) - p^2 - q^2\right]\epsilon^{\frac{1}{2}}(\xi)\right\} \tag{16}$$

if

$$W^{(1),(2)}(\xi) = \epsilon^{-\frac{1}{2}}(\xi)V^{(1),(2)}(\xi). \tag{17}$$

The primes indicate the derivative of the function with respect to ξ.

III. THE SINUSOIDALLY STRATIFIED PLASMA MEDIUM

If the free electron density distribution is assumed to have a sinusoidal stratification as follows:

$$n(z) = n_0\left(1 - \delta\cos\frac{2\pi z}{d}\right), \tag{18}$$

where n_0 is the average electron density, δ is a known constant with $0 \leq \delta \leq 1$, and d denotes the period of the sinusoidal variation, then the dielectric constant of this plasma medium is

$$\epsilon(\xi) = \epsilon_0\left\{\left(1 - \frac{\omega_{p0}^2}{\omega^2}\right) + \frac{\omega_{p0}^2}{\omega^2}\delta\cos 2\xi\right\} \tag{19}$$

and

$$\omega_{p0}^2 = n_0e^2/m\epsilon_0.$$

Substituting Eq. (19) into (14) gives

$$\lambda^{\mathrm{TE}}(\xi) = \left(\frac{k_0d}{\pi}\right)^2 - \left(\frac{k_{p0}d}{\pi}\right)^2 - \left(\frac{sd}{\pi}\right)^2 - \left(\frac{wd}{\pi}\right)^2 + \left(\frac{k_{p0}d}{\pi}\right)^2\delta\cos 2\xi, \tag{20}$$

with

$$k_{p0}^2 = \omega_{p0}^2\mu_0\epsilon_0,$$

for the transverse electric wave; substituting Eq. (19) into (16) gives

$$\lambda^{\mathrm{TM}}(\xi) = -\frac{2(\omega_{p0}/\omega)^2\delta\cos 2\xi}{1 - (\omega_{p0}/\omega)^2(1 - \delta\cos 2\xi)} - \frac{3(\omega_{p0}/\omega)^2\delta^2\sin^2 2\xi}{[1 - (\omega_{p0}/\omega)^2(1 - \delta\cos 2\xi)]^2} + \left(\frac{k_0d}{\pi}\right)^2\left[1 - \left(\frac{\omega_{p0}}{\omega}\right)^2(1 - \delta\cos 2\xi)\right] - \left(\frac{pd}{\pi}\right)^2 - \left(\frac{qd}{\pi}\right)^2, \tag{21}$$

for the transverse magnetic wave. Since, according to Eqs. (20) and (21), $\lambda^{\mathrm{TE}}(\xi)$ or $\lambda^{\mathrm{TM}}(\xi)$ is an even periodic function, it can, therefore, be represented by a

Fourier cosine series

$$\lambda^{\mathrm{TE,TM}}(\xi) = \theta_0^{\mathrm{TE,TM}} + 2\sum_{n=1}^{\infty}\theta_n^{\mathrm{TE,TM}}\cos 2n\xi, \quad (22)$$

where

$$\theta_0^{\mathrm{TE}} = \left(\frac{k_0 d}{\pi}\right)^2 - \left(\frac{k_{p0}d}{\pi}\right)^2 - \left(\frac{sd}{\pi}\right)^2 - \left(\frac{wd}{\pi}\right)^2, \quad (23a)$$

$$\theta_1^{\mathrm{TE}} = \tfrac{1}{2}\delta(k_{p0}d/\pi)^2, \quad (23b)$$

$$\theta_n^{\mathrm{TE}} = 0, \quad n \geq 2, \quad (23c)$$

and

$$\theta_0^{\mathrm{TM}} = \left(\frac{k_0 d}{\pi}\right)^2 - \left(\frac{k_{p0}d}{\pi}\right)^2 - \left(\frac{pd}{\pi}\right)^2 - \left(\frac{qd}{\pi}\right)^2 - \left[\frac{1}{(1-\Lambda^2)^{\frac{1}{2}}} - 1\right], \quad (24a)$$

$$\theta_1^{\mathrm{TM}} = \tfrac{1}{2}\delta\left(\frac{k_{p0}d}{\pi}\right)^2 + \frac{4b^3 - 2b}{b^2 - 1}, \quad (24b)$$

$$\theta_n^{\mathrm{TM}} = \frac{(3n+1)b^{n+2} - (3n-1)b^n}{b^2 - 1}, \quad n \geq 2, \quad (24c)$$

with

$$b = \Lambda^{-1} - \Lambda^{-1}(1-\Lambda^2)^{\frac{1}{2}}, \quad (24d)$$

$$\Lambda = \delta/(1 - \omega^2/\omega_{p0}^2). \quad (24e)$$

It should be noted that the above Fourier series representation for the TM case converges absolutely when

$$0 \leq |\delta/(1 - \omega^2/\omega_{p0}^2)| < 1 \quad (25)$$

with $0 \leq \delta \leq 1$.

IV. SOLUTIONS OF HILL'S EQUATION

Substituting (22) into (13), one obtains

$$\left[\frac{d^2}{d\xi^2} + \theta_0^{\mathrm{TE,TM}} + 2\sum_{n=1}^{\infty}\theta_n^{\mathrm{TE,TM}}\cos 2n\xi\right]W^{(1),(2)}(\xi) = 0 \quad (26)$$

which is the general form of Hill's equation.[8,9] It is known that two types of solutions of Hill's equations exist: one called the stable type and the other called the unstable type. In order to have propagating waves in the z direction, only the stable type is allowed.

With the help of Floquet's Theorem,[7-9] the solutions of Hill's equation can be expressed in the following form:

$$W^{(1),(2)}(\xi) = e^{\pm i\nu\xi}\sum_{n=-\infty}^{\infty}C_n(\nu)e^{\pm 2in\xi}, \quad (27)$$

where ν, the characteristic exponent, and $C_n(\nu)$ are as yet unknown. After substituting (27) into (26) and simplifying, one obtains the following recursion relations:

$$-(\nu + 2n)^2 C_n + \sum_{m=-\infty}^{\infty}\theta_m^{\mathrm{TE,TM}}C_{n-m} = 0, \quad n = \cdots -2, -1, 0, 1, 2, \cdots, \quad (28)$$

with $\theta_{-m}^{\mathrm{TE,TM}} = \theta_m^{\mathrm{TE,TM}}$. It is understood that $\nu = \nu^{\mathrm{TE}}$ and $C_n = C_n^{\mathrm{TE}}$ when θ_m^{TE} are used, while $\nu = \nu^{\mathrm{TM}}$ and $C_n = C_n^{\mathrm{TM}}$ when θ_m^{TM} are used. Equation (28) is a set of an infinite number of homogeneous linear algebraic equations in C_n. For a nontrivial solution to exist the characteristic number ν and the coefficients θ_m must satisfy the characteristic equation of the Hill equation[9]:

$$\sin^2\frac{\pi\nu}{2} = \Delta(0)\sin^2\frac{\pi(\theta_0^{\mathrm{TE,TM}})^{\frac{1}{2}}}{2}. \quad (29)$$

$\Delta(0)$ is the determinant of the matrix $[M]$ whose elements are

$$M_{mm} = 1,$$

$$M_{mn} = \frac{-\theta_{m-n}^{\mathrm{TE,TM}}}{4m^2 - \theta_0^{\mathrm{TE,TM}}}, \quad m \neq n. \quad (30)$$

The characteristic number ν can be obtained from (29).

Real values of ν yield stable solutions to Hill's equation, while complex values of ν produce unstable solutions. Physically speaking the stable solutions correspond to modulated propagating waves, and the unstable solutions correspond to damped or growing waves. For the present problem, the fields for the growing waves do not satisfy the radiation condition at infinity, hence they must be omitted.

Numerical computation has been carried out for (29). The values for the infinite determinant $\Delta(0)$ were obtained by the successive approximation method.[10] In other words, computations were carried out for a 3×3 determinant, a 5×5 determinant, etc., until the desired accuracy was reached. It was found (numerically) that the infinite determinant converges quite rapidly within the present region of interest. For example, for small values of δ and k_{p0}, such that $\delta \leq 0.1$ and $(k_{p0}d/\pi)^2 \leq 0.5$, at no time was any determinant greater than 7×7 required to achieve an accuracy of three significant figures. For large values of δ and k_{p0}, no determinants of order greater than 15×15 were required to obtain the desired accuracy.

V. PROPAGATION CHARACTERISTICS OF TE WAVES

Returning now to the problem of obtaining the propagation characteristics of waves in a sinusoidally

[8] J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer-Verlag, Berlin, 1954).

[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953).

[10] L. Kantorovich and V. Krylov, *Approximate Methods of Higher Analysis* (Interscience Publishers, New York, 1958).

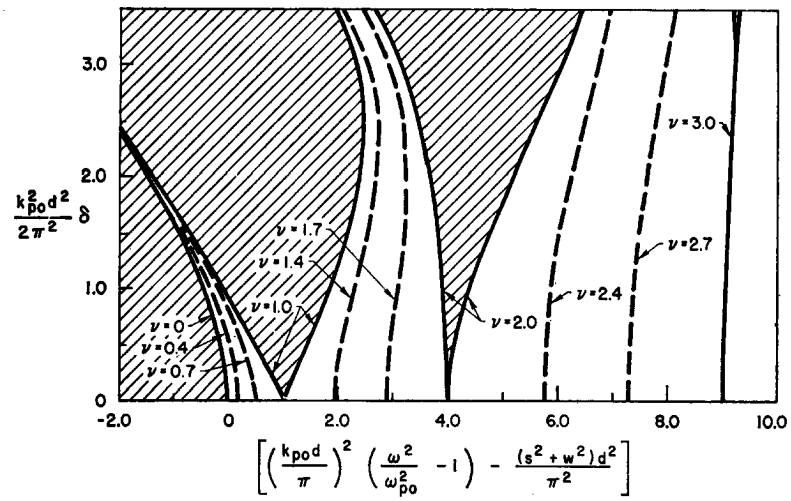


FIG. 1. Stability chart for Mathieu's equation from which the propagation characteristics of TE waves may be determined. Unstable regions are shaded.

stratified plasma medium, we note, upon substituting Eq. (22) for TE waves into Eq. (13), that the resultant differential equation for TE waves is the Mathieu equation, a special case of Hill's equation. It is customary to express the results of computation for the characteristic exponents of Mathieu functions in terms of a "stability diagram."[9,11] Figure 1 shows the stability diagram which was obtained from previously tabulated values for the characteristic exponents of Mathieu functions. The unshaded areas are the "stable regions" wherein ν is purely real; the shaded areas are the "unstable regions" wherein ν is complex.

A. Infinite Region Filled With Sinusoidally Stratified Plasma

The transverse electric field components of a TE wave in an infinite medium filled with sinusoidally stratified plasma can be obtained from Eqs. (5) and (6):

$$E_x^{\mathrm{TE}} = \sum_{n=-\infty}^{\infty} iwC_n^{\mathrm{TE}} e^{isx} e^{iwy} e^{i(\nu+2n)\pi z/d}, \tag{31}$$

$$E_y^{\mathrm{TE}} = \sum_{n=-\infty}^{\infty} -isC_n^{\mathrm{TE}} e^{isx} e^{iwy} e^{i(\nu+2n)\pi z/d}, \tag{32}$$

where the coefficients C_n^{TE} can be determined from Eq. (28) in terms of C_0^{TE}. The propagation constant β is related to ν by the equation $\beta = \nu\pi/d$. C_0^{TE} is obtained from a normalization condition. All magnetic field components may be found from Maxwell's equations.

Unlike the case of a TE wave propagating in an infinite homogeneous plasma in which β is simply related to s and w by the following:

$$\beta^2 = \omega^2\mu_0\epsilon_0(1 - \omega_p^2/\omega^2) - (s^2 + w^2), \tag{33}$$

where ω_p is the plasma frequency of the homogeneous plasma, the propagation constant β for the inhomogeneous case is related to s and w through the stability diagram given by Fig. 1. Real values of β as a function of real values of $(s^2 + w^2)$ for fixed values of δ, $k_0 d$, and $k_{p0}d$ are shown in Fig. 2. It is recalled that complex values of β indicate the presence of damped waves (i.e., nonpropagating waves). s and w are taken to be real. The unshaded regions in these figures indicate the regions in which β is real (i.e., regions in which propagating waves may exist). Equation (33) is also plotted in Fig. 2 with $\omega_p = \omega_{p0}$. As one can see, the propagation characteristics of waves in the inhomogeneous plasma are significantly modified from those in the homogeneous plasma. The presence of stop-band and pass-band regions in Fig. 2 is characteristic of wave propagation in periodic structures.

B. Waveguide Filled Longitudinally With Sinusoidally Stratified Plasma

It is assumed that a rectangular waveguide of dimensions h_1 and h_2 is filled completely with an

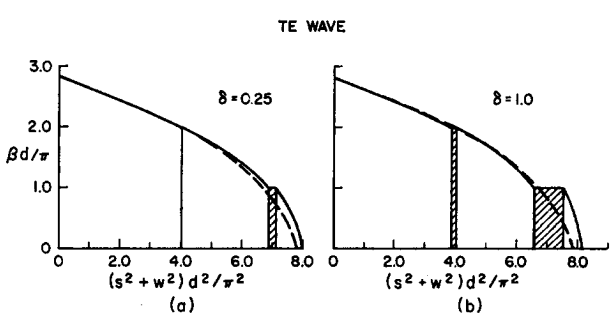


FIG. 2. The propagation constant $\beta d/\pi$ as a function of $(s^2 + w^2)\,d^2/\pi^2$ with $(k_{p0}d/\pi^2) = 1.0$ and $(k_0 d/\pi)^2 = 9.0$. Stop bands are shaded. The dashed line represents the behavior of the propagation constant for an equivalent homogeneous plasma medium with $\omega_p = \omega_{p0}$ (TE wave).

[11] T. Tamir, Math. Computation 16, 100 (1962).

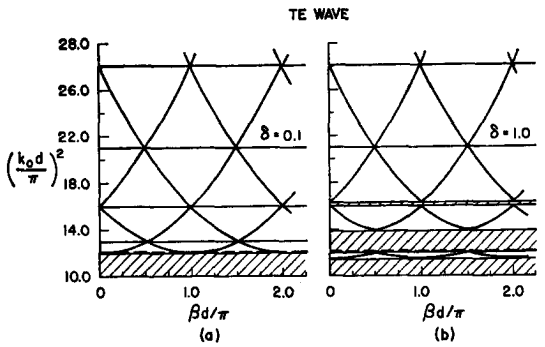


FIG. 3. Frequency as a function of propagation constant with $(k_{p0}d/\pi)^2 = 2.0$ and $(\gamma d/\pi)^2 = 10.0$. Stopbands are shaded. Dashed line indicates the cutoff frequency for a waveguide filled with an equivalent homogeneous plasma medium with $\omega_p = \omega_{p0}$ (TE wave).

inhomogeneous plasma medium, which varies sinusoidally in the longitudinal direction. The general expressions for the transverse electric field components of a TE wave are

$$E_x^{\mathrm{TE}} = \sum_{m=1}^{\infty}\sum_{r=1}^{\infty}\sum_{n=-\infty}^{\infty} - C_n^{\mathrm{TE}} \frac{r\pi}{h_2} \times \cos\frac{m\pi x}{h_1} \sin\frac{r\pi y}{h_2} e^{i(\nu+2n)\pi z/d}, \quad (34)$$

$$E_y^{\mathrm{TE}} = \sum_{m=1}^{\infty}\sum_{r=1}^{\infty}\sum_{n=-\infty}^{\infty} C_n^{\mathrm{TE}} \frac{m\pi}{h_1} \times \sin\frac{m\pi x}{h_1} \cos\frac{r\pi y}{h_2} e^{i(\nu+2n)\pi z/d}, \quad (35)$$

with $\beta = \nu\pi/d$. The C_n^{TE} are arbitrary constants and C_n^{TE} can be determined from Eq. (28) in terms of C_0^{TE}. Expressions for the magnetic field components can easily be derived from Maxwell's equations.

The dispersion relations expressed in terms of the ω–β diagrams can be found from Fig. 1 for fixed values of δ, $k_{p0}d$, and $s^2 + w^2$, where

$$(\gamma^{\mathrm{TE}})^2 = s^2 + w^2 = \left(\frac{m\pi}{h_1}\right)^2 + \left(\frac{r\pi}{h_2}\right)^2. \quad (36)$$

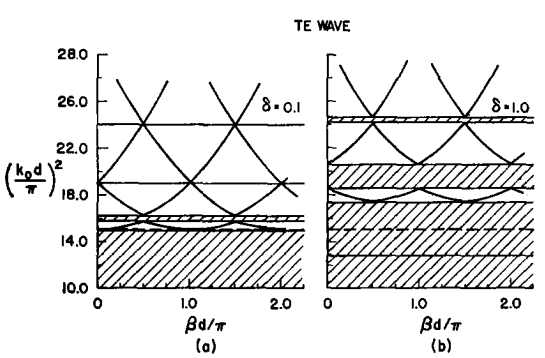


FIG. 4. Frequency as a function of propagation constant with $(k_{p0}d/\pi)^2 = 5.0$ and $(\gamma d/\pi)^2 = 10.0$. Stop bands are shaded. Dashed line indicates the cutoff frequency for a waveguide filled with an equivalent homogeneous plasma medium with $\omega_p = \omega_{p0}$ (TE wave).

The ω–β diagrams are given for $\delta = 0.1$, 1.0, $(k_{p0}d/\pi)^2 = 2.0$, 5.0 and $(s^2 + w^2)(d/\pi)^2 = 5.0$, 10.0 in Figs. 3 and 4. The pass-band and stop-band characteristics can clearly be seen. Furthermore, an increase in δ results in a corresponding increase in the bandwidth of the stop bands. Also shown in these figures is the cutoff frequency for an identical waveguide filled with homogeneous plasma which has a plasma frequency equal to ω_{p0}. It is noted that the cutoff frequency for the equivalent homogeneous case is above the cutoff frequency of the first pass band for the inhomogeneous case.

VI. PROPAGATION CHARACTERISTICS OF TM WAVES

Unlike the TE case, the TM case requires the solution of the more complex Hill equation. It is more straightforward to obtain the propagation characteristics of TM waves directly from Eq. (29) than to use the stability diagram, as described earlier for TE waves.

It is recalled that, in order that Eq. (12) for TM waves may be put in the form of the Hill equation, a very important limitation on the ratio ω/ω_{p0} must be satisfied, i.e., from Eq. (25),

$$0 \leq \frac{1}{|1 - \omega^2/\omega_{p0}^2|} < \frac{1}{\delta} \quad (37)$$

with $0 \leq \delta \leq 1$. A sketch of the above relation is given in Fig. 5. The shaded area indicates the forbidden region in which the relation (37) is not satisfied. It was found numerically that, for moderate values of $(p^2 + q^2)(d/\pi)^2$, the lower cutoff frequency of the first pass band occurs at frequencies considerably above ω_{p0}.

A. Infinite Region Filled With Sinusoidally Stratified Plasma

The transverse magnetic-field components of a TM wave in an infinite medium filled with sinusoidally stratified plasma can be obtained from Eqs. (7)

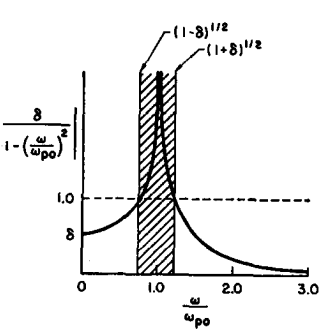


FIG. 5. A plot of Eq. (37). Shaded region indicates the region in which the convergence of the series expansion breaks down with $|\delta| \leq 1.0$.

and (8):

$$H_x^{\mathrm{TM}} = \sum_{n=-\infty}^{\infty} iqC_n^{\mathrm{TM}} e^{ipx} e^{iqy} e^{i(\nu+2n)\pi z/d} \times \left[1 - \left(\frac{\omega_{p0}}{\omega}\right)^2 \left(1 - \delta \cos\frac{2\pi z}{d}\right)\right]^{\frac{1}{2}}, \quad (38)$$

$$H_y^{\mathrm{TM}} = \sum_{n=-\infty}^{\infty} -ipC_n^{\mathrm{TM}} e^{ipx} e^{iqy} e^{i(\nu+2n)\pi z/d} \times \left[1 - \left(\frac{\omega_{p0}}{\omega}\right)^2 \left(1 - \delta \cos\frac{2\pi z}{d}\right)\right]^{\frac{1}{2}}, \quad (39)$$

with $\beta = \nu\pi/d$. The coefficients C_n^{TM} can be determined from Eq. (28) in terms of C_0^{TM}. The electric field components can be found from Maxwell's equations.

Real values of β as a function of real values of $(p^2 + q^2)$ for fixed values of δ, $k_0 d$, and $k_{p0} d$ are shown in Fig. 6. Again, the pass-band and stop-band features are apparent in this figure. Results are rather similar to those given in Fig. 2 for TE wave, although quantitatively the results are different.

B. Waveguide Filled Longitudinally With Sinusoidally Stratified Plasma

The general expressions for the transverse magnetic-field components of a TM wave in a rectangular waveguide of dimensions h_1 and h_2 filled longitudinally with an inhomogeneous plasma medium are

$$H_x^{\mathrm{TM}} = \sum_{m=1}^{\infty}\sum_{r=1}^{\infty}\sum_{n=-\infty}^{\infty} C_n^{\mathrm{TM}} \sin\frac{m\pi x}{h_1} \cos\frac{r\pi y}{h_2} e^{-(\nu+2n)\pi z/d} \times \frac{r\pi}{h_2}\left[1 - \left(\frac{\omega_{p0}}{\omega}\right)^2 \left(1 - \delta\cos\frac{2\pi z}{d}\right)\right]^{\frac{1}{2}}, \quad (40)$$

$$H_y^{\mathrm{TM}} = \sum_{m=1}^{\infty}\sum_{r=1}^{\infty}\sum_{n=-\infty}^{\infty} C_n^{\mathrm{TM}} \cos\frac{m\pi x}{h_1} \sin\frac{r\pi y}{h_2} e^{i(\nu+2n)\pi z/d} \times \left(-\frac{m\pi}{h_1}\right)\left[1 - \left(\frac{\omega_{p0}}{\omega}\right)^2 \left(1 - \delta\cos\frac{2\pi z}{d}\right)\right]^{\frac{1}{2}}, \quad (41)$$

where the coefficients C_n^{TM} can be determined from Eq. (28) in terms of C_0^{TM}. The electric-field components can be found from Maxwell's equations.

Solving Eq. (29) for fixed values of δ, $k_{p0} d$, and $(p^2 + q^2)$, where

$$(\gamma^{\mathrm{TM}})^2 = p^2 + q^2 = \left(\frac{m\pi}{h_1}\right)^2 + \left(\frac{r\pi}{h_2}\right)^2,$$

yields the propagation constant β as a function of ω as shown in Figs. 7 and 8. Again, the propagation characteristics of TM and TE waves are similar. However, the bandwidths, center frequencies of the stop bands, and the waveguide cutoff frequencies vary considerably for the two modes. Also noted from the definitions (23) and (24) is that $\theta_n^{\mathrm{TM}} \simeq \theta_n^{\mathrm{TE}}$ provided that

$$\frac{\delta^2}{[(\omega/\omega_{p0})^2 - 1]^2} \to 0.$$

In other words, if $[(\omega/\omega_{p0})^2 - 1]^2 \gg \delta^2$, the propagation characteristics of TE and TM waves are almost the same. Otherwise, the results for TE and TM waves will be quite different. For example, in the lower frequency ranges, the bandwidths of the stop bands for the TM waves are greater than the bandwidths of

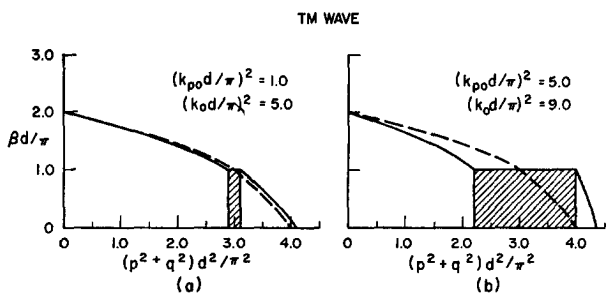


FIG. 6. The propagation constant $\beta d/\pi$ as a function of $(p^2 + q^2)\, d^2/\pi^2$ with $\delta = 0.25$. Stop bands are shaded. The dashed line represents the behavior of the propagation constant for an equivalent homogeneous plasma medium with $\omega_p = \omega_{p0}$ (TM wave).

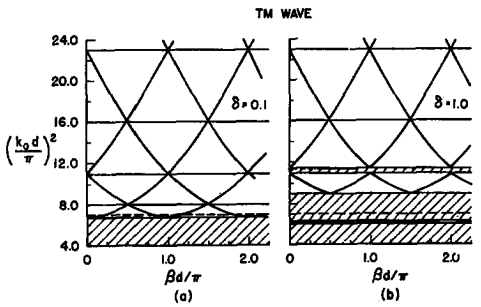


FIG. 7. Frequency as a function of propagation constant with $(k_{p0} d/\pi)^2 = 2.0$ and $(\gamma d/\pi)^2 = 5.0$. Stop bands are shaded. Dashed line indicates the cutoff frequency for a waveguide filled with an equivalent homogeneous plasma with $\omega_p = \omega_{p0}$ (TM wave).

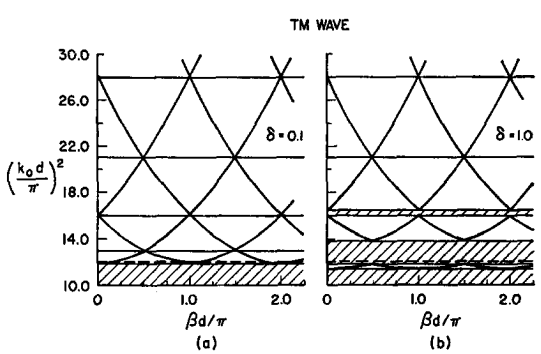


FIG. 8. Frequency as a function of propagation constant with $(k_{p0} d/\pi)^2 = 2.0$ and $(\gamma d/\pi)^2 = 10.0$. Stop bands are shaded. Dashed line indicates the cutoff frequency for a waveguide filled with an equivalent homogeneous plasma with $\omega_p = \omega_{p0}$ (TM wave).

the corresponding stop bands for the TE waves, and the waveguide cutoff frequencies for the TM waves are lower than those of the TE waves. (Compare Fig. 8 with Fig. 3.) As the frequency is increased beyond the second pass band, the bandwidths of the stop bands are approximately the same.

The ω–β diagrams given by Figs. 3 and 4 for TE waves and Figs. 7 and 8 for TM waves are also applicable for waveguides of arbitrary cross-section. For example, in a circular waveguide of radius ρ_0, $\gamma^{TE,TM}$ is given by $\Gamma_{mr}^{TE,TM}/\rho_0$, where $\Gamma_{mr}^{TE,TM}$ are, respectively, the rth roots of the equations

$$J'_m(\Gamma_{mr}^{TE}) = 0$$

and

$$J_m(\Gamma_{mr}^{TM}) = 0.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Expansion of Lie Groups and Representations of $SL(3, C)$

N. MUKUNDA
Tata Institute of Fundamental Research, Bombay, India

(Received 9 July 1968)

We show that in any unitary representation of $SL(3, C)$ belonging to the principal series, the generators can be expressed as functions of the generators of a suitably chosen unitary representation of the nonsemisimple group $SU(3) \times T_8$. Both the nondegenerate and the degenerate series are considered. Since the representations of $SU(3) \times T_8$ in an $SU(3)$ basis are known, this provides a solution to the multiplicity problem in setting up the $SL(3, C)$ representations in an $SU(3)$ basis.

1. INTRODUCTION

Recent developments indicate that unitary infinite-dimensional representations of certain noncompact Lie groups may be usefully employed in describing the physics of elementary particles.[1] For such purposes, the primary need is to have the relevant representations in as explicit and manageable a form as possible.

In setting up the representations of a semisimple noncompact Lie group, one has to decide from physical considerations which complete commuting set of operators is to be used to provide labels for the various states that appear in the representation. In some cases, this set will be associated with a compact subgroup of the original group, in others with a noncompact one. Even when a complete commuting set has been found, the problem of finding the possible eigenvalues of the operators in this set, and then of evaluating the matrix elements of all the generators of the group, can be an almost insoluble one. This difficulty is not necessarily related to the noncompactness of the Lie group one is interested in. For example, in the use of the SU_3 group in the nuclear shell model, the relevant subgroup is the R_3 subgroup corresponding to space rotations; and one wishes to obtain the SU_3 representations in a basis of angular-momentum eigenstates, rather than eigenstates of the $SU_2 \otimes U$ subgroup of SU_3.[2] (It is the latter that is relevant in the use of SU_3 in elementary particle physics.) And the difficulties of obtaining the SU_3 representations in an R_3 basis are well known.

We are concerned with the problem of obtaining the unitary infinite-dimensional representations of a semisimple noncompact Lie group in a basis made up of unitary finite-dimensional representations of the maximal compact subgroup. Here the situation is as follows. For the groups of the form $SU(n, 1)$ or $SO(n, 1)$ [pseudo-unitary and pseudo-orthogonal groups in $(n + 1)$ complex or real dimensions, respectively], it is known that in any unitary irreducible representation (UIR) of the noncompact group, each UIR of the maximal compact subgroup appears at most once.[3] [The maximal compact subgroups in the above cases are $SU(n) \otimes U(1)$ and $SO(n)$, respectively.] Thus these groups are "multiplicity-free," and this property is closely related to a similar property for the compact groups $SU(n + 1)$ and $SO(n + 1)$ in the reduction of their UIR's under $SU(n) \otimes U(1)$ and

[1] As an example, see the proceedings of the session on "Infinite Representations of Particles," in *Proceedings of the 1967 International Conference on Particles and Fields, University of Rochester, Rochester, N.Y.* (Interscience Publishers, Inc., New York, 1967).

[2] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128 (1958); V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960), **23**, 177 (1961); G. Racah in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1964).

[3] UIR's of these groups for all n have been obtained by U. Ottoson, Commun. Math. Phys. **8**, 228 (1968); **10**, 114 (1968). See also L. C. Biedenharn, *Noncompact Groups in Particle Physics*, Y. Chow, Ed. (W. A. Benjamin, Inc., New York, 1966).

the corresponding stop bands for the TE waves, and the waveguide cutoff frequencies for the TM waves are lower than those of the TE waves. (Compare Fig. 8 with Fig. 3.) As the frequency is increased beyond the second pass band, the bandwidths of the stop bands are approximately the same.

The ω–β diagrams given by Figs. 3 and 4 for TE waves and Figs. 7 and 8 for TM waves are also applicable for waveguides of arbitrary cross-section. For example, in a circular waveguide of radius ρ_0, $\gamma^{TE,TM}$ is given by $\Gamma_{mr}^{TE,TM}/\rho_0$, where $\Gamma_{mr}^{TE,TM}$ are, respectively, the rth roots of the equations

$$J'_m(\Gamma_{mr}^{TE}) = 0$$

and

$$J_m(\Gamma_{mr}^{TM}) = 0.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Expansion of Lie Groups and Representations of $SL(3, C)$

N. MUKUNDA
Tata Institute of Fundamental Research, Bombay, India

(Received 9 July 1968)

We show that in any unitary representation of $SL(3, C)$ belonging to the principal series, the generators can be expressed as functions of the generators of a suitably chosen unitary representation of the nonsemisimple group $SU(3) \times T_8$. Both the nondegenerate and the degenerate series are considered. Since the representations of $SU(3) \times T_8$ in an $SU(3)$ basis are known, this provides a solution to the multiplicity problem in setting up the $SL(3, C)$ representations in an $SU(3)$ basis.

1. INTRODUCTION

Recent developments indicate that unitary infinite-dimensional representations of certain noncompact Lie groups may be usefully employed in describing the physics of elementary particles.[1] For such purposes, the primary need is to have the relevant representations in as explicit and manageable a form as possible.

In setting up the representations of a semisimple noncompact Lie group, one has to decide from physical considerations which complete commuting set of operators is to be used to provide labels for the various states that appear in the representation. In some cases, this set will be associated with a compact subgroup of the original group, in others with a noncompact one. Even when a complete commuting set has been found, the problem of finding the possible eigenvalues of the operators in this set, and then of evaluating the matrix elements of all the generators of the group, can be an almost insoluble one. This difficulty is not necessarily related to the noncompactness of the Lie group one is interested in. For example, in the use of the SU_3 group in the nuclear shell model, the relevant subgroup is the R_3 subgroup corresponding to space rotations; and one wishes to obtain the SU_3 representations in a basis of angular-momentum eigenstates, rather than eigenstates of the $SU_2 \otimes U$ subgroup of SU_3.[2] (It is the latter that is relevant in the use of SU_3 in elementary particle physics.) And the difficulties of obtaining the SU_3 representations in an R_3 basis are well known.

We are concerned with the problem of obtaining the unitary infinite-dimensional representations of a semisimple noncompact Lie group in a basis made up of unitary finite-dimensional representations of the maximal compact subgroup. Here the situation is as follows. For the groups of the form $SU(n, 1)$ or $SO(n, 1)$ [pseudo-unitary and pseudo-orthogonal groups in $(n + 1)$ complex or real dimensions, respectively], it is known that in any unitary irreducible representation (UIR) of the noncompact group, each UIR of the maximal compact subgroup appears at most once.[3] [The maximal compact subgroups in the above cases are $SU(n) \otimes U(1)$ and $SO(n)$, respectively.] Thus these groups are "multiplicity-free," and this property is closely related to a similar property for the compact groups $SU(n + 1)$ and $SO(n + 1)$ in the reduction of their UIR's under $SU(n) \otimes U(1)$ and

[1] As an example, see the proceedings of the session on "Infinite Representations of Particles," in *Proceedings of the 1967 International Conference on Particles and Fields, University of Rochester, Rochester, N.Y.* (Interscience Publishers, Inc., New York, 1967).

[2] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128 (1958); V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960), **23**, 177 (1961); G. Racah in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1964).

[3] UIR's of these groups for all n have been obtained by U. Ottoson, Commun. Math. Phys. **8**, 228 (1968); **10**, 114 (1968). See also L. C. Biedenharn, *Noncompact Groups in Particle Physics*, Y. Chow, Ed. (W. A. Benjamin, Inc., New York, 1966).

$SO(n)$, respectively. However, in the cases of groups of the form $SU(p, q)$ or $SO(p, q)$ with p and q greater than unity, or the unimodular linear groups $SL(n, R)$ and $SL(n, C)$ for $n \geq 3$, this multiplicity-free property is generally lost; in other words, in general a UIR of one of these groups contains a given finite-dimensional UIR of the maximal compact subgroup more than once.[4] [There are, of course, special representations which are multiplicity-free.] In the multiplicity-free cases, the Casimir operators and internal labels associated with the maximal compact subgroup do form a complete commuting set, and provide a sufficient number of discrete quantum numbers to label the states of the entire UIR. And the UIR's of $SU(n, 1)$ and of $SO(n, 1)$ have been obtained by this method. When there is a multiplicity problem, however, the maximal compact subgroup by itself cannot supply a complete commuting set of operators, and the problem of taking care of the multiplicity is difficult. In fact there seem to be very few, if any, examples of noncompact groups whose UIR's involve multiplicity, and whose UIR's have been completely worked out in a basis of UIR's of the maximal compact subgroup.

In this paper we describe a new approach to the solution of this multiplicity problem, based on the idea of "expansion" of Lie groups. Group contraction was introduced by Inönü and Wigner; it is a process by which representations of a given group yield, via a limiting procedure, representations of another group.[5] Loosely defined, group expansion is the inverse process to group contraction. We treat in detail the case of the group $SL(3, C)$; however, this group already shows many of the features to be expected in principle in the most general case, and the methods we are suggesting have been adequately demonstrated.

Before we proceed with the details, it is worthwhile to outline the way in which the multiplicity problem is solved. The group $SL(3, C)$ is the group of all unimodular complex matrices in three dimensions. Its generators are sixteen in number; they may be separated into the eight generators of the maximal compact subgroup $SU(3)$, and eight others to which we refer as the "noncompact generators." Using tensor notation with respect to $SU(3)$, we can write the former as J^α_β, and the latter as K^α_β ($\alpha, \beta = 1, 2, 3$). In a UIR of $SL(3, C)$ these generators obey the following Hermiticity and tracelessness properties (repeated indices are to be summed over):

$$(J^\alpha_\beta)^\dagger = J^\beta_\alpha; \quad (K^\alpha_\beta)^\dagger = K^\beta_\alpha; \quad J^\alpha_\alpha = K^\alpha_\alpha = 0; \quad (1.1)$$

and the following commutation relations:

$$[J^\alpha_\beta, J^\lambda_\mu] = \delta^\alpha_\mu J^\lambda_\beta - \delta^\lambda_\beta J^\alpha_\mu, \quad (1.2a)$$

$$[J^\alpha_\beta, K^\lambda_\mu] = \delta^\alpha_\mu K^\lambda_\beta - \delta^\lambda_\beta K^\alpha_\mu, \quad (1.2b)$$

$$[K^\alpha_\beta, K^\lambda_\mu] = \delta^\lambda_\beta J^\alpha_\mu - \delta^\alpha_\mu J^\lambda_\beta. \quad (1.2c)$$

Starting with suitable families of UIR's of $SL(3, C)$, one obtains via group contraction, UIR's of another Lie group, $SU(3) \times T_8$, which is not semisimple but has a semidirect product structure. $SU(3) \times T_8$ also has sixteen generators; eight of them are the J^α_β generating the subgroup $SU(3)$, and the remaining eight form an *Abelian set* transforming under $SU(3)$ as a tensor belonging to the 8-dimensional adjoint representation of $SU(3)$. Calling these P^α_β, we have the following structure for $SU(3) \times T_8$:

$$(P^\alpha_\beta)^\dagger = P^\beta_\alpha; \quad P^\alpha_\alpha = 0; \quad (1.3a)$$

$$[J^\alpha_\beta, P^\lambda_\mu] = \delta^\alpha_\mu P^\lambda_\beta - \delta^\lambda_\beta P^\alpha_\mu; \quad (1.3b)$$

$$[P^\alpha_\beta, P^\lambda_\mu] = 0. \quad (1.3c)$$

Equation (1.2a) continues to hold, Eq. (1.3b) is the analog of Eq. (1.2b), and Eq. (1.3c) is the "contracted version" of Eq. (1.2c). Now it turns out that for many classes of representations, the multiplicity structure of $SU(3)$ representations is the same for $SL(3, C)$ and $SU(3) \times T_8$. To be specific, corresponding to UIR's of $SL(3, C)$ of the principal series, there are UIR's of $SU_3 \times T_8$ obtained from the former by group contraction, with the property that the spectrum of $SU(3)$ representations is the same in both cases and is not affected by the contraction procedure. We can now take advantage of the fact that *the multiplicity problem for* UIR's *of* $SU_3 \times T_8$ (and any other such semidirect product in which the semisimple part is compact and the invariant part is Abelian) *is completely solved*, to split the $SL(3, C)$ problem into two parts. First, we find a basis for the Hilbert space of a UIR of $SU(3) \times T_8$, made up of states belonging to definite representations of $SU(3)$, in which appropriate labels have been introduced to take care of the multiplicity problem. The way to do this has been shown by Goebel in connection with the group-theoretical formulation of strong-coupling theory; from Goebel's work we also know the matrix elements of the $SU_3 \times T_8$ generators P^α_β in the $SU(3)$ basis in explicit form.[6] Having done this, we now try to express the

[4] For an example of this situation, see A. Kihlberg, V. F. Müller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).

[5] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953). See also E. J. Saletan, J. Math. Phys. **2**, 1 (1961).

[6] C. Goebel, Phys. Rev. Letters **16**, 1130 (1966). See also C. Goebel in *Noncompact Groups in Particle Physics*, Y. Chow, Ed. (W. A. Benjamin, Inc., New York, 1966); and T. Cook and B. Sakita, J. Math. Phys. **8**, 708 (1967).

$SL(3, C)$ generators K_β^α as (polynomial) functions of the $SU(3) \times T_8$ generators J_β^α and P_β^α. It is this process which we refer to as group expansion. When this has been done, we can completely work out the matrix elements of K_β^α in the $SU(3)$ basis, knowing the matrix elements of J_β^α and P_β^α in this basis.

This kind of expansion of groups has been attempted in the past for *multiplicity-free* groups. For example, it has been shown that one can express the generators of $SO(n, 1)$ as (polynomial) functions of generators of $E(n)$, the Euclidean group in n dimensions; similarly one can pass from the inhomogeneous unitary group $IU(n)$ to the pseudo-unitary group $SU(n, 1)$.[7] Many new features appear, however, when one encounters groups with multiplicity problems.

The material of this paper is arranged as follows. In Sec. 2, we describe the principal series of UIR's of $SL(3, C)$, in the form in which they are given by Gel'fand and Naimark. This class of UIR's of $SL(3, C)$ itself consists of two distinct subclasses, namely, the principal nondegenerate series and the principal degenerate series. It is only in the former that one meets the multiplicity problem, the latter consisting of multiplicity-free UIR's. However, for completeness and for exhibiting the structure of the expansion formulas, we consider both classes. In Sec. 3, we describe the structure of all the UIR's of $SU(3) \times T_8$, as obtained by Goebel's method. These UIR's can be usefully grouped into three families; two of them alone are contracted forms of the two subclasses of principal-series UIR's of $SL(3, C)$. Sections 4 and 5 are devoted to the construction of the expansion formulas expressing the $SL(3, C)$ generators as functions of the $SU(3) \times T_8$ generators, in the principal nondegenerate series and the principal degenerate series, respectively. For comparison with the case of $SL(3, C)$, we have described in Appendix A the way in which the generators of $SL(2, C)$ can be written as functions of the generators of the nonsemisimple group $SU(2) \times T_3$ [i.e., the Euclidean group $E(3)$ in 3-dimensional space]. Appendix B contains an account of Goebel's method for obtaining UIR's of the groups of the semidirect product form $C \times T$, where C is an arbitrary compact semisimple group, and T denotes an Abelian set of generators transforming irreducibly according to some given UIR of C. The problem of treating the supplementary series of UIR's of $SL(3, C)$, and the purely algebraic question of computing the matrix elements of the $SL(3, C)$ generators using the expansion formulas, will both be taken up elsewhere.

2. PRINCIPAL SERIES UIR'S OF $SL(3, C)$

The UIR's of the noncompact groups $SL(n, c)$ have been given by Gel'fand and Naimark.[8] In this section, we summarize their construction of the principal nondegenerate series and the principal degenerate series of UIR's of $SL(3, C)$. We utilize these constructions later on.

Each element $g \in SL(3, C)$ corresponds to a complex 3×3 unimodular matrix:

$$g \to \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix}, \quad \det |g_{ij}| = +1, \quad (2.1)$$

and vice versa. The Lie algebra of $SL(3, C)$ is 16-dimensional, being spanned by the eight generators of the compact subgroup $SU(3)$, and a set of eight "noncompact" generators. The former generate the subset of elements g which correspond to unimodular *unitary* matrices, while the latter generate the subset of elements g corresponding to unimodular *Hermitian* matrices. In the defining 3-dimensional representation of $SL(3, C)$, one may identify the $SU(3)$ generators J_j and the "noncompact" ones K_j, with the λ-matrices of Gell-Mann[9]:

$$J_j \to \tfrac{1}{2}\lambda_j, \quad K_j \to \tfrac{1}{2}i(\lambda_j). \quad (2.2)$$

Identified in this way, the J_j and K_j in a UIR of $SL(3, C)$ would be linear self-adjoint operators. We find it more convenient, as far as the derivations of the expansion formulas are concerned, to deal with the J's and K's in tensor form, the components J_β^α being given by[10]

$$\begin{aligned} &J_1^1 = J_3 + (3)^{-\frac{1}{2}}J_8, \quad J_1^2 = J_1 + iJ_2, \quad J_1^3 = J_4 + iJ_5, \\ &J_2^1 = J_1 - iJ_2, \quad J_2^2 = -J_3 + (3)^{-\frac{1}{2}}J_8, \quad J_2^3 = J_6 + iJ_7, \\ &J_3^1 = J_4 - iJ_5, \quad J_3^2 = J_6 - iJ_7, \quad J_3^3 = -2(3)^{-\frac{1}{2}}J_8. \end{aligned} \quad (2.3)$$

Similar relations define K_β^α. The commutation relations among J_β^α and K_β^α, and the Hermiticity properties valid in a UIR of $SL(3, C)$, have been given in Eqs. (1.1) and (1.2).

We now describe first the principal nondegenerate series of UIR's, and then the principal degenerate series.

[7] A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965); M. Y. Han, *ibid.* **42B**, 367 (1966); J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966); A. Böhm, *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1966); R. Hermann, Commun. Math. Phys. **2**, 155 (1966); A. Chakrabarti, "A Class of Representations of the $IU(n)$ Algebra and Deformation to $U(n, 1)$," Preprint No. A110.1267, Centre de Physique Theorique de l'Ecole Polytechnique, Paris, 1967.

[8] I. M. Gel'fand and M. A. Naimark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie-Verlag, Berlin, 1957).

[9] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962).

[10] J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).

Principal Nondegenerate Series

Define two subgroups Z and K of $SL(3, C)$ as follows. The subgroup Z consists of all elements z corresponding to matrices of the form

$$z \to \begin{pmatrix} 1 & 0 & 0 \\ \gamma & 1 & 0 \\ \beta & \alpha & 1 \end{pmatrix} \tag{2.4}$$

with arbitrary complex α, β, γ. The subgroup K consists of all elements k corresponding to matrices of the form

$$k \to \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ 0 & k_{22} & k_{23} \\ 0 & 0 & k_{33} \end{pmatrix}, \quad k_{11}\,k_{22}\,k_{33} = +1. \tag{2.5}$$

Then (almost) every element $g \in SL(3, C)$ can be expressed in a unique way as the product of one element in K and one element in Z:

$$g = kz, \quad k \in K, \quad z \in Z \tag{2.6}$$

(Those g that cannot be so expressed may be ignored. For details, see Ref. 8.)

A UIR of the principal nondegenerate series is labeled by four parameters $(m_2, m_3, \rho_2, \rho_3)$. Of these, m_2 and m_3 are integers (positive, negative, or zero), while ρ_2 and ρ_3 are two arbitrary real numbers. For given values of these parameters, the UIR is explicitly constructed as follows. We introduce in the Hilbert space $\mathcal{H}(m_2, m_3, \rho_2, \rho_3)$ of this UIR a basis made up of nonnormalizable vectors, one corresponding to each element z in the subgroup Z defined above.[11] In other words, we have a basis labeled by three continuous complex variables α, β, γ (alternatively, six real ones):

$$|z\rangle = |\alpha, \beta, \gamma\rangle \tag{2.7}$$

normalized according to

$$\begin{aligned} \langle z_1 \,|\, z_2\rangle &\equiv \langle \alpha_1\beta_1\gamma_1 \,|\, \alpha_2\beta_2\gamma_2\rangle \\ &= \delta(\alpha_1 - \alpha_2)\delta(\beta_1 - \beta_2)\delta(\gamma_1 - \gamma_2), \\ \delta(\alpha_1 - \alpha_2) &\equiv \delta(\mathrm{Re}\,\alpha_1 - \mathrm{Re}\,\alpha_2)\delta(\mathrm{Im}\,\alpha_1 - \mathrm{Im}\,\alpha_2), \text{ etc.} \end{aligned} \tag{2.8}$$

Given an element $z \in Z$, and any element $g \in SL(3, C)$, the product zg can be written in a unique way as the product of one element in K and one in Z:

$$zg = kz_g. \tag{2.9}$$

[11] This is a physicist's way of saying that the Hilbert space $\mathcal{H}(m_2, m_3, \rho_2, \rho_3)$ consists of all functions $f(\alpha, \beta, \gamma)$ which are square-integrable, i.e.,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\,\mathrm{Re}\,\alpha\, d\,\mathrm{Im}\,\alpha \cdots d\,\mathrm{Im}\,\gamma\, |f(\alpha, \beta, \gamma)|^2 < \infty.$$

The matrix elements of k and z_g are easily expressed in terms of those of z and g. If we write the matrix corresponding to g^{-1} as

$$g^{-1} \to \begin{pmatrix} G_{11} & -G_{21} & G_{31} \\ -G_{12} & G_{22} & -G_{32} \\ G_{13} & -G_{23} & G_{33} \end{pmatrix}, \tag{2.10}$$

then we have

$$\begin{aligned} k_{11} &= [G_{11} + \gamma G_{21} + (\alpha\gamma - \beta)G_{31}]^{-1}, \\ k_{12} &= k_{33}^{-1}[G_{21} + \alpha G_{31}], \\ k_{13} &= g_{13}, \\ k_{22} &= k_{33}^{-1}[G_{11} + \gamma G_{21} + (\alpha\gamma - \beta)G_{31}], \\ k_{23} &= \gamma g_{13} + g_{23}, \\ k_{33} &= \beta g_{13} + \alpha g_{23} + g_{33}, \\ \alpha_g &= k_{33}^{-1}[\beta g_{12} + \alpha g_{22} + g_{32}], \\ \beta_g &= k_{33}^{-1}[\beta g_{11} + \alpha g_{21} + g_{31}], \\ \gamma_g &= k_{11}[G_{12} + \gamma G_{22} + (\alpha\gamma - \beta)G_{32}]. \end{aligned} \tag{2.11}$$

Now, the unitary operator $U(g)$ that represents the element g in the UIR $(m_2, m_3, \rho_2, \rho_3)$ is given by specifying its action on a general basis vector:

$$U(g)\,|z\rangle = \mu(g, z)\,|z_g\rangle,$$

$$\mu(g, z) = |k_{22}|^{-2+i\rho_2}\,|k_{33}|^{-4+i\rho_3}\left(\frac{k_{22}^*}{k_{22}}\right)^{\frac{1}{2}m_2}\left(\frac{k_{33}^*}{k_{33}}\right)^{\frac{1}{2}m_3}. \tag{2.12}$$

The manner in which the UIR $(m_2, m_3, \rho_2, \rho_3)$ reduces under $SU(3)$ is determined solely by the discrete parameters (m_2, m_3). Let us refer to the finite-dimensional UIR's of $SU(3)$ as $D(\lambda_1, \lambda_2)$ in the highest weight notation.[10] Within a UIR of $SU(3)$, "weight" means an eigenvalue pair for the two commuting operators $(H_1, H_2) = [\frac{1}{2}(3)^{-\frac{1}{2}}(J_1^1 - J_2^2), -\frac{1}{2}J_3^3]$. [In terms of isospin and hypercharge labels as used in particle physics, $H_1 = (3)^{-\frac{1}{2}}I_3$ and $H_2 = \frac{1}{2}Y$.] Then, the UIR $(m_2, m_3, \rho_2, \rho_3)$ of $SL(3, C)$ contains a given UIR $D(\lambda_1, \lambda_2)$ of $SU(3)$ as many times as this UIR of $SU(3)$ contains linearly independent weight vectors corresponding to the weight $[\frac{1}{2}(3)^{-\frac{1}{2}}m_2, \frac{1}{6}(2m_3 - m_2)]$.[12] This describes the way in which the Hilbert space $\mathcal{H}(m_2, m_3, \rho_2, \rho_3)$ decomposes under $SU(3)$. It is clear that in general if a UIR of $SU(3)$ is contained in the space, it is contained many times.

Principal Degenerate Series

The principal degenerate series of UIR's of $SL(3, C)$ is constructed with the aid of two subgroups of $SL(3, C)$. Let us define the subgroup Z' to consist of

[12] A. Böhm, Phys. Rev. **158**, 1408 (1967).

all elements z' corresponding to matrices of the form

$$z' \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta & \alpha & 1 \end{pmatrix} \tag{2.13}$$

with arbitrary complex α, β. Next, we define the subgroup K' made up of all elements k' corresponding to matrices of the form

$$k' \to \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ 0 & 0 & k_{33} \end{pmatrix}, \quad (k_{11}k_{22} - k_{12}k_{21})k_{33} = 1. \tag{2.14}$$

Then, (almost) every element $g \in SL(3, C)$ can be expressed in a unique way as the product of one element in K' and one in Z':

$$g = k'z', \quad k' \in K', \quad z' \in Z'. \tag{2.15}$$

(Once again, one may ignore the set of those elements g that cannot be so decomposed.)

A UIR of the principal degenerate series is characterized by two parameters (m_3, ρ_3), where m_3 is any integer (positive, negative, or zero), and ρ_3 is an arbitrary real number. The explicit construction of these UIR's proceeds as follows. In the Hilbert space $\mathcal{H}(m_3, \rho_3)$, we introduce a basis of nonnormalizable vectors labeled by the elements z' of the subgroup Z' of $SL(3, C)$.[13] That is, we have a basis labeled by the two complex variables α, β (alternatively, four real ones):

$$|z'\rangle = |\alpha, \beta\rangle \tag{2.16}$$

normalized according to

$$\langle z_1' \,|\, z_2' \rangle \equiv \langle \alpha_1\beta_1 \,|\, \alpha_2\beta_2 \rangle = \delta(\alpha_1 - \alpha_2)\delta(\beta_1 - \beta_2). \tag{2.17}$$

Given an element $z' \in Z'$, and any $g \in SL(3, C)$, we write the product $z'g$ in a unique way in the following form:

$$z'g = k'z_g', \quad k' \in K', \quad z_g' \in Z'. \tag{2.18}$$

The matrix elements of k' and z_g' are given in terms of z' and g by:

$$\begin{aligned} k_{11} &= k_{33}^{-1}(\alpha G_{32} + G_{22}), \\ k_{12} &= k_{33}^{-1}(\alpha G_{31} + G_{21}), \\ k_{13} &= g_{13}, \\ k_{21} &= k_{33}^{-1}(-\beta G_{32} + G_{12}), \\ k_{22} &= k_{33}^{-1}(-\beta G_{31} + G_{11}), \\ k_{33} &= \beta g_{13} + \alpha g_{23} + g_{33}, \\ \alpha_g &= k_{33}^{-1}(\beta g_{12} + \alpha g_{22} + g_{32}), \\ \beta_g &= k_{33}^{-1}(\beta g_{11} + \alpha g_{21} + g_{31}). \end{aligned} \tag{2.19}$$

[13] Once again, this is equivalent to saying that $\mathcal{H}(m_3, \rho_3)$ consists of all square-integrable functions of two complex variables α, β.

Now, the unitary operator $U(g)$ corresponding to any $g \in SL(3, C)$ acts in the following way:

$$\begin{aligned} U(g)\,|z'\rangle &= \lambda(g, z')\,|z_g'\rangle, \\ \lambda(g, z') &= |k_{33}|^{-3+i\rho_3}(k_{33}^*/k_{33})^{\frac{1}{2}m_3}. \end{aligned} \tag{2.20}$$

As for the reduction of the UIR (m_3, ρ_3) under $SU(3)$, it turns out to be multiplicity free, and independent of ρ_3. The UIR $D(\lambda_1, \lambda_2)$ of $SU(3)$ is contained in the UIR (m_3, ρ_3) of $SL(3, C)$ if and only if $D(\lambda_1, \lambda_2)$ contains a state which is an eigenstate of H_2 with eigenvalue $\frac{1}{3}m_3$ and which is invariant under the $SU(2)$ subgroup generated by J_1^2, J_2^1, and $\frac{1}{2}(J_1^1 - J_2^2)$. If so, then $D(\lambda_1, \lambda_2)$ appears exactly once in the UIR (m_3, ρ_3) of $SL(3, C)$. In the language of particle physics, (m_3, ρ_3) contains, once each, those UIR's $D(\lambda_1, \lambda_2)$ of $SU(3)$ that contain an isospin singlet with hypercharge $Y = \frac{2}{3}m_3$, and no others.

3. UIR'S OF THE GROUP $SU(3) \times T_8$

The Hermiticity and commutation properties of the elements J_β^α, P_β^α which generate the Lie algebra of the group $SU(3) \times T_8$ are contained in Eqs. (1.1)–(1.3). Here we write the UIR's of this group in an $SU(3)$ basis, making explicit the solution of the multiplicity problem. We need to specify the way we label states belonging to definite UIR's (λ_1, λ_2) of $SU(3)$. The operators J_1, J_2, J_3, and J_8 [see (2.3)] generate an $SU(2) \otimes U(1)$ subgroup of $SU(3)$. Referring to the eigenvalues of $(J_1)^2 + (J_2)^2 + (J_3)^2$, J_3, and $2(3)^{-\frac{1}{2}}J_8$ as $I(I + 1)$, M, and Y, respectively, the basis states for the UIR (λ_1, λ_2) of $SU(3)$ can be written as[14]

$$|\lambda_1\lambda_2;\, IMY;\, \alpha\rangle. \tag{3.1}$$

The symbol α denotes whatever labels may be necessary to distinguish multiple occurrences of one and the same UIR (λ_1, λ_2) of $SU(3)$ in a UIR of $SL(3, C)$. (It will be specified in more detail shortly.) α is an $SU(3)$-invariant label, being unchanged by the application of the $SU(3)$ generators J_β^α to the state (3.1). The way in which the J_β^α change the labels IMY is known and standardized.[15]

We must next arrange the components of P_β^α in such a way that they transform like, and are also labeled like, the basis states of the (1, 1) (i.e., adjoint) representation of $SU(3)$. This is done by the following

[14] The values of I, M, Y within the UIR (λ_1, λ_2) of $SU(3)$ may be parametrized as follows: for each pair of integers f, g obeying the inequalities $\lambda_1 + \lambda_2 \geq f \geq \lambda_2 \geq g \geq 0$, we have once the (I, Y) multiplet with $I = \frac{1}{2}(f - g)$, $Y = f + g - \frac{2}{3}(\lambda_1 + 2\lambda_2)$. For given I, of course, $M = -I, -I + 1, \cdots, I - 1, I$. See A. J. Macfarlane, E. C. G. Sudarshan, and C. Dullemond, Nuovo Cimento **30**, 845 (1963).

[15] L. C. Biedenharn, Phys. Letters **3**, 69 (1962).

identifications[16]:

$$P_{110} = -(6)^{-\frac{1}{2}}P_1^2, \quad P_{100} = \tfrac{1}{2}(3)^{-\frac{1}{2}}(P_1^1 - P_2^2),$$
$$P_{1-10} = (6)^{-\frac{1}{2}}P_2^1,$$
$$P_{000} = \tfrac{1}{2}P_3^3, \quad P_{\frac{1}{2}\frac{1}{2}1} = (6)^{-\frac{1}{2}}P_1^3, \quad P_{\frac{1}{2}-\frac{1}{2}1} = (6)^{-\frac{1}{2}}P_2^3,$$
$$P_{\frac{1}{2}\frac{1}{2}-1} = -(6)^{-\frac{1}{2}}P_2^3, \quad P_{\frac{1}{2}-\frac{1}{2}-1} = (6)^{-\frac{1}{2}}P_1^3. \quad (3.2)$$

We have labeled the components of P, like the states in (3.1), as P_{IMY}, where IMY go over the set of values appropriate to the (1, 1) representation of $SU(3)$.

The UIR's of $SU(3) \times T_8$ are given in a Hilbert space in which the following states form an orthonormal basis:

$$|\lambda_1\lambda_2; IMY; I^{(0)}M^{(0)}Y^{(0)}\rangle. \quad (3.3)$$

(λ_1, λ_2) denotes a UIR of $SU(3)$; for given (λ_1, λ_2), the range of values of IMY is known[17]; $I^{(0)}M^{(0)}Y^{(0)}$ constitute the multiplicity label α of (3.1). In principle, $I^{(0)}M^{(0)}Y^{(0)}$ can go over the same range of values as IMY, though in each case they will be restricted to some subset of this range. In this basis, the matrix elements of P_{IMY} can be expressed via the Wigner–Eckart theorem in terms of Clebsch–Gordan (CG) coefficients for $SU(3)$, and reduced matrix elements.[18] We have the following formula valid for all UIR's of $SU(3) \times T_8$:

$$\begin{aligned}\langle\lambda_1'\lambda_2'; I'M'Y'; I^{(0)\prime}M^{(0)\prime}Y^{(0)\prime}| &\\ \times P_{I''M''Y''}|\lambda_1\lambda_2; IMY; I^{(0)}M^{(0)}Y^{(0)}\rangle &\\ = \sum_\gamma C(\lambda_1\lambda_2\, 1\, 1\, \lambda_1'\lambda_2'\gamma; IMY\, I''M''Y''\, I'M'Y') &\\ \times \langle\lambda_1'\lambda_2'; I^{(0)\prime}M^{(0)\prime}Y^{(0)\prime}\| P \|\lambda_1\lambda_2; I^{(0)}M^{(0)}Y^{(0)}\rangle_\gamma, & \quad (3.4a)\end{aligned}$$

$$\begin{aligned}\langle\lambda_1'\lambda_2'; I^{(0)\prime}M^{(0)\prime}Y^{(0)\prime}\| P \|\lambda_1\lambda_2; I^{(0)}M^{(0)}Y^{(0)}\rangle_\gamma &\\ = [d(\lambda_1\lambda_2)/d(\lambda_1'\lambda_2')]^{\frac{1}{2}} \sum_{I^{(0)\prime\prime}=0,1} p_{I^{(0)\prime\prime}} &\\ \times C(\lambda_1\lambda_2\, 1\, 1\, \lambda_1'\lambda_2'\gamma; I^{(0)}M^{(0)}Y^{(0)} &\\ \times I^{(0)\prime\prime}\, 0\, 0\, I^{(0)\prime}M^{(0)\prime}Y^{(0)\prime}). & \quad (3.4b)\end{aligned}$$

The quantities p_0 and p_1 are two arbitrary real numbers. The symbol γ is needed because the reduction of the direct product $(1, 1) \times (\lambda_1\lambda_2)$ of UIR's of $SU(3)$ generally yields the UIR $(\lambda_1\lambda_2)$ *twice*; correspondingly, we have two independent orthogonal sets of CG coefficients, and also two independent reduced matrix elements, in the case $(\lambda_1'\lambda_2') = (\lambda_1\lambda_2)$. The symbol $d(\lambda_1\lambda_2)$ denotes the dimensionality of the UIR $(\lambda_1\lambda_2)$ of $SU(3)$.[19] Formula (3.4) specifies completely the matrix elements of P in every UIR of $SU(3) \times T_8$. Different classes of UIR's are distinguished by stating the ranges of $\lambda_1\lambda_2 I^{(0)}M^{(0)}Y^{(0)}$ and the restrictions on p_0 and p_1. (Notice that in any case, P has nonzero matrix elements only if $M^{(0)\prime} = M^{(0)}$, $Y^{(0)\prime} = Y^{(0)}$.) We now give these details, dividing the UIR's into three types.

Type I: This type consists of UIR's with $p_1 \neq 0$, $\pm(3)^{\frac{1}{2}}p_0$ (p_0 may or may not be zero). $M^{(0)}$ and $Y^{(0)}$ are kept fixed at any allowed values, while $I^{(0)}$ varies. Having chosen $M^{(0)}$ and $Y^{(0)}$, we can describe the representation space as follows. Each UIR $(\lambda_1\lambda_2)$ of $SU(3)$ occurs as often as it contains states IMY with $M = M^{(0)}$ and $Y = Y^{(0)}$; these several appearances of $(\lambda_1\lambda_2)$ are distinguished by the label $I^{(0)}$, and states with different values of $I^{(0)}$ are orthogonal. Thus, a UIR of Type I is completely specified by $(M^{(0)}, Y^{(0)}, p_0, p_1)$. The $SU(3)$ content of such a UIR is clearly identical with the $SU(3)$ content of a principal nondegenerate series UIR $(m_2m_3\rho_2\rho_3)$ of $SL(3, C)$ for which

$$m_2 = 2M^{(0)}, \quad 2m_3 - m_2 = 3Y^{(0)}. \quad (3.5)$$

Type II: We now set $p_1 = 0$, $p_0 \neq 0$. Then (3.4) shows that for nonzero matrix elements, $I^{(0)}$, $M^{(0)}$, and $Y^{(0)}$ should all be unchanged. We define UIR's of Type II by the choice $I^{(0)} = M^{(0)} = 0$, $Y^{(0)} =$ any fixed allowed value. Thus, a Type II UIR is specified by $(Y^{(0)}, p_0)$. Any UIR $(\lambda_1\lambda_2)$ of $SU(3)$ appears once if it contains the state $IMY = (0, 0, Y^{(0)})$, and is absent otherwise. The $SU(3)$ content of such a UIR is identical with the $SU(3)$ content of a principal degenerate series UIR (m_3, ρ_3) of $SL(3, C)$,[20] for which

$$2m_3 = 3Y^{(0)}. \quad (3.6)$$

Type III: We still maintain $p_1 = 0$, $p_0 \neq 0$, but consider the cases $I^{(0)} \neq 0$. As a matter of convenience we choose $M^{(0)} = I^{(0)}$ (the representation is independent of how $M^{(0)}$ is chosen). As before, $Y^{(0)}$ may be assigned any fixed allowed value. So a Type III UIR of $SU(3) \times T_8$ is specified by $(I^{(0)}, Y^{(0)}, p_0)$ with $I^{(0)} \neq 0$. This UIR contains the UIR $(\lambda_1\lambda_2)$ of $SU(3)$ once, if $(\lambda_1\lambda_2)$ contains a state IMY with $I = M = I^{(0)}$, $Y = Y^{(0)}$, and not otherwise. We see that there is no member of the principal series UIR's of $SL(3, C)$ whose $SU(3)$ content coincides exactly with that of a Type III UIR of $SU(3) \times T_8$.

We have separated the UIR's of $SU(3) \times T_8$ in the case $p_1 = 0$ into Types II and III precisely because

[16] N. Mukunda and L. K. Pandit, J. Math. Phys. **6**, 746 (1965); J. G. Kuriyan, D. Lurié, and A. J. Macfarlane, *ibid.* **6**, 722 (1965).

[17] See footnote 14.

[18] The notation used here for Clebsch–Gordan coefficients is that of J. G. Kuriyan, D. Lurié, and A. J. Macfarlane, J. Math. Phys. **6**, 722 (1965). The CG coefficients for an octet operator, which are what we need, are given in detail in this paper.

[19] Explicitly, $d(\lambda_1, \lambda_2) = \frac{1}{2}(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_1 + \lambda_2 + 2)$.

[20] The $SU(3)$ content of the UIR (m_3, ρ_3) of $SL(3, C)$ is very simply stated. Each UIR (λ_1, λ_2) for which $\lambda_2 - \lambda_1 = m_3$ appears once.

only in one case do we have an $SU(3)$ decomposition identical with that for some principal degenerate series UIR of $SL(3, C)$.

4. ANALYSIS OF PRINCIPAL NONDEGENERATE SERIES UIR'S

Our aim in this section is to show that the generators of $SL(3, C)$ in any UIR $(m_2, m_3, \rho_2, \rho_3)$ of the principal nondegenerate series can be expressed as functions of the generators of a suitably chosen UIR of $SU(3) \times T_8$ (belonging to Type I). The construction of the UIR's of $SL(3, C)$ described in Sec. 2 has the unsatisfactory feature that even if we consider an element g belonging to the maximal compact subgroup $SU(3)$, the right-hand side of Eq. (2.12) has a dependence on the continuous parameters ρ_2 and ρ_3. This dependence can be removed by working with a new set of basis states differing from the old ones by suitably chosen phase factors. We define the new basis vectors as follows:

$$|z) = \exp\,[i\phi(z)]\,|z\rangle,$$
$$\phi(z) = \tfrac{1}{2}(\rho_2 - \rho_3)\ln(1 + |\alpha|^2 + |\beta|^2) - \tfrac{1}{2}\rho_2 \ln(1 + |\gamma|^2 + |\beta - \alpha\gamma|^2). \quad (4.1)$$

(The symbol z stands for the three complex parameters α, β, and γ.) The new states obey the same orthonormality conditions as the old ones:

$$(z_2 \,|\, z_1) = \delta(\alpha_2 - \alpha_1)\delta(\beta_2 - \beta_1)\delta(\gamma_2 - \gamma_1). \quad (4.2)$$

The action of $U(g)$ on these new vectors is easily computed. It reads:

$$U(g)\,|z) = \exp\,[i\phi(z) - i\phi(z_g)]\mu(g, z)\,|z_g). \quad (4.3)$$

The quantity $\mu(g, z)$ is defined in Eq. (2.12), while the parameters defining z_g are given in Eq. (2.11). Now one may verify that if g is an element of $SU(3)$, then the right-hand side of (4.3) does not depend at all on the continuous parameters ρ_2 and ρ_3.

We are now in a position to compute the $SL(3, C)$ generators, by considering g in (4.3) to be an element close to the identity. Each generator appears as a linear differential operator in the variables α, β, γ. We express everything in terms of α, β, γ and their complex conjugates α^*, β^*, γ^*. Partial differentiation with respect to α and α^* is defined as follows: if $\alpha = x + iy$ is the separation of α into real and imaginary parts, then

$$2\frac{\partial}{\partial\alpha} = \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}, \quad 2\frac{\partial}{\partial\alpha^*} = \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}. \quad (4.4)$$

The rule for Hermitian conjugation is

$$\left(\frac{\partial}{\partial\alpha}\right)^\dagger = -\frac{\partial}{\partial\alpha^*}, \quad \left(\frac{\partial}{\partial\alpha^*}\right)^\dagger = -\frac{\partial}{\partial\alpha}. \quad (4.5)$$

Similar rules hold with respect to β and γ. We list now some of the $SU(3)$ generators J_β^α:

$$J_1^1 = \tfrac{1}{3}(m_2 + m_3) + \beta\frac{\partial}{\partial\beta} - \beta^*\frac{\partial}{\partial\beta^*} + \gamma\frac{\partial}{\partial\gamma} - \gamma^*\frac{\partial}{\partial\gamma^*},$$
$$J_1^2 = -\tfrac{1}{2}m_2\gamma - \gamma + \beta\frac{\partial}{\partial\alpha} - \alpha^*\frac{\partial}{\partial\beta^*} - \gamma^2\frac{\partial}{\partial\gamma} - \frac{\partial}{\partial\gamma^*},$$
$$J_1^3 = \tfrac{1}{2}m_2\alpha\gamma - \tfrac{1}{2}m_3\beta - \beta + (\alpha\gamma - \beta) - \alpha\beta\frac{\partial}{\partial\alpha} - \beta^2\frac{\partial}{\partial\beta} - \frac{\partial}{\partial\beta^*} + \gamma(\alpha\gamma - \beta)\frac{\partial}{\partial\gamma},$$
$$J_2^2 = \tfrac{1}{3}(m_3 - 2m_2) + \alpha\frac{\partial}{\partial\alpha} - \alpha^*\frac{\partial}{\partial\alpha^*} - \gamma\frac{\partial}{\partial\gamma} + \gamma^*\frac{\partial}{\partial\gamma^*},$$
$$J_2^3 = \tfrac{1}{2}(m_2 - m_3)\alpha - \alpha - \alpha^2\frac{\partial}{\partial\alpha} - \frac{\partial}{\partial\alpha^*} - \alpha\beta\frac{\partial}{\partial\beta} + (\alpha\gamma - \beta)\frac{\partial}{\partial\gamma}. \quad (4.6)$$

The four remaining operators J_β^α can be obtained from those listed by using the tracelessness and Hermiticity properties given in Eq. (1.1).

Turning our attention next to the "noncompact" generators K_β^α, the expressions here are somewhat more involved than for J_β^α. Each K_β^α consists of two parts, one involving the parameters ρ_2 and ρ_3 (and not containing differential operators), and another involving differential operators. We find it convenient to use s and ρ_2 in place of ρ_3 and ρ_2, where

$$\rho_3 = (s + 1)\rho_2. \quad (4.7)$$

We write K_β^α in the form

$$K_\beta^\alpha = \rho_2 p_\beta^\alpha + K_\beta^{(0)\alpha}, \quad (4.8)$$

where the second term on the right has no ρ_2 or s dependence, and where p_β^α and $K_\beta^{(0)\alpha}$ individually obey the same tracelessness and Hermiticity conditions as K_β^α. As in (4.6), it is sufficient to write down just five components of p_β^α and of $K_\beta^{(0)\alpha}$. These are:

$$p_1^1 = (s/3A)(2\beta\beta^* - \alpha\alpha^* - 1) + \tfrac{1}{3}B^{-1}[(\alpha\gamma - \beta)(\alpha^*\gamma^* - \beta^*) + \gamma\gamma^* - 2],$$
$$p_1^2 = (s/A)\beta\alpha^* + B^{-1}\gamma,$$
$$p_1^3 = (s/A)\beta + B^{-1}(\beta - \alpha\gamma),$$
$$p_2^2 = (s/3A)(2\alpha\alpha^* - \beta\beta^* - 1) + \tfrac{1}{3}B^{-1}[(\alpha\gamma - \beta)(\alpha^*\gamma^* - \beta^*) - 2\gamma\gamma^* + 1],$$
$$p_2^3 = (s/A)\alpha + B^{-1}(\alpha\gamma - \beta)\gamma^*,$$
$$A = 1 + |\alpha|^2 + |\beta|^2, \quad B = 1 + |\gamma|^2 + |\alpha\gamma - \beta|^2, \quad (4.9)$$

and

$$K_1^{(0)1} = -2i - i\left(\beta\frac{\partial}{\partial\beta} + \beta^*\frac{\partial}{\partial\beta^*} + \gamma\frac{\partial}{\partial\gamma} + \gamma^*\frac{\partial}{\partial\gamma^*}\right),$$

$$K_1^{(0)2} = \frac{i}{2}m_2\gamma + i\gamma - i\left(\beta\frac{\partial}{\partial\alpha} + \alpha^*\frac{\partial}{\partial\beta^*} + \frac{\partial}{\partial\gamma^*} - \gamma^2\frac{\partial}{\partial\gamma}\right),$$

$$K_1^{(0)3} = -\frac{i}{2}m_2\alpha\gamma + \frac{i}{2}m_3\beta + i(\beta - \alpha\gamma) + i\beta + i\left(\alpha\beta\frac{\partial}{\partial\alpha} + \beta^2\frac{\partial}{\partial\beta} - \frac{\partial}{\partial\beta^*} - \gamma(\alpha\gamma - \beta)\frac{\partial}{\partial\gamma}\right),$$

$$K_2^{(0)2} = -i\left(\alpha\frac{\partial}{\partial\alpha} + \alpha^*\frac{\partial}{\partial\alpha^*} - \gamma\frac{\partial}{\partial\gamma} - \gamma^*\frac{\partial}{\partial\gamma^*}\right),$$

$$K_2^{(0)3} = \frac{i}{2}(m_3 - m_2)\alpha + i\alpha + i\left(\alpha^2\frac{\partial}{\partial\alpha} - \frac{\partial}{\partial\alpha^*} + \alpha\beta\frac{\partial}{\partial\beta} + (\beta - \alpha\gamma)\frac{\partial}{\partial\gamma}\right). \tag{4.10}$$

As was to be expected, the $SU(3)$ generators J_β^α do not have any ρ_2 or s dependence at all. Knowing the commutation rules obeyed by J_β^α and K_β^α, we see immediately that if we "contract" K_β^α to p_β^α (keeping s fixed), so that

$$p_\beta^\alpha = \operatorname*{Lt}_{\rho_2\to\infty}\frac{1}{\rho_2}K_\beta^\alpha, \tag{4.11}$$

then J_β^α and p_β^α will obey the commutation rules of $SU(3) \times T_8$. In this process of contraction, the $SU(3)$ content of the representation space has not been affected at all. Thus it is plausible that via contraction we have obtained a Type I UIR of $SU(3) \times T_8$, whose $SU(3)$ content is identical with that of the original $SL(3, C)$ representation. We cannot yet be sure that this is the case, since we do not know whether J_β^α and p_β^α form an irreducible set of operators. However, we now show that we can in fact express K_β^α explicitly as a function of J_β^α and p_β^α: this will then be a proof that we are dealing here with a Type I UIR of $SU(3) \times T_8$.

The process of establishing that K_β^α is a function of J_β^α and p_β^α is a rather tedious one. The main problem is that the expressions we have given in (4.6), (4.9), and (4.10) for these operators are unwieldy, and one needs a more compact way of writing them. To this end, let us introduce six auxiliary complex variables z_α, ω^α ($\alpha = 1, 2, 3$), their complex conjugates, and the following rules of Hermitian conjugation:

$$\left(\frac{\partial}{\partial z_\alpha}\right)^\dagger = -\frac{\partial}{\partial z_\alpha^*}, \quad \left(\frac{\partial}{\partial\omega^\alpha}\right)^\dagger = -\frac{\partial}{\partial\omega^{\alpha*}}. \tag{4.12}$$

In terms of these independent variables, consider the following operators:

$$\bar{J}_\beta^\alpha = z_\beta\frac{\partial}{\partial z_\alpha} - z_\alpha^*\frac{\partial}{\partial z_\beta^*} - \omega^\alpha\frac{\partial}{\partial\omega^\beta} + \omega^{\beta*}\frac{\partial}{\partial\omega^{\alpha*}} - \tfrac{1}{3}\delta_\beta^\alpha\left(z_\gamma\frac{\partial}{\partial z_\gamma} - z_\gamma^*\frac{\partial}{\partial z_\gamma^*} - \omega^\gamma\frac{\partial}{\partial\omega^\gamma} + \omega^{\gamma*}\frac{\partial}{\partial\omega^{\gamma*}}\right),$$

$$\bar{p}_\beta^\alpha = s\frac{z_\beta z_\alpha^*}{z_\gamma z_\gamma^*} - \frac{\omega^\alpha\omega^{\beta*}}{\omega^\gamma\omega^{\gamma*}} - \tfrac{1}{3}(s-1)\delta_\beta^\alpha,$$

$$\bar{K}_\beta^\alpha = \rho_2\bar{p}_\beta^\alpha - i\left(z_\beta\frac{\partial}{\partial z_\alpha} + z_\alpha^*\frac{\partial}{\partial z_\beta^*} - \omega^\alpha\frac{\partial}{\partial\omega^\beta} - \omega^{\beta*}\frac{\partial}{\partial\omega^{\alpha*}}\right) + \frac{i}{3}\delta_\beta^\alpha\left(z_\gamma\frac{\partial}{\partial z_\gamma} + z_\gamma^*\frac{\partial}{\partial z_\gamma^*} - \omega^\gamma\frac{\partial}{\partial\omega^\gamma} - \omega^{\gamma*}\frac{\partial}{\partial\omega^{\gamma*}}\right). \tag{4.13}$$

By construction these expressions obey the same tracelessness and Hermiticity conditions as J_β^α, p_β^α, and K_β^α. Further, it is quite easy to check that on the one hand $\bar{J}$ and $\bar{K}$ obey the $SL(3, C)$ commutation rules, while $\bar{J}$ and $\bar{p}$ obey those of $SU(3) \times T_8$. (These statements are true, whatever be the values of ρ_2 and s.) Now it is comparatively easy to handle $\bar{J}$, $\bar{p}$, and $\bar{K}$. However, under certain conditions, these are nothing but the generators J, p, and K given earlier! Let us proceed to demonstrate this fact. First of all, one can see that the variables χ and χ^*, where

$$\chi = z_\alpha\omega^\alpha, \tag{4.14}$$

commute with $\bar{J}_\beta^\alpha$, $\bar{p}_\beta^\alpha$, *and* $\bar{K}_\beta^\alpha$. Next, let us use in place of the independent variables z_α, ω^α, the following six quantities (and their complex conjugates):

$$\alpha = z_2/z_3, \quad \beta = z_1/z_3, \quad z_3; \qquad \gamma = -\omega^2/\omega^1, \quad \chi, \quad \omega^1. \tag{4.15}$$

These are independent variables. One can now express all the operators $\bar{J}$, $\bar{p}$, $\bar{K}$ in terms of α, β, z_3, γ, χ, and ω^1 (and their complex conjugates). Then the fact that χ and χ^* commute with $\bar{J}$ and $\bar{K}$ means that when this is done, *there will be no terms in* $\bar{J}$ *and* $\bar{K}$ *involving* $(\partial/\partial\chi)$ *and* $(\partial/\partial\chi^*)$. Therefore, if after expressing everything in terms of the new variables, we set χ and χ^* identically equal to zero, the resulting linear differential operators will continue to obey the *same* commutation rules as $\bar{J}$, $\bar{p}$, and $\bar{K}$. Imagine now that χ and χ^* have been equated to zero. Then one finds (and this is no surprise) that the only dependence on z_3, ω^1, z_3^*, and ω^{1*} in any of these operators is through the quantities

$$E = z_3\frac{\partial}{\partial z_3}, \quad E^* = z_3^*\frac{\partial}{\partial z_3^*}, \quad F = \omega^1\frac{\partial}{\partial\omega^1}, \quad F^* = \omega^{1*}\frac{\partial}{\partial\omega^{1*}}. \tag{4.16}$$

Once more, it follows that the commutation rules

among $\bar{J}$, $\bar{p}$, and $\bar{K}$ will not be affected at all, if each of the operators E, E^*, F, F^* is replaced by a numerical quantity. To make contact with the known expressions for J, p, and K, one has to set[21]

$$E \to -1 + \tfrac{1}{2}(m_2 - m_3), \quad E^* \to -1 - \tfrac{1}{2}(m_2 - m_3),$$
$$F \to -1 - \tfrac{1}{2}m_2, \quad F^* \to -1 + \tfrac{1}{2}m_2. \quad (4.17)$$

Let us indicate by an asterisk any equation which is valid only when $\chi = \chi^* = 0$ and E, E^*, F, F^* have been replaced as above. Then we have

$$\begin{aligned} \bar{J}^\alpha_\beta &\stackrel{*}{=} J^\alpha_\beta, \\ \bar{p}^\alpha_\beta &\stackrel{*}{=} p^\alpha_\beta, \\ \bar{K}^\alpha_\beta &\stackrel{*}{=} K^\alpha_\beta. \end{aligned} \quad (4.18)$$

As a consequence of these remarks, it is enough to find a relationship among $\bar{J}$, $\bar{K}$, and $\bar{p}$, in order to obtain one among J, K, and p. Further, we need not keep track of terms which have χ or χ^* as a coefficient.

The task of expressing $\bar{K}$ as a function of $\bar{J}$ and $\bar{p}$ is comparatively easy. It is convenient to regard $\bar{K}$, $\bar{p}$, and $\bar{J}$ as 3×3 matrices, the subscript (superscript) being identified as the column (row) index. We begin by defining the square of the matrix $\bar{p}$:

$$\begin{aligned} q^\alpha_\beta &= (\bar{p}\bar{p})^\alpha_\beta = \bar{p}^\alpha_\gamma \bar{p}^\gamma_\beta \\ &\stackrel{*}{=} \frac{s(s+2)}{3}\frac{z_\beta z^*_\alpha}{z_\gamma z^*_\gamma} + \frac{(2s+1)}{3}\frac{\omega^\alpha \omega^{\beta *}}{\omega^\gamma \omega^{\gamma *}} + \frac{(s-1)^2}{9}\delta^\alpha_\beta. \end{aligned} \quad (4.19)$$

We next compute certain matrix products involving $\bar{p}$, q, and $\bar{J}$. We use the following abbreviations:

$$\begin{aligned} z_\gamma z^*_\gamma &= |z|^2, \quad \omega^\gamma \omega^{\gamma *} = |\omega|^2; \\ R^\alpha_\beta &= \omega^\alpha z_\beta z^*_\gamma \frac{\partial}{\partial \omega^\gamma} + z^*_\alpha \omega^{\beta *} z_\gamma \frac{\partial}{\partial \omega^{\gamma *}}; \\ S^\alpha_\beta &= \omega^\alpha z_\beta \omega^{\gamma *} \frac{\partial}{\partial z_\gamma} + z^*_\alpha \omega^{\beta *} \omega^\gamma \frac{\partial}{\partial z^*_\gamma}. \end{aligned} \quad (4.20)$$

By staring long and hard at the expressions involved, one sees that it is advantageous to compute the following combinations[22]:

$$\begin{aligned} (\bar{p}\bar{J} - \bar{J}\bar{p})^\alpha_\beta &\stackrel{*}{=} 3p^\alpha_\beta + \frac{s}{|z|^2}R^\alpha_\beta - \frac{1}{|\omega|^2}S^\alpha_\beta \\ &- s\left(z_\beta \frac{\partial}{\partial z_\alpha} + z^*_\alpha \frac{\partial}{\partial z^*_\beta}\right) + \left(\omega^\alpha \frac{\partial}{\partial \omega^\beta} + \omega^{\beta *}\frac{\partial}{\partial \omega^{\alpha *}}\right) \\ &- 2s\frac{z_\beta z^*_\alpha}{|z|^2} + 2\frac{\omega^\alpha \omega^{\beta *}}{|\omega|^2}, \end{aligned} \quad (4.21a)$$

$$\begin{aligned} (q\bar{J} - \bar{J}q)^\alpha_\beta &\stackrel{*}{=} 3q^\alpha_\beta - \tfrac{2}{3}(s^2 + s + 1)\delta^\alpha_\beta \\ &+ \frac{1}{3}\frac{s(s+2)}{|z|^2}R^\alpha_\beta + \frac{1}{3}\frac{(2s+1)}{|\omega|^2}S^\alpha_\beta \\ &- \frac{1}{3}s(s+2)\left(z_\beta \frac{\partial}{\partial z_\alpha} + z^*_\alpha \frac{\partial}{\partial z^*_\beta}\right) \\ &- \frac{1}{3}(2s+1)\left(\omega^\alpha \frac{\partial}{\partial \omega^\beta} + \omega^{\beta *}\frac{\partial}{\partial \omega^{\alpha *}}\right) \\ &- \frac{2}{3}s(s+2)\frac{z_\beta z^*_\alpha}{|z|^2} - \frac{2}{3}(2s+1)\frac{\omega^\alpha \omega^{\beta *}}{|\omega|^2}, \end{aligned} \quad (4.21b)$$

$$\begin{aligned} (q\bar{J}\bar{p} - \bar{p}\bar{J}q)^\alpha_\beta &\stackrel{*}{=} -\frac{2}{3}(s^2 + s + 1)p^\alpha_\beta \\ &- \frac{s(s+1)}{|z|^2}R^\alpha_\beta + \frac{s(s+1)}{|\omega|^2}S^\alpha_\beta \\ &- \frac{1}{9}(s-1)^2(\bar{p}\bar{J} - \bar{J}\bar{p})^\alpha_\beta - \frac{1}{3}(s-1)(q\bar{J} - \bar{J}q)^\alpha_\beta. \end{aligned} \quad (4.21c)$$

We compare these expressions with $\bar{K}^\alpha_\beta$ as written down in Eq. (4.13) and try to express $\bar{K}$ as a linear combination of the above expressions. Explicitly we set

$$\begin{aligned} \bar{K}^\alpha_\beta - \rho_2 \bar{p}^\alpha_\beta &\stackrel{*}{=} ia(\bar{p}\bar{J} - \bar{J}\bar{p})^\alpha_\beta + ib(q\bar{J} - \bar{J}q)^\alpha_\beta \\ &+ ic[q\bar{J}\bar{p} - \bar{p}\bar{J}q + \tfrac{1}{9}(s-1)^2 \\ &\times (\bar{p}\bar{J} - \bar{J}\bar{p}) + \tfrac{1}{3}(s-1)(q\bar{J} - \bar{J}q)]^\alpha_\beta \\ &+ \text{terms involving } \bar{p}^\alpha_\beta, q^\alpha_\beta, \text{ and } \delta^\alpha_\beta. \end{aligned} \quad (4.22)$$

Comparing the coefficients of $z_\beta(\partial/\partial z_\alpha)$, $\omega^\alpha(\partial/\partial\omega^\beta)$, R^α_β, and S^α_β, we get:

$$\begin{aligned} as + \tfrac{1}{3}bs(s+2) &= 1, \\ a - \tfrac{1}{3}b(2s+1) &= 1, \\ as + \tfrac{1}{3}bs(s+2) - cs(s+1) &= 0, \\ -a + \tfrac{1}{3}b(2s+1) + cs(s+1) &= 0. \end{aligned} \quad (4.23)$$

These four equations are consistent with one another, and give a unique solution for a, b, and c:

$$\begin{aligned} a &= (s^2 + 4s + 1)/3s(s+1), \\ b &= (1-s)/s(s+1), \\ c &= 1/s(s+1). \end{aligned} \quad (4.24)$$

We are now in a position to insert these values of a, b, c into (4.22), and convert (4.22) into a relation that holds between the generators K^α_β, p^α_β, and J^α_β given in Eqs. (4.6) to (4.10). Instead of expressing this relationship in terms of q, we prefer to write it in

[21] This replacement of E, E^*, F, F^* by numbers corresponds to the statement that one reaches unitary representations of $SL(3, C)$ by starting with finite-dimensional nonunitary tensorial representations, and then making the rank of the tensor nonintegral.

[22] One is led to these expressions by noticing that the differential operators in $\bar{K}^\alpha_\beta$ appear with an explicit factor of i, whereas such factors are not present in either $\bar{p}^\alpha_\beta$ or $\bar{J}^\alpha_\beta$, and combining this with the form of the Hermiticity condition on $\bar{K}^\alpha_\beta$.

terms of the traceless part of q,[23] namely, in terms of

$$k^\alpha_\beta = q^\alpha_\beta - \tfrac{1}{3}\delta^\alpha_\beta q^\gamma_\gamma \stackrel{*}{=} p^\alpha_\gamma p^\gamma_\beta - \tfrac{1}{8}\delta^\alpha_\beta p^\delta_\gamma p^\gamma_\delta. \quad (4.25)$$

Straightforward algebra then gives:

$$K^\alpha_\beta = \rho_2 p^\alpha_\beta + \frac{2}{9} i \frac{(s^2+4s+1)}{s(s+1)}(pJ - Jp)^\alpha_\beta + \frac{2}{3} i \frac{(1-s)}{s(s+1)}(kJ - Jk)^\alpha_\beta + \frac{i}{s(s+1)}(kJp - pJk)^\alpha_\beta. \quad (4.26)$$

We have thus succeeded in expressing the $SL(3, C)$ generators K^α_β as functions of the $SU(3) \times T_8$ generators J^α_β and p^α_β. [The $SU(3)$ part is carried over unchanged from one group to the other, of course.] It remains to identify the parameters that define the $SU(3) \times T_8$ representation generated by J^α_β and p^α_β. In Sec. 3, we have seen that a UIR of $SU(3) \times T_8$ belonging to Type I is specified by the four parameters $(M^{(0)}, Y^{(0)}, p_0, p_1)$. Since we already know the $SU(3)$ content of the $SL(3, C)$ representation, in terms of m_2 and m_3, we must have

$$M^{(0)} = \tfrac{1}{2}m_2, \quad Y^{(0)} = \tfrac{1}{3}(2m_3 - m_2). \quad (4.27)$$

To determine p_0, and p_1, we note that, using the equality

$$\bar{p}^\alpha_\beta \stackrel{*}{=} p^\alpha_\beta, \quad (4.28)$$

we can quite easily compute the invariants associated with p^α_β of Eq. (4.9). We find

$$\mathrm{Tr}\,(pp) \equiv p^\alpha_\beta p^\beta_\alpha = \tfrac{2}{3}(s^2 + s + 1),$$
$$\mathrm{Tr}\,(ppp) \equiv p^\alpha_\beta p^\beta_\gamma p^\gamma_\alpha = \tfrac{1}{9}(2s+1)(s^2 + s - 2). \quad (4.29)$$

On the other hand, we can also compute these same invariants in the general $SU(3) \times T_8$ UIR set up in Sec. 3. In the UIR $(M^{(0)}, Y^{(0)}, p_0, p_1)$, we find

$$\mathrm{Tr}\,(PP) = 6(p_0^2 + p_1^2),$$
$$\mathrm{Tr}\,(PPP) = 6p_0(p_0^2 - 3p_1^2). \quad (4.30)$$

[Here, one must use the numerical relationship between P^α_β written as a matrix, and the same quantities written in the form P_{IMY}; this is given in Eq. (3.2).] However, equating (4.29) and (4.30) does not uniquely determine p_0 and p_1, and one has to compute the Casimir operators of $SL(3, C)$ and the remaining ones of $SU(3) \times T_8$.[24] Thereupon, one finds the following values for p_0 and p_1 in terms of s:

$$p_0 = \tfrac{1}{6}(2s+1), \quad p_1 = -\tfrac{1}{2}(3)^{-\frac{1}{2}}. \quad (4.31)$$

Putting all the pieces together, we can state the final result as follows: In order to obtain the generators J^α_β, K^α_β of a principal nondegenerate series UIR $(m_2, m_3, \rho_2, \rho_3)$ of $SL(3, C)$, we start with a Type I UIR $(M^{(0)}, Y^{(0)}, p_0, p_1)$ of $SU(3) \times T_8$, with $M^{(0)}$ and $Y^{(0)}$ being determined by m_2 and m_3 by (4.27), and p_0 and p_1 given in terms of $s = (\rho_3/\rho_2) - 1$ by (4.31). In this UIR of $SU(3) \times T_8$ we can construct the "noncompact" $SL(3, C)$ generators K^α_β explicitly in terms of the $SU(3) \times T_8$ generators J^α_β and p^α_β by means of the formula (4.26). The $SU(3)$ generators J^α_β are the same for both groups. Since in Sec. 3 we have shown how one can construct any $SU(3) \times T_8$ UIR in an $SU(3)$ basis, with a complete solution to the multiplicity problem, and with explicit expressions for the matrix elements of p^α_β in such a basis, one can in principle use (4.26) to compute the matrix elements of K^α_β in an $SU(3)$ basis.

5. ANALYSIS OF PRINCIPAL DEGENERATE SERIES UIR'S

We now turn to the UIR's of $SL(3, C)$ of the principal degenerate series, and show how their generators can be constructed in terms of the $SU(3) \times T_8$ generators belonging to UIR's of Type II. As we have stated in Sec. 1, in this case there is no multiplicity problem at all. However, the results are interesting in the sense that the expansion formula relating K^α_β to J^α_β and p^α_β is quite different in appearance from (4.26).

Consider a UIR (m_3, ρ_3) of $SL(3, C)$, belonging to the principal degenerate series. Its explicit construction is given in Sec. 2. As in the previous section, we begin by defining a new set of basis vectors differing from the old ones by a phase factor:

$$|z'\rangle = \exp\,[i\psi(z')]\,|z'\rangle,$$
$$\psi(z') = -\tfrac{1}{2}\rho_3 \ln\,(1 + |\alpha|^2 + |\beta|^2). \quad (5.1)$$

(The symbol z' stands for the two complex variables α, β.) The orthonormality relation is unchanged:

$$(z'_2 \,|\, z'_1) = \delta(\alpha_2 - \alpha_1)\delta(\beta_2 - \beta_1). \quad (5.2)$$

The unitary operator $U(g)$ acts as follows:

$$U(g)\,|z'\rangle = \exp\,[i\psi(z') - i\psi(z'_g)]\lambda(g, z')\,|z'_g\rangle. \quad (5.3)$$

The parameters corresponding to z'_g, and the function $\lambda(g, z')$, have been defined in Eqs. (2.19) and (2.20).

[23] It is easy to see that if J^α_β and p^α_β are the generators of the Type I UIR $(M^{(0)}, Y^{(0)}, p_0, p_1)$ of $SU(3) \times T_8$, then J^α_β and

$$k^\alpha_\beta = p^\alpha_\gamma p^\gamma_\beta - \tfrac{1}{3}\delta^\alpha_\beta p^\delta_\gamma p^\gamma_\delta$$

are the generators of the UIR $(M^{(0)}, Y^{(0)}, p_0^2 - p_1^2, -2p_0p_1)$ of $SU(3) \times T_8$. (If $p_0 = 0$, $\pm\sqrt{3}\,p_1$, then the set J^α_β, k^α_β is reducible.)

[24] One has to equate the values of the Casimir operators of $SU(3) \times T_8$ to the "contracted" values of those of $SL(3, C)$.

With the choice of basis states, the right-hand side of (5.3) is independent of ρ_3 when g belongs to $SU(3)$.

We now follow closely the steps of Sec. 4. First we list the $SU(3)$ generators as linear differential operators:

$$
\begin{aligned}
J_1^1 &= \frac{1}{3} m_3 + \beta \frac{\partial}{\partial \beta} - \beta^* \frac{\partial}{\partial \beta^*}, \\
J_1^2 &= \beta \frac{\partial}{\partial \alpha} - \alpha^* \frac{\partial}{\partial \beta^*}, \\
J_1^3 &= -\frac{1}{2} m_3 \beta - \frac{3}{2} \beta - \alpha\beta \frac{\partial}{\partial \alpha} - \beta^2 \frac{\partial}{\partial \beta} - \frac{\partial}{\partial \beta^*}, \\
J_2^2 &= \frac{1}{3} m_3 + \alpha \frac{\partial}{\partial \alpha} - \alpha^* \frac{\partial}{\partial \alpha^*}, \\
J_2^3 &= -\frac{1}{2} m_3 \alpha - \frac{3}{2} \alpha - \alpha^2 \frac{\partial}{\partial \alpha} - \frac{\partial}{\partial \alpha^*} - \alpha\beta \frac{\partial}{\partial \beta}.
\end{aligned} \tag{5.4}
$$

The noncompact generators K_β^α split into ρ_3 dependent and ρ_3 independent parts:

$$K_\beta^\alpha = \rho_3 p_\beta^\alpha + K_\beta^{(0)\alpha}. \tag{5.5}$$

We list p and $K^{(0)}$ next:

$$
\begin{aligned}
&p_1^1 = \tfrac{1}{3}A^{-1}(2\beta\beta^* - \alpha\alpha^* - 1), \quad p_1^2 = A^{-1}\beta\alpha^*, \\
&p_1^3 = A^{-1}\beta, \quad p_2^2 = \tfrac{1}{3}A^{-1}(2\alpha\alpha^* - \beta\beta^* - 1), \\
&p_2^3 = A^{-1}\alpha; \quad A = 1 + |\alpha|^2 + |\beta|^2,
\end{aligned} \tag{5.6}
$$

and

$$
\begin{aligned}
K_1^{(0)1} &= -i - i\left(\beta \frac{\partial}{\partial \beta} + \beta^* \frac{\partial}{\partial \beta^*}\right), \\
K_1^{(0)2} &= -i\left(\beta \frac{\partial}{\partial \alpha} + \alpha^* \frac{\partial}{\partial \beta^*}\right), \\
K_1^{(0)3} &= \frac{i}{2} m_3 \beta + \frac{3}{2} i\beta + i\left(\alpha\beta \frac{\partial}{\partial \alpha} + \beta^2 \frac{\partial}{\partial \beta} - \frac{\partial}{\partial \beta^*}\right), \\
K_2^{(0)2} &= -i - i\left(\alpha \frac{\partial}{\partial \alpha} + \alpha^* \frac{\partial}{\partial \alpha^*}\right), \\
K_2^{(0)3} &= \frac{i}{2} m_3 \alpha + \frac{3}{2} i\alpha + i\left(\alpha^2 \frac{\partial}{\partial \alpha} - \frac{\partial}{\partial \alpha^*} + \alpha\beta \frac{\partial}{\partial \beta}\right).
\end{aligned} \tag{5.7}
$$

From the components of J, K, and p that we have listed, the remaining ones may be obtained by using the Hermiticity relations and tracelessness. Here again we may regard p_β^α as the "contracted form" of K_β^α, so that

$$p_\beta^\alpha = \operatorname*{Lt}_{\rho_3 \to \infty} \frac{1}{\rho_3} K_\beta^\alpha. \tag{5.8}$$

Therefore, since J_β^α is independent of ρ_3, J_β^α, and p_β^α are the generators of a UIR of $SU(3) \times T_8$ of Type II [we already know the $SU(3)$ content of the space of the UIR]. We go on to show that K_β^α can be reconstructed in terms of J_β^α and p_β^α.

The first step is to introduce more symmetric expressions for the generators. Introduce three auxiliary complex variables z_α and their complex conjugates, and set

$$
\begin{aligned}
\bar{J}_\beta^\alpha &= z_\beta \frac{\partial}{\partial z_\alpha} - z_\alpha^* \frac{\partial}{\partial z_\beta^*} - \frac{1}{3}\delta_\beta^\alpha\left(z_\gamma \frac{\partial}{\partial z_\gamma} - z_\gamma^* \frac{\partial}{\partial z_\gamma^*}\right), \\
\bar{p}_\beta^\alpha &= \frac{z_\beta z_\alpha^*}{z_\gamma z_\gamma^*} - \frac{1}{3}\delta_\beta^\alpha, \\
\bar{K}_\beta^\alpha &= \rho_3 \bar{p}_\beta^\alpha - i\left(z_\beta \frac{\partial}{\partial z_\alpha} + z_\alpha^* \frac{\partial}{\partial z_\beta^*}\right) \\
&\quad + \frac{i}{3}\delta_\beta^\alpha\left(z_\gamma \frac{\partial}{\partial z_\gamma} + z_\gamma^* \frac{\partial}{\partial z_\gamma^*}\right).
\end{aligned} \tag{5.9}
$$

It is easy to check that $\bar{J}$ and $\bar{p}$ obey the commutation relations of $SU(3) \times T_8$, while $\bar{J}$ and $\bar{K}$ obey those of $SL(3, C)$, for any value of ρ_3. If now we rewrite these operators in terms of the new variables

$$\alpha = z_2/z_3, \quad \beta = z_1/z_3, \quad z_3, \tag{5.10}$$

and their complex conjugates, then the variables z_3 and z_3^* and the operators $\partial/\partial z_3$ and $\partial/\partial z_3^*$ appear only in the following combinations:

$$z_3(\partial/\partial z_3), \quad z_3^*(\partial/\partial z_3^*). \tag{5.11}$$

Therefore, if after expressing $\bar{J}$, $\bar{p}$, and $\bar{K}$ in terms of the new variables we were to replace the quantities in (5.11) by pure numbers, this would not alter the commutation relations obeyed by $\bar{J}$, $\bar{p}$, and $\bar{K}$. We make the choice

$$z_3 \frac{\partial}{\partial z_3} \to -\frac{3}{2} - \frac{1}{2} m_3, \quad z_3^* \frac{\partial}{\partial z_3^*} \to -\frac{3}{2} + \frac{1}{2} m_3, \tag{5.12}$$

and indicate by an asterisk any equation valid only after (5.12) has been used. Then we find

$$\bar{J}_\beta^\alpha \stackrel{*}{=} J_\beta^\alpha, \quad \bar{p}_\beta^\alpha \stackrel{*}{=} p_\beta^\alpha, \quad \bar{K}_\beta^\alpha \stackrel{*}{=} K_\beta^\alpha, \tag{5.13}$$

where the J, p, and K operators are listed in Eqs. (5.4) to (5.7). Thus it suffices to find a way to express $\bar{K}$ in terms of $\bar{p}$ and $\bar{J}$.

For this purpose, we compute the expression $\bar{p}\bar{J} - \bar{J}\bar{p}$, so that

$$(\bar{p}\bar{J} - \bar{J}\bar{p})_\beta^\alpha \stackrel{*}{=} -\delta_\beta^\alpha - \left(z_\beta \frac{\partial}{\partial z_\alpha} + z_\alpha^* \frac{\partial}{\partial z_\beta^*}\right). \tag{5.14}$$

This leads immediately to the desired relation among $\bar{K}$, $\bar{p}$, and $\bar{J}$:

$$\bar{K}_\beta^\alpha \stackrel{*}{=} \rho_3 \bar{p}_\beta^\alpha + i(\bar{p}\bar{J} - \bar{J}\bar{p})_\beta^\alpha. \tag{5.15}$$

That is, the $SL(3, C)$ generators K_β^α of (5.5) are expressible in terms of the $SU(3) \times T_8$ generators J_β^α, p_β^α of (5.4) and (5.6) by means of the formula

$$K_\beta^\alpha = \rho_3 p_\beta^\alpha + i(pJ - Jp)_\beta^\alpha. \tag{5.16}$$

It remains to identify the UIR of $SU(3) \times T_8$ that one must use in order to obtain the $SL(3, C)$ generators for a given UIR (m_3, ρ_3) of $SL(3, C)$. We have seen in Sec. 3 that a Type II UIR of $SU(3) \times T_8$ is labeled in the form $(Y^{(0)}, p_0)$. Comparing the $SU(3)$ contents involved, we deduce the equation

$$Y^{(0)} = \tfrac{2}{3}m_3. \tag{5.17}$$

To fix p_0, we note that in the UIR $(Y^{(0)}, p_0)$ of $SU(3) \times T_8$, we have

$$\mathrm{Tr}\,(PP) = 6p_0^2, \tag{5.18}$$

while from (5.6) and (5.13) we have

$$\mathrm{Tr}\,(pp) = \tfrac{2}{3}. \tag{5.19}$$

Comparing (5.18) and (5.19) we conclude that in order to obtain the generators of the principal degenerate series UIR (m_3, ρ_3) of $SL(3, C)$, we start with the generators J, p of the Type II UIR $(\frac{2}{3}m_3, \frac{1}{3})$ of $SU(3) \times T_8$, and in this UIR construct K according to (5.16). Then J and K are the required generators of $SL(3, C)$.

6. CONCLUSION

We have demonstrated that given any unitary representation of $SL(3, C)$ belonging to the principal series (either nondegenerate or degenerate), one may express the corresponding infinitesimal generators as functions of a suitably chosen unitary representation of $SU(3) \times T_8$. Since we know how to handle the representations of the latter group in an $SU(3)$ basis, we have achieved a solution of the multiplicity problem for $SL(3, C)$.

"Expansion formulas" of this kind, expressing the generators of a semisimple noncompact Lie group as functions of the generators of a related nonsemisimple Lie group, have been used in the past.[25] For example, formulas showing how one may construct the generators of the group $SO(n, 1)$ in terms of those of $E(n)$ have been derived. First, one notices that in these cases, the UIR's of $SO(n, 1)$ are multiplicity-free in their reduction with respect to $SO(n)$. Perhaps related to this is the fact that the formulas expressing the $SO(n, 1)$ generators in terms of those of $E(n)$ are *universal*, that is, representation-independent. Namely, given only the commutation relations obeyed by the $E(n)$ generators, one is able to construct functions of these generators, obeying the commutation relations of the group $SO(n, 1)$. The functional form of the $SO(n, 1)$ generators as expressed in terms of the $E(n)$ generators, as well as the fact that they obey the $SO(n, 1)$ commutation relations, are independent of the particular representation of $E(n)$ that one starts with. Second, one can carry out the construction of the $SO(n, 1)$ generators starting with *any* unitary irreducible representation of $E(n)$ and using the universal formulas. This fact is of course a consequence of the first property. Both these properties are in direct contrast to the case of $SL(3, C)$. First of all, we find that the formulas expressing K_β^α as functions of J_β^α and P_β^α have quite different forms in the nondegenerate UIR's on the one hand and the degenerate UIR's on the other. Thus, if one starts with a UIR of $SU(3) \times T_8$ with generators J_β^α and P_β^α and does not specify which type of UIR one has, but depends only on the commutation relations obeyed by them, it turns out to be impossible to construct functions K_β^α of J_β^α and P_β^α (at least, functions involving J_β^α no more than linearly!) which will have the commutation relations of $SL(3, C)$. In other words, if we take the Lie algebra of $SU(3) \times T_8$, and write down the most general operator function of J_β^α and P_β^α (which is at most linear in J_β^α) having the right $SU(3)$ transformation properties, we obtain an expression involving a certain number of numerical coefficients. If we now demand that these operators obey the $SL(3, C)$ commutation rules, we get a large number of algebraic equations for the coefficients; *and these equations turn out to be inconsistent!* Thus, the reason why the expressions we have obtained in each case for K_β^α in terms of J_β^α and P_β^α do have the correct commutation properties is that in each kind of representation of $SU(3) \times T_8$ there always exist many operator identities among J_β^α and P_β^α, in addition to the Lie-algebra relations. These additional relations differ from one type of representation to another and cannot be specified in advance. Second, assuming that the list of principal series UIR's of $SL(3, C)$ given by Gel'fand and Naimark is complete, there is a whole class of UIR's of $SU(3) \times T_8$ in which we are unable to form functions having the $SL(3, C)$ commutation rules (or so it seems at present!). This class of UIR's is what we call Type III. At any rate, one can say that on contracting principal series UIR's of $SL(3, C)$ in the way we have prescribed, one ends up with $SU(3) \times T_8$ representations of Types I and II only. We conjecture that these properties, which distinguish $SL(3, C)$ from $SO(n, 1)$, for example, are intimately related to the existence of the multiplicity problem for $SL(3, C)$.

We have here been able to derive the principal series UIR's of $SL(3, C)$ from the $SU(3) \times T_8$ UIR's.

[25] See Ref. 7.

We expect that once the matrix elements of the $SL(3, C)$ generators in the $SU(3)$ basis are known, then by a process of analytic continuation in the parameters of the representation one will arrive at the supplementary series of UIR's of $SL(3, C)$. This certainly is the case for $SO(n, 1)$.[26]

When one examines the expansion formulas a bit closely, it seems unlikely that the matrix elements of K_β^α in the $SU(3)$ basis will have any simple algebraic appearance. For instance, they do not seem to be square roots of rational functions of the various parameters involved. They seem rather to be sums of such expressions. An analogous situation was encountered in the problem of constructing $SU(3)$ representations in an $R(3)$ basis.[27]

An unfortunate feature of the present work must be admitted. We have derived the expansion formulas by starting with the $SL(3, C)$ representations in some form and then by algebraic manipulations showing that the requisite relations hold among the $SL(3, C)$ and $SU(3) \times T_8$ generators. Such an approach would be valueless if one's aim was to obtain the $SL(3, C)$ representations starting with just the representations of the far simpler $SU(3) \times T_8$ structure. Hopefully, the experience gained herein will help us reverse this situation.

In any case, one can see that similar solutions to the multiplicity problems exist for all the $SL(n, C)$ groups. It would be most interesting to develop similar techniques to deal with the groups $SL(3, R)$, $O(3, 2)$, and $O(4, 2)$ since till now it has been practically impossible to deal with the multiplicity problems in these cases.

ACKNOWLEDGMENTS

The author would like to thank Professor A. Böhm and Professor E. C. G. Sudarshan for much encouragement and help, and Dr. S. M. Vaidya for a careful reading of the manuscript.

Most of this work was done while the author was a member of the Physics Department of Syracuse University during the academic year 1966–67. For the hospitality of that institution, he extends his thanks to Professor N. Ginsburg and Professor E. C. G. Sudarshan.

APPENDIX A

The principal series of UIR's of $SL(2, C)$ may be obtained from the UIR's of $E(3)$ by an "expansion formula." We briefly outline the method, so that one may compare this with the $SL(3, C)$ case.

The $SL(2, C)$ and $E(3)$ algebras are spanned by J_k, K_k, and J_k, P_k $(k = 1, 2, 3)$ respectively. The $SL(2, C)$ commutation rules are

$$[J_k, J_l] = i\epsilon_{klm}J_m, \tag{A1a}$$

$$[J_k, K_l] = i\epsilon_{klm}K_m, \tag{A1b}$$

$$[K_k, K_l] = -i\epsilon_{klm}J_m. \tag{A1c}$$

For $E(3)$, (A1a) remains valid, while (A1b) and (A1c) are replaced by:

$$[J_k, P_l] = i\epsilon_{klm}P_m,$$
$$[P_k, P_l] = 0. \tag{A2}$$

The principal series of UIR's of $SL(2, C)$ is parametrized in the form $\{j_0, \rho\}$, where j_0 assumes one of the values $0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, \infty$, while ρ is any real number. j_0 specifies the "lowest" representation of $SU(2)$ (smallest spin) present in the UIR of $SL(2, C)$.[28] The UIR's of $E(3)$ are labeled in the form $\{j_0, p\}$ where once again j_0 is the lowest representation of $SU(2)$ present, and $|p|$ is the magnitude of P.[29] Explicitly one has:

$$SL(2, C): \quad K_jK_j - J_jJ_j = 1 + \rho^2 - j_0^2, \quad K_jJ_j = -\rho j_0;$$
$$E(3): \quad P_jP_j = p^2, \quad P_jJ_j = -pj_0. \tag{A3}$$

Let us for simplicity work with UIR's of $E(3)$ in which P_j is normalized so that $p = +1$. Then given the generators J_j, P_j of such a representation, one finds that the operators

$$K_j = \rho P_j + \tfrac{1}{2}\epsilon_{jkl}(J_kP_l - P_kJ_l) \tag{A4}$$

obey both (A1b) and (A1c), for any value of the (real) parameter ρ. The main point in verifying this statement is that one needs only to use the commutation relations of $E(3)$, namely (A2), without specifying the UIR of $E(3)$ involved (apart of course from the condition $p = +1$). One finds easily that the UIR of $SL(2, C)$ obtained this way is $\{j_0, \rho\}$. Thus by starting with a suitable UIR of $E(3)$ (choice of j_0) and choosing ρ in (A4) appropriately, one obtains all the principal series UIR's of $SL(2, C)$. (A4) can be rewritten as

$$K_j = (\rho + i)P_j - \epsilon_{jkl}P_kJ_l. \tag{A5}$$

In tensor notation, this becomes somewhat similar to the $SL(3, C)$ expansion formulas; writing

$$K_1^1 = -K_2^2 = K_3,$$
$$K_1^2 = K_1 + iK_2,$$
$$K_2^1 = K_1 - iK_2,$$

[26] J. G. Kuriyan, N. Mukunda, and E. C. G. Sudarshan, Commun. Math. Phys. **8**, 204 (1968).

[27] See the comments by G. Racah, Ref. 2.

[28] For details, see I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (The Macmillan Co., New York, 1963).

[29] For representations of $E(3)$, see W. Pauli, "Continuous Groups in Quantum Mechanics," in *Ergebnisse der exakten Naturwissenschaften* (Springer-Verlag, Berlin, 1965), Vol. 37.

and similarly for P and J, (A5) looks as follows:

$$K^{\alpha}_{\beta} = \rho P^{\alpha}_{\beta} + \tfrac{1}{2}i(PJ - JP)^{\alpha}_{\beta}. \tag{A6}$$

APPENDIX B

We describe in this appendix a technique, due to Goebel,[6] for obtaining the UIR's of a class of groups which have a semidirect product structure. The main virtues of this method are its simplicity, the fact that one obtains all the UIR's, an explicit method for labeling basis vectors, and explicit expressions for matrix elements of the generators.

We consider groups of the form $C \times T$: C denotes any compact simple Lie group, and T stands for a set of Abelian generators transforming according to some irreducible finite-dimensional unitary representation of C. (The $\times$ sign denotes the semidirect product.) We denote elements of C by Ω, $\Omega', \cdots$, and we use the symbol j to run over the set of all inequivalent UIR's of C. Thus in the UIR j of C, the element Ω is represented by a unitary matrix which we write as

$$\Omega \to (\Omega)^{j}_{mn}. \tag{B1}$$

(It is assumed that in each UIR a definite way to label the states has been adopted.) These matrices obey the group composition law in the form

$$\sum_{m'} (\Omega)^{j}_{mm'}(\Omega')^{j}_{m'm''} = (\Omega\Omega')^{j}_{mm''} \tag{B2}$$

and are unitary:

$$\sum_{m} (\Omega)^{j*}_{mm'}(\Omega)^{j}_{mm''} = \delta_{m'm''}. \tag{B3}$$

We denote the UIR that T belongs to by λ and label the components of T by $\alpha, \beta, \cdots$. Looking for a UIR of $C \times T$ amounts to finding unitary operators $U(\Omega)$, and operators T^{λ}_{α}, obeying the following conditions:

$$U(\Omega)U(\Omega') = U(\Omega\Omega'), \tag{B4a}$$

$$U(\Omega)T^{\lambda}_{\alpha}U(\Omega^{-1}) = (\Omega)^{\lambda}_{\beta\alpha}T^{\lambda}_{\beta}, \tag{B4b}$$

$$[T^{\lambda}_{\alpha}, T^{\lambda}_{\beta}] = 0. \tag{B4c}$$

We assume that the UIR λ is equivalent to its complex conjugate. In that case, we have to impose appropriate Hermiticity conditions on T^{λ}_{α}.

Equations (B4) are solved in the following manner. We introduce the Hilbert space $\mathcal{H}$ of all square integrable functions $f(\Omega)$ on C. (We use, of course, the left- and right-translation invariant integration over C, normalized so that C has unit volume.) A nonnormalizable basis for $\mathcal{H}$ consists of the vectors

$$|\Omega\rangle;\ \langle\Omega' \,|\, \Omega\rangle = \delta(\Omega', \Omega), \tag{B5}$$

the delta function being normalized so that

$$\int \delta(\Omega', \Omega)\, d\Omega = 1. \tag{B6}$$

The operators $U(\Omega)$ are defined via left translation as

$$U(\Omega)\,|\Omega'\rangle = |\Omega\Omega'\rangle, \tag{B7}$$

so that they are unitary, and obey (B4a). The states $|\Omega\rangle$ are chosen to be simultaneous eigenstates of T^{λ}_{α}, and one sets

$$T^{\lambda}_{\alpha}\,|\Omega\rangle = \sum_{\beta} t_{\beta}(\Omega^{-1})^{\lambda}_{\beta\alpha}\,|\Omega\rangle. \tag{B8}$$

Here t_{β} is any set of constants, as many in number as there are components of T, and obeying reality conditions that parallel the Hermiticity conditions on T. One can easily check that (B4b) is valid. Since the T^{λ}_{α} are simultaneously diagonal, (B4c) is obviously satisfied.

Equations (B7) and (B8) provide us with a class of unitary representations of $C \times T$, different members of this class being distinguished by the choice of t_{β}. These representations may be reducible. At any rate, one can see easily that every UIR of $C \times T$ is contained in one of the representations constructed above. This follows from two facts: first, if the numerical tensor t_{β} has a continuous "little group," this "little group" is certainly compact and hence the Peter–Weyl theorem applies to it; second, in the representation constructed above, we have in fact chosen the regular representation of this "little group."

It is easy to write (B7) and (B8) in a basis made up of UIR's of C. For this we use the Peter–Weyl theorem for C, which says that the set of all the matrix elements (B1) for all j spans the Hilbert space $\mathcal{H}$. Correspondingly we may introduce a discrete basis in $\mathcal{H}$:

$$|j, m; n\rangle = (d_j)^{\frac{1}{2}} \int d\Omega(\Omega^{-1})^{j}_{mn}\,|\Omega\rangle. \tag{B9}$$

Here, d_j denotes the dimensionality of the UIR j of C. The representation matrices of C obey the usual completeness and orthogonality relations:

$$\sum_{jmn} d_j(\Omega)^{j}_{mn}(\Omega')^{j*}_{mn} = \delta(\Omega', \Omega), \tag{B10a}$$

$$\int d\Omega(\Omega)^{j'*}_{m'n'}(\Omega)^{j}_{mn} = d_j^{-1}\delta_{j'j}\delta_{m'm}\delta_{n'n}. \tag{B10b}$$

Using all these properties, one establishes that the states $|jmn\rangle$ form an orthonormal basis, namely,

$$\langle j'm'n' \,|\, jmn\rangle = \delta_{j'j}\delta_{m'm}\delta_{n'n}, \tag{B11}$$

and that they transform under $U(\Omega)$ as follows:

$$U(\Omega)\,|jmn\rangle = \sum_{m'} (\Omega)^{j}_{m'm}\,|jm'n\rangle. \tag{B12}$$

This last equation makes explicit the fact that $\mathcal{H}$ contains each UIR j of C as often as its dimension d_j. The label n, which is invariant under application of the unitary operators $U(\Omega)$, labels this multiplicity.

We can now compute the matrix elements of T^λ_α in the new basis. We find

$$\langle j'm'n'|\, T^\lambda_\alpha \,|jmn\rangle = (d_{j'}d_j)^{\frac{1}{2}} \sum_\beta t_\beta \int d\Omega (\Omega)^{j'*}_{n'm'} (\Omega)^j_{nm} (\Omega)^\lambda_{\beta\alpha}. \quad \text{(B13)}$$

The integral appearing here can be evaluated and expressed in terms of Clebsch–Gordan (CG) coefficients of the group C. We denote these, in analogy to the familiar $SU(2)$ case, by

$$C(jj'j''r \mid mm'm''). \quad \text{(B14)}$$

We have an index r appearing along with the "final" representation j'' to take care of the fact that the direct product of the UIR's j and j' may contain the UIR j'' several times. Correspondingly we may have several linearly independent sets of CG coefficients for fixed j, j', and j''. We assume that for different values of r they are orthogonal. Explicitly, we have

$$\sum_{mm'} C(jj'j''r \mid mm'm'')C(jj'j'''s \mid mm'm''') = \delta_{j''j'''}\delta_{m''m'''}\delta_{rs}. \quad \text{(B15)}$$

We then find

$$\langle j'm'n'|\, T^\lambda_\alpha \,|jmn\rangle = [d_j/d_{j'}]^{\frac{1}{2}} \sum_{\beta,r} t_\beta C(j\lambda j'r \mid m\alpha m')C(j\lambda j'r \mid n\beta n'). \quad \text{(B16)}$$

The first CG coefficient on the right-hand side of (B16) is the one demanded by the Wigner–Eckart theorem. The second one is essentially the reduced matrix element of T:

$$\langle j'n' \| \, T^\lambda \, \| jn\rangle_r = [d_j/d_{j'}]^{\frac{1}{2}} \sum_\beta t_\beta C(j\lambda j'r \mid n\beta n'). \quad \text{(B17)}$$

This completes the evaluation of the matrix elements of T in the basis made up of UIR's of C.

The next important thing we need to know is that two choices of t_β related in a certain way give rise to unitarily equivalent representations of $C \times T$. Let us define a set of operators $V(\Omega)$ in $\mathcal{H}$, so that

$$V(\Omega)\,|\Omega'\rangle = |\Omega'\Omega^{-1}\rangle. \quad \text{(B18)}$$

Clearly they are unitary, produce another representation of C, and commute with the $U(\Omega)$:

$$U(\Omega)V(\Omega') = V(\Omega')U(\Omega). \quad \text{(B19)}$$

If we start with a representation of $C \times T$ afforded by the operators $U(\Omega)$ and T^λ_α (determined by a definite choice of t_β), we can obtain a unitarily equivalent representation by performing a unitary transformation with the operator $V(\Omega_0)$, where Ω_0 is any fixed element of C. By (B19) this transformation does not affect $U(\Omega)$ at all. On the other hand, we find

$$V(\Omega_0^{-1})T^\lambda_\alpha V(\Omega_0)\,|\Omega\rangle = \sum_\beta t'_\beta (\Omega^{-1})_{\beta\alpha}\,|\Omega\rangle,$$
$$t'_\beta = \sum_\gamma (\Omega_0)^\lambda_{\gamma\beta} t_\gamma. \quad \text{(B20)}$$

Thus, the representations of $C \times T$ corresponding to t_β and t'_β are unitarily equivalent. This allows us to choose t_β in the most convenient form, without really altering the representation. We are allowed to assume the vanishing of many components of t_β, if this can be achieved by a transformation of the form that takes us from t_β to t'_β.

In the case of $SU(3) \times T_8$, the representation λ of $SU(3)$ is the octet or adjoint representation. Correspondingly, t_β can be written in form t_{IMY} [see Eq. (3.2)]. It is easily verified that by an appropriate choice of Ω_0 in (B20), we can arrange to have the only nonvanishing components of t_{IMY} to be t_{100} and t_{000}. This is the basis for the classification of UIR's of $SU(3) \times T_8$ given in Sec. 3.

Statistical Properties of Polycrystalline Dielectrics

JOHN E. MOLYNEUX
Department of Mechanical and Aerospace Sciences, College of Engineering and Applied Science
University of Rochester, Rochester, New York

(Received 6 September 1967)

In this paper we consider the determination of the two-point electric field correlation tensor in a random polycrystalline medium subjected to a constant average electric field. The medium is supposed to be statistically homogeneous and isotropic and to be composed of crystals all of the same kind. For media whose principal permittivities do not differ greatly from one another, we employ perturbation techniques to linearize the governing equations and derive explicit expressions for both the correlation function and the cross correlation of the electric field and permittivity tensor. We also determine the effective permittivity in this limit. For media with arbitrary principal permittivities, we derive bounds on the effective permittivity which depend on certain two point correlation functions and which reduce to our perturbation solution in the limit of small differences in principal permittivities.

INTRODUCTION

In a statistical description of a polycrystal subjected to an applied electric field, the permittivity tensor ϵ_{ij} of the medium is characterized by the sequence of correlation functions

$$\langle \epsilon_{i_1 j_1}(x_1)\epsilon_{i_2 j_2}(x_2)\cdots\epsilon_{i_n j_n}(x_n)\rangle, \quad n = 1, 2, \cdots, \quad i_k, j_k = 1, 2, 3, \quad (1)$$

and the field E_i by the moments

$$\langle E_{i_1}(x_1)E_{i_2}(x_2)\cdots E_{i_n}(x_n)\rangle \quad (2)$$

and

$$\langle E_{i_1}(x_1)\cdots E_{i_n}(x_n) \times \epsilon_{i_{n+1}j_{n+1}}(x_{n+1})\cdots\epsilon_{i_{n+m}j_{n+m}}(x_{n+m})\rangle, \quad (3)$$

where brackets denote ensemble averages. This paper is concerned only with a study of the two-point correlation functions $\langle E_i(x_1)E_j(x_2)\rangle$ and $\langle E_i(x_1)\epsilon_{jk}(x_2)\rangle$. Because of the statistical nonlinearity of the problem, each of these functions must be determined by the solution of an infinite set of coupled partial differential equations. We consider a medium whose crystals have principal permittivities ϵ_i which satisfy the inequality

$$\left[\left(\tfrac{2}{45}\right)\sum_{1\le i<j\le 3}(\epsilon_i - \epsilon_j)^2\right]^{\frac{1}{2}} \ll \tfrac{1}{3}\sum_{i=1}^{3}\epsilon_i.$$

In this case we can employ perturbation techniques similar to those used in Beran and Molyneux[1] to linearize the problem and to determine the statistical moments of interest. We also determine the effective permittivity of the medium and find that it depends explicitly on the form of the two-point permittivity correlation tensor. This result is contrary to the small perturbation limit of the bounds derived by Hashin and Shtrikman.[2] We are thus lead to a careful analysis of their work which reveals that they have made implicit assumptions about the statistics of the medium and, thus, that their results are not valid in general. In the course of this analysis we derive upper and lower bounds on the effective permittivity.

BASIC EQUATIONS

We refer the polycrystalline medium to a fixed set of Cartesian axes and assume that the relation between the displacement vector $D = (D_1, D_2, D_3)$ and the electric field vector $E = (E_1, E_2, E_3)$ at a point $x = (x_1, x_2, x_3)$ is given by

$$D_i(x) = \epsilon_{ij}(x)E_j(x), \quad (4)$$

where the permittivity tensor ϵ_{ij} is symmetric and we sum over repeated indices. We assume that the medium is infinite, statistically homogeneous, and isotropic, and composed of randomly oriented crystals all of which have the same principal permittivities ϵ_i, $i = 1, 2, 3$. In this case, it is easy to show that the average value of ϵ_{ij} is equal to $3^{-1}(\epsilon_1 + \epsilon_2 + \epsilon_3)\delta_{ij}$ where δ_{ij} is the Kronecker delta. Thus we can write ϵ_{ij} in the form $\epsilon_{ij} = 3^{-1}(\epsilon_1 + \epsilon_2 + \epsilon_3)\delta_{ij} + \epsilon'_{ij}$, where ϵ'_{ji} is the fluctuating part of the permittivity, or as

$$\epsilon_{ij} = \langle\epsilon\rangle[\delta_{ij} + \alpha\theta_{ij}], \quad (5)$$

where $\langle\epsilon\rangle = 3^{-1}(\epsilon_1 + \epsilon_2 + \epsilon_3)$,

$$\alpha = \langle\epsilon\rangle^{-1}[\max_{i,j}\langle(\epsilon'_{ij})^2\rangle]^{\frac{1}{2}},$$

and $\theta_{ij} = (\langle\epsilon\rangle\alpha)^{-1}\epsilon'_{ij}$ has zero average and zero trace. As is well known, the static electric field is governed by the equations

$$\frac{\partial}{\partial x_k}\epsilon_{kj}E_j = 0 \quad (6)$$

[1] M. Beran and J. Molyneux, Nuovo Cimento **30**, 1406 (1963).

[2] Z. Hashin and S. Shtrikman, Phys. Rev. **130**, 129 (1963).

and

$$\delta_{ijk}\frac{\partial}{\partial x_j}E_k = 0, \tag{7}$$

where δ_{ijk} is the alternating tensor of Levi-Civita. Equation (7) shows that E is the gradient of a scalar function, i.e.,

$$E_i = -\frac{\partial\varphi}{\partial x_i}. \tag{8}$$

Combining Eqs. (5), (6), and (8), we have

$$\frac{\partial^2\varphi}{\partial x_k\partial x_k} + \alpha\frac{\partial}{\partial x_k}\theta_{kl}\frac{\partial\varphi}{\partial x_l} = 0. \tag{9}$$

We split φ into its average and fluctuating parts by letting

$$\varphi = \langle\varphi\rangle + \varphi', \quad \frac{\partial\langle\varphi\rangle}{\partial x_i} = -\langle E_i\rangle, \quad \frac{\partial\varphi'}{\partial x_i} = -E_i', \tag{10}$$

where $\langle E_i\rangle$ is the constant average electric field and E_i' is the fluctuating field. Substituting Eq. (10) into Eq. (9), we find that

$$\frac{\partial^2\varphi'}{\partial x_k\partial x_k} - \alpha\langle E_l\rangle\frac{\partial\theta_{kl}}{\partial x_k} + \alpha\frac{\partial}{\partial x_k}\theta_{kl}\frac{\partial\varphi'}{\partial x_l} = 0. \tag{11}$$

SMALL PERTURBATION RESULTS

By the same reasoning which leads to the expression for $\langle\epsilon_{ij}\rangle$, it may be shown that

$$\langle\epsilon_{ij}\epsilon_{kl}\rangle = \langle\epsilon^2\rangle\delta_{ij}\delta_{kl} + (15)^{-1}\beta \times [-3^{-1}\delta_{ij}\delta_{kl} + 2^{-1}(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl})], \tag{12}$$

where $\beta = (\epsilon_1 - \epsilon_2)^2 + (\epsilon_1 - \epsilon_3)^2 + (\epsilon_2 - \epsilon_3)^2$. Hence, the parameter $\alpha = \langle\epsilon\rangle^{-1}[(45)^{-1}2\beta]^{\frac{1}{2}}$ appearing in Eq. (11) will be small if

$$\left[(45)^{-1}2\sum_{1\le i<j\le 3}(\epsilon_i - \epsilon_j)^2\right]^{\frac{1}{2}} \ll 3^{-1}\sum_{i=1}^{3}\epsilon_i.$$

We assume this is the case and write φ' as the formal series

$$\varphi'(x) = \sum_{n=1}^{\infty}\alpha^n\varphi^{(n)}(x). \tag{13}$$

Substituting Eq. (13) into Eq. (11) and equating powers of α, we find

$$\frac{\partial^2\varphi^{(1)}}{\partial x_m\partial x_m} = \langle E_l\rangle\frac{\partial\theta_{ml}}{\partial x_m}, \tag{14}$$

$$\frac{\partial^2\varphi^{(n)}}{\partial x_m\partial x_m} = -\frac{\partial}{\partial x_k}\theta_{kl}\frac{\partial\varphi^{(n-1)}}{\partial x_l}, \quad n = 2, 3, \cdots. \tag{15}$$

In principle, all terms in the perturbation series for φ' may be calculated from Eqs. (14) and (15). We shall consider only the lowest-order terms. Multiplying Eq. (14) by $\theta_{jk}(x')$ and averaging, we obtain

$$\nabla^2\langle\varphi^{(1)}(x)\theta_{jk}(x+r)\rangle = -\langle E_l\rangle\frac{\partial}{\partial r_m}\langle\theta_{ml}(x)\theta_{jk}(x+r)\rangle, \tag{16}$$

where $r = x' - x$, $\nabla^2 = \partial^2/\partial r_m\partial r_m$, and by statistical homogeneity both the permittivity correlation tensor and the cross correlation between the potential and the permittivity depend on r only. We find an equation for the potential correlation function by multiplying Eq. (14) by $\varphi^{(1)}(x')$ and averaging. The result is

$$\nabla^2\langle\varphi^{(1)}(x)\varphi^{(1)}(x+r)\rangle = -\langle E_l\rangle\frac{\partial}{\partial r_m}\langle\varphi^{(1)}(x)\theta_{lm}(x-r)\rangle. \tag{17}$$

Thus, to the order of approximation considered, both correlation functions of interest satisfy Poisson's equation. We require that they be analytic for all r and vanish as $|r| \to \infty$. By well-known potential theory arguments, the solutions of Eqs. (16) and (17) will be unique and, because of the conditions imposed at infinity, will be the particular solutions of these equations. Since the source term on the right-hand side of Eq. (16) is a contracted product of a third-order homogeneous tensor with the average electric field, it is natural to assume that

$$\langle\varphi^{(1)}(x)\theta_{jk}(x+r)\rangle = \Psi_{jkl}(r)\langle E_l\rangle. \tag{18}$$

Indeed, using the Green's function for Poisson's equation in infinite space, it may be shown that the cross-correlation function must be of this form. The function Ψ_{jkl} satisfies

$$\nabla^2\Psi_{jkl} = -\frac{\partial}{\partial r_m}C_{mljk}, \tag{19}$$

where $C_{ijkl}(r) = \langle\theta_{ij}(x)\theta_{kl}(x+r)\rangle$. Arguments similar to those given above lead to the form

$$\langle\varphi^{(1)}(x)\varphi^{(1)}(x+r)\rangle = \Phi_{kl}(r)\langle E_k\rangle\langle E_l\rangle \tag{20}$$

for the potential correlation function. The equation satisfied by Φ_{ij} is

$$\nabla^2\Phi_{ij} = -\frac{\partial}{\partial r_k}\Psi_{kij}(-r). \tag{21}$$

It is easy to show that the electric field correlation functions of interest are obtained from Φ_{ij} and Ψ_{ijk} by the formulas

$$\langle E_i'(x)E_j'(x+r)\rangle = -\alpha^2\frac{\partial^2}{\partial r_i\partial r_j}\Phi_{kl}(r)\langle E_k\rangle\langle E_l\rangle \tag{22}$$

and

$$\langle E_i'(x)\epsilon_{jk}'(x+r)\rangle = \langle\epsilon\rangle\alpha^2\frac{\partial}{\partial r_i}\Psi_{jkl}(r)\langle E_l\rangle. \tag{23}$$

Since the medium is assumed to be isotropic as well as homogeneous, the permittivity correlation function must be an isotropic tensor of the fourth order. From the symmetry of θ_{ij} and the fact that $C_{ijkl}(r) = C_{klij}(-r)$, it follows that this tensor must have the form[3]

$$\begin{aligned} C_{ijkl} = {} & Pr_ir_jr_kr_l + Q[r_ir_j\delta_{kl} + r_kr_l\delta_{ij}] \\ & + R[r_ir_k\delta_{jl} + r_ir_l\delta_{jk} + r_jr_k\delta_{il} + r_jr_l\delta_{ik}] \\ & + S\delta_{ij}\delta_{kl} + T[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}], \end{aligned} \tag{24}$$

where P, Q, etc., are functions of $|r|^2$ only. The vanishing of the trace of the θ_{ij} implies the conditions

$$|r|^2 P + 3Q + 4R = 0, \tag{25}$$

$$|r|^2 Q + 3S + 2T = 0, \tag{26}$$

and, using Eq. (12), we see that

$$S(0) = -3^{-1}2T(0) = 2^{-1}. \tag{27}$$

Now let us solve Eq. (19). Since the source term, $\partial C_{mljk}/\partial r_m$, is an isotropic tensor of the third order, the same must be true for Ψ_{jkl}. Taking into account the required symmetries of the indices, we see that Ψ_{jkl} must have the form

$$\Psi_{jkl}(r) = \psi_1 r_j r_k r_l + \psi_2[r_j\delta_{kl} + r_k\delta_{jl}] + \psi_3 r_l\delta_{jk}, \tag{28}$$

where ψ_ν, $\nu = 1, 2, 3$ are functions of $|r|^2$ only, and we determine the particular solution by requiring ψ_ν to satisfy Eq. (19). When Eq. (28) is substituted into Eq. (19), we find

$$\begin{aligned} & \{\psi_1'' + 8\rho^{-1}\psi_1' + C_1\}r_jr_kr_l \\ & \quad + \{\psi_2'' + 4\rho^{-1}\psi_2' + 2\psi_1 + C_2\}[r_j\delta_{kl} + r_l\delta_{jk}] \\ & \quad + \{\psi_3'' + 4\rho^{-1}\psi_2' + 2\psi_1 + C_3\}r_l\delta_{jk} = 0, \end{aligned} \tag{29}$$

where $\rho = |r|$, ψ_ν and C_ν are functions of ρ only, $' = d/d\rho$, and

$$C_1 = \rho P' + 6P + \rho^{-1}Q^{-1} + 2\rho^{-1}R', \tag{30}$$

$$C_2 = Q + 5R + \rho R' + \rho^{-1}T', \tag{31}$$

$$C_3 = 4Q + 2R + \rho Q' + \rho^{-1}S'. \tag{32}$$

Since Eq. (29) must hold for all j, k, and l, we obtain three ordinary differential equations (of the Euler type) for $\psi_\nu(\rho)$ by setting each of the three curly brackets equal to zero. The solutions of these equations which are finite at the origin and vanish at infinity are given by

$$\begin{aligned} \psi_1(\rho) = {} & \tfrac{2}{7}\Big[\rho^{-7}\int_0^\rho \xi^6(Q - R)\,d\xi \\ & - 2\int_\rho^\infty (3Q + 4R)\xi^{-1}\,d\xi\Big], \end{aligned} \tag{33}$$

$$\begin{aligned} \psi_2(\rho) = {} & -\tfrac{2}{35}\rho^{-5}\int_0^\rho \xi^6(Q - R)\,d\xi \\ & + \tfrac{1}{15}\rho^{-3}\int_0^\rho (3\xi^4 Q - 6\xi^4 R + 5\xi^3 T')\,d\xi \\ & + \tfrac{4}{35}\rho^2\int_\rho^\infty (3Q + 4R)\xi^{-1}\,d\xi \\ & - \tfrac{3}{15}\int_\rho^\infty \xi(Q - R)\,d\xi - \tfrac{1}{3}T(\rho), \end{aligned} \tag{34}$$

$$\begin{aligned} \psi_3(\rho) = {} & -\tfrac{2}{35}\rho^{-5}\int_0^\rho \xi^6(Q - R)\,d\xi \\ & - \tfrac{2}{15}\rho^{-3}\int_0^\rho (\xi^4 Q - 2\xi^4 R + \tfrac{5}{3}\xi^3 T')\,d\xi \\ & + \tfrac{4}{35}\rho^2\int_\rho^\infty (3Q + 4R)\xi^{-1}\,d\xi \\ & + \tfrac{2}{15}\int_\rho^\infty \xi(Q - R)\,d\xi + \tfrac{2}{9}T(\rho). \end{aligned} \tag{35}$$

The cross correlation between the electric field and permittivity tensor may be obtained by differentiation. Equation (23) becomes

$$\begin{aligned} & \langle E_i'(x)\epsilon_{jk}'(x + r)\rangle \\ & \quad = \langle\epsilon\rangle\alpha^2\langle E_l\rangle\{\rho^{-1}\psi_1' r_i r_j r_k r_l \\ & \qquad + \psi_1[r_k r_l\delta_{ij} + r_j r_l\delta_{ik} + r_j r_k\delta_{il}] \\ & \qquad + \rho^{-1}\psi_2'[r_i r_j\delta_{kl} + r_i r_k\delta_{jl}] + \rho^{-1}\psi_3' r_i r_l\delta_{jk} \\ & \qquad + \psi_3\delta_{il}\delta_{jk} + \psi_2[\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl}]\}. \end{aligned} \tag{36}$$

The solution of Eq. (21) is obtained in a similar way. Since $\Phi_{ij}(r)$ is an isotropic tensor of the second order, it has the form

$$\Phi_{ij}(r) = Fr_ir_j + G\delta_{ij}, \tag{37}$$

where F and G are functions of $\rho^2 = |r|^2$ only. It is found that these functions satisfy the equations

$$F'' + 6\rho^{-1}F' = [\rho\psi_1' + 5\psi_1 + \rho^{-1}\psi_2' + \rho^{-1}\psi_3'] \equiv A, \tag{38}$$

$$G'' + 2\rho^{-1}G' + 2F = [\rho\psi_2' + 4\psi_2 + \psi_3] \equiv B. \tag{39}$$

The solutions of Eqs. (38) and (39) having the required

[3] G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge University Press, New York, 1953).

behavior at $\rho = 0$ and ∞ are

$$F(\rho) = -5^{-1}\left[\rho^{-5}\int_0^{\rho}\xi^6 A\,d\xi + \int_{\rho}^{\infty}\xi A\,d\xi\right], \tag{40}$$

$$\begin{aligned} G(\rho) = &-\left[\rho^{-1}\int_0^{\rho}\xi^2 B\,d\xi + \int_{\rho}^{\infty}\xi B\,d\xi\right] \\ &- 3^{-1}\left[\rho^{-1}\int_0^{\rho}\xi^4 A\,d\xi + \int_{\rho}^{\infty}\xi^3 A\,d\xi\right] \\ &+ (15)^{-1}\left[\rho^{-3}\int_0^{\rho}\xi^6 A\,d\xi + \rho^2\int_{\rho}^{\infty}\xi A\,d\xi\right]. \end{aligned} \tag{41}$$

Using Eq. (22), it is found that the electric-field correlation function is given by

$$\begin{aligned} &\langle E_i'(x)E_j'(x+r)\rangle \\ &\quad = -\alpha^2\langle E_k\rangle\langle E_l\rangle\{\rho^{-2}(F'' - \rho^{-1}F')r_i r_j r_k r_l \\ &\qquad + \rho^{-1}F'[r_i r_l\delta_{jk} + r_i r_k\delta_{jl} + r_j r_k\delta_{il} \\ &\qquad + r_j r_l\delta_{ik} + r_k r_l\delta_{ij}] + \rho^{-2}(G'' - \rho^{-1}G')r_i r_j\delta_{kl} \\ &\qquad + F[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}] + \rho^{-1}G'\delta_{ij}\delta_{kl}\}, \end{aligned} \tag{42}$$

and in particular we find

$$\begin{aligned} \langle E_i'E_j'\rangle = &-\alpha^2\langle E_k\rangle\langle E_l\rangle \\ &\times \left\{\tfrac{2}{5}\left[\tfrac{1}{9}T(0) - \tfrac{1}{3}\int_0^{\infty}\xi(7Q + 11R)\,d\xi\right]\right. \\ &\times [\delta_{ik}\delta_{jl} - \tfrac{1}{3}\delta_{ij}\delta_{kl}] \\ &\left. + \tfrac{2}{9}\left[\tfrac{5}{3}T(0) + \int_0^{\infty}\xi(Q - R)\,d\xi\right]\delta_{ij}\delta_{kl}\right\}. \end{aligned} \tag{43}$$

SUMMARY OF PERTURBATION RESULTS

The behavior of $\langle E_i'(x)\epsilon_{jk}'(x+r)\rangle$ and $\langle E_i'(x)E_j'(x+r)\rangle$ for large values of $|r|$ is similar to the scalar permittivity case already considered in Beran and Molyneux.[1] If it is assumed that the functions $T(|r|)$ and $S(|r|)$ decay as $|r|^{-\sigma}$, $\sigma > 0$ as $|r| \to \infty$, then Q and R will fall off as $|r|^{-(\sigma+2)}$ and P will approach zero as $|r|^{-(\sigma+4)}$. It is then found that both correlation functions behave as $|r|^{-\sigma}$ if $0 < \sigma \leq 3$ and as $|r|^{-3}$ if $\sigma > 3$. This behavior is due to the fact that the governing equations are essentially equivalent (for $\sigma > 3$) to Poisson's equation with a dipole source.

The effective permittivity, ϵ_{eff}, may be calculated from the results already given. This quantity is defined by the equation $\epsilon_{\text{eff}}\langle E_i\rangle = \langle D_i\rangle$, where $\langle D_i\rangle$ is the average displacement field. Thus,

$$\begin{aligned} \epsilon_{\text{eff}}\langle E_i\rangle &= \langle\epsilon_{ij}E_j\rangle = \langle\epsilon\rangle\langle E_i\rangle + \langle\epsilon_{ij}'E_j'\rangle \\ &= [\langle\epsilon\rangle + \langle\epsilon\rangle\alpha^2(4\psi_2(0) + \psi_3(0))]\langle E_i\rangle \end{aligned} \tag{44}$$

and, using Eqs. (34), (35), and (27), we obtain

$$\epsilon_{\text{eff}} = \langle\epsilon\rangle\left[1 - \tfrac{5}{6}\alpha^2 + \tfrac{2}{3}\alpha^2\int_0^{\infty}\xi(R - Q)\,d\xi\right]. \tag{45}$$

It is seen that the effective permittivity depends explicitly on the form of the correlation tensor $C_{ijkl}(r)$. Now the bounds on ϵ_{eff} obtained in Ref. 2 coincide in the limit of small perturbations and equal the first two terms of Eq. (45). This fact leads to the conjecture that an implicit assumption about the form of the permittivity correlation tensor has been made in Ref. 2. In the next section we show that this is true and, therefore, that these bounds are only valid for special types of polycrystals. We shall also derive a set of bounds which are valid for arbitrary polycrystals and which reduce to Eq. (45) in the small perturbation limit.

REMARKS ON THE EFFECTIVE PERMITTIVITY

If $u = (u_1, u_2, u_3)$ is a spatially stationary random function, an ensemble average of the form $F[u] = \langle f(u(x))\rangle$, where f is an ordinary function of three variables, will be independent of x and can be considered as a functional of u. It is possible, therefore, to formulate statistical variational principles based on functionals of this type. We begin this section by proving an extension of a variational principle given in Ref. 2 to a statistical principle which is applicable to the class of media dealt with in this paper.

Consider a statistically homogeneous and isotropic medium composed of randomly oriented crystals each of which has the same principal permittivities. Let the medium be subjected to a constant average electric field $\langle E\rangle$. Let $\mathcal{M}$ be the class of all random vector functions $C = (C_1, C_2, C_3)$ defined on the medium such that any C belonging to $\mathcal{M}$ satisfies the conditions:

(i) C is a mean-square differentiable (see, for example, Ref. 4).

(ii) The n-point correlations of C with itself, any other element of $\mathcal{M}$, or with the permittivity tensor are functions only of the $3(n-1)$-dimensional vector which specifies the configuration of the points, and are sufficiently smooth to allow formal interchange of averaging and differentiation.

Define on $\mathcal{M}$ the functional

$$U[C] = \epsilon_0\langle E\rangle^2 - \langle\pi_{kl}^{-1}C_kC_l - 2C_k\langle E_k\rangle - C_kA_k\rangle, \tag{46}$$

where ϵ_0 is a constant, π_{ij}^{-1} satisfies $\pi_{ik}^{-1}\pi_{kj} = \delta_{ij}$, $\pi_{ij} = \epsilon_{ij} - \epsilon_0\delta_{ij}$,[5] and A is that element of $\mathcal{M}$ which

[4] M. S. Bartlett, *An Introduction to Stochastic Processes* (Cambridge University Press, New York, 1955).

[5] At each point x in the medium, the principal axes for $\pi_{ij}(x)$ coincide with those of the permittivity tensor $\epsilon_{ij}(x)$. Hence, the principal values for $\pi_{ij}(x)$ at x are $\epsilon_i - \epsilon_0$, $i = 1, 2, 3$, where ϵ_i are the (nonrandom) principal permittivities of the crystals comprising the medium. Thus π_{ij}^{-1} exists for all $\epsilon_0 \neq \epsilon_i$, $i = 1, 2, 3$.

satisfies

$$\epsilon_0 \frac{\partial A_k}{\partial x_k} = -\frac{\partial C_k}{\partial x_k}, \tag{47}$$

$$\delta_{ijk} \frac{\partial A_k}{\partial x_j} = 0, \tag{48}$$

$$\langle A \rangle = 0. \tag{49}$$

It can be shown that the vector A is uniquely determined almost everywhere by Eqs. (47), (48), and (49). Indeed, suppose that there were two fields $A^{(1)}$ and $A^{(2)}$ satisfying these equations. Then $A^{(3)} \equiv A^{(1)} - A^{(2)}$ would have vanishing divergence, curl, and average. It would then follow that

$$\langle A_i^{(3)} A_i^{(3)} \rangle = \left\langle A_i^{(3)} \frac{\partial \psi}{\partial x_i} \right\rangle = \left\langle \frac{\partial}{\partial x_i} (A_i^{(3)} \psi) \right\rangle - \left\langle 4 \frac{\partial A_i^{(3)}}{\partial x_i} \right\rangle$$
$$= \frac{\partial}{\partial x_i} \langle A_i^{(3)} \psi \rangle = 0. \tag{50}$$

The first equality in Eq. (50) follows from the fact that curl $A^{(3)} = 0$ so that $A^{(3)} = \text{grad } \psi$. The second equality is simply an identity. The third equality follows from the fact that div $A^{(3)} = 0$ and that $A^{(3)}$ is an element of $\mathcal{M}$ so that averaging and differentiation commute. The last equality follows from statistical homogeneity. Since $\langle A^{(3)} \rangle = 0$, it follows from Eq. (50) that the dispersion of $A^{(3)}$ is zero and therefore that $A^{(3)}$ vanishes almost everywhere. The functional U defined by Eq. (46) is stationary for $C_i = \pi_{ik} E_k$ where E is the actual electric field in the medium which satisfies Eqs. (6) and (7). This stationary value is an absolute minimum (maximum) when ϵ_0 is greater (less) than the largest (smallest) principal permittivity.

It may be shown that for $C_i = \pi_{ik} E_k$, Eqs. (47)–(49) imply that $A = E'$, the actual fluctuating field in the medium. Thus, if we let $C_i = \pi_{ik} E_k + c_i$, we find $A_i = E_i' + a_i$ where a is determined from c by Eqs. (47) through (49). Substitution into (46) gives

$$U[\pi_{ik} E_k + c_i] = U[\pi_{ik} E_k] - \langle c_k E_k' \rangle + \langle a_k \pi_{kl} E_l \rangle$$
$$- \langle \pi_{kl}^{-1} c_k c_l \rangle + \langle c_k a_k \rangle. \tag{51}$$

Since $E' = -\text{grad } \varphi'$,

$$\langle c_k E_k' \rangle = -\left\langle c_k \frac{\partial \varphi'}{\partial x_k} \right\rangle = -\left\langle \frac{\partial}{\partial x_k} (\varphi' c_k) - \varphi' \frac{\partial c_k}{\partial x_k} \right\rangle$$
$$= -\frac{\partial}{\partial x_k} \langle \varphi' c_k \rangle + \left\langle \varphi' \frac{\partial c_k}{\partial x_k} \right\rangle = \left\langle \varphi' \frac{\partial c_k}{\partial x_k} \right\rangle, \tag{52}$$

where we use statistical homogeneity and the fact that averaging and differentiation commute. From the fact that $\langle a \rangle = 0$, we obtain

$$\langle a_k \pi_{kl} E_l \rangle = \langle a_k \epsilon_{kl} E_l \rangle - \epsilon_0 \langle a_k E_k' \rangle. \tag{53}$$

It follows from Eq. (48) that $a_k = -\partial \psi / \partial x_k$, and we can manipulate both terms in Eq. (53) as we have done in (52). The first term on the right vanishes because of Eq. (6) and the second term becomes $\langle \varphi' (\partial c_k / \partial x_k) \rangle$ because of (47). Similar considerations give

$$\langle c_k a_k \rangle = -\epsilon_0 \langle a_k a_k \rangle \tag{54}$$

and, therefore,

$$U[\pi_{ik} E_k + c_i] = U[\pi_{ik} E_k]$$
$$+ \{ -\langle \pi_{kl}^{-1} c_k c_l \rangle - \epsilon_0 \langle a_k a_k \rangle \}, \tag{55}$$

where, according to (54), the term in the curly brackets may also be written in the form

$$-\langle (\pi_{kl}^{-1} + \epsilon_0^{-1} \delta_{kl}) c_k c_l \rangle + \epsilon_0^{-1} \langle (\epsilon_0 a_k + c_k)(\epsilon_0 a_k + c_k) \rangle.$$

The proof is concluded by finding the principal values of π_{kl}^{-1} and using the well-known properties of quadratic forms.

In order to compute U for a given trial function, it is necessary to find $\langle A_k C_k \rangle$. We let $A_i = \partial \psi / \partial x_i$ so as to satisfy Eq. (48), then multiply Eq. (47) by $C_j(x')$, and obtain (after setting $r = x' - x$ and using homogeneity)

$$\nabla^2 \langle \psi(x) C_j(x + r) \rangle = \epsilon_0^{-1} \frac{\partial}{\partial r_i} \langle C_i(x) C_j(x + r) \rangle. \tag{56}$$

Using the Green's function for Poisson's equation in infinite space to solve Eq. (56), we find

$$\langle \psi(x) C_j(x + r) \rangle$$
$$= (4\pi\epsilon_0)^{-1} \int |r - s|^{-1} \frac{\partial}{\partial s_i} \langle C_i(x) C_j(x + s) \rangle \, ds. \tag{57}$$

Finally, taking the partial derivative of Eq. (57) with respect to x_j and then setting $r = 0$, we have the result

$$\langle A_k C_k \rangle = \langle A_k C_k' \rangle = (4\pi\epsilon_0)^{-1} \int |r|^{-1} \frac{\partial^2}{\partial r_i \partial r_j} V_{ij}(r) \, dr, \tag{58}$$

where $C_k' = C_k - \langle C_k \rangle$ and $V_{ij}(r) = \langle C_i'(x) C_j'(x + r) \rangle$. Consider the function

$$C_i(x) = [3\epsilon_0 \langle E_j \rangle + \langle C_j \rangle] \gamma_{ij}(x), \tag{59}$$

where γ_{ij} is defined by the equation $\gamma_{ij}(3\epsilon_0 \pi_{jk}^{-1} + \gamma_{jk}) = \delta_{ik}$. It may be shown that γ_{ij} is a symmetric tensor having the same principle axes as the permittivity tensor and having principal values $\gamma_i = (\epsilon_i - \epsilon_0)(\epsilon_i + 2\epsilon_0)^{-1}$, $i = 1, 2, 3$. Thus, $\langle \gamma_{ij} \rangle = \langle \gamma \rangle \delta_{ij}$ with

$$\langle \gamma \rangle = 3^{-1} \sum_{i=1}^{3} \gamma_i.$$

Letting γ'_{ij} be the fluctuating part of γ_{ij}, we find that

$$\langle C_i \rangle = \frac{3\epsilon_0 \langle \gamma \rangle}{1 - \langle \gamma \rangle} \langle E_i \rangle, \tag{60}$$

$$C'_i = \frac{3\epsilon_0}{1 - \langle \gamma \rangle} \gamma'_{ij} \langle E_j \rangle. \tag{61}$$

We assume that C is an element of the class $\mathcal{M}$ and can be used to determine bounds on ϵ_{eff}. We introduce the parameter $\alpha_0 \equiv [\max_{i,j} \langle (\gamma'_{ij})^2 \rangle]^{\frac{1}{2}} \langle \gamma \rangle^{-1}$ and define ω_{ij} as $(\alpha_0 \langle \gamma \rangle)^{-1} \gamma'_{ij}$. Then from Eq. (61) we find that

$$V_{ij}(r) = \left[\frac{3\epsilon_0 \langle \gamma \rangle}{1 - \langle \gamma \rangle}\right]^2 \alpha_0^2 C^{(0)}_{ikjl}(r) \langle E_k \rangle \langle E_l \rangle, \tag{62}$$

where, by the same reasoning that leads to Eq. (24),

$$\begin{aligned} C^{(0)}_{ijkl}(r) &\equiv \langle \omega_{ij}(x) \omega_{kl}(x + r) \rangle \\ &= P^{(0)} r_i r_j r_k r_l + Q^{(0)} [r_i r_j \delta_{kl} + r_k r_l \delta_{ij}] \\ &\quad + R^{(0)} [r_i r_k \delta_{jl} + r_i r_l \delta_{jk} + r_j r_k \delta_{il} + r_j r_l \delta_{ik}] \\ &\quad + S^{(0)} \delta_{ij} \delta_{kl} + T^{(0)} [\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}], \end{aligned} \tag{63}$$

with $P^{(0)}$, $Q^{(0)}$, etc., functions of $|r|$ only which satisfy relations (25) through (27). It may be shown that

$$\alpha_0^2 \langle \gamma \rangle^2 C^{(0)}_{ijkl}(r) = \sum_{p,q=1}^{3} \langle a_{pi}(x) a_{pj}(x) a_{qk}(x + r) a_{ql}(x + r) \rangle \gamma_p \gamma_q - \langle \gamma_{ij} \gamma_{kl} \rangle,$$

where $a_{ij}(x)$ is the direction cosine of the ith principal axis of the crystal at x relative to the x_j direction. Thus, all the functions in Eq. (63) can be calculated from the known statistical properties of the medium. Substituting for $V_{ij}(r)$ from the above expression and performing the integrations in Eq. (58), we find that

$$\langle A_k C_k \rangle = -(3\epsilon_0)^{-1} \langle C'_i C'_i \rangle + \epsilon_0 \langle E \rangle^2 J, \tag{64}$$

where

$$J = \frac{2}{3}\left[\frac{3\langle \gamma \rangle}{1 - \langle \gamma \rangle}\right] \alpha_0^2 \int_0^\infty \xi (R^{(0)} - Q^{(0)})\, d\xi. \tag{65}$$

Hence, we have

$$U[C] = \epsilon_0 \langle E \rangle^2 \left\{ \frac{1 + 2\langle \gamma \rangle}{1 - \langle \gamma \rangle} + J \right\}. \tag{66}$$

By the definition of the effective permittivity, it follows that $U[\pi_{ij} E_j] = \langle \epsilon_{kj} E_j \rangle \langle E_k \rangle = \epsilon_{\text{eff}} \langle E \rangle^2$. Thus, if it is assumed that $\epsilon_1 < \epsilon_2 < \epsilon_3$, Eq. (66) gives the bounds[6]

$$\epsilon_1 \left\{ \frac{1 + 2\langle \gamma \rangle}{1 - \langle \gamma \rangle} + J \right\}_{\epsilon_0 = \epsilon_1} < \epsilon_{\text{eff}} < \epsilon_3 \left\{ \frac{1 + 2\langle \gamma \rangle}{1 - \langle \gamma \rangle} + J \right\}_{\epsilon_0 = \epsilon_3}$$

where the values of ϵ_0 have been chosen so as to make the upper and lower bounds coincide in the small perturbation limit. It may be shown that, in this limit, Eq. (67) gives the same expression for ϵ_{eff} as Eq. (45).

The integral J which appears in Eq. (67) is absent from the bounds given in Ref. 2 which are also derived using the trial function of Eq. (59). The reason for this discrepancy is that Hashin and Shtrikman have assumed $\langle A_k C_k \rangle = (3\epsilon_0)^{-1} \langle C'_k C'_k \rangle$, and, as is seen from Eq. (64), this formula is true only for media in which the integral J vanishes. Thus, the results they have given are valid only for such media. Furthermore, the remark made in Ref. 2 that the actual value of U is approached most closely by the trial function of Eq. (59) is not valid unless $J = 0$. The problem of obtaining better bounds on ϵ_{eff} is currently being considered by the author.

We note in conclusion that the results derived in this paper apply as well to the analysis of random media with tensor electrical conductivity and other mathematically analogous problems.

[6] The tensor $\pi_{ij}(x)$ appearing in the variational principle (45) is not invertible for $\epsilon_0 = \epsilon_1$ or ϵ_3. However, $U[C]$ as given by Eq. (66) is continuous for these values of ϵ_0. Hence, the bounds given by Eq. (67) can be obtained by passage to the limit in Eq. (66).

Analytic Representations of Two-Point Functions with Noncanonical Light-Cone Singularities. I

JERZY LUKIERSKI
Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland

(Received 1 August 1968)

We introduce the classes of analytic functions describing by means of its boundary values the sets of commutator functions with singularities $\delta^{(k)}(x^2)$ and $(x^2)_+^{-k}$ $(k = 1, 2, \cdots)$ on the light cone. Their relation with the Källen–Lehmann representation having nonintegrable spectral functions is given. The generalized wave-renormalization constants, measuring noncanonical singularities, are introduced. The formulas exhibit in a clear way the light-cone behavior and provide a proper scheme for studying the equal-time limits.

1. INTRODUCTION

The interaction modifies the canonical light-cone singularity of free two-point functions. This modification is described formally by the wave-renormalization constant Z_3^{-1}. Such description, however, becomes mathematically meaningless if Z_3^{-1} is infinite. Treating the results of perturbation theory as a guide, one should expect that Z_3^{-1} in all four-dimensional nontrivial examples of interacting local fields is infinite, and it is necessary to look for another method of describing the short-distance behavior of two-point functions.

The renormalization factor Z_3^{-1} describes the modification of the free field functions. New approach should therefore introduce the methods of description of short-distance singularities without any reference to the free field solutions. In this paper we introduce a class of analytic functions, constructed in accordance with the analytic properties of VEV in the Wightman scheme.[1] These functions, similarly as in Källen–Lehmann spectral formulas, can be used for construction of spectral representation with integrable spectral functions, describing two-point functions with noncanonical singularities.

Using the boundary prescriptions[1–3] one can relate our analytic functions with the two-point functions characterized by some standard nonintegrable Källen–Lehmann spectral functions. Some particular examples of such two-point functions have been studied in the framework of distribution theory by Steinmann,[4] Guttinger,[5] Pfaffelhuber,[6] Rieckers and Guttinger[7], and Vladimirov.[8] In this paper we introduce larger classes of such functions, sufficient for introduction of the spectral decompositions of any commutator function having the singularities $\delta^{(k)}(x^2)$ and $(x^2)_+^{-k}$ $(k = 1, 2, \cdots)$, and we use the tools of the theory of analytic functions, that make all operations unique and well defined. (It has to be mentioned, however, that not all problems of Lorentz-invariant distributions can be solved by such approach. In a full treatment of Lorentz-invariant distributions the analytic methods have to be supplemented with the discussion of so-called Gårding mapping[9] of invariant four-dimensional distributions and the discussion of inverse Gårding mapping. For an extensive treatment of these mappings see Ref. 7.)

In recent years one of the most fashionable subjects in axiomatic field theory is the study of inequivalent representations of canonical commutation relations.[10,11] It has to be stressed, however, that such a method is justified if the wave-renormalization constant Z_3^{-1} is finite. In relativistic quantum field theory, therefore, one cannot escape from the conclusion that the interaction modifies the algebraic structure of the equal-time limits. Because the correct way of calculating the equal-time limit leads to one-to-one correspondence between the equal-time singularities and the singularities of the four-dimensional commutator function on the light cone, we conclude that the study of light-cone singularities for at least the lowest Green's functions represents a program of classification of interactions in the Wightman scheme. (The best example is provided by the free field case where the delta singularity on the light cone implies the canonical commutation relations. The argument

[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[2] R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics and All That* (W. A. Benjamin, Inc., New York, 1965).
[3] G. Källen, in *Proceedings of the MIT Conference on the Mathematical Theory of Elementary Particles, September 1965* (MIT Press, Cambridge, 1966).
[4] O. Steinmann, J. Math. Phys. **4**, 583 (1963).
[5] W. Guttinger, Fortschr. Physik **14**, 483 (1966).
[6] E. Pfaffelhuber, University of Munich thesis, 1966.
[7] A. Rieckers and W. Guttinger, Commun. Math. Phys. 7, 190 (1968).
[8] V. S. Vladimirov, *Methods of the Theory of Analytic Functions of Several Complex Variables* (Nauka, Moscow, 1964) (in Russian).
[9] L. Gårding, Nuovo Cimento Suppl. **14**, 45 (1959).
[10] M. Guenin, *Lectures at the Boulder 1966 Summer Institute* (Gordon and Breach Science Publishers, New York, 1968).
[11] A. S. Wightman, review talk at Rochester Conference on the Theory of Fields and Particles, August 1967.

can be extended also to large class of other light-cone singularities. The ambiguity consisting of the presence of distributions with support $x = 0$ is nonphysical.) In this paper we discuss the two-point functions; the case of three-point function will be treated in another publication.

In this paper we discuss only the example of scalar neutral Wightman field, but the results can be easily extended to the nonscalar fields. Particularly interesting is the case of vector field, because of recent discussions about the validity of current-algebra assumptions in the general framework of relativistic quantum field theory. In our considerations we assume that only the ultraviolet divergences are present, i.e., the Källen–Lehmann spectral function is locally integrable. It is interesting to mention that the infrared divergences can be excluded formally by means of the Wightman postulate of positiveness of metric in the space of states.[12]

The analytic representations, describing commutator functions with the singularities $\delta^{(k)}(x^2)$ $(k = 0, 1, 2, \cdots)$ and $(x^2)_+^{-l}$ $(l = 1, 2, \cdots)$ and depending on the continuous parameter m analogous to the mass variable in the free field case, are introduced in Sec. 2. In Sec. 3 we present the connection between some standard nonintegrable spectral functions and our analytic representations. In Sec. 4 we introduce the numerical parameters—wave-renormalization constants, unrenormalized mass and generalized wave-renormalization constants—as characteristics of the light-cone singularity. In the last section some general remarks about the noncanonical singularities are given.

2. THE CLASSES OF ANALYTIC REPRESENTATIONS OF NONCANONICAL TWO-POINT FUNCTIONS

In the Wightman formalism, all two-point functions (two-point VEV, commutator functions, causal propagator, etc.) are the distribution-valued boundary values of an analytic function $G(z^2)$ $(z^2 = z_\mu z^\mu$, $z_\mu = x_\mu + iy_\mu)$,[1-3] holomorphic in a whole complex z^2 plane $(z^2 = s + iu)$ except the points along the positive real axis $(u = 0;\ s \geq 0)$. Such an analytic function $G(z^2)$ is characterized by its discontinuity across the cut

$$\xi(s) = (2\pi i)^{-1}\{G(s + i0) - G(s - i0)\}. \quad (2.1)$$

[The distributions as boundary values of analytic functions are extensively discussed in Refs. 5, 8, and 13–15. The analytic function $G(z^2)$ is called an analytic representation, generating the distribution $\xi(s)$. For a large class of distributions $\xi(s)$ one can write for $G(z^2)$ a Cauchy representation, leading to dispersion relations in coordinate space.[16,17]]

Using the boundary prescriptions for the two-point VEV[1] one gets the following formula for the commutator function:

$$\begin{aligned} G(x) &= i\langle 0|\ [\varphi(x/2),\ \varphi(-x/2)]\ |0\rangle \\ &= 2\pi\epsilon(x_0)\xi(x^2). \end{aligned} \quad (2.2)$$

Because the commutator function is a tempered distribution, we see that

$$\xi(x^2) \subset S'(R_+) \quad (2.3)$$

and $\xi(x^2)$ should be real. This last condition implies that one can write

$$G(z^2) = \tilde{G}(z^2) + F(z^2), \quad (2.4)$$

where $\tilde{G}(z^2)$ has a real discontinuity (2.1) and satisfies the condition

$$\tilde{G}^*(z^2) = \tilde{G}(z^{2*}). \quad (2.5)$$

The function $F(z^2)$ is an entire function in z^2 complex plane. One can say that the function $\tilde{G}(z^2)$ determines the algebraic structure of the theory and $F(z^2)$ depends only on the representation. [This determination is, of course, a partial one. Only for generalized free field, $\tilde{G}(z^2)$ determines the algebraic structure completely.] Finally, using the temperedness assumption for the two-point VEV we see that the function $G(s \pm iO)$ should be bounded for large positive as well as negative values of s by a polynomial.

The example of an analytic function, satisfying all requirements mentioned above is provided by the analytic continuation of free field VEV, defined as follows:

$$G_0(z^2; m^2) = \frac{m^2}{8\pi i}\,\frac{H_1^{(1)}(mz)}{mz}\,. \quad (2.6)$$

The discontinuity (2.1) along the positive real axis is described by the function [the analytic representation of $\delta(s)$ is $-(z^2)^{-1}$ and $\theta(s)$ is generated by $-\ln(-z^2)$ (see Refs. 5 and 13)]:

$$\xi(s; m^2) = \frac{1}{4\pi^2}\left\{\delta(s) - \theta(s)\,\frac{m^2}{2}\,\frac{J_1(m\lambda)}{m\lambda}\right\}, \quad (2.7)$$

where $s = \lambda^2$ and the light-cone behavior is determined by the singularity of $G_0(z^2; m^2)$ near the

[12] D. A. Dubin and J. Tarski, J. Math. Phys. 7, 574 (1966).
[13] H. J. Bremermann and L. Durand, J. Math. Phys. 2, 240 (1961).
[14] H. J. Bremermann, *Distributions, Complex Variables and Fourier Transforms* (Addison-Wesley Publ. Co., Reading, Mass., 1965).
[15] F. Constantinescu, Commun. Math. Phys. 7, 225 (1968).
[16] A. E. Szabat, Tr. Fiz. Inst. Akad. Nauk SSSR 29, 151 (1965).
[17] J. Lukierski, Lecture at the 5th Winter School of Theoretical Physics, February 1968, University of Wrocław preprint No. 169 to be published in Proceedings of 5th Winter School, Wrocław University Press.

point $z = 0$:

$$G_0(z^2; m^2) = -\frac{1}{4\pi}\frac{1}{z^2} + \text{less singular terms.} \quad (2.8)$$

We see that the light-cone singularity is mass-independent, and because it leads to the canonical commutation relations it will be called a canonical singularity. The interacting fields have, however, different singularities on the light cone. In the following we shall assume that the only singular point for the distribution (2.1) is the light cone $s = 0$ and that the distribution $\xi(s)$ for large s is described by a real function, satisfying the condition

$$|\xi(s)| < A/s^{\frac{3}{4}}. \quad (2.9)$$

The first requirement means that the main part of the perturbation propagates without delay along the light cone and the relation (2.9) determines the asymptotic behavior of the action with very large delay time. [The value $\frac{3}{4}$ of the inverse power in (2.9) can be justified by the requirement of positive-definiteness of the metric in the space of physical states.]

We shall consider in this paper the following two classes of light-cone singularities:

$$\xi_k(s) = \delta^{(k)}(s), \quad k = 0, 1, 2, \cdots, \quad (2.10a)$$

and

$$\xi_l(s) = s_+^{-l}, \quad l = 1, 2, 3, \cdots. \quad (2.10b)$$

The results can be generalized also to more general cases, particularly to the case when l is a continuous index.

A. $\delta^{(k)}(s)$ $(k = 0, 1, 2, \cdots)$

The simplest generalization of the formula (2.6) is to introduce other Hankel functions of the first kind with the argument mz. It can be easily shown that only such combination of Bessel functions and Neumann functions satisfy the temperedness assumption for spacelike distances (z^2 negative, Im z real and positive). We introduce the following family of analytic functions which can be used for the description of the analytically continued VEV:

$$G_n(z^2; m^2) = \left(\frac{d}{dz^2}\right)^n G_0(z^2; m^2). \quad (2.11)$$

Using the formula

$$\left(\frac{d}{z\,dz}\right)^n \left\{\frac{H_1^{(1)}(z)}{z}\right\} = (-1)^n \frac{H_{n+1}^{(1)}(z)}{z^{n+1}}, \quad (2.12)$$

one gets the following result:

$$G_n(z^2; m^2) = \frac{(-1)^n}{4\pi i}\left(\frac{m^2}{2}\right)^{n+1} \frac{H_{n+1}^{(1)}(mz)}{(mz)^{n+1}}. \quad (2.13)$$

Using the expression for the Hankel function $H_n^{(1)}(mz)$, one gets

$$G_{n-1}(z^2; m^2) = \frac{(-1)^n}{4\pi^2}\left(\frac{m^2}{2}\right)^n \left\{[2\gamma - \ln 4 + \ln(-m^2z^2)] \times \frac{J_n(mz)}{(mz)^n} - \frac{1}{\pi}\sum_{k=0}^{n-1}\frac{a_{n;k}}{(mz)^{2(n-k)}} - \frac{1}{\pi}\sum_{k=0}^{\infty} b_{n;k}(mz)^{2k}\right\} \quad (2.14)$$

where $\gamma = 0.577$ (Euler constant) and where

$$a_{n;k} = 2^{n-2k}\frac{(n-k-1)!}{k!}, \qquad b_{n;k} = \frac{(-1)^k}{2^{n+2k}}\frac{C_{k+n} + C_n}{k!\,(k+n)!}, \qquad C_s = 1 + \tfrac{1}{2} + \cdots + \frac{1}{s}, \quad C_0 = 0, \quad (2.15)$$

and the relation $\ln(-m^2z^2) = 2\ln mz + i\pi$ has been used. We obtain the following discontinuity:

$$\xi_{n-1}(s; m^2) = \frac{1}{2\pi i}\{G_{n-1}(s + i0; m^2) - G_{n-1}(s - i0; m^2)\} = \frac{1}{4\pi^2}\sum_{k=0}^{n-1}\frac{(-1)^k}{4^k}\frac{m^{2k}}{k!}\delta^{(n-k-1)}(s) + \left(-\frac{m^2}{2}\right)^n \frac{\theta(s)}{4\pi^2}\frac{J_n(m\lambda)}{(m\lambda)^n}. \quad (2.16)$$

The leading light-cone singularity is mass-independent and equal to $(1/4\pi^2)\delta^{(n)}(s)$. Putting $m = 0$, one gets

$$G_n(z^2; 0) = \frac{(-1)^{n+1}}{4\pi^2}\frac{n!}{(z^2)^{n+1}}, \quad (2.17)$$

leading to the result obtained in Refs. 5–8:

$$\xi_n(s; 0) = \frac{1}{4\pi^2}\delta^{(n)}(s). \quad (2.18)$$

The formula (2.11) can be easily generalized. We introduce

$$G_{n;r}(z^2; m^2) = (z^2)^r G_n(z^2; m^2), \quad (2.19)$$

where $-n \leq r \leq n$. The leading singularity on the light cone for $\xi_{n;r}(s; m^2)$ is independent on m^2 and proportional to $\delta^{(n-r)}(s)$. One gets particularly interesting class by the following choice:

$$G_{2k-1;k-1}(z^2; m^2) = -\frac{1}{4\pi i}\left(\frac{m^2}{4}\right)^k \frac{H_{2k}^{(1)}(mz)}{z^2}, \quad k = 1, 2, 3, \cdots, \quad (2.20)$$

with leading singularity on the light cone described by the limit $m^2 \to 0$:

$$G_{2k-1;k-1}(z^2; 0) = \frac{1}{4\pi^2} \frac{(2k-3)!}{(z^2)^{k+1}}. \quad (2.21)$$

It will be seen in Sec. 3 that the functions (2.20) correspond to the case of polynomial behavior of the Källen–Lehmann spectral function.

B. s_+^{-l} $(l = 1, 2, \cdots)$

In order to describe the singularities (2.10b) for $l = 1$ we introduce the following analytic function:

$$\tilde{G}_1(z^2; m^2) = \frac{1}{4\pi i} \frac{H_0^{(1)}(mz)}{z^2}. \quad (2.22)$$

[It can be mentioned that one obtains (2.22) from (2.19) by putting $n = 0$, $r = -2$.]

Using the formula

$$H_0^{(1)}(mz) = \frac{i}{\pi}[\gamma + \tfrac{1}{2} \ln(-m^2z^2) - \ln 2]J_0(mz) + \frac{2i}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} J_{2n}(mz), \quad (2.23)$$

one gets

$$\xi_1(s; m^2) = (2\pi i)^{-1}\{\tilde{G}_1(s + i0; m^2) - \tilde{G}_1(s - i0; m^2)\} = (4\pi^2)^{-1}S_+^{-1}J_0(m\lambda) - (4\pi^2)^{-1}(\ln m^2 - 2\gamma + \ln 4)\delta(s). \quad (2.24)$$

We see that the function (2.19) does not allow us to perform the limit $m^2 \to 0$. The logarithmic term, which becomes infinite with vanishing m^2, occurs in the solutions of derivative–coupling models in two dimensions[18,19] and four dimensions.[20,21] In order to get only the singularity s_+^{-1} one should subtract the following "counterterm" from the function (2.19):

$$[\ln m^2 + 2(\gamma - \ln 2)]G_0(z^2; \mu^2), \quad (2.25)$$

where the mass μ^2 is, in the general case, not related to the mass m^2 and, particularly, can be chosen equal to zero.

In order to get the singularities (2.10b) with $l = 1, 2, 3, \cdots$, one should introduce the following analytic functions:

$$\tilde{G}_l(z^2; m^2) = \left(\frac{d}{dz^2}\right)^{l-1} \tilde{G}_1(z^2; m^2). \quad (2.26)$$

Using the formula

$$\left(\frac{d}{z\,dz}\right)^n H_0^{(1)}(z) = (-1)^n \frac{H_n^{(1)}(z)}{z^n}, \quad (2.27)$$

one gets

$$\tilde{G}_l(z^2; m^2) = \frac{1}{4\pi i}(-1)^{l-1}\frac{(l-1)!}{(z^2)^l} \times \sum_{n=0}^{l-1} \frac{1}{n!}\left(\frac{mz}{2}\right)^n H_n^{(1)}(mz). \quad (2.28)$$

The formula (2.28) implies the following leading light-cone singularities:

$$\xi_l(s; m^2) = (4\pi^2)^{-1}(-1)^{l-1}(l-1)!\, S_+^{-l} - (4\pi^2)^{-1}\{\ln m^2 + 2(\gamma - \ln 2) + C_{l-1}\}\delta^{(l-1)}(s) + O(s_+^{-l+1}). \quad (2.29)$$

Introducing suitable counterterms [compare with (2.25)] one can cancel out all terms with delta functions.

Another way of introducing the light-cone singularities of type (2.10b) is to multiply the functions $G_n(z^2; m^2)$ by the function $\ln(-z^2)$. Let us consider, for example,

$$G_0^{\ln}(z^2; m^2) = \ln(-z^2)G_0(z^2; m^2). \quad (2.30)$$

The discontinuity of (2.30) is given by the formula

$$\xi_0^{\ln}(s; m^2) = \frac{1}{4\pi^2}\left\{s_+^{-1} - \ln s_+ \frac{m^2}{2}\frac{J_1(m\lambda)}{m\lambda}\right\}. \quad (2.31)$$

The differentiation of (2.30) with respect to z^2 leads to analytic functions with the discontinuity $\xi(s)$ having singularities only of the type (2.10b).

3. THE ANALYTIC REPRESENTATION AND NONINTEGRABLE KÄLLEN–LEHMANN SPECTRAL FUNCTIONS

The two-point functions are usually described by means of the spectral function $\rho(\kappa^2)$ which represents a Lorentz-invariant four-dimensional Fourier transform of the distribution $\xi(x^2)$. In this section we shall find the analytic representations corresponding to some standard choices of nonintegrable spectral functions.

Let us write the Källen–Lehmann spectral representation in complex coordinate space

$$G(z^2) = \int_0^{\infty} \rho(\kappa^2)G_0(z^2; \kappa^2)\, d\kappa^2. \quad (3.1)$$

We consider (3.1) for z^2 off the real axis, i.e., for $z = \lambda + i\eta$ where z lies in upper half-plane ($\eta > 0$). The function $G_0(z^2; \kappa^2)$ behaves for large κ like $e^{-\kappa\eta}$ and the integral exists for all locally integrable $\rho(\kappa^2) \subset S'(R_+)$.

[18] B. Schroer, Fortschr. Physik **11**, 1 (1963).
[19] A. S. Wightman, Lecture notes at Cargese Summer School, 1964 (published in Russian in 1968).
[20] J. Lukierski, Bull. Acad. Polon. Sci. **16**, 219 (1968).
[21] J. Lukierski, Lecture at Varna Seminar on Elementary Particles, May 1968, University of Wrocław preprint No. 167.

Now we introduce the following operator, acting on complex variables z_μ:

$$\frac{\partial}{\partial z_\mu}\frac{\partial}{\partial z^\mu} f(z^2) \overset{\text{DEF}}{=} \Box_z f(z^2) = -4\left(\frac{d}{dz^2}\right)^2 [z^2 f(z^2)]. \quad (3.2)$$

Using the identity

$$z\left(\frac{d}{z\,dz}\right)^2 zH_1^{(1)}(mz) + m^2 H_1^{(1)}(mz) = 0, \quad (3.3)$$

one gets

$$(\Box_2 - \kappa^2)G_0(z^2; \kappa^2) = 0 \quad (3.4)$$

and

$$\Box_z^n G(z^2) = \int_0^\infty (\kappa^2)^l \rho(\kappa^2) G_0(z^2; \kappa^2)\, d\kappa^2. \quad (3.5)$$

Using the formula (3.5) one can always relate the spectral function $\rho(\kappa^2)$ which is locally integrable and belongs to $S'(R_+)$ with a function having a Hankel transform on the real axis. We introduce, in general, for $\eta > 0$:

$$\tilde{g}(z) = \int_0^\infty g(\kappa) H_1^{(1)}(\kappa z)(\kappa z)^{\frac{1}{2}}\, d\kappa. \quad (3.6)$$

If $g(\kappa) \subset L_1(0, \infty)$, the transform (3.6) necessarily exists also if $\eta = 0$.[22] Using (2.6) and (3.6) one can write (3.1) as follows:

$$G(z^2) = (4\pi i z^{\frac{3}{2}})^{-1}\tilde{g}(z), \quad (3.7)$$

where

$$g(\kappa) = \rho(\kappa^2)\kappa^{\frac{3}{2}}. \quad (3.8)$$

Our method of determining the analytic representation for nonintegrable spectral functions is based on the following two steps:

(a) We take from the tables of integral transforms (see, for example, Ref. 23) the Hankel transform (3.6) for

$$g_{\text{reg}}(\kappa) = \rho_{\text{reg}}(\kappa^2)\kappa^{\frac{3}{2}} = \rho(\kappa^2)\kappa^{\frac{3}{2}}/(\kappa^2)^n, \quad (3.9)$$

where n is chosen sufficiently large.

(b) We use the formula (3.5).

We see, therefore, that every spectral function with $g(\kappa)$ having the real Hankel transform (3.6) generates the family of analytic representations for all two-point functions with the spectral functions of the form $g(\kappa)(\kappa^2)^n$.

We shall consider below two such families.

[22] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, 1937).

[23] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill Book Co., New York, 1954), Vol. 2.

A. $\rho(\kappa^2) = \theta(\kappa^2 - m^2)(\kappa^2)^k$ ($k = 0, 1, 2, \cdots$ **and** $k = -1$)

Let us consider first $k = -1$. From (3.8) it follows that one should find the Hankel transform (3.6) with $g(\kappa) = \theta(\kappa - m)\kappa^{-\frac{1}{2}}$. One gets[23]

$$\Delta_{-1}(z^2; m^2) = \frac{1}{4\pi i z^{\frac{3}{2}}}\int_m^\infty \kappa^{-\frac{1}{2}} H_1^{(1)}(\kappa z)(\kappa z)^{\frac{1}{2}}\, d\kappa$$
$$= \frac{1}{4\pi i z^2} H_0^{(1)}(mz). \quad (3.10)$$

We see, therefore, that the analytic function (2.22) describes the two-point functions characterized by the logarithmically divergent wave-renormalization constant. Because

$$\Delta_{-1}(z^2; m^2) = \int_{m^2}^\infty \frac{d\kappa^2}{\kappa^2} G_0(z^2; \kappa^2), \quad (3.11)$$

it is clear the origin of the term $\ln m^2$ in (2.24), describing the infrared divergence of the wave-renormalization constant.

One gets the results for $k = 0, 1, \cdots$ using the relation (3.5). We have

$$\Delta_k(z^2; m^2) = \int_{m^2}^\infty d\kappa^2 (\kappa^2)^k G_0(z^2; \kappa^2) = \Box_z^{k+1}\Delta_{-1}(z^2; m^2). \quad (3.12)$$

Using the formula

$$\Box_z \equiv -4z^2\left(\frac{d}{dz^2}\right)^2 - 8\frac{d}{dz^2},$$

one gets for the most interesting cases $k = 0$ and $k = 1$:

$$\Delta_0(z^2; m^2) = -\frac{m^2}{4\pi i z^2} H_2^{(1)}(mz) \quad (3.13a)$$

and

$$\Delta_1(z^2; m^2) = -\frac{m^2}{4\pi i z^2} H_4^{(1)}(mz) + \frac{m^3}{i\pi z^3} H_3^{(1)}(mz). \quad (3.13b)$$

We see that for $k = -1$ and $k = 0$ we obtained up to some constant factor the analytic functions (2.20). For $k \geq 1$ one gets also some additional terms proportional to the functions $G_{k+1;0}(mz)$, $G_{k+2;1}(mz)$, $\cdots$, $G_{2k;k-1}(mz)$. All these functions contribute to the leading light-cone singularity, which can be obtained if $m \to 0$. For example, for $k = 1$ we get, using the formula

$$H_n^{(1)}(z) \underset{z=0}{\sim} -(i/\pi)(\tfrac{1}{2}z)^{-n}(n-1)!,$$

that

$$\Delta_1(z^2; 0) = \frac{8}{\pi^2}\frac{1}{z^4} \quad (3.14)$$

and both terms in (3.13b) contribute to result (3.14).

B. $\rho(\kappa^2) = \theta(\kappa^2 - m^2)(\kappa^2)^k \ln \kappa^2$ $(k = -1, 0, +1, \cdots)$

First we consider the case $k = -1$. One gets after somewhat tricky calculations the following estimate at $z = 0$:

$$\Delta_{-1}^{\log}(z^2; m^2) = \frac{1}{4\pi i z^{\frac{3}{2}}} \int_m^\infty \frac{\ln \kappa^2}{\kappa^{\frac{1}{2}}} H_1^{(1)}(\kappa z)(\kappa z)^{\frac{1}{2}}\, d\kappa = \frac{3}{8\pi^2} \frac{\ln^2 z^2}{z^2} + O\left(\frac{\ln z^2}{z^2}\right) \tag{3.15}$$

and the discontinuity (2.1) has a following leading singularity:

$$\xi_{-1}^{\log}(s; m^2) = (3/8\pi)\, i\, s \cdot s_+^{-1} + O(s_+^{-1}). \tag{3.16}$$

(A complete formula for $\Delta_{-1}^{\log}$ will be given in the second part of this paper.) Applying the operator (3.2), one gets

$$\Delta_k^{\log}(z^2; m^2) = \int_{m^2}^\infty (\kappa^2)^k \ln \kappa^2 G_0(z^2; \kappa^2)\, d\kappa^2 = \frac{(-1)^k \cdot 3 \cdot 4^{k-1}(k!)^2(k+1)}{\pi^2 (z^2)^{k+2}} \ln z^2 + O\left(\frac{1}{z^{2(k+2)}}\right) \tag{3.17}$$

and one can check easily that the leading singularity is of type (2.10b).

Finally, it should be stressed that, even when

$$\int \rho(\kappa^2)\kappa^{\frac{1}{2}}\, d\kappa^2 < \infty \tag{3.18}$$

which assures that the Hankel transform (3.6) exists for $\eta = 0$, the function $G_0(z^2; \kappa^2)$ cannot be expanded under the integral into the powers series, because the coefficients of consecutive powers will not exist. If we, nevertheless, use such a method, it is easy to see that the differentiation (3.5) will not produce any noncanonical singularities. Indeed, the terms occurring in power expansion of $G_0(z^2; m^2)$ are z^{-2}, $\ln z^2 (z^2)^k$, $(k = 0, 1, \cdots)$, and $(z^2)^k$. Because

$$\begin{aligned} \Box_z z^{-2} &= 0, \\ \Box_z (z^2)^k &= -4k(k+1)(z^2)^{k-1}, \\ \Box_z \ln z^2 (z^2)^k &= -4\{k(z^2)^{k-1} + 2(k+1)(z^2)^k \\ &\quad + k(k+1) \ln z^2 (z^2)^{k-1}\}, \end{aligned} \tag{3.19}$$

the differentiation (3.5) will again reproduce only the terms occurring in $G_0(z^2; m^2)$. We see, therefore, that noncanonical terms can be easily lost if we use unjustified mathematics.

4. REMARKS ABOUT THE RENORMALIZATION PROCEDURE

It has been mentioned in the introduction that the wave-renormalization constant measure the modification of free field singularities. One introduces the cutoff-dependent wave-renormalization constant $Z_3^{-1}(\Lambda^2)$, where

$$Z_3^{-1}(\Lambda^2) = \int_0^{\Lambda^2} \rho(\kappa^2)\, d\kappa^2, \tag{4.1}$$

and one studies the limit

$$Z_3^{-1} = \lim_{\Lambda^2 \to \infty} Z_3^{-1}(\Lambda^2). \tag{4.2}$$

Different types of infinities correspond to different types of noncanonical singularities.

The wave-renormalization constant can be, however, defined by means of the analytic representation $G(z^2)$ as follows:

$$Z_3^{-1} = \lim_{z^2 \to 0} \frac{G(z^2)}{G_0(z^2)}, \tag{4.3}$$

where the function $G_0(z^2)$ can be characterized by any mass. Similarly, one can introduce the unrenormalized mass parameter m_0^2 by means of the following limit:

$$m_0^2 = \lim_{z^2 \to 0} \frac{\Box G(z^2)}{G_0(z^2)}. \tag{4.4}$$

It is easy to see that for the free field $G(z^2) = G_0(z^2; m^2)$ one gets $m_0^2 = m^2$ and in the general case

$$m_0^2 = \lim_{\Lambda^2 \to \infty} \int_0^{\Lambda^2} \kappa^2 \rho(\kappa^2)\, d\kappa^2. \tag{4.5}$$

These two characteristics of light-cone singularities come out from the comparison with the free field case. It is possible, however, to introduce generalized wave-renormalization constants $Z_{3;n}^{-1}$ describing the light-cone singularity compared with the singularities of the analytic functions $G_n(z^2; m^2)$ for $z^2 \to 0$.

We define particularly, $Z_{3;0}^{-1} \equiv Z_3^{-1}$,

$$Z_{3;n}^{-1} = \lim_{z^2 \to 0} \frac{G(z^2)}{G_n(z^2)}, \quad n = 0, 1, 2, \cdots. \tag{4.6}$$

It follows from the postulate of positive metric in the space of physical states that $Z_{3;0}^{-1} \geq 1$. If $Z_{3;0}^{-1} = \infty$, it can be found, however, for all such n, that $Z_{3;n}^{-1} < \infty$.

5. CONCLUSIONS

In this paper we have introduced a new class of basic two-point functions describing noncanonical two-body forces, more singular in static approximation than the Yukawa term. This modification of the $1/r$ singularity is caused by the exchange of infinite number of quanta with very large momenta. In the usual approach, such a process leads to ultraviolet divergences and the necessity of infinite renormalization.

In our approach we introduce some objects, characterized by the continuous mass spectrum and formed out of infinite number of quanta. These objects are chosen in such a way that the "one-particle exchange" approximates in a correct way the short-distance singularity for complete two-body forces. (In Ref. 24 the free field with polynomial spectral functions have been called the "inverse multipole field.")

The presence of noncanonical forces modifies the interaction at very small distances in such a way that the notion of charge and mass for these distances are not valid. Indeed, the charge and mass can be defined only under the assumption that the interaction has in static approximation the Yukawa form. One defines the unrenormalized parameters as follows:

$$e_0^2 = \lim_{r\to 0} V(r)r, \tag{5.1}$$

$$m_0^2 = \lim_{r\to 0} \frac{\partial^2}{\partial r^2}[V(r)r]. \tag{5.2}$$

The formulas (5.1) and (5.2) give infinite results because the Yukawa law for very small [we call them submicroscopic (see Ref. 17, Sec. 5)] distances is modified.

If we consider two-body forces we can always split them into two parts: with $1/r$ singularity (canonical terms) and with the singularity stronger than $1/r$ (noncanonical terms). The submicroscopic distances are defined by the requirement that the effects of noncanonical terms cannot be neglected. The validity of perturbation expansion is strictly connected with the neglecting of physical importance of noncanonical terms and cannot be used for submicroscopic distances. Using first orders of the perturbation theory one can guess, however, that the submicroscopic distances in QED are indeed beyond the range of physical measurements. One can calculate[17,25] that the noncanonical terms, occurring in the second order of perturbation theory in QED, can be neglected if

$$(\alpha/3\pi)\ln(\Lambda/M_e) \ll 1, \tag{5.3}$$

where M_e denotes the electron mass and $\Lambda = 1/a$ describes the cutoff parameter corresponding to the penetration distance a. Using the value

$$\alpha = e^2/4\pi \sim 1/137,$$

one obtains

$$\ln(\Lambda/M_e) \ll 1000. \tag{5.4}$$

We see easily from (5.4) that it is not possible to detect by scattering experiments the modification of the Coulomb-law singularity and, particularly, the Pauli–Villars regularization procedure, removing noncanonical terms, can be used. (For the demonstration how the regularization procedure removes noncanonical terms see Refs. 26, 27.) To the contrary, it is easy to check that the estimate for strong interactions leads to the range of submicroscopic distances overlapping with the values of scattering parameters in present high energy experiments. We see, therefore, that the conventional perturbation expansion cannot be used and some other approximations, using, perhaps, the propagators introduced in this paper, should be developed.

ACKNOWLEDGMENTS

The author is indebted to Professor R. Haag, Professor G. Källen, and Professor R. F. Streater for helpful discussions at the 5th Winter School of Theoretical Physics in Karpacz.

[24] J. Lukierski, Nuovo Cimento **49**, 312 (1967).

[25] M. Gell-Mann and F. Low, Phys. Rev. **95**, 1300 (1954).

[26] G. Källen, Schladming Lecture, 1968, and Karpacz Winter School Lecture, 1968.

[27] B. Hemprecht, Nuovo Cimento **47A**, 770 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Determination of Invariant Amplitudes from Experimental Observables*

MICHAEL J. MORAVCSIK AND WING-YIN YU
Department of Physics and Institute of Theoretical Science
University of Oregon, Eugene, Oregon

(Received 26 July 1968)

Necessary and sufficient conditions are formulated for deciding whether a set of $2n-1$ bilinear products of n complex amplitudes determines these amplitudes with no continuum of ambiguity or not. The conditions can be translated into very simple geometrical prescriptions which in most cases provide quick and easy practical tests for such decisions.

I. MOTIVATION

In any quantum-mechanical system, the experimentally measurable quantities are bilinear in the amplitudes. On the other hand, theories of quantum-mechanical phenomena almost always predict the amplitudes themselves. From them, it is generally easy to compute the experimental observables and, thus, comparison between experiment and theory can be made at the level of these observables. On the other hand, when a reliable theory is missing and the task is to present the mass of experimental information in a form which is optimally economical and most suitable for quick comparison with theories to come, the obvious choice for such a form is the amplitudes themselves. In that case, one has to face the problem of determining the amplitudes from the measured bilinear combinations.

In particular, in elementary particle physics, one would like to be able to determine experimentally the amplitudes in the S matrix of a particle reaction. At low energies, a partial wave decomposition is often used, but, in general, the task is to determine the various invariant amplitudes from the measured cross sections, polarizations, spin correlations, and other observables. Although, with experiments at ever higher energies and the development of techniques for sophisticated measurements of spinwise complicated observables, this problem is becoming more and more central, it has not been solved in other than a few special cases.

The problem has several parts, some purely mathematical in nature, others also involving the physical relationship between observables and amplitudes. In the present paper we will consider one purely mathematical part of this problem. It is probably the least straightforward part, since it involves the analysis of the bilinear nature of the relationship between amplitudes and observables.

* Work supported by the U.S. Atomic Energy Commission.

II. THE PROBLEM

The question can be posed quite simply.

We have n complex amplitudes a_i $(i = 1, \cdots, n)$, with one over-all phase arbitrary. We then consider the n^2 bilinear products of these amplitudes. They will be referred to as bits, and will be either $|a_i|^2$ (denoted by Q_i), or $\mathrm{Re}\, a_i a_j^*$ (denoted by R_{ij}), or $\mathrm{Im}\, a_i a_j^*$ (denoted by I_i^j). The number of Q_i's is n, and the numbers of R_{ij} and I_i^j (for $i < j$) each is $\frac{1}{2}n(n-1)$. (Note that $R_{ji} = R_{ij}$ and $I_j^i = -I_i^j$.)

The problem we want to solve is the description of the properties of those sets of bits from which the amplitudes can be determined unambiguously. In particular, if we start with just a few bits, there will be a continuum of sets of amplitudes which yield those bits. As we increase the number of bits, the extent of the ambiguity decreases. At some point, there will cease to be a continuum of solutions for the amplitudes, although there might still exist a finite, discrete ambiguity. With the addition of further bits those ambiguities can also be eliminated.

In this paper we will give necessary and sufficient conditions for a set of bits to determine the amplitudes with no continuum of ambiguity. The problem is equivalent to finding a set of $2n-1$ functionally independent bits.

III. THE RESULTS

It is very helpful to introduce, in order to facilitate the bookkeeping, a geometrical analogy to our bits and amplitudes. We will denote each amplitude by a point and each bit by a line connecting the corresponding amplitudes. Thus, Q_i will be denoted by a loop starting from and returning to the point i, while R_{ij} will be represented by a solid line connecting the points i and j. Similarly, I_i^j will be represented by a dashed line connecting i and j. (See Fig. 1.) In this notation, for example, the three amplitudes with Q_1, Q_2, R_{12}, I_2^3, and I_1^3 given would form a pattern like the last one in the second row of Fig. 2.

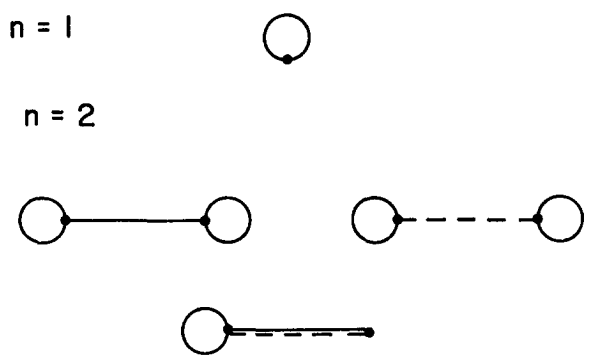


FIG. 1. Amplitude-bit patterns for one and two amplitudes. All of these diagrams determine the amplitudes with no continuum of ambiguity.

A procedure for testing for a continuum of ambiguities should be described by referring to an example. Take three amplitudes a_1, a_2, and a_3 and assume that Q_1, Q_2, R_{12}, I_2^3, and I_1^3 are given. With this information, is there a continuum of ambiguity in the three amplitudes? Denote Re a_i by u_i, Im a_i by w_i, and assume that we make the transformation $u_i \to u_i + \Delta u_i$, $w_i \to w_i + \Delta w_i$. Since one of the three amplitudes can be taken to be real, choose a_1 real or $w_1 = 0$. Then from Q_1 we have $\Delta u_1 = 0$, from R_{12} we obtain $\Delta u_2 = 0$, and from Q_2 we get $\Delta w_2 = 0$, while I_1^3 gives $\Delta w_3 = 0$. Finally, I_2^3 yields $\Delta u_3 = 0$, thus showing that it is not possible to move away from a solution for a_1, a_2, and a_3 if Q_1, Q_2, R_{12}, I_2^3, and I_1^3 are given. There are, of course, a number of equivalent procedures.

Definition: We say a set of bits *belongs to* a set of amplitudes if the bits have only those subscripts and superscripts which appear on the amplitudes (i.e., the bilinear combinations contain only members of that set).

Example: The bit set $\{Q_1, Q_3, R_{12}, R_{23}, I_1^3\}$ belongs to the amplitude set $\{a_1, a_2, a_3\}$. On the other hand, the bit set $\{Q_1, R_{12}, R_{34}, I_1^2\}$ does not belong to the amplitude set $\{a_1, a_2, a_3\}$ because of the presence of R_{34}.

Theorem 1: n amplitudes are determined up to an over-all arbitrary phase factor, with no continuum of ambiguity, by a set of $2n - 1$ bits only if this set is such that the bit subset belonging to any subset of k amplitudes ($k = 1, \cdots, n$) contains no more than $2k - 1$ bits. This condition will be referred to as condition A.

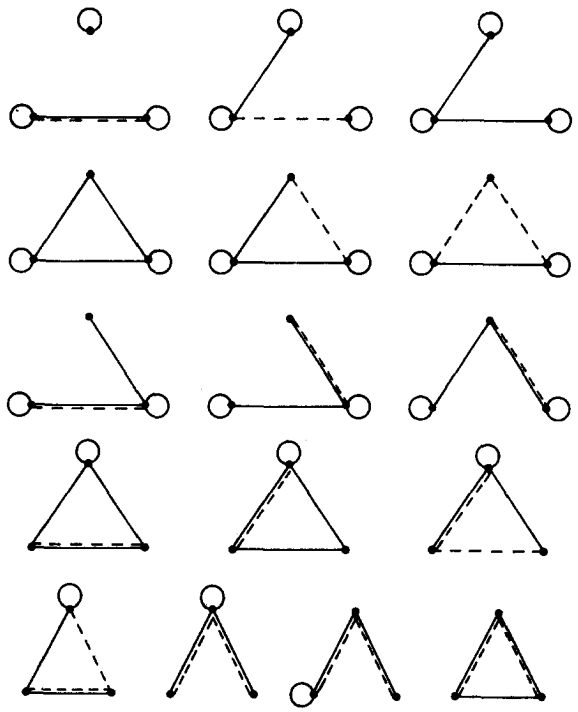


FIG. 2. Amplitude-bit patterns for three amplitudes. All of these diagrams determine the amplitudes with no continuum of ambiguity, except the first one in the first row and the first one in the third row. Only about half of the patterns are depicted; the rest can be obtained from these by interchanging solid and dashed lines.

Proof: It is equivalent to show that when the condition A is not satisfied, the set of n amplitudes will not be determined up to an over-all arbitrary phase factor with no continuum of ambiguity. Consider a set of n amplitudes with a subset of k amplitudes having m bits belonging to this subset and $m > 2k - 1$. Since these m bits are functions of the $2k - 1$ variables,[1] u_i and w_i of the k amplitudes, $m > 2k - 1$ implies that the m bits are not mutually functionally independent. It follows that the set of $2n - 1$ bits is not a functionally independent set. But we know that n amplitudes are determined up to an over-all arbitrary phase factor with no continuum of ambiguity by a set of $2n - 1$ bits if and only if this set is a functionally independent set of bits belonging to the n amplitudes. Q.E.D.

Theorem 2: The condition A is necessary and sufficient for the case where the set of $2n - 1$ bits has a geometrical representation which can be constructed from a "good" diagram by the successive addition of a point and two bits.

[1] There are $2k - 1$ and not $2k$ variables because the over-all phase factor is arbitrary and so we can set one amplitude to be real without loss of generality.

By a "good" diagram we mean a geometrical representation of a set of bits which will yield solutions with no continuum of ambiguity for the set of amplitudes it belongs to.

Proof: Consider a diagram of n points constructed by the successive addition of r points to a "good" diagram containing s points ($n = r + s$). We shall use induction on r.

Obviously the theorem is true for $r = 0$. Let us assume that the theorem holds for the successive addition of $r - 1$ points, making a diagram of $n - 1$ points. Let us add then one more point, and correspondingly two more bits. These two bits can be added in three different ways. They can both connect two of the old, $n - 1$ points. In this case, these $n - 1$ amplitudes will have more than $2n - 3$ bits belonging to them. On the other hand, in that case the nth amplitude is completely undetermined. The second possibility is that one of the new bits connects two old points, and the other the new point with an old one. In this case, the $n - 1$ old points will have $2n - 2$ bits belonging to them, and, on the other hand, the new point will still have a continuum of ambiguity since only one bit will tie down its real *and* imaginary parts. Thus, so far our theorem is satisfied. Finally, the third possibility is that *both* new bits connect some old point and the new point (or one of them can be Q_i for the new point). In that case no subset of points will have more than $2k - 1$ bits belonging to it, but on the other hand the new point will be fixed with no continuum of ambiguity by the two new bits connecting with it. Q.E.D.

Observe that, in the proofs of Theorems 1 and 2, knowledge of the particular functional forms of the bits is not required. It follows that the theorems are true for a much larger class of functions than the bilinear type of relationship considered in the present context.

The condition A would be a sufficient condition in general as well if there were no relations among the bits inherent to the particular functional forms of the bits.

We shall state here another necessary condition, which arises from the particular functional forms of the bits.

Theorem 3: A loopless diagram is "good" only if for each point there exists a path whereby it is possible to return to the same point after traversing an odd number of bits.[2] This condition will be called condition B.

[2] M. J. Moravcsik, Experimental Restrictions in the Determination of Invariant Amplitudes, Phys. Rev. **170**, 1440 (1968). Second category, situation *E*.

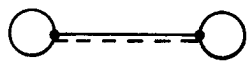


FIG. 3. An amplitude-bit pattern which, if it appears inside a larger pattern, is evidence for a continuum of ambiguity. For details, see the text.

Theorems 2 and 3 are not criteria of most general applicability but nevertheless they are simple and easy to apply.

There are a number of obvious ways of having a continuum of ambiguities for n amplitudes with $2n - 1$ bits, which are all special cases of these theorems. For instance, if the pattern consists of two or more disconnected subpatterns, the relative phases of these subsets are undetermined. In that case, one of these subsets (populations k_i and $\sum_i k_i = n$) must have more than $2k_i - 1$ bits (since the sum of the numbers of bits in the subsets is $2n - 1$). If the pattern contains a one-legged terminal (i.e., a loopless point which is attached to the rest of the diagram by only one line), it clearly contains a continuum of ambiguity since the point at the end of this terminal has only one bit to tie down its real and imaginary parts. But then the remaining $n - 1$ points have $2n - 2$ bits belonging to them. Finally, the appearance of the subpattern shown in Fig. 3 clearly exhibits one superfluous bit and this is a trivial special case of Theorem 1. Here the superfluity arises from the relationship $Q_iQ_j = (R_{ij})^2 + (I_i^j)^2$.

IV. A PRESCRIPTION

We summarize here a practical procedure suggested by these results.

If you want to know whether a certain set of $2n - 1$ bilinear products of n complex amplitudes determines these amplitudes (up to an over-all arbitrary phase factor) with no continuum of ambiguity or not (i.e., whether the set is "good" or "bad"), use the following prescription:

1. Draw n points, representing the n amplitudes.
2. Draw a line connecting two of the points for each bilinear product containing the amplitudes corresponding to those points.
3. Draw a loop starting from and returning to a given point for the absolute value squared of an amplitude corresponding to that point.
4. For each subset of the set of n points, count the number of lines starting from and ending on members of that subset.
5. If condition A is not satisfied, the diagram can be rejected as "bad." If condition A is satisfied, for a loopless diagram proceed to 6; otherwise proceed to 7.

6. If condition B is not satisfied, the diagram can be rejected as "bad." If condition B is satisfied, proceed to 7.

7. Use Theorem 2 if applicable. If not applicable, proceed to 8.

8. Write down the Jacobian of the bits with respect to the $2n - 1$ variables, u_i and w_i of a_i. The n amplitudes will be determined up to an over-all arbitrary phase factor with no continuum of ambiguity if and only if $|J| \neq 0$.

The Jacobian approach to the problem through functional dependence remains a most powerful method and provides the only necessary and sufficient condition formulated so far. However, the test whether $|J| \neq 0$ is tedious. The criteria afforded by Theorems 1, 2, and 3 serve to reduce considerably the number of cases that need to be subjected to the $|J| \neq 0$ test.

ACKNOWLEDGMENT

We are indebted to Dr. Hans Rosdolsky for pointing out, through a counter example, that the first version of our manuscript contained an incomplete proof.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Hamilton–Dirac Theory of Hamilton's Equations

ROBERT G. CAWLEY
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

(Received 24 October 1968)

It is shown that in a very general way two distinct canonical formalisms can be used to describe a classical system. No corresponding nonuniqueness is introduced into the canonical quantization procedure if the Dirac bracket correspondence to the quantum-mechanical commutators is employed.

I. INTRODUCTION

The present paper is an observation concerning an aspect of the uniqueness of the canonical formalism. A system of differential equations derivable from a Hamilton stationary action principle can be converted to a larger system of first-order equations by an order-reduction procedure involving the use of Legendre transformations active in the velocities. This can be managed even when the Lagrangian involves time derivatives of the coordinates of arbitrary finite order. Furthermore, the form of the first-order equations is canonical.

Restrictions on these procedures have to do with the possibility that the equations defining the generalized momenta cannot be solved for the velocities. In the case that these equations cannot be inverted, a Hamiltonian formalism can still be introduced and by the same procedures, but it suffers from the defect of being nonunique; famous examples in which this situation appears are provided by the canonical theories of the Schrödinger, Dirac, and Maxwell fields. A generalized Hamiltonian dynamics treating systems containing constraints (vanishing functions of coordinates and momenta) was invented by Dirac[1] to handle such problems as these.

In the usual Hamilton–Jacobi theory a class of Hamiltonians is generated from any given one by means of canonical transformations, but, owing to the canonical invariance of the theory, these all describe the same system. The Hamiltonian then possesses the uniqueness of the equivalence class modulo the canonical group and this is the uniqueness of the canonical formalism. Because of their ability to deal with canonical constraints, the Dirac techniques allow for a more flexible approach to the construction of a Hamiltonian theory; one can vary the size of the phase space as well as its coordinate system. This can be done by varying the choice of active and passive variables in the Legendre transformation giving the Hamiltonian.

Consider the example of the Schrödinger field for which a Lagrangian (density) is

$$\mathfrak{L} = +i\hbar\psi^*\dot{\psi} - \frac{\hbar^2}{2m}\nabla\psi^* \cdot \nabla\psi - V\psi^*\psi.$$

[1] P. A. M. Dirac, *Belfer Graduate School of Science Monograph Series, No. 2: Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964).

6. If condition B is not satisfied, the diagram can be rejected as "bad." If condition B is satisfied, proceed to 7.

7. Use Theorem 2 if applicable. If not applicable, proceed to 8.

8. Write down the Jacobian of the bits with respect to the $2n - 1$ variables, u_i and w_i of a_i. The n amplitudes will be determined up to an over-all arbitrary phase factor with no continuum of ambiguity if and only if $|J| \neq 0$.

The Jacobian approach to the problem through functional dependence remains a most powerful method and provides the only necessary and sufficient condition formulated so far. However, the test whether $|J| \neq 0$ is tedious. The criteria afforded by Theorems 1, 2, and 3 serve to reduce considerably the number of cases that need to be subjected to the $|J| \neq 0$ test.

ACKNOWLEDGMENT

We are indebted to Dr. Hans Rosdolsky for pointing out, through a counter example, that the first version of our manuscript contained an incomplete proof.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Hamilton–Dirac Theory of Hamilton's Equations

ROBERT G. CAWLEY
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

(Received 24 October 1968)

It is shown that in a very general way two distinct canonical formalisms can be used to describe a classical system. No corresponding nonuniqueness is introduced into the canonical quantization procedure if the Dirac bracket correspondence to the quantum-mechanical commutators is employed.

I. INTRODUCTION

The present paper is an observation concerning an aspect of the uniqueness of the canonical formalism. A system of differential equations derivable from a Hamilton stationary action principle can be converted to a larger system of first-order equations by an order-reduction procedure involving the use of Legendre transformations active in the velocities. This can be managed even when the Lagrangian involves time derivatives of the coordinates of arbitrary finite order. Furthermore, the form of the first-order equations is canonical.

Restrictions on these procedures have to do with the possibility that the equations defining the generalized momenta cannot be solved for the velocities. In the case that these equations cannot be inverted, a Hamiltonian formalism can still be introduced and by the same procedures, but it suffers from the defect of being nonunique; famous examples in which this situation appears are provided by the canonical theories of the Schrödinger, Dirac, and Maxwell fields. A generalized Hamiltonian dynamics treating systems containing constraints (vanishing functions of coordinates and momenta) was invented by Dirac[1] to handle such problems as these.

In the usual Hamilton–Jacobi theory a class of Hamiltonians is generated from any given one by means of canonical transformations, but, owing to the canonical invariance of the theory, these all describe the same system. The Hamiltonian then possesses the uniqueness of the equivalence class modulo the canonical group and this is the uniqueness of the canonical formalism. Because of their ability to deal with canonical constraints, the Dirac techniques allow for a more flexible approach to the construction of a Hamiltonian theory; one can vary the size of the phase space as well as its coordinate system. This can be done by varying the choice of active and passive variables in the Legendre transformation giving the Hamiltonian.

Consider the example of the Schrödinger field for which a Lagrangian (density) is

$$\mathfrak{L} = +i\hbar\psi^*\dot{\psi} - \frac{\hbar^2}{2m}\nabla\psi^* \cdot \nabla\psi - V\psi^*\psi.$$

[1] P. A. M. Dirac, *Belfer Graduate School of Science Monograph Series, No. 2: Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964).

Here one can introduce $\pi = \partial\mathfrak{L}/\partial\dot{\psi}$ and $\pi^* = \partial\mathfrak{L}/\partial\dot{\psi}^*$, finding a four-dimensional phase space; or one can keep $\dot{\psi}^*$ passive, introducing only π, and find a two-dimensional phase space as ψ^* can then be eliminated in favor of π. In the first procedure one finds two second-class constraints, $\pi + i\hbar\psi^* = 0$ and $\pi^* = 0$, and the elimination procedure of Dirac, involving the Dirac bracket in place of the Poisson bracket, yields a canonical theory the same as that based on the two-dimensional phase space. The canonical formalism based on the Dirac bracket is displaying a kind of uniqueness here having to do with the fact that its use corresponds to freezing a canonically conjugate pair and projecting the system description to a two-dimensional subspace of the original phase space. The conditions under which Dirac's elimination procedure will do this have been given recently by Mukunda and Sudarshan.[2]

Uniqueness of this kind is actually important for the passage to a quantum theory of a given classical system, as it is possible to envision formalisms which mix up coordinates and their conjugate momenta in a way which cannot happen with canonical transformations on a *given* space. Consider for example a system described by the four independent canonical coordinates q_1, q_2, p_1, and p_2; the transformation sending these into the set $Q_1 = q_1$, $Q_2 = p_1$, $P_1 = q_2$, $P_2 = p_2$ is not canonical because the Poisson brackets are not preserved. In passing to a quantum theory of such a system, the identification of the fundamental classical Poisson brackets with the corresponding quantum mechanical commutators can never lead to an inconsistency involving the uncertainty relation.

The present paper shows a general and quite direct way in which one can always construct two equivalent Hamiltonian theories which have the property that the conjugate variables q, p of one scheme are the independent coordinates of the other. But the second theory (we will call it the "stretched" theory) has constraints and it will be shown that these can always be eliminated by the Dirac procedure. Furthermore, while q, p are independent coordinates of the stretched theory with respect to the Poisson brackets, they are canonically conjugate coordinates with respect to the Dirac bracket.

In Sec. II, after presenting a brief description of certain aspects of the Dirac theory, we construct the "stretched" formalism and show how the Dirac elimination procedure recovers the original canonical scheme. The cases where canonical constraints are absent and present are both treated. In Sec. III we give a short summary.

[2] N. Mukunda and E. C. G. Sudarshan, J. Math. Phys. **9**, 411 (1968).

II. CONSTRUCTION OF FORMALISM

We consider a system having N degrees of freedom represented by the coordinates q which obey equations derivable from Hamilton's principle. Lagrange's function $L = L(q, \dot{q})$ is assumed to be expressible as a function of the coordinates q and their velocities $\dot{q}$, as indicated. The Hamilton formulation results from applying Legendre's transformation[3] to $L(q, \dot{q})$ with the q's passive:

$$H = p_n\dot{q}_n - L(q, \dot{q}), \tag{1}$$

where

$$p_n = \partial L/\partial\dot{q}_n, \tag{2}$$

and a summation on n is understood in Eq. (1). The Hamiltonian can be expressed as a function of the q's and p's, as is seen by noting that the most general variation of the right side of Eq. (1) is

$$\dot{q}_n\delta p_n - \frac{\partial L}{\partial q_n}\delta q_n + \left(p_n - \frac{\partial L}{\partial\dot{q}_n}\right)\delta\dot{q}_n = \dot{q}_n\delta p_n - \frac{\partial L}{\partial q_n}\delta q_n, \tag{3}$$

by Eq. (2). The determination of the most general $H(q, p)$ may not be unique, however. For example, it may happen that only l of the velocities can be got from Eq. (2) with the remaining $N - l$ determinations replaced by constraint equations of the form

$$\varphi_m(q, p) = 0. \tag{4}$$

In this instance, we may proceed in the manner of Dirac[1] by introducing the extended Hamiltonian and eliminating the second-class constraints by replacing the Poisson bracket

$$[\xi, \eta]_P = \frac{\partial\xi}{\partial q_n}\frac{\partial\eta}{\partial p_n} - \frac{\partial\xi}{\partial p_n}\frac{\partial\eta}{\partial q_n} \tag{5}$$

with the Dirac bracket

$$[\xi, \eta]_D = [\xi, \eta]_P - [\xi, \varphi_s]_P c_{ss'}[\varphi_{s'}, \eta]_P, \tag{6}$$

where $\{\varphi_s\}$ is an irreducible set of second class constraints and the $c_{ss'}$ are given through

$$c_{ss'}[\varphi_{s'}, \varphi_{s''}] = \delta_{ss''}. \tag{7}$$

In the absence of constraints, there are no $c_{ss'}$ and the Dirac bracket reduces to the Poisson bracket.

The cases where constraints of the type given by Eq. (4) are present and where they are absent will both be treated. We do the latter first.

[3] A quite readable exposition of this beautiful theory may be found in C. P. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, Ontario, Canada, 1949), p. 161 ff.

A. No Constraints

In this case the Hamiltonian is uniquely determined and we start from the fact[4] that the Hamilton equations can be found not only from the properties of Legendre's transformation but also from a variation principle based on the canonical integral. Thus, solving Eq. (1), one has

$$L(q, \dot{q}) = p_n \dot{q}_n - H(q, p) \equiv \mathfrak{L}(q, p; \dot{q}, \dot{p}) \quad (8)$$

and, since arbitrary variations of *p produce no change in* $\mathfrak{L}$, we can replace Hamilton's principle for $L(q, \dot{q})$ with

$$\delta I[q, p] = \delta \int_{t_1}^{t_2} dt \mathfrak{L}(q, p; \dot{q}, \dot{p}) = 0, \quad (9)$$

the changes of I being generated now by independent variations of q's and p's, of the same variety as are used in Hamilton's principle, i.e., those for which δq, δp vanish at t_1 and t_2 and for which $[\delta, d/dt] = 0$.

We consider this approach in the light of a ("stretched") Lagrangian theory and set about constructing a Hamilton theory from it by applying a Legendre transformation to $\mathfrak{L}(q, p; \dot{q}, \dot{p})$. This gives us a stretched Hamiltonian

$$\mathcal{H} = r_n \dot{q}_n + s_n \dot{p}_n - \mathfrak{L}(q, p; \dot{q}, \dot{p}), \quad (10)$$

where

$$r_n = \partial \mathfrak{L}/\partial \dot{q}_n \quad \text{and} \quad s_n = \partial \mathfrak{L}/\partial \dot{p}_n. \quad (11)$$

Proceeding as in Eq. (3) and with the same *caveat* as before concerning uniqueness, we see that $\mathcal{H}$ must be expressible as a function of the $4N$ coordinates q, p, r, and s. But here we must pay closer attention to questions of uniqueness, for constraint equations there unavoidably are, and in fact $2N$ of them. They are provided explicitly by Eqs. (11) and (8); one finds

$$\varphi_{1m} = s_m \approx 0, \quad (12a)$$

$$\varphi_{2m} = r_m - p_m \approx 0, \quad (12b)$$

and[5] these are primary constraints. Variation of Eq. (10) in the presence of the constraints (12) leads to the stretched canonical system,

$$\frac{\partial \mathcal{H}_{\mathfrak{T}}}{\partial r_n} \approx \dot{q}_n, \qquad \frac{\partial \mathcal{H}_{\mathfrak{T}}}{\partial s_n} \approx \dot{p}_n,$$

$$-\frac{\partial \mathcal{H}_{\mathfrak{T}}}{\partial q_n} \approx \dot{r}_n, \qquad -\frac{\partial \mathcal{H}_{\mathfrak{T}}}{\partial p_n} \approx \dot{s}_n, \quad (13)$$

where $\mathcal{H}_{\mathfrak{T}}$ is given by

$$\mathcal{H}_{\mathfrak{T}} = \mathcal{H} + u_{1m}\varphi_{1m} + u_{2m}\varphi_{2m}$$
$$= \mathcal{H} + u_{1m}s_m + u_{2m}(r_m - p_m), \quad (14)$$

and the u's are Lagrange multipliers to Eqs. (11). Poisson brackets are now defined with respect to q and p as coordinates, and r and s as (respectively) conjugate momenta.

Secondary constraints sometimes arise from the consistency conditions $\dot{\varphi}_m \approx [\varphi_m, \mathcal{H}_{\mathfrak{T}}]_{\mathfrak{T}} \approx 0$; in the present example there are none. (We use script suffixes to denote the fact that the brackets are defined with respect to the stretched phase space.) One gets instead

$$-\frac{\partial \mathcal{H}}{\partial p_m} + u_{2m} \approx 0, \quad (15a)$$

$$-\frac{\partial \mathcal{H}}{\partial q_m} - \frac{\partial \mathcal{H}}{\partial s_m} - u_{1m} \approx 0, \quad (15b)$$

which merely serve to determine the u's. Furthermore, since

$$[\varphi_{1m}, \varphi_{2n}]_{\mathfrak{T}} = \delta_{mn} \quad (16a)$$

and

$$[\varphi_{1m}, \varphi_{1n}]_{\mathfrak{T}} = 0, \quad (16b)$$

$$[\varphi_{2m}, \varphi_{2n}]_{\mathfrak{T}} = 0, \quad (16c)$$

all the constraints are second class and in the terminology of Ref. 1 the total and extended Hamiltonians are the same. Also, using again the notation of Ref. (1),

$$\mathcal{H}_{\mathfrak{T}} = \mathcal{H}', \quad (17)$$

so the Hamiltonian is uniquely determined by the consistency conditions.

We can get $\mathcal{H}$, which must be independent of the velocities, by combining Eqs. (8) and (10) and using Eq. (12); thus,

$$\mathcal{H}(qprs) = H(q, p). \quad (18)$$

So Eqs. (15) now give

$$u_{1m} = \mathfrak{U}_{1m}(qprs) = -\frac{\partial \mathcal{H}}{\partial q_m} - \frac{\partial \mathcal{H}}{\partial s_m} = -\frac{\partial H}{\partial q_m}, \quad (19a)$$

$$u_{2m} = \mathfrak{U}_{2m}(qprs) = \frac{\partial \mathcal{H}}{\partial p_m} = \frac{\partial H}{\partial p_m}, \quad (19b)$$

so by Eqs. (14) and (17) we arrive at the unique (strong) stretched Hamiltonian,

$$\mathcal{H}_{\mathfrak{U}}(qprs) = \mathcal{H}_{\mathfrak{T}} = H(q, p) - \frac{\partial H}{\partial q_n} s_n + \frac{\partial H}{\partial p_n}(r_n - p_n). \quad (20)$$

It may be verified that this satisfies Eqs. (13) when, in accord with the consistency conditions, the constraints are time-independent.

The classical system under consideration is equally well represented by the canonical formalism based on $H(q, p)$, with q and p canonically conjugate, and by that based on $\mathcal{H}_{\mathfrak{U}}(qprs)$ where q and p are coordinates whose conjugate momenta are r and s. Suppose now

[4] Reference 3, p. 168.

[5] The wavy equals sign $\approx$ denotes weak equality in the sense of Dirac, Ref. 1.

we want to quantize this system. If we adopt the procedure based on $H(q,p)$, we will identify the canonical commutators with the Poisson brackets. We would hope that the quantization based on the stretched formalism would result in a theory indistinguishable from that based on $H(q,p)$. But blind quantization through the Poisson bracket relations would tell us in the one formalism that q and p are canonically conjugate dynamical variables obeying an uncertainty relation, while in the second (stretched) scheme they emerge as independent coordinates enjoying simultaneous measurability privileges. The problem is that this does not take account of the constraints. The Dirac quantization scheme is an attempt to do this and we can use the present example to test its consistency. In the Dirac scheme the fundamental approach is through the Dirac brackets, together with the elimination of the second class constraints.

To this end we recall that all of the $2N$ constraints of the stretched formalism are second class. From Eqs. (7) and (16), we find the matrix $c_{ss'}$ has the block diagonal form

$$\|c_{ss'}\| = \begin{pmatrix} 0 & -1 & & & & \\ 1 & 0 & & & \mathbf{O} & \\ & & 0 & -1 & & \\ & & 1 & 0 & & \\ & & & & 0 & -1 \\ & \mathbf{O} & & & 1 & 0 \end{pmatrix}, \quad (21)$$

where pairs of rows and columns are labeled by values of n and members of each pair by 1 and 2 corresponding to the enumeration of the constraints. Combining Eqs. (6), (12), and (21), we find, suspending the summation convention,

$$[\xi, \eta]_{\mathfrak{D}} = [\xi, \eta]_{\mathfrak{F}} + [\xi, s_m]_{\mathfrak{F}}[r_m - p_m, \eta]_{\mathfrak{F}} - [\xi, r_m - p_m]_{\mathfrak{F}}[s_m, \eta]_{\mathfrak{F}}, \quad (22)$$

so that for the fundamental Dirac brackets we have

$$\begin{aligned} &[q_m, s_n]_{\mathfrak{D}} = 0, \quad [q_m, p_n]_{\mathfrak{D}} = \delta_{mn}, \quad [q_m, r_n]_{\mathfrak{D}} = \delta_{mn}, \\ &[q_m, q_n]_{\mathfrak{D}} = 0, \quad [r_m, p_n]_{\mathfrak{D}} = 0, \quad [r_m, r_n]_{\mathfrak{D}} = 0, \\ &[p_m, s_n]_{\mathfrak{D}} = 0, \quad [p_m, p_n]_{\mathfrak{D}} = 0, \quad [s_m, s_n]_{\mathfrak{D}} = 0, \\ &\qquad\qquad\qquad\qquad\qquad [r_m, s_n]_{\mathfrak{D}} = 0, \end{aligned} \quad (23)$$

while the remaining brackets can be found by application of the relation $[\xi, \eta]_{\mathfrak{D}} = -[\eta, \xi]_{\mathfrak{D}}$.

The elimination procedure of Dirac can now be implemented. The N conjugate pairs (p, s), whose rather uninspiring time development is given by $p = r$ and $s = 0$, are eliminated in favor of the q's and r's, and the fundamental Dirac bracket relations,

$$[q_m, r_n]_{\mathfrak{D}} = \delta_{mn}, \quad [q_m, q_n]_{\mathfrak{D}} = [r_m, r_n]_{\mathfrak{D}} = 0, \quad (24)$$

assume the role of the fundamental Poisson brackets of the usual theory. For the new Hamiltonian we have $\mathcal{H}_{\mathfrak{U}}(qprs) \to \tilde{\mathcal{H}}(q, r)$, where, by Eq. (20),

$$\mathcal{H}_{\mathfrak{U}}(q, p, r, s) = \tilde{\mathcal{H}}(q, r) = H(q, r). \quad (25)$$

Equations (24) and (25) show that the distinction between quantization of this system through the Dirac prescription based on the stretched formulation and the usual prescription (also Dirac's) based on the "unstretched" formulation is no more than alphabetical: the use of r in place of p for the same object.

B. Constraints Present in the Unstretched Theory

We assume now that the Hamilton formulation for the equations describing the system must be carried out in the face of constraints of the type given by Eq. (4):

$$\Phi_k(q, p) = 0. \quad (26)$$

The Dirac theory shows that the total Hamiltonian,

$$H_T = H + w_k\Phi_k, \quad (27)$$

where H is given by Eq. (1), satisfies the usual Hamilton equations weakly. Actually Eq. (27) may not represent a single Hamiltonian, but a class of Hamiltonians differing from one another by arbitrary linear combinations of first-class constraints. This nonuniqueness springs from the fact that the state of the system at a given time does not necessarily determine a unique set of coordinates.

It can be shown that the entire physical theory based on H_T, the equations of motion, the nonuniqueness, everything, can be recovered from the action principle

$$\begin{aligned} \delta I &= \delta\int_{t_1}^{t_2} dt\mathfrak{L}(q, p; \dot{q}, \dot{p}) \\ &= \delta\int_{t_1}^{t_2} dt(p_n\dot{q}_n - H(q, \dot{q})) = 0, \end{aligned} \quad (28)$$

with Eq. (26) as subsidiary conditions on the variations. We are precisely where we were before now except for the presence of the constraints. We can pass as before to the stretched Hamiltonian (10) and have only to include Eq. (26) together with Eq. (12) to recover Eqs. (13) with $\mathcal{H}_{\mathfrak{G}}$ replaced by

$$\mathcal{H}^t_{\mathfrak{G}} = \mathcal{H} + \lambda_k\Phi_k + u_{1m}\varphi_{1m} + u_{2m}\varphi_{2m}, \quad (29)$$

where λ_k are Lagrange multipliers to Eq. (26). We note that Eq. (18) still holds.

Turning to the consistency conditions, those for φ_{1m} and φ_{2m} determine the u's uniquely when the λ's

are specified. The results are correctly given by Eqs. (15) if $\mathcal{H}$ is replaced by

$$\mathcal{H}^t \equiv \mathcal{H} + \lambda_k \Phi_k. \tag{30}$$

Hence, substituting into Eq. (29), we find

$$\mathcal{H}^t_{\mathcal{G}} = \mathcal{H}^t - \frac{\partial \mathcal{H}^t}{\partial q_n} s_n + \frac{\partial \mathcal{H}^t}{\partial p_n}(r_n - p_n). \tag{31}$$

To handle the unstretched constraints $\Phi_k = 0$, we introduce

$$\Phi^S_k = \Phi_k - \frac{\partial \Phi_k}{\partial q_n} s_n + \frac{\partial \Phi_k}{\partial p_n}(r_n - p_n), \tag{32}$$

define

$$\mathcal{H}_{\mathcal{G}} \equiv \mathcal{H} - \frac{\partial \mathcal{H}}{\partial q_n} s_n + \frac{\partial \mathcal{H}}{\partial p_n}(r_n - p_n), \tag{33}$$

and insert into Eq. (31) to get

$$\mathcal{H}^t_{\mathcal{G}} = \mathcal{H}_{\mathcal{G}} + \lambda_k \Phi^S_k. \tag{34}$$

The consistency conditions on $\Phi_k = 0$ are given with the help of Eq. (34) by

$$\dot{\Phi}_k \approx [\Phi_k, \mathcal{H}^t_{\mathcal{G}}]_{\mathcal{F}} \approx [\Phi_k, \mathcal{H}_{\mathcal{G}}]_{\mathcal{F}} + \lambda_{k'}[\Phi_k, \Phi^S_{k'}] \approx 0. \tag{35}$$

The first term, using Eq. (18), is

$$\begin{aligned}[\Phi_k, \mathcal{H}_{\mathcal{G}}]_{\mathcal{F}} &= \left[\Phi_k, H - \frac{\partial H}{\partial q_n} s_n + \frac{\partial H}{\partial p_n}(r_n - p_n)\right]_{\mathcal{F}} \\ &= -\frac{\partial H}{\partial q_n}[\Phi_k, s_n] + \frac{\partial H}{\partial p_n}[\Phi_k, r_n]_{\mathcal{F}} \\ &= -\frac{\partial H}{\partial q_n}\frac{\partial \Phi_k}{\partial p_n} + \frac{\partial H}{\partial p_n}\frac{\partial \Phi_k}{\partial q_n} = [\Phi_k, H]_P.\end{aligned} \tag{36a}$$

Similarly,

$$[\Phi_k, \Phi^S_{k'}]_{\mathcal{F}} = [\Phi_k, \Phi^S_{k'}]_P, \tag{36b}$$

so combining Eqs. (35) and (36) we find the conditions $\dot{\Phi}_k \approx 0$ become

$$[\Phi_k, H + \lambda_{k'}\Phi_{k'}]_P \approx 0, \tag{37}$$

so the remaining, secondary constraints are found in the same way as in the unstretched theory and the λ_k in the same way as the w_k of Eq. (27) to precisely the extent that the w_k *are* determined. Hence we may, without loss of generality, take

$$H + \lambda_k \Phi_k = H_T, \tag{38}$$

and by Eqs. (18) and (30),

$$\mathcal{H}^t = H_T, \tag{39}$$

and finally,[6]

$$\mathcal{H}^t_{\mathcal{G}} = H_T - \frac{\partial H_T}{\partial q_n} s_n + \frac{\partial H_T}{\partial p_n}(r_n - p_n). \tag{40}$$

A partial elimination of the second-class constraints can be effected, singling out $\varphi_{1m} = 0$ and $\varphi_{2m} = 0$ for attack and leaving any others there may be untouched. To do this we need a "truncated" Dirac bracket where, in Eq. (6), s only runs over the indices on the φ's. The matrix (21) is unchanged and the relations (22), (23), and (24) are all the same if $[\ ]_{\mathcal{D}}$ is replaced by $[\ ]_{\mathcal{D}'}$, prime for truncated. The truncated Dirac bracket of the stretched formalism assumes the role of the Poisson bracket of the unstretched formalism and we have[6b]

$$\mathcal{H}^t_{\mathcal{G}} = H_T \quad \text{and} \quad \mathcal{H}^e_{\mathcal{G}} = H_E. \tag{41}$$

III. SUMMARY AND CONCLUSION

The fact that Hamilton's equations are the Euler–Lagrange equations of an action principle with respect to a "stretched" *configuration* space, which is in fact just the phase space of the Hamilton theory, enables us to regard the canonical equations in the context of a stretched Lagrangian theory. From this new theory is constructed a canonical formalism, the stretched formalism. Owing to the special form of the "kinetic energy," i.e., $p_n\dot{q}_n$, and the independence of the stretched Lagrangian of the $\dot{p}$'s, exactly $2N$ second-class constraints arise. Elimination of these constraints in the Dirac fashion, through the introduction of the Dirac bracket, recovers the original canonical formalism completely and no ambiguities can survive quantization if the Dirac bracket correspondence to the fundamental commutators is employed.

---

[6] (a) $\mathcal{H}^t_{\mathcal{G}}$ is only unique modulo the first-class primary constraints. Based on $H_T = H' + v_a\Phi_a$, Eq. (40) gives $\mathcal{H}^t_{\mathcal{G}} = \mathcal{H}' + v_a\Phi^S_a$, where $\mathcal{H}'$ is determined by the consistency conditions, the v_a are arbitrary and the Φ^S_a are given as in Eq. (32). The "stretched constraints" Φ^S_a can be shown to be first class if and only if Φ_a is first class in the unstretched theory; in fact all the Φ's satisfy the strong equations, $[\Phi^S_l, \Phi^S_{l'}]_{\mathcal{F}} = [\Phi_l, \Phi_{l'}]_P$. (b) Passage to the formalism of the extended Hamiltonian ($\mathcal{H}^e_{\mathcal{G}}$) can now be accomplished by replacing H_T in Eq. (40) by $H_E = H_T + v_{a'}\Phi_{a'}$, the $\Phi_{a'}$ being all those first-class secondary constraints generated by forming *unstretched* Poisson brackets of the Φ_a with one another!

Measure-Theoretical Description of Klein–Gordon Multiparticle States

EDUARD PRUGOVEČKI
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada*

(Received 5 June 1967)

It is shown that a multiparticle state constructed from solutions of the Klein–Gordon equation can be also described by a family of complex measures having simple properties. These measures are the expectation values for that state of products of projectors corresponding to the spectral decompositions of certain self-adjoint operators. The results obtained are used in discussing the problem of defining quantized free fields at a point.

1. INTRODUCTION

On previous occasions[1–3] we have shown how one can describe the physical states in nonrelativistic quantum mechanics by families of measures, obeying some simple rules. Such a description starts with the premise that we have at hand an unambiguous physical situation, in which observables are represented in the Hilbert-space formalism by self-adjoint operators. A measure is then attached to each combination consisting of a state and an n-tuple of observables.

In relativistic quantum mechanics, when dealing with the Klein–Gordon equation or Dirac equation, well-known difficulties appear[4] when one attempts to introduce observables as fundamental as the position observables. It is, therefore, interesting to reconsider the description of pure states in terms of families of measures. However, as we are not dealing any more with a physically clear-cut situation, we adjust the nature and properties of these measures to achieve the simplest possible formal properties. As the description of states with arbitrarily great numbers of identical particles is of particular interest, we consider the Hilbert space which is the direct sum of one, two, etc., particle states. We also allow for the presence of antiparticles in order to be able to apply the obtained results (Appendix) to the problem of defining free fields at a point.

Our considerations are carried out for the case of spin-zero particles obeying the Klein–Gordon equation. The treatment of the case of spin-$\frac{1}{2}$ particles obeying the Dirac equation could proceed along the same lines, but with a certain increased intricacy in the notation.

* Present address: Department of Mathematics, University of Toronto, Toronto 5, Canada.

[1] E. Prugovečki, J. Math. Phys. 7, 1054 (1966).

[2] E. Prugovečki, J. Math. Phys. 7, 1070 (1966).

[3] E. Prugovečki, J. Math. Phys. 7, 1680 (1966).

[4] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. 21, 400 (1949).

2. THE HILBERT SPACE OF PURE STATES

Consider particles of spin zero described by wave-functions $f(x)$ obeying the Klein–Gordon equation

$$(\Box - m^2)f(x) = 0, \quad \Box = \frac{\partial^2}{\partial x_\nu \partial x^\nu}, \tag{2.1}$$

expressed in the metric

$$x^2 = g_{\mu\nu}x^\mu x^\nu = (x^0)^2 - \mathbf{x}^2. \tag{2.2}$$

The Hilbert space $\mathcal{H}^{(1)}$ of all pure one-particle states can be obtained in the following way:

Denote with $K^{(1)}$ the function space of all positive energy solutions of (2.1), i.e., of all functions $f(x)$ which can be written in the form[5]

$$\begin{aligned} f(x) &= \frac{\sqrt{2}}{(2\pi)^{\frac{3}{2}}}\int d^4k e^{-ik\cdot x}\delta(k^2 - m^2)\theta(k^0)\hat{f}(k) \\ &= (2\pi)^{-\frac{3}{2}}\int \frac{d^3\mathbf{k}}{[2(\mathbf{k}^2 + m^2)]^{\frac{1}{2}}} \\ &\quad \times \exp\{i[\mathbf{k}\mathbf{x} - x^0(\mathbf{k}^2 + m^2)^{\frac{1}{2}}]\}\hat{f}_E(\mathbf{k}), \end{aligned} \tag{2.3}$$

where $\hat{f}_E(\mathbf{k})$ is a function for which

$$\begin{aligned} &\int \frac{d^3\mathbf{k}}{(\mathbf{k}^2 + m^2)^{\frac{1}{2}}}|f_E(\mathbf{k})|^2 < +\infty, \\ &f_E(\mathbf{k}) = \hat{f}[\mathbf{k}, (\mathbf{k}^2 + m^2)^{\frac{1}{2}}]. \end{aligned} \tag{2.4}$$

We introduce in $K^{(1)}$ the inner product

$$\begin{aligned} (f \mid g) &= i\int_{t=t_0} f^*(x)\overleftrightarrow{\frac{\partial}{\partial x^0}}g(x)\, d^3\mathbf{x} \\ &= \int \frac{d^3\mathbf{k}}{(\mathbf{k}^2 + m^2)^{\frac{1}{2}}}\hat{f}^*_E(\mathbf{k})g_E(\mathbf{k}). \end{aligned} \tag{2.5}$$

The Hilbert space $\mathcal{H}^{(1)}$ is the completion of $K^{(1)}$ with respect to the norm

$$\|f\| = [(f|f)]^{\frac{1}{2}}. \tag{2.6}$$

[5] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson and Co., Elmsford, N.Y., 1961), Secs. 3 and 6.

According to the established practice, if we have antiparticles which are distinct from the considered particles, we assume that the Hilbert state $\overline{\mathcal{H}}^{(1)}$ of all states of systems consisting of antiparticles is constructed from the space $\bar{K}^{(1)}$ of the negative-energy solutions of (2.1) in the same fashion as $\mathcal{H}^{(1)}$ was constructed from $K^{(1)}$. The general element of $\bar{K}^{(1)}$ will be

$$f(x) = \frac{\sqrt{2}}{(2\pi)^{\frac{3}{2}}} \int d^4k e^{ik\cdot x} \delta(k^2 - m^2)\theta(-k_0)\tilde{f}(k), \quad (2.7)$$

where $\tilde{f}[\mathbf{k}, -(\mathbf{k}^2 + m^2)^{\frac{1}{2}}]$ satisfies (2.4).

In order to construct the Hilbert space of vectors representing pure states with arbitrary numbers of particles and antiparticles, consider the Hilbert spaces

$$\hat{\mathcal{H}}^{(m,n)} = (\mathcal{H}^{(1)})^{\otimes m} \otimes (\overline{\mathcal{H}}^{(1)})^{\otimes n}, \quad m, n = 1, 2, \cdots, \quad (2.8)$$

which are the closures of the linear manifolds spanned by all the vectors of the form

$$f_1 \otimes \cdots \otimes f_m \otimes g_1 \otimes \cdots \otimes g_n,$$
$$f_1, \cdots, f_m \in \mathcal{H}^{(1)}, \quad g_1, \cdots, g_n \in \overline{\mathcal{H}}^{(1)}. \quad (2.9)$$

In $\hat{\mathcal{H}}^{(m,n)}$ we can introduce the symmetrizer

$$S^{(m,n)} = \frac{1}{m!\, n!} \sum_{\sigma \in G_m} \sum_{\tau \in G_m} [\sigma] \otimes \overline{[\tau]}, \quad (2.10)$$

where G_k denotes the group of all permutations of k objects, while $[\sigma] \otimes \overline{[\tau]}$, when applied on (2.9), performs the permutation σ on the first m indices, and the permutation τ on the last n indices. If we write

$$\mathcal{H}^{(m,n)} = [S^{(m,n)}\Phi, \Phi \in \hat{\mathcal{H}}^{(m,n)}], \quad (2.11)$$

then

$$\mathcal{H} = \bigoplus_{m,n=0}^{\infty} \mathcal{H}^{(m,n)} \quad (2.12)$$

is the Hilbert space of all pure physical states that we were searching. Note that for $m = n = 0$ we have introduced an extra space $\mathcal{H}^{(0,0)}$, which will be required to be one-dimensional, and consequently generated by a single normalized vector Ω, which will be called the vacuum state.

3. STATES DESCRIBED BY FAMILIES OF MEASURES

As is well known,[6] the j-coordinate position operator in $\mathcal{H}^{(1)}$ cannot be defined by the mapping

$$f(x) \to x^j f(x) \quad (3.1)$$

because such a mapping leads us outside $\mathcal{H}^{(1)}$, as is seen from the relation

$$(\Box - m^2)[x^\nu f(x)] = 2\partial^\nu f(x). \quad (3.2)$$

Instead we have the operators Q^j introduced by Newton and Wigner[7]

$$(Q^j f)(x) = x^j f(x) + \frac{1}{8\pi} \int \frac{\exp(-m|\mathbf{x} - \mathbf{y}|)}{|\mathbf{x} - \mathbf{y}|} \frac{\partial f(\mathbf{y})}{\partial y^j} d^3\mathbf{y}, \quad (3.3)$$

which do lead, however, to noncovariant localized states.

From (2.3) it is, however, clear that for fixed $x^0 = t$ a state $\Psi \in K^{(1)}$ is given completely if $\tilde{f}_E(\mathbf{k})$ is known, or equivalently if the function

$$\tilde{\psi}_t(\mathbf{k}; f) = (\mathbf{k}^2 + m^2)^{-\frac{1}{4}} \exp[-it(\mathbf{k} + m^2)^{\frac{1}{2}}]\tilde{f}_E(\mathbf{k}), \quad f \in \mathcal{H}^{(1)}, \quad (3.4)$$

is given. This suggests the one-to-one linear mapping

$$\tilde{f}_E(\mathbf{k}) \to (M_t^{(1,0)} f_E)(\mathbf{k}) = \tilde{\psi}_t(\mathbf{k}; f) \quad (3.5)$$

of $K^{(1)}$ into a Hilbert space $H^{(1)}$ of square-integrable functions $\tilde{\psi}(\mathbf{k})$, with an inner product

$$\langle \Psi_1 \mid \Psi_2 \rangle = \int d^3\mathbf{k} \tilde{\Psi}_1^*(\mathbf{k})\tilde{\Psi}_2(\mathbf{k}). \quad (3.6)$$

This mapping is bounded because

$$\langle M_t f \mid M_t f \rangle = \int d^3\mathbf{k} \, |\tilde{\psi}_t(\mathbf{k}; f)|^2 = \int \frac{d^3k}{(\mathbf{k}^2 + m^2)^{\frac{1}{2}}} |\tilde{f}_E(\mathbf{k})|^2 = (f \mid f), \quad (3.7)$$

and consequently its domain of definition can be extended in a unique fashion to the entire $\mathcal{H}^{(1)}$. Furthermore, it obviously maps $\mathcal{H}^{(1)}$ *onto* $H^{(1)}$.

A similar mapping of $\mathcal{H}^{(1)}$ onto $H^{(1)}$

$$\tilde{g}_E(\mathbf{k}) \to (M_t^{(0,1)} g_E)(\mathbf{k}) = \tilde{\psi}_t(\mathbf{k}; g), \quad g \in \mathcal{H}^{(1)} \quad (3.8)$$

can be introduced, where

$$\psi_t(\mathbf{k}; g) = (\mathbf{k} + m^2)^{-\frac{1}{4}} \exp[it(\mathbf{k} + m^2)^{\frac{1}{2}}]g_E(-\mathbf{k}), \quad (3.9)$$

for which the same remarks apply.

We can also map $\hat{\mathcal{H}}^{(m,n)}$ onto

$$\hat{H}^{(m+n)} = (H^{(1)})^{\otimes m+n} \quad (3.10)$$

by means of

$$M_t^{(m,n)} = (M_t^{(1,0)})^{\otimes m} \otimes (M_t^{(0,1)})^{\otimes n}. \quad (3.11)$$

Furthermore, if we define in $H^{(m,n)}$ a symmetrizer analogous to (2.10), such that

$$T^{(m,n)} = \frac{1}{m!\, n!} \sum_{\sigma \in G_m} \sum_{\tau \in G_n} [\sigma] \otimes [\tau], \quad (3.12)$$

[6] A. S. Wightman, Rev. Mod. Phys. 34, 845 (1962), Appendix I.

[7] P. R. Halmos, *Measure Theory* (D. Van Nostrand, Inc., Princeton, N.J., 1950).

where $[\sigma]$ acts on $\hat{H}^{(m)}$ and $[\tau]$ on $\hat{H}^{(n)}$, we can define

$$H^{(m,n)} = \{T^{(m,n)}\Psi, \Psi \in \hat{H}^{(m,n)}\} \tag{3.13}$$

and obviously have $\mathcal{K}^{(m,n)}$ mapped onto $H^{(m,n)}$ by $M^{(m,n)}$. Hence, if we take $H^{(0,0)} = \mathcal{K}^{(0,0)}$, the operator

$$M_t = \mathbf{1} \oplus M^{(1,0)} \oplus M^{(0,1)} \oplus \overset{\infty}{\underset{m,n=1}{\oplus}} M^{(m,n)} \tag{3.14}$$

provides a linear mapping of $\mathcal{K}$ onto H,

$$H = \underset{m,n=0}{\oplus} H^{(m,n)}, \tag{3.15}$$

which evidently has an inverse.

On $H^{(1)}$ we can define the "energy–momentum" operators

$$\begin{aligned}(P^j\psi)\tilde{}\,(\mathbf{k}) &= k^j\tilde{\psi}(\mathbf{k}), \quad j = 1, 2, 3,\\ (P^0\psi)\tilde{}\,(\mathbf{k}) &= (\mathbf{k}^2 + m^2)^{\frac{1}{2}}\tilde{\psi}(\mathbf{k}),\end{aligned} \tag{3.16}$$

which are obviously essentially self-adjoint, and hence their completions, when denoted by the same symbol, have the spectral decompositions

$$P^\nu = \int_{-\infty}^{-\infty} \lambda\, dF_\nu(I_\lambda), \quad I_\lambda = \{u: -\infty < u \leq \lambda\}. \tag{3.17}$$

For the Borel set B, the projectors $F_\nu(B)$, $\nu = 1, 2, 3$ have the particularly simple representation

$$\begin{aligned}(F_\nu(B)\psi)\tilde{}\,(\mathbf{k}) &= \chi_B(\mathbf{k})\tilde{\psi}(\mathbf{k}),\\ \chi_B(\mathbf{k}) &= \begin{cases}1, & k^\nu \in B,\\ 0, & k^\nu \notin B.\end{cases}\end{aligned} \tag{3.18}$$

We define the "position" operators X^j by first introducing the integral transforms

$$\psi(\mathbf{x}) = \frac{(2\pi)^{-\frac{3}{2}}}{\sqrt{2}} \int \frac{\exp i\mathbf{kx}}{(\mathbf{k}^2 + m^2)^{\frac{1}{4}}}\, \tilde{\psi}(\mathbf{k})\, d^3\mathbf{k} \tag{3.19}$$

and then defining

$$(X^j\psi)(\mathbf{x}) = x^j\psi(\mathbf{x}), \quad j = 1, 2, 3. \tag{3.20}$$

As the transform (3.19) has the inverse

$$\tilde{\psi}(\mathbf{k}) = \sqrt{2}\,(2\pi)^{-\frac{3}{2}} \int \exp(-i\mathbf{kx})(\mathbf{k}^2 + m^2)^{\frac{1}{4}}\psi(\mathbf{x})\, d^3\mathbf{x}, \tag{3.21}$$

(3.20) defines indeed a unique linear operator with everywhere dense domain in $H^{(1)}$, which can be seen to be essentially self-adjoint. Consequently, we have the spectral decompositions

$$X^j = \int_{-\infty}^{+\infty} \lambda\, dE_j(I_\lambda). \tag{3.22}$$

The reason for this particular choice of definition for the "position" operator becomes clear at the end of this section.

By standard methods we can build from the projection $E_j(S), j = 1, 2, 3$, and $F_\nu(S), \nu = 0, \cdots, 3$, the projection valued measures[6] $E^{(1)}(B)$ and $F^{(1)}(B)$ on $\mathbb{R}^3$ and $\mathbb{R}^4$ respectively, which are such that for $B = B_1 \times B_2 \times B_3$ and $B = B_0 \times \cdots \times B_3$, respectively, (where $B_0, \cdots, B_3$ are Borel sets on $\mathbb{R}^1$) they reduce to the products

$$\begin{aligned}E^{(1)}(B_1 \times B_2 \times B_3) &= E_1(B_1)E_2(B_2)E_3(B_3),\\ F^{(1)}(B_0 \times B_1 \times B_2 \times B_3) &= F_0(B_0)F_1(B_1)F_2(B_2)F_3(B_3).\end{aligned} \tag{3.23}$$

Denote in general by $\mathfrak{B}^n$ the family of Borel sets on $\mathbb{R}^n$. To Borel sets

$$\begin{aligned}\mathbf{R} &= R_1 \times \cdots \times R_n \in (\mathfrak{B}^3)^{\otimes n},\\ \mathbf{S} &= S_1 \times \cdots \times S_n \in (\mathfrak{B}^4)^{\otimes n},\end{aligned} \tag{3.24}$$

we can assign, respectively, the following projectors on $\hat{H}^{(n)}$:

$$\begin{aligned}\hat{E}^{(n)}(\mathbf{R}) &= E^{(1)}(R_1)\cdots E^{(n)}(R_n),\\ \hat{F}^{(n)}(\mathbf{S}) &= F^{(1)}(S_1)\cdots F^{(n)}(S_n).\end{aligned} \tag{3.25}$$

The above projectors do not leave, however, the space $H^{(n,0)}$ invariant. Consequently, we introduce the operators

$$\begin{aligned}&E_0^{(m,n)}(\mathbf{R}' \times \mathbf{R}'')\\ &= \frac{1}{m!\,n!} \sum_{\substack{(i_1,\cdots,i_m)\in G_m\\ (j_1,\cdots,j_n)\in G_n}} \hat{E}^{(m+n)}(R'_{i_1} \times \cdots \times R'_{i_m} \times R'_{i_m} \times R''_{j_1} \times \cdots \times R''_{j_n}),\\ &F_0^{(m,n)}(\mathbf{S}' \times \mathbf{S}'')\\ &= \frac{1}{m!\,n!} \sum_{\substack{(i_1,\cdots,i_m)\in G_m\\ (j_1,\cdots,j_n)\in G_n}} \hat{F}^{(m+n)}(S'_{i_1} \times \cdots \times S'_{i_m} \times S''_{j_1} \times \cdots \times S''_{j_n}),\end{aligned} \tag{3.26}$$

where the summation is taken over all permutations. These operators are not, in general, projectors, but they leave $H^{(m,n)}$ invariant.

Let $P^{(m,n)}$ denote the projector which projects a vector $\Psi \in H$ into its component $\Psi^{(m,n)}$ in $H^{(m,n)}$. Define

$$\begin{aligned}E^{(m,n)}(\mathbf{R}) &= E_0^{(m,n)}(\mathbf{R})P^{(m,n)}, \quad E^{(0,0)}(R) = P^{(0,0)},\\ F^{(m,n)}(\mathbf{S}) &= F_0^{(m,n)}(\mathbf{S})P^{(m,n)}, \quad F^{(0,0)}(S) = P^{(0,0)},\end{aligned} \tag{3.27}$$

and consider, for fixed Ψ, $\|\Psi\| = 1$, the family

$$\begin{aligned}&W_\Phi^{(m,n,k)}(\mathbf{R}_1, \mathbf{S}_1, \cdots, \mathbf{R}_k, \mathbf{S}_k; t)\\ &= \langle\Psi|\, E^{(m,n)}(\mathbf{R}_1)F^{(m,n)}(\mathbf{S}_1)\cdots E^{(m,n)}(\mathbf{R}_k)F^{(m,n)}(\mathbf{S}_k)\,|\Psi\rangle\\ &\qquad (k = 0, 1, 2, \cdots, m, n = 0, 1, 2, \cdots)\end{aligned} \tag{3.28}$$

of set functions. As is shown later in Theorem 2, Sec. 5, the family (3.28) provides a unique description of the Hilbert ray generated by Ψ. Consequently, due to the existence of the mapping (3.14) which has an inverse, it also provides a unique description of the ray generated by the corresponding vector

$$\Phi = M_t^{-1}\Psi \in \mathcal{H}, \tag{3.29}$$

i.e., of the pure physical state described by Φ.

The set functions (3.28) have $(\mathcal{B}^3)^{\otimes k(m+n)}$ as a domain of definition. From (3.23) and (3.26) we can immediately see that they are σ-additive[7] with respect to each of its arguments from $\mathcal{B}^3$. Since $\mathcal{B}^{3k(m+n)}$ is the Boolean σ-algebra generated by $(\mathcal{B}^3)^{\otimes k(m+n)}$, we can compute easily that each of the set functions (3.28) can be extended in a unique manner to a complex measure over $\mathcal{B}^{3k(m+n)}$.

For given m and n, the measures $W_\Phi^{(k)}$ are not independent of one another. As we can see immediately from (3.26) that

$$E_0^{(m,n)}(\mathbb{R}^{3(m+n)}) = F_0^{(m,n)}(\mathbb{R}^{4(m+n)}) = \mathbf{1}, \tag{3.30}$$

we have

$$W_\Phi^{(m,n,k)}(\mathbf{R}_1, \mathbf{S}_1, \cdots, \mathbf{R}_k, \mathbf{S}_k; t) = W_\Phi^{(m,n,k+1)}(\mathbf{R}_1, \mathbf{S}_1, \cdots, \mathbf{R}_k, \mathbf{S}_k, \mathbb{R}^{3(m+n)}, \mathbb{R}^{4(m+n)}; t). \tag{3.31}$$

Consider now

$$W_\Phi^{(m,n,k)}(\cdots, \mathbf{R}, \cdots; t) = \langle \Psi_1 | E^{(m,n)}(\mathbf{R}) | \Psi_2 \rangle, \tag{3.32}$$

where $\Psi_1, \Psi_2 \in H^{(m,n)}$ are obtained by applying on Ψ in (3.28), respectively, the adjoint of the operator preceding $E^{(m,n)}(\mathbf{R})$ and the operator succeeding $E^{(m,n)}(\mathbf{R})$.

In order to keep the notation simple, we take that $\Psi_\kappa^{(m,n)} = 0$, $\kappa = 1, 2$, for $n \neq 0$, i.e., that there are no antiparticles present in the considered states. However, the same considerations apply to the general case.

We can express $\Psi^{(m,0)}$ as functions in the $\mathbf{k}$-space

$$\tilde{\Psi}_\kappa^{(m,0)}(\mathbf{k}^m), \quad \mathbf{k}^m = (\mathbf{k}_1, \cdots, \mathbf{k}_m) \tag{3.33}$$

or in the $\mathbf{x}$-space

$$\Psi_\kappa^{(m,0)}(\mathbf{x}^m) = 2^{-m/2}(2\pi)^{-3m/2}\int \exp(i\mathbf{k}^m\mathbf{x}^m) \times \prod_{\nu=1}^m (\mathbf{k}^2 + m^2)^{-\frac{1}{4}}\tilde{\Psi}^{(m,0)}(\mathbf{k}^s)\, d\mathbf{k}^s,$$
$$\mathbf{x}^m = (\mathbf{x}_1, \cdots, \mathbf{x}_s), \quad \mathbf{k}^m\mathbf{x}^m = \mathbf{k}_1\mathbf{x}_1 + \cdots + \mathbf{k}_m\mathbf{x}_m. \tag{3.34}$$

By taking into account the symmetry properties of (3.33) and (3.34) under permutations of their arguments, we get from (3.26) and (3.27) that

$$\langle \Psi_1 | E^{(m,0)}(\mathbf{R}) | \Psi_2 \rangle = \langle \Psi_1^{(m,0)} | \hat{E}^{(m,0)}(\mathbf{R}) | \Psi_2^{(m,0)} \rangle = 2^{-m}(2\pi)^{-3m}\int d\mathbf{k}^m \int_{\mathbf{R}} d\mathbf{x}^m \int_{\mathbf{R}} d\mathbf{y}^m \exp[i\mathbf{k}^m(\mathbf{x}^m - \mathbf{y}^m)] \times \prod_{\nu=1}^m (\mathbf{k}_\nu^2 + m^2)^{\frac{1}{2}}\overline{\Psi_1^{(m,0)}(\mathbf{x}^m)}\Psi_2^{(m,0)}(\mathbf{y}^m). \tag{3.35}$$

From (3.4) and (3.19) we can easily derive that for given t

$$\Psi^{(m,n)}(\mathbf{x}^m, \mathbf{y}^n) = \Phi^{(m,n)}(\mathbf{x}^m, \mathbf{y}^n), \quad \Phi = M_t^{-1}\Psi, \tag{3.36}$$

where $\mathbf{x}^m$, when entering as argument in $\Phi^{(s,0)}(\mathbf{x}^m)$, stands for $\mathbf{x}_1, t, \cdots, \mathbf{x}_m, t$

$$\Phi^{(m,0)}(\mathbf{x}^m) = \Phi^{(m,0)}(\mathbf{x}_1, t, \cdots, \mathbf{x}_m, t); \tag{3.37}$$

and similarly for $\mathbf{y}^n$.

From (3.35), (3.36), and (3.4) we get

$$\langle \Psi_1 | E^{(m,0)}(\mathbf{R}) | \Psi_2 \rangle = i^m \int_{\mathbf{R}} d\mathbf{x}^m \overline{\Phi_1^{(m,0)}(\mathbf{x}^m)} \frac{\overleftrightarrow{\partial}}{\partial x_1^0} \cdots \frac{\overleftrightarrow{\partial}}{\partial x_m^0} \Phi_2^{(m,0)}(\mathbf{x}^m). \tag{3.38}$$

Similarly, for

$$W_\Phi^{(m,0,k)}(\cdots, \mathbf{S}, \cdots; t) = \langle \Psi_1 | F^{(m,0)}(\mathbf{S}) | \Psi_2 \rangle, \tag{3.39}$$

we get

$$\langle \Psi_1 | F^{(m,0)}(\mathbf{S}) | \Psi_2 \rangle = \int_{\mathbf{S}} d\mathbf{k}^m \left(\prod_{\nu=1}^m k_\nu^0\right)^{-1} \overline{\tilde{\Phi}_{1E}^{(m,0)}(\mathbf{k}^m)}\tilde{\Phi}_{2E}^{(m,0)}(\mathbf{k}^m), \quad k_\nu^0 = (\mathbf{k}_\nu^2 + m^2)^{\frac{1}{2}}, \tag{3.40}$$

where

$$\Phi^{(m,0)}(\mathbf{x}^m) = (2\pi)^{-3m/2} \times \int d\mathbf{k}^m \left(\prod_{\nu=1}^m k_\nu^2 + m^2\right)^{-\frac{1}{2}} \exp[i\mathbf{k}^m\mathbf{x}^m] \times \exp\left[-it\sum_{\nu=1}^m (\mathbf{k}_\nu^2 + m^2)^{\frac{1}{2}}\right]\tilde{\Phi}_E(\mathbf{k}^m). \tag{3.41}$$

4. MEASURE THEORETICAL DESCRIPTION OF STATES IN $\hat{H}$

We prove in this section the following theorem:

Theorem 1: A Hilbert space $\hat{H}_1^{(r)}$ with inner product $\langle \Psi_1 | \Psi_2 \rangle_1$ can be constructed from all the families of set functions

$$V_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k) = \langle \Psi | \hat{E}^{(r)}(\mathbf{R}_1) \cdots \hat{F}^{(r)}(\mathbf{S}_k) | \Psi \rangle \times \mathbf{R}_1, \cdots, \mathbf{R}_k \in (\mathcal{B}^3)^{\otimes r}, \quad \mathbf{S}_1, \cdots, \mathbf{S}_k \in (\mathcal{B}^4)^{\otimes r}, \quad k = 1, 2, \cdots, \tag{4.1}$$

corresponding to all $\Psi \in \hat{H}^{(r)}$, and projectors $\hat{E}^{(r)}(\mathbf{R})$ and $\hat{F}_1^{(r)}(\mathbf{S})$ can be assigned to each $\mathbf{R} \in (\mathfrak{B}^3)^{\otimes m+n}$ and $\mathbf{S} \in (\mathfrak{B}^4)^{\otimes m+n}$ in such a manner that for each $\Psi \in H^{(r)}$

$$V_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_n) = \langle \Psi' | \hat{E}_1^{(r)}(\mathbf{R}_1) \cdots \hat{F}_1^{(r)}(\mathbf{S}_k) | \Psi' \rangle \tag{4.2}$$

for some $\Psi' \in \hat{H}_1^{(r)}$. An isometric mapping $\mathfrak{U}$ of $\hat{H}^{(r)}$ onto $\hat{H}_1^{(r)}$ exists for which

$$\hat{E}_1^{(r)}(\mathbf{R}) = \mathfrak{U}\hat{E}^{(r)}(\mathbf{R})\mathfrak{U}^{-1}, \quad \hat{F}_1^{(r)}(\mathbf{S}) = \mathfrak{U}\hat{F}^{(r)}(\mathbf{S})\mathfrak{U}^{-1}. \tag{4.3}$$

We will establish the above theorem by demonstrating two lemmas. Since we consider in the next two lemmas a Hilbert space $\hat{H}^{(r)}$ for fixed r, we simplify the notation by replacing $\hat{H}^{(r)} = (H^{(1)})^{\otimes r}$ by $\hat{H}$, $\hat{E}^{(r)}(\mathbf{R})$ by $\hat{E}(\mathbf{R})$ and $\hat{F}^{(r)}(\mathbf{S})$ by $\hat{F}(\mathbf{S})$.

Lemma 1: For each $\Psi \in \hat{H}$, consider the family V_Ψ of all set functions

$$V_\Psi^{(k)}(\mathbf{R}_1 \times \mathbf{S}_1 \times \cdots \times \mathbf{R}_k \times \mathbf{S}_k) = \langle \Psi | \hat{E}(\mathbf{R}_1)\hat{F}(\mathbf{S}_1) \cdots \hat{E}(\mathbf{R}_k)\hat{F}(\mathbf{S}_k) | \Psi \rangle$$
$$\mathbf{R}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^3)^{\otimes r}, \quad k = 1, 2, \cdots, \tag{4.4}$$

where

$$\hat{E}(\mathbf{R}) = \hat{E}^1(R^{(1)}) \otimes \cdots \otimes \hat{E}^r(R^{(r)}),$$
$$\hat{F}(\mathbf{R}) = \hat{F}^1(R^{(1)}) \otimes \cdots \otimes \hat{F}^r(R^{(r)}), \tag{4.5}$$
$$\mathbf{R} = R^{(1)} \times \cdots \times R^{(r)} \in (\mathfrak{B}^3)^{\otimes r},$$

with $\hat{E}^i(R^{(i)})$ and $\hat{F}^i(R^{(i)})$ acting on the ith Hilbert space in the direct product $\hat{H} = (H^{(1)})^{\otimes r}$, and is explicitly defined by (3.23) with $R^{(i)} = B_1 \times B_2 \times B_3$ and $B_0 = \mathbb{R}^1$. From the family

$$\mathcal{S} = \{V_\Psi, \Psi \in \hat{H}\} \tag{4.6}$$

a Hilbert space H_1 with inner product $\langle \Psi_1 | \Psi_2 \rangle_1$ can be constructed, and projectors $\hat{E}_1(\mathbf{R})$ and $\hat{F}_1(\mathbf{R})$ acting on $\hat{H}_1$ can be attached to each $\mathbf{R} \in (\mathfrak{B}^3)^{\otimes r}$ in such a manner that for each Ψ we have

$$V_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k) = \langle \Psi' | \hat{E}(\mathbf{R}_1) \cdots \hat{F}_1(\mathbf{S}_k) | \Psi' \rangle_1,$$
$$\mathbf{R}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^3)^{\otimes r}, \quad k = 1, 2, \cdots, \tag{4.5}$$

for some vector $\Psi' \in \hat{H}_1$.

Proof: Consider the following $6r$ self-adjoint operators acting on $\hat{H}$ and defined by

$$X_j^{(i)} = \mathbf{1} \otimes \cdots \otimes X^j \otimes \cdots \otimes \mathbf{1},$$
$$P_j^{(i)} = \mathbf{1} \otimes \cdots \otimes P^j \otimes \cdots \otimes \mathbf{1},$$
$$i = 1, \cdots, r, \quad j = 1, 2, 3, \tag{4.6}$$

where the ith factors in the above direct products are the operators (3.22) and (3.17), respectively, acting on the ith $\hat{H}^{(1)}$ factor in the tensor product Hilbert space $\hat{H} = (H^{(1)})^{\otimes r}$. We shall consider these $6r$ operators, which constitute the set $\mathcal{O}_0$ of fundamental observables introduced in Ref. 1, Sec. 2. To each $\Psi \in \hat{H}$, $\|\Psi\| = 1$, we can assign a family P_Ψ of complex probability measures[1]

$$P_\Psi^{\alpha_1, \cdots \alpha_\nu}(B), \quad B \in \mathfrak{B}^\nu, \quad \alpha_1, \cdots, \alpha_\nu \in \mathcal{O}_0, \quad \nu = 1, 2, \cdots, \tag{4.7}$$

generated by the set function

$$P_\Psi^{\alpha_1; \cdots; \alpha_\nu}(B_1 \times \cdots \times B_\nu) = \langle \Psi | E_{\alpha_1}(B_1) \cdots E_{\alpha_\nu}(B_\nu) | \Psi \rangle,$$
$$B_1, \cdots, B_\nu \in \mathfrak{B}^1. \tag{4.8}$$

In the above expression, if α is the observable represented by the operator α, which can be any one of the $6r$ operators (4.6), then $E_\alpha(B)$, $B \in \mathfrak{B}^1$, is the spectral measure belonging to that operator

$$\alpha = \int_{-\infty}^{+\infty} \lambda \, dE_\alpha(I_\lambda), \quad I_\lambda = (-\infty, \lambda]. \tag{4.9}$$

The families (4.7) of all complex probability measures belonging to all normalized vectors from $\hat{H}$ obey (cf. Ref. 1, Sec. 2.2) all the ten axioms of Refs. 1 and 2. Consequently a Hilbert space $\hat{H}$ with an inner product $\langle \Psi_1 | \Psi_2 \rangle_1$ can be built as outlined in Ref. 2 (cf. Sec. 3.3, Theorem 3), and to each $\alpha \in \mathcal{O}_0$ and $B \in \mathfrak{B}^1$ a projector $E'_\alpha(B)$ on $\hat{H}_1$ can be assigned in such a manner that for each normalized $\Psi \in \hat{H}$

$$P_\Psi^{\alpha_1, \cdots \alpha_\nu}(B_1 \times \cdots \times B_\nu) = \langle \Psi' | E'_{\alpha_1}(B_1) \cdots E'_{\alpha_\nu}(B_\nu) | \Psi' \rangle$$
$$\alpha_1, \cdots, \alpha_\nu \in \mathcal{O}_0, \quad \nu = 1, 2, \cdots, \tag{4.10}$$

for some $\Psi' \in \hat{H}_1$.

On the other hand, we have

$$V_\Psi^{(k)}(\mathbf{R}_1 \times \mathbf{S}_1 \times \cdots \times \mathbf{R}_k \times \mathbf{S}_k) = P^{\hat{\alpha}_1; \hat{\alpha}_2; \cdots; \hat{\alpha}_{2k-1}; \hat{\alpha}_{2k}}(\mathbf{R}_1 \times \mathbf{S}_1 \times \cdots \times \mathbf{R}_k \times \mathbf{S}_k) \tag{4.11}$$

with

$$\hat{\alpha}_{2\mu-1} = (X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, \cdots, X_1^{(r)}, X_2^{(r)}, X_3^{(r)})$$
$$\hat{\alpha}_{2\mu} = (P_1^{(1)}, P_2^{(1)}, P_3^{(1)}, \cdots, P_1^{(r)}, P_2^{(r)}, P_3^{(r)}),$$
$$\mu = 1, \cdots k. \tag{4.12}$$

Owing to simple properties of complex probability measures (cf. Ref. 1, Axiom I), we can recover any of the set functions (4.8) from the set function (4.11), and in turn, according to well-known measure-theoretical theorems,[7] any of the measures in (4.7) can be recovered from the corresponding set function (4.8). This fact, in conjunction with the remarks leading to (4.10), establishes the validity of the lemma.

Lemma 2: There is a unitary mapping U of H onto H_1 which is such that for all $\alpha \in \mathcal{O}_0 \equiv \{X_1^{(1)}, \cdots, X_3^{(r)}, \cdots, P_3^{(r)}\}$,

$$E'_\alpha(B) = UE_\alpha(B)U^{-1}, \quad B \in \mathfrak{B}^1.$$

Proof: Consider the following operators acting on $H^{(1)}$ [cf. the notation in (3.17) and (3.27)]:

$$\Pi^j = P^j[(P^1)^2 + (P^2)^2 + (P^3)^2 + m^2]^{\frac{1}{2}} = \int_{-\infty}^{+\infty}\int\int k^j(\mathbf{k}^2 + m^2)^{\frac{1}{2}}\, dF^{(1)}(\mathbb{R}^1 \times I_{\mathbf{k}}), \quad j = 1, 2, 3, \tag{4.13}$$

$$I_{\mathbf{k}} = \{\boldsymbol{\lambda}: -\infty < \lambda^j \leq k^j, j = 1, 2, 3\}, \tag{4.14}$$

which are obviously (essentially) self-adjoint. When Π^j acts on a vector $\Psi \in H^{(1)}$ which can be represented by a wavefunction (3.19), we get

$$(\Pi^j\Psi)(\mathbf{x}) = -i(\partial\psi/\partial x_j). \tag{4.15}$$

Consequently we can derive

$$[X^j, \Pi^{j'}]\Psi = i\delta_{jj'}\Psi \tag{4.16}$$

for vectors Ψ from the common domain of definition of X^j and $P^{j'}$.

The corresponding operators

$$\Pi_j^{(i)} = P_j^{(i)}[(P_1^{(i)})^2 + (P_2^{(i)})^2 + (P_3^{(i)})^2 + m^2]^{\frac{1}{2}} \tag{4.17}$$

acting on $\hat{H}$ satisfy the commutation relations

$$[X_j^{(i)}, \Pi_{j'}^{(i')}]\Psi = i\delta_{ii'}\delta_{jj'}\Psi, \quad i, i' = 1, \cdots, r, \tag{4.18}$$

for Ψ from the common domain of definition of X^j and $P^{j'}$. The $6r$ operators $X_j^{(i)}$ and $\Pi_j^{(i)}$ provide a representation of the commutation relations, which has to be irreducible; in fact, if it were reducible, since we have that

$$P_j^{(i)} = \Pi_j^{(i)}[g(\Pi_1^{(i)2} + \Pi_2^{(i)2} + \Pi_3^{(i)2}) + m^2]^{-\frac{1}{2}}, \tag{4.19}$$

where $g(\eta)$ is the positive root of the cubic equation

$$\xi^3 + m^2\xi^2 - \eta = 0, \tag{4.20}$$

then we could conclude that $X_j^{(i)}$ and $P_j^{(i)}$ leave a nontrivial linear submanifold of $\hat{H}$ invariant, which is obviously false.

In $\hat{H}_1$ we can introduce the operators

$$\int_{-\infty}^{+\infty} \lambda\, dE'_\alpha(I_\lambda) = \begin{cases} X_j^{(i)}, & \alpha = X_j^{(i)}, \\ P_j^{(i)}, & \alpha = P_j^{(i)}, \end{cases}$$
$$\Pi_j'^{(i)} = P_j'^{(i)}[(P_j'^{(i)})^2 + (P_2'^{(i)})^2 + (P_3'^{(i)})^2 + m^2]^{\frac{1}{2}}. \tag{4.21}$$

For any normalized $\Psi \in \hat{H}$, write

$$\Psi_1 = E_{\alpha_\mu}(B'_\mu)\cdots E_{\alpha_1}(B'_1)\Psi,$$
$$\Psi_2 = E_{\beta_1}(B''_1)\cdots E_{\beta_\nu}(B''_\nu)\Psi,$$
$$\alpha_1, \cdots, \alpha_\mu, \quad \beta_1, \cdots, \beta_\nu \in \mathcal{O}_0. \tag{4.22}$$

If the above vectors are from the common domain of definition of $X_j^{(i)}$ and $\Pi_j^{(i)}$, then we obtain immediately from (4.10), after carrying out appropriate integrations, that the vectors

$$\Psi'_1 = E'_{\alpha_\mu}(B'_\mu)\cdots E'_{\alpha_1}(B'_1)\Psi',$$
$$\Psi'_2 = E'_{\beta_1}(B''_1)\cdots E'_{\beta_\nu}(B''_\nu)\Psi' \tag{4.23}$$

are from the common domain of $X_j'^{(i)}$ and $\Pi_j'^{(i)}$, and that

$$\langle\Psi'_1|\,[X_j'^{(i)}, \Pi_l'^{(k)}]\,|\Psi'_2\rangle_1 = \langle\Psi_1|\,[X_j^{(i)}, \Pi_l^{(k)}]\,|\Psi_2\rangle = i\delta_{ik}\delta_{jl}\langle\Psi_1\,|\,\Psi_2\rangle = i\delta_{ik}\delta_{jl}\langle\Psi'_1\,|\,\Psi'_2\rangle_1. \tag{4.24}$$

From the way $\hat{H}_1$ is constructed (cf. Ref. 2, Sec. 3.3, Theorem 3) we can infer that (4.24) can be true only if

$$[X_j'^{(i)}, \Pi_l'^{(k)}]\Psi' = i\delta_{ik}\delta_{jl}\Psi' \tag{4.25}$$

for vectors Ψ' from the common domain of definition of $X_j'^{(i)}$ and $P_l'^{(k)}$. Furthermore, we must have that (4.24) represents an irreducible representation of the commutation relations; otherwise, from the relations

$$\langle\Psi'_1|\,X_j'^{(i)}\,|\Psi'_2\rangle = \langle\Psi_1|\,X_j^{(i)}\,|\Psi_2\rangle,$$
$$\langle\Psi'_1|\,\Pi_j'^{(i)}\,|\Psi'_2\rangle = \langle\Psi_1|\,\Pi_j^{(i)}\,|\Psi_2\rangle, \tag{4.26}$$

which are readily derivable from (4.10), we could infer that (4.18) is a reducible representation of the commutation relations, which we have seen not to be true.

According to the well-known von Neumann's theorem,[8] there is a unitary operator U mapping $\hat{H}$ onto $\hat{H}_1$ for which

$$X_j'^{(i)} = UX_j^{(i)}U^{-1}, \quad \Pi_j'^{(i)} = U\Pi_j^{(i)}U^{-1}. \tag{4.27}$$

We also have, for any $B \in \mathfrak{B}^1$,

$$E'_\alpha(B) = UE_\alpha(B)U^{-1}, \quad \alpha \in \{X_1^{(1)}, X_2^{(1)}, \cdots, X_3^{(r)}\}, \tag{4.28}$$

and

$$A_j'^{(i)}(B) = UA_j^{(i)}(B)U^{-1}, \tag{4.29}$$

where $A_j^{(i)}(B)$ and $A_j'^{(i)}(B)$ are the spectral measures corresponding to the spectral decompositions of $\Pi_j^{(i)}$ and $\Pi_j'^{(i)}$, respectively:

$$\Pi_j^{(i)} = \int_{-\infty}^{+\infty} \lambda\, dA_j^{(i)}(I_\lambda), \quad \Pi_j'^{(i)} = \int_{-\infty}^{+\infty} \lambda\, dA_j'^{(i)}(I_\lambda). \tag{4.30}$$

In order to establish that (4.28) is also true for $\alpha \in \{P_1^{(1)}, \cdots P_3^{(r)}\}$, introduce the projectors $A^{(i)}(B)$ and $A'^{(i)}(B)$, $B \in \mathfrak{B}^3$, which for $B = B_1 \times B_2 \times B_3 \in (\mathfrak{B}^1)^{\otimes 3}$ become

$$A^{(i)}(B_1 \times B_2 \times B_3) = A_1^{(i)}(B_1)A_2^{(i)}(B_2)A_3^{(i)}(B_3),$$
$$A'^{(i)}(B_1 \times B_2 \times B_3) = A_1'^{(i)}(B_1)A_2'^{(i)}(B_2)A_3'^{(i)}(B_3). \tag{4.31}$$

[8] J. von Neumann, Math. Ann. **104**, 570 (1931).

It is easy to derive from (4.28) that

$$A'^{(i)}(B) = UA^{(i)}(B)U^{-1}, \quad B \in \mathfrak{B}^3. \tag{4.32}$$

On the other hand, we get that for $\alpha = P_j^{(i)}$, due to (4.19),

$$E_\alpha(B) = A^{(i)}[f_j^{-1}(B)], \quad E'_\alpha(B) = A'^{(i)}[f_j^{-1}(B)], \tag{4.33}$$

with

$$f_j(\mathbf{k}) = k^j[g(|\mathbf{k}|^2) + m^2]^{-\frac{1}{4}}, \quad j = 1, 2, 3. \tag{4.34}$$

From (4.32) and (4.33) it follows immediately that (4.28) is true for any $\alpha = P_j^{(i)}$, and consequently the lemma is proved.

5. MEASURE-THEORETICAL DESCRIPTION OF STATES IN H

We prove the following theorem, which has been mentioned at the end of Sec. 3.

Theorem 2: Consider any vector Ψ belonging to $H^{(m,n)}$, and the family of set functions

$$W_\Psi^{(k)}(\mathbf{R}_1 \times \mathbf{S}_1 \times \cdots \times \mathbf{R}_k \times \mathbf{S}_k) = \langle \Psi | E^{(m,n)}(\mathbf{R}_1)F^{(m,n)}(\mathbf{S}_1) \cdots E^{(m,n)}(\mathbf{R}_k)F^{(m,n)}(\mathbf{S}_k) | \Psi \rangle, \quad \mathbf{R}_1, \cdots, \mathbf{R}_k \in (\mathfrak{B}^3)^{\otimes m+n}, \quad \mathbf{S}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^4)^{\otimes m+n}. \tag{5.1}$$

If the family (5.1) is given for each $\Psi \in H^{(m,n)}$, then a Hilbert space H_1 with inner product $\langle \Psi' | \Psi'' \rangle_1$ can be constructed from this set of families of set functions, and to each $R \in (\mathfrak{B}^3)^{\otimes m+n}$ and $S \in (\mathfrak{B}^4)^{\otimes m+n}$ bounded operators $E_1(\mathbf{R})$ and $F_1(\mathbf{R})$ acting on H_1 can be assigned in such a manner that for given $\Psi \in H^{(m,n)}$ there is a $\Psi' \in H_1$, for which

$$W_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k) = \langle \Psi' | E_1(\mathbf{R}_1) \cdots F_1(\mathbf{S}_k) | \Psi' \rangle_1, \quad \mathbf{R}_1, \cdots, \mathbf{R}_k \in (\mathfrak{B}^3)^{\otimes m+n}, \quad \mathbf{S}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^4)^{\otimes m+n}, \quad k = 1, 2 \cdots. \tag{5.2}$$

Furthermore, if a vector $\Psi_1 \in H^{(m,n)}$ satisfies the equality

$$\langle \Psi_1 | E^{(m,n)}(\mathbf{R}_1) \cdots F^{(m,n)}(\mathbf{S}_k) | \Psi_1 \rangle = W_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k) \tag{5.3}$$

for all $\mathbf{R}_1, \cdots \mathbf{R}_k \in (\mathfrak{B}^3)^{\otimes m+n}$, $\mathbf{S}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^4)^{\otimes m+n}$ and all $k = 1, 2, \cdots$, it follows that

$$\Psi_1 = a\Psi, \quad |a| = 1, \tag{5.4}$$

for some complex number a.

In order to establish Theorem 2, we have to prove a result similar to Theorem 1, but for the set functions (5.1) rather than for (4.1). Now, the W functions (5.1) can be explicitly expressed, owing to (3.26), in terms of the V functions (4.1), but the converse is not true in general. However, another kind of relation can be established. In order to avoid complicating the notation, consider the case of the Hilbert space $H = H^{(m,0)}$ and $\hat{H} = \hat{H}^{(m,0)}$.

The domain of definition of the projector-valued set functions $E(\mathbf{R})$, $R \in (\mathfrak{B}^3)^{\otimes m}$, and $F(\mathbf{S})$, $S \in (\mathfrak{B}^4)^{\otimes m}$, can be extended[6] to arbitrary elements of $\mathfrak{B}^{3m}$ and $\mathfrak{B}^{4m}$, respectively, in such a manner that the resulting projector-valued set functions will be projector-valued measures. Accordingly, the domain of definition of $V_\Psi^{(k)}(B_1 \times B_2 \times \cdots \times B_{2k})$ can be extended in a unique manner to arbitrary $B_1, B_3, \cdots, B_{2k-1} \in \mathfrak{B}^{3m}$, $B_2, B_4, \cdots, B_{2k} \in \mathfrak{B}^{4m}$.

Denote by $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}$ the sets

$$\tilde{\mathbf{R}} = \bigcup_{(k_1, \cdots, k_n) \in G_n} R_{k_1} \times \cdots \times R_{k_n} \in \mathfrak{B}^{3n},$$
$$\tilde{\mathbf{S}} = \bigcup_{(k_1, \cdots, k_n) \in G_n} S_{k_i} \times \cdots \times S_{k_n} \in \mathfrak{B}^{4n}, \tag{5.5}$$

in case the $\mathbf{R}$ and $\mathbf{S}$ are the sets appearing in (3.24). By $\tilde{\mathfrak{B}}^{3n}$ and $\tilde{\mathfrak{B}}^{4n}$ we denote the Boolean σ-algebras generated by all the sets $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}$, respectively.

In the case that $\mathbf{R}$ and $\mathbf{S}$ are of the form

$$\mathbf{R} = R_1 \times \cdots \times R_m \in (\mathfrak{B}^3)^{\otimes m}, \quad R_i \cap R_j = \varnothing, \quad i \neq j,$$
$$\mathbf{S} = S_1 \times \cdots \times S_m \in (\mathfrak{B}^4)^{\otimes m}, \quad S_i \cap S_j = \varnothing, \quad i \neq j, \tag{5.6}$$

we see immediately from (3.26) that

$$E_0^{(m,0)}(\mathbf{R}) = \frac{1}{m!}\hat{E}^{(m)}(\tilde{\mathbf{R}}),$$
$$F_0^{(m,0)}(\mathbf{S}) = \frac{1}{m!}\hat{F}^{(m)}(\tilde{\mathbf{S}}). \tag{5.7}$$

Consequently we have, whenever $\mathbf{R}_1, \cdots, \mathbf{S}_k$ satisfy the condition (5.6), that

$$V_\Psi^{(k)}(\tilde{\mathbf{R}}_1 \times \tilde{\mathbf{S}}_1 \times \cdots \times \tilde{\mathbf{R}}_k \times \tilde{\mathbf{S}}_k) = (m!)^{2k} W_\Psi^{(k)}(\mathbf{R}_1 \times \mathbf{S}_1 \times \cdots \times \mathbf{R}_k \times \mathbf{S}_k). \tag{5.8}$$

On the other hand, we have the following lemma:

Lemma 3: From the set $\tilde{\mathfrak{S}}$ of all families $\tilde{V}_\Psi$, $\Psi \in H^{(m,0)}$, of set functions, where $\tilde{V}_\Psi$ for a given $\Psi \in H^{(m,0)}$ denotes the family of all set functions

$$V_\Psi^{(k)}(\tilde{\mathbf{R}}_1 \times \cdots \times \tilde{\mathbf{S}}_k), \quad \tilde{\mathbf{R}}_1, \cdots, \tilde{\mathbf{R}}_k \in \mathfrak{B}^{3m}, \quad \tilde{\mathbf{S}}_1, \cdots, \tilde{\mathbf{S}}_k \in \mathfrak{B}^{4m} \quad k = 1, 2, \cdots, \tag{5.9}$$

a Hilbert space $H_1^{(m,0)}$ with inner product $\langle \Psi' | \Psi'' \rangle$ can be constructed, and projectors $E_1(B')$, and $F_1(B''_1$ on $H_1^{(m,0)}$ can be attached to each $B' \in \mathfrak{B}^{3m}$ and each

$B'' \in \mathfrak{B}^{4m}$ in such a manner, that for each $\Psi \in H^{(m,0)}$ we have

$$V_\Psi^{(k)}(\tilde{\mathbf{R}}_1 \times \cdots \times \tilde{\mathbf{S}}_k) = \langle \Psi' | E_1(\tilde{\mathbf{R}}_1) \cdots F_1(\tilde{\mathbf{S}}_k) | \Psi' \rangle \tag{5.10}$$

for some vector $\Psi' \in H_1^{(m,0)}$. Furthermore, if another $\Psi'' \in H^{(m,0)}$ satisfies (5.10), then $\Psi'' = a\Psi'$ for some complex number a, $|a| = 1$.

Proof: We define a projector measure $\tilde{E}(B)$ on $\tilde{\mathfrak{B}}^{3m}$ by requiring that $\tilde{E}(B) = E(B)$ for $B \in \tilde{\mathfrak{B}}^{3m}$. In a similar manner, we can define $\tilde{F}(B)$ on $\mathfrak{B}^{4m}$. These projection-valued measures can be looked upon as the spectral decompositions of $3m$- and $4m$-dimensional[3] observables, respectively. The results of Refs. 1 and 2 have been generalized to such cases in Ref. 3. The existence of a Hilbert space H_1 which can be built from the complex probability measures

$$P_\Psi^{A_1;\cdots;A_s}(B_1 \times \cdots \times B_s) = \langle \Psi | A_1(B_1) \cdots A_s(B_s) | \Psi \rangle, \tag{5.11}$$

where A_i can stand for either $\tilde{E}$ or $\tilde{F}$, has been proven in Ref. 3, Sec. 5. The same construction yields the existence of the projectors $E_1(B')$ and $F_1(B'')$ for each $B' \in \mathfrak{B}^{3m}$ and $B'' \in \mathfrak{B}^{4m}$, which are such that

$$P^{A_1;\cdots;A_s}(B_1 \times \cdots \times B_s) = \langle \Psi' | A_1'(B_1) \cdots A_s'(B_s) | \Psi' \rangle \tag{5.12}$$

for some $\Psi \in H_1$, where A_k' stands for E_1 or F_1 when A_k stands for E or F, respectively. On the other hand, one can show that all the measures (5.11) are uniquely determined by the family of set functions (5.9); the proof of this statement parallels the proof of Lemma 1.

By following step by step the construction of H_1 as described in Ref. 3, Secs. 4 and 5, one can easily establish that there is a linear mapping

$$\Psi \to \Psi = L\Psi, \quad \Psi \in H, \quad \Psi \in H_1, \tag{5.13}$$

of H onto H_1, which satisfies (5.12) for every $\Psi \in H$. If some $\Psi \in H$ is mapped into the zero element of H_1, i.e., if $L\Psi_1 = 0$, then we have accordingly that

$$P_\Psi^{A_1;\cdots;A_s}(B_1 \times \cdots \times B_s) \equiv 0,$$
$$A_1, \cdots, A_s \in \{\hat{E}, \hat{F}\}, \quad s = 1, 2, \cdots, \tag{5.14}$$

and, consequently, that

$$V_{\Psi_1}^{(k)}(\tilde{\mathbf{R}}_1 \times \cdots \times \tilde{\mathbf{S}}_k) \equiv 0, \quad k = 1, 2, \cdots. \tag{5.15}$$

However, (5.15) implies that $\Psi_1 = \mathbf{0}$. Consequently, L has an inverse and the lemma is proved.

We can immediately infer from the above lemma that Theorem 2 is true by taking into consideration that (5.8) is true whenever $\mathbf{R}_1, \cdots, \mathbf{S}_k$ satisfy (5.6), and by employing the lemma proved above in combination with the following lemma.

Lemma 4: The set function $W_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k)$, $\Psi \in H^{(m,n)}$, is uniquely determined for arbitrary $\mathbf{R}_1, \cdots, \mathbf{R}_k \in (\mathfrak{B}^3)^{\otimes m+n}, \mathbf{S}_1, \cdots, \mathbf{S}_k \in (\mathfrak{B}^4)^{\otimes m+n}$, if its values on sets $\mathbf{R}_1, \cdots, \mathbf{S}_k$ obeying (5.6) are given.

Proof: In order to avoid unnecessarily involved notation, consider the case when $n = 0$, which possesses all the essential features of the general case. Write

$$W_\Psi^{(k)}(\mathbf{R}_1 \times \cdots \times \mathbf{S}_k) = \langle \Psi | E_0^{(m,0)}(\mathbf{R}_1) | \Psi_1 \rangle$$
$$\Psi_1 = F_0^{(m,0)}(\mathbf{S}_1) \cdots E_0^{(m,0)}(\mathbf{R}_k) F_0^{(m,0)}(\mathbf{S}_k)\Psi. \tag{5.16}$$

Consider the set function

$$Z(R^{(1)}, \cdots, R^{(m)}) = \langle \Psi | E_0^{(m,0)}(\mathbf{R}) | \Psi_1 \rangle \tag{5.17}$$

for arguments

$$\mathbf{R} = R^{(1)} \times \cdots \times R^{(m)}, \quad R^{(1)}, \cdots, R^{(m)} \in \mathfrak{B}^3. \tag{5.18}$$

We immediately infer from (3.26) that the set function (5.17) is σ-additive in each of its arguments $R^{(1)}, \cdots R^{(m)}$. Consequently, its values for any $\mathbf{R}$ in (5.18) are uniquely determined if its values on the family $\mathfrak{J}^3$ of intervals on $\mathbb{R}^3$

$$R^{(1)} = I^{(1)}, \cdots, R^{(m)} = I^{(m)} \in \mathfrak{J}^3, \tag{5.19}$$

are given, because $\mathfrak{B}^3$ is the Boolean σ-algebra generated by $\mathfrak{J}^3$.

For arbitrarily given $I^{(1)}, \cdots, I^{(m)} \in \mathfrak{J}^3$, we can split each of these intervals into

$$I^{(s)} = \bigcup_{i_s=1}^{N_s} J_{i_s}^{(s)}, \quad J_i^{(s)} \cap J_j^{(s)} = \varnothing, \quad i \neq j, \tag{5.20}$$

and consequently

$$Z(I^{(1)}, \cdots, I^{(m)})$$
$$= \sum_{i_1=1}^{N_1} \cdots \sum_{i_m=1}^{N_m} \langle \Psi | E^{(m,0)}(J_{i_1}^{(1)} \times \cdots \times J_{i_m}^{(m)}) | \Psi_1 \rangle. \tag{5.21}$$

We want to prove that in (5.21) the contributions of terms in which two or more $J_{i_s}^{(s)}$ have points in common can be made arbitrarily small in the limit of finer and finer partitions (5.20). Consider first the case of all contributions when $J_{i_1}^{(i)}$ and $J_{i_2}^{(2)}$ have common points for certain values of i_1 and i_2. If we express $E_0^{(m,0)}$ in terms of $E^{(m,0)}$ by means of (3.26), the first term corresponding to the identity permutation of the indices $i_1, \cdots, i_m$ is

$$\frac{1}{m!} \sum_{i_1=1}^{N_1} \cdots \sum_{i_m=1}^{N_m} \langle \Psi | E^{(m,0)}(J_{i_1}^{(1)} \times \cdots \times J_{i_m}^{(m)}) | \Psi_1 \rangle$$
$$= \frac{1}{m!} \sum_{i_1=1}^{N_1} \cdots \sum_{i_m=1}^{N_m} \langle \Psi | E_1(J_{i_1}^{(1)}) \cdots E_m(J_{i_m}^{(m)}) | \Psi_1 \rangle, \tag{5.22}$$

where, if we follow the notation introduced in (4.9),

$$E_i(B_1 \times B_2 \times B_3) = E_{\alpha_1}(B_1)E_{\alpha_2}(B_2)E_{\alpha_3}(B_3),$$
$$\alpha_1 = X_1^{(i)}, \quad \alpha_2 = X_2^{(i)}, \quad \alpha_3 = X_3^{(k)},$$
$$B_1, B_2, B_3 \in \mathcal{B}^1. \quad (5.23)$$

The contribution to (5.22) of terms for which $J_{i_1}^{(1)}$ and $J_{i_2}^{(2)}$ have points in common can be written as

$$\frac{1}{m!}\sum_{i_1,i_2}\langle\Psi| E_1(J_{i_1}^{(1)})E_2(J_{i_2}^{(2)}) \times \sum_{i_3=1}^{N_3}\cdots\sum_{i_m=1}^{N_m} E_3(J_{i_3}^{(3)})\cdots E_m(J_{i_m}^{(m)})|\Psi_1\rangle,$$
$$(i_1, i_2) \in \{(r, s): J_r^{(1)} \cap J_s^{(2)} \neq \varnothing\}. \quad (5.24)$$

With the help of the Schwarz–Cauchy inequality, we derive that the absolute value of (5.24) does not exceed

$$\frac{1}{m!}\sum_{i=1}^{N_1}\|E_1(J_i^{(1)})\Psi\| \cdot \|E_2(K_i)\Psi_2\|, \quad (5.25)$$

where

$$K_i = \bigcup_{i_2 \in \mathcal{F}_i} J_{i_2}^{(2)}, \quad \mathcal{F}_i = \{i_2: J_i^{(1)} \cap J_{i_2}^{(2)} \neq \varnothing\},$$
$$\Psi_2 = E_3(I_3)\cdots E_m(I_m)\Psi_1. \quad (5.26)$$

As we have, due to the mutual orthogonality of the vectors $E_1(J_1^{(1)})\Psi, \cdots, E_1(J_{N_1}^{(1)})\Psi$, that

$$\sum_{i=1}^{N_1}\|E_1(J_i^{(1)})\Psi\| \leq \|\Psi\|^{-1}\sum_{i=1}^{N_1}\|E_1(J_i^{(1)})\Psi\|^2 = \|\Psi\|^{-1}\|E_1(I_1)\Psi\|^2, \quad (5.27)$$

we see that (5.24) is certainly smaller than

$$\frac{1}{m!}\|\Psi\|\,\|E_2(K_{i_0})\Psi_2\|,$$
$$\|E_2(K_{i_0})\Psi_2\| = \sup_{i=1,\cdots,N_1}\|E_2(K_i)\Psi_2\|. \quad (5.28)$$

The above expression becomes arbitrarily small in the limit of finer and finer partitions of I_1 and I_2; namely, in that limit the diameter of K_{i_0} will converge to zero, and consequently $\|E_2(K_{i_0})\Psi_2\|$ goes to zero due to the fact that $X_1^{(2)}$, $X_2^{(2)}$, and $X_3^{(2)}$ have a continuous spectrum.

There is a finite number of pairs (S_1, S_2), $S_1 > S_2$, $S_1, S_2 = 1, \cdots m$. The contribution of terms in the sum (5.22) corresponding to sets $J_i^{(S_1)}$ and $J_l^{(S_2)}$ with points in common can be estimated by the same procedure which led to (5.28). This can be done in a number of successive steps, corresponding to an arbitrarily chosen sequence of pairs (S_1, S_2), $S_1 > S_2$, which starts with the already treated case $S_1 = 1$, $S_2 = 2$. At each step (S_1, S_2), at the stage (5.24), i.e., when we estimate

$$\frac{1}{m!}\sum_{i_{s1},i_{s2}}\langle\Psi| E_{S_1}(J_{i_{s_1}}^{(S_1)})E_{S_2}(J_{i_{s_2}}^{(S_2)}) \times \sum_{i_1}'\cdots\sum_{i_m}' E_1(J_1^{(1)})\cdots E_m(J_{i_m}^{(m)})|\Psi_1\rangle \quad (5.29)$$

(where the primes in $\sum'$ denote that the summation over i_S is not taken when $S = S_1$ or $S = S_2$), the summation over i_S ($S \neq S_1$, $S \neq S_2$) is not in general from 1 to N_s because some of those terms might have been taken into account at an earlier stage. Thus we arrive at the upper bound

$$\frac{1}{m!}\sum_{i=1}^{N_{s_1}}\|E_{s_1}(J_i^{(S_1)})\Psi\| \cdot \|E_{s_1}(K_i^{(S_2)})\Psi_{s_2}'\|, \quad (5.30)$$

with

$$\Psi_{S_2}' = \sum_{i_1}'\cdots\sum_{i_m}' E_1(J_{i_1}^{(1)})\cdots E_m(J_{i_m}^{(m)})\Psi_1 \quad (5.31)$$

for the absolute value of (5.29). As it is easy to see that

$$\|E_{s_2}\Psi_{s_2}'\|^2 \leq \sum_{i_1=1}^{N_1}{}'\cdots\sum_{i_m=1}^{N_m}{}'\|E_{s_2}E_1(J_{i_1}^{(1)})\cdots E_m(J_{i_m})\Psi_1\|^2 = \|E_{s_2}\Psi_{s_2}'\|^2, \quad (5.32)$$

where

$$\Psi_{s_2}' = \sum_{i_1=1}^{N_1}{}'\cdots\sum_{i_m=1}^{N_m}{}' E_1(J_{i_1}^{(1)})\cdots E_m(J_{i_m}^{(m)})\Psi_1, \quad (5.33)$$

we arrive at an upper bound of the absolute value of (5.29) which is analogous to (5.28). Thus we come to the same final conclusion; namely that (5.29) vanishes in the limit of finer and finer partitions of I_{s_1} and I_{s_2}.

Thus we arrive at the conclusion that

$$\langle\Psi|\hat{E}^{(m,0)}(I^{(1)} \times \cdots \times I^{(m)})|\Psi_1\rangle = \lim \sum_{i_1,\cdots i_m}'\langle\Psi|\hat{E}^{(m,0)}(J_{i_1}^{(1)} \times \cdots \times J_{i_m}^{(m)})|\Psi_1\rangle, \quad (5.34)$$

where the prime in the above $\sum'$ indicates that the summation is taken only over those m-tuples $i_1, \cdots i_m$ for which the intervals $J_{i_1}^{(1)}, \cdots J_{i_m}^{(m)}$ are disjoint, and the limit is for finer and finer partitions (5.20). As we have that

$$Z(I^{(1)}, \cdots, I^{(m)}) = \frac{1}{m!}\sum_{(r_1,\cdots,r_m)\in G}\langle\Psi|\hat{E}^{(m,0)}(I^{(r)} \times \cdots \times I^{(r_m)})|\Psi_1\rangle, \quad (5.35)$$

the same conclusion (5.34) applies to each term in the above sum. Consequently, we have

$$Z(I^{(1)}, \cdots, I^{(m)}) = \lim \sum_{i_1,\cdots,i_m}'\langle\Psi|E^{(m,0)}(J_{i_1}^{(1)})\cdots E^{(m,0)}(J_{i_m}^{(m)})|\Psi_1\rangle, \quad (5.36)$$

where the lim and $\sum'$ have the same meaning as in (5.34).

The validity of the lemma now follows from the fact that the same procedure which has been applied to $E_0^{(m,0)}(\mathbf{R}_1)$ in $W_\Psi^{(k)}$ can be applied in the same form to $F_0^{(m,0)}(\mathbf{S}_1), \cdots, E_0^{(m,0)}(\mathbf{R}_k), F_0^{(m,0)}(\mathbf{S}_k)$. Thus the lemma is proved.

ACKNOWLEDGMENTS

The author would like to express his gratitude to the Theoretical Physics Institute of the University of Alberta in Edmonton for the award of a fellowship, and to thank Professor A. B. Bhatia for the kind hospitality extended to him at the Institute.

APPENDIX: FREE FIELDS AT A POINT

The Fock-space method, as originally introduced[9] and essentially used since the earliest papers on quantum field theory,[10] employs the concept of free field defined at a point. Such fields, though referred to as linear operators, are not at all operators on the Hilbert space of states with arbitrary number of particles, because they certainly lead outside that space when applied to any vectors. This can be easily seen from the following:

The standard way of introducing free fields is by first choosing an orthonormal basis

$$\begin{aligned} &f_1, f_2, \cdots, \quad (f_1 \,|\, f_j) = \delta_{ij}, \\ &g_1, g_2, \cdots, \quad (g_i \,|\, g_j) = \delta_{ij}, \\ &\tilde{f}_i(\mathbf{k}, (\mathbf{k}^2 + m^2)^{\frac{1}{2}}) = \tilde{g}_i(-\mathbf{k}, -(\mathbf{k}^2 + m^2)^{\frac{1}{2}}) \end{aligned} \tag{A1}$$

in $\mathfrak{H}^{(1)}$ and $\bar{\mathfrak{H}}^{(1)}$, respectively, and then a corresponding basis in $\mathfrak{H}$:

$$\Phi_{\substack{\{m_1,m_2,\cdots\} \\ \{n_1,n_2,\cdots\}}} = S^{(m,n)} f_{i_1} \otimes \cdots \otimes f_{i_m} \otimes g_{j_1} \otimes \cdots \otimes g_{j_n}, \tag{A2}$$

where among the indices $i_1, \cdots, i_m$ and $j_1, \cdots, j_m$ the index r appears m_r and n_r times, respectively. If we denote with a_r and a_r^* (b_r and b_r^*) the annihilation and creation operators of a particle (antiparticle) in the rth state of (A1), then we can write the field "operators" as

$$\begin{aligned} &\varphi(x) = \varphi_p(x) + \varphi_a^*(x), \quad \varphi^*(x) = \varphi_p^*(x) + \varphi_a(x), \\ &\varphi_p(x) = \sum_{r=1}^{\infty} f_r(x) a_r, \qquad \varphi_p^*(x) = \sum_{r=1}^{\infty} f_r^*(x) a_r^*, \\ &\varphi_a(x) = \sum_{r=1}^{\infty} g_r(x) b_r, \qquad \varphi_a^*(x) = \sum_{r=1}^{\infty} g_r^*(x) b_r^*. \end{aligned} \tag{A3}$$

Note that a_r, a_r^*, b_r and b_r^*, as well as $\varphi_p(x)$ and $\varphi_a(x)$, are indeed linear operators with the same domain of definition which includes the open linear manifold spanned by all the vectors of the form (A2), and which is everywhere dense in $\mathfrak{H}$. On the other hand, $\varphi_p^*(x)$ and $\varphi_a^*(x)$ are *not* taking any vector, except the zero vector, into an element of $\mathfrak{H}$, and consequently they are *not* operators in the accepted sense of the word. However, when appearing in matrix elements, they can be adequately treated in a formal manner so as to give meaningful results—a fact which is the basis of any perturbation field-theoretical calculation.

As is well known, the remedy leading to meaningful operators is simple and consists in the "smearing" of $\varphi(x)$ with some function $h(x)$ from an appropriately chosen space of functions,[11] i.e., in considering the operators

$$\varphi(h) = \sum_{r=1}^{\infty} (f_r^* \,|\, h) a_r + (g_r \,|\, h) b_r^* \tag{A4}$$

for which, e.g.,

$$\|\varphi_a^*(h)\Omega\|^2 = \sum_{i=1}^{\infty} |(g_i \,|\, h)|^2 = \|h\|^2 < +\infty. \tag{A5}$$

However, in conventional quantum field theory the considered interactions are of the local type and expressed in terms of products of field "operators" at a point. In investigating the possibility of introducing relativistic quantum fields at a point, the natural starting point could be to try first to formulate a mathematically meaningful concept of free field at a point, which would share with the conventional concept the formal features which are derivable on physical grounds. Since according to the earlier remark such a free field could not possibly be a linear operator on a Hilbert space, there seem to be two main types of approaches to the problem.

One possibility is to consider as the space of all pure physical states a linear space which would be "larger" than a Hilbert space, but on which it would still be possible to retain (with, perhaps, straightforward generalizations) the conventional physical interpretation.

The nested Hilbert spaces which have been suggested[12] some time ago are providing a structure of this type and have been recently used in order to define fields as operators on such spaces.[13]

We now apply (unfortunately with only partial success) the results of this paper to the second alternative, in which one keeps essentially a space of pure physical states which is equivalent to a Hilbert space, but drops the requirement that a free field be a linear operator. The main justification for such an approach is that in all the calculations in which the formalism of free fields is used, one is never concerned with the

[9] V. Fock, Z. Physik **75**, 622 (1932).
[10] P. M. Dirac, Proc. Roy. Soc. (London) **A114**, 431 (1927).
[11] A. S. Wightman and L. Gårding, Arkiv Fysik **28**, 129 (1964).
[12] A. Grossmann, J. Math. Phys. **5**, 1025 (1964); Commun. Math. Phys. **2**, 1 (1966).
[13] A. Grossmann, Commun. Math. Phys. **3**, 203 (1907).

field "operators" themselves, but rather with matrix elements of products of such "operators." These matrix elements are defined by formal manipulations designed to give a meaning to expressions which, strictly speaking, are meaningless as long as the fields are really considered to be operators. Hence the formalism is built in terms of "expectation values" of products of fields taken at *points* of space-time.

First we formally derive some relations for these "matrix elements."

By using the fact[5] that we can decompose the symmetrizer (2.10) in the following fashion:

$$S^{(m,n)} = \frac{1}{m!} S^{(m-1,n)}_{\{1,\cdots,Z-1,Z+1,\cdots,m\}} \sum_{Z=1}^{m} [1, Z], \quad \text{(A6)}$$

where $S^{(m-1,n)}_{\{1,\cdots\}}$ is again a symmetrizer given by the formula (2.10) but acting only in the indices appearing in the curly brackets, while $[1, Z]$ is an operator which transposes the indices 1 and Z, we can easily derive that

$$\begin{aligned}
&(\varphi_p(x)\Phi_{\substack{\{m_1,\cdots,m_r,\cdots\}\\ \{n_1,\cdots,n_r,\cdots\}}})^{(m,n)}(x_1,\cdots,x_m;\cdots,x_{m+n}) \\
&= \sum_{r=1}^{\infty} (m_r)^{\frac{1}{2}} f_r(x) \Phi^{(m,n)}_{\substack{\{m_1,\cdots,m_r-1,\cdots\}\\ \{n_1,\cdots,n_r,\cdots\}}}(x_1,\cdots,x_m;\cdots,x_{m+n}) \\
&= (m+1)^{\frac{1}{2}} \Phi^{(m+1,n)}_{\substack{\{m_1,\cdots,m_r,\cdots\}\\ \{n_1,\cdots,n_r,\cdots\}}}(x, x_1,\cdots,x_m;\cdots,x_{m+n}).
\end{aligned} \quad \text{(A7)}$$

A similar formula holds for $\varphi_a(x)$.

With the help of (A5), it can be immediately derived that more generally for $r = 1, 2, \cdots$ and any $\Phi_1, \Phi_2 \in K$:

$$\begin{aligned}
&\langle\Phi_1| :\varphi^*_p(x_1)\varphi_p(x_1)\cdots\varphi^*_p(x_r)\varphi_p(x_r): |\Phi_2\rangle \\
&= \sum_{m=0}^{\infty} i^{m+n} \binom{m+r}{r} \\
&\quad \times \int \Phi_1^{*(m+r,n)}(x_1,\cdots,x_r,y_1,\cdots,y_m,\cdots,y_{m+n}) \\
&\quad \times \frac{\overleftrightarrow{\partial}}{\partial y_1^0}\cdots\frac{\overleftrightarrow{\partial}}{\partial y_1^0}\cdots\Phi_2^{(m+r,n)} \\
&\quad \times (x_1,\cdots,x_2,y_1,\cdots,y_m,\cdots,y_{m+n}) \\
&\quad \times d^3\mathbf{y}_1\cdots d^3\mathbf{y}_m\cdots d^3\mathbf{y}_{m+n}.
\end{aligned} \quad \text{(A8)}$$

A similar expression holds for the antiparticle fields $\varphi_a(x)$ and $\lambda^*_a(x)$.

We desire to replace the description of a pure state, given in terms of the families (3.28), and more conventially by the vector Φ in (3.29), by a description in terms of set functions assigned to quantities which we could call "fields." Such quantities are required to justify the name of "field" given to them by being related in simple manner to the field "operators" (A7).

We assign to each space-region (open set in $\mathbb{R}^3$) Δ and instant t a field $\rho(\Delta, t)$ by relating to $\rho(\Delta, t)$ the measure on $\mathbb{R}^1$:

$$\begin{aligned}
&N^{(t)}_\Phi(\rho(\Delta,t) \mid B) \\
&= \begin{cases} \sum_0^\infty \langle\Psi| E^{(i+1,0)}(\Delta\times\mathbb{R}^{3i})|\Psi\rangle, & 1\in B,\ -1\notin B, \\ 1, & -1\notin B, \\ \sum_{j=0}^\infty \langle\Psi| E^{(0,j+1)}(\Delta\times\mathbb{R}^{3j})|\Psi\rangle, & 1\notin B,\ -1\in B, \end{cases}
\end{aligned} \quad \text{(A9)}$$

which is obviously concentrated at the points 1 and -1. More generally, we define

$$\begin{aligned}
&N_\Psi(\rho(\Delta_1,t),\cdots,\rho(\Delta_r,t) \mid B_1\times\cdots\times B_r) \\
&= \sum_{m,n=0}^{\infty} \sum_{k=0}^{r} \sum_{(i_1,\cdots i_n)\in G_n} \theta(B_{i_1},\cdots,B_{i_k};B_{i_{k+1}},\cdots,B_{i_n}) \\
&\quad \times \langle\Psi| E^{(m+k,n+r-k)}(\Delta_{i_1}\times\cdots\times\Delta_{i_k} \\
&\quad \times \mathbb{R}^{3m}\times\Delta_{i_{k+1}}\times\cdots\times\Delta_{i_n}\times\mathbb{R}^{3n})|\Psi\rangle,
\end{aligned} \quad \text{(A10)}$$

where $B_1, \cdots, B_r$ are Borel sets in $\mathbb{R}^1$ and

$$\begin{aligned}
&\theta(B_1,\cdots,B_k;B_{k+1},\cdots,B_r) \\
&= \begin{cases} \binom{m+k}{k}\binom{n+r-k}{r-k}, & 1\in B_1\cap\cdots\cap B_k, \\ & -1\in B_{k+1}\cap\cdots\cap B_r, \\ 0, & 1\notin B_1\cap\cdots\cap B_k \\ & \text{or}\quad -1\notin B_{k+1}\cap\cdots\cap B_r. \end{cases}
\end{aligned} \quad \text{(A11)}$$

We can easily derive the necessary and sufficient condition for the convergence of the series (A8) for all choices of $\Delta_1, \cdots, \Delta_r$, $B_1, \cdots, B_r$, $r = 1, 2, \cdots$. In order to do that, note that for any $\Psi \in H^{(m,n)}$ we get from (3.26) and (3.27)

$$|\langle\Psi| E^{(m,n)}(\mathbf{R})|\Psi\rangle| \le \|\Psi\|, \quad \Psi\in H^{(m,n)}. \quad \text{(A12)}$$

In particular, we have

$$E^{(m,n)}(\mathbb{R}^{3(m+n)}) = \mathbf{1} \quad \text{(A13)}$$

and, consequently, in that case the equality sign is valid in (A12).

Any element Ψ of H can be decomposed in the form

$$\Psi = \sum_{m,n=0}^{\infty} C_{mn}\Psi^{(m,n)}, \quad \Psi^{(m,n)}\in H^{(m,n)},$$
$$\|\Psi^{(m,n)}\| = 1, \quad \text{(A14)}$$

and we have

$$\sum_{m,n=0}^{\infty} |C_{mn}|^2 = \langle\Psi \mid \Psi\rangle. \quad \text{(A15)}$$

From (A12) and (A13), we can immediately infer that the above-mentioned condition is

$$\sum_{m,n=0}^{\infty}\binom{m+k}{k}\binom{n+k-k}{r-k}|C_{mn}|^2<+\infty,$$
$$k=0,\cdots,r,\quad n=1,2,\cdots. \tag{A16}$$

Therefore, the definition (A10) is not meaningful for all $\Psi\in H$. However, this definition is certainly meaningful for all Ψ from the linear manifold

$$\mathfrak{L}=\bigcup_{m,n=0}^{\infty}\left(\bigoplus_{i=0}^{m}\bigoplus_{j=0}^{n}H^{(i,j)}\right), \tag{A17}$$

which is everywhere dense in H.

We note that for each $\Psi\in\mathfrak{L}$ we can recover from (A10) all the set functions

$$\langle\Psi|\,E^{(m,n)}(\Delta)\,|\Psi\rangle,\quad m,n=0,1,2,\cdots, \tag{A18}$$

for all the argument values

$$\Delta=\Delta_1\times\cdots\times\Delta_{m+n}, \tag{A19}$$

where $\Delta_1,\cdots,$ are regions. This we achieve by solving the equations

$$\begin{aligned}&N_\Psi(\rho(\Delta_1,t),\cdots\rho(\Delta_r,t),\cdots\rho(\Delta_{r+s},t)\,|\,\{1\}\\&\quad\times\cdots\times\{1\}\times\cdots\times\{-1\})\\&=\sum_{i,j=0}^{\infty}\binom{i+r}{r}\binom{j+s}{s}\langle\Psi|\,E^{(j+r,j+s)}(\Delta_1\times\cdots\times\Delta_r\\&\quad\times\mathbb{R}^{3i}\times\Delta_{r+1}\times\cdots\times\Delta_{r+s}\times\mathbb{R}^{3s})\,|\Psi\rangle,\end{aligned} \tag{A20}$$

which follow directly from (A10).

Now, the set functions (3.28) are σ-additive in each of the separate arguments with domain of values $\mathfrak{B}^3$ entering in the $\mathbf{R}$'s and $\mathbf{S}$'s according to (3.24). Since the family of all open regions in $\mathbb{R}^3$ generates a Boolean σ-algebra which is identical with $\mathfrak{B}^3$, we can conclude on basis of well-known measure-theoretical theorems[7] that the set functions (A18) defined only on sets (A19) uniquely determine the set functions

$$\langle\Psi|\,E^{(m,n)}(\mathbf{R})\,|\Psi\rangle,\quad \mathbf{R}\in(\mathfrak{B}^3)^{\otimes(m+n)},$$
$$m,n=0,1,2,\cdots. \tag{A21}$$

However, these are only the set functions of the lowest order from the family (3.28) necessary for a unique description of a physical state. In order to enlarge accordingly the family (A10), we will introduce in addition to the fields $\rho(\Delta)$ also the global energy momentum $p=(p^0,p^1,p^2,p^3)$.

To make the notation more compact, define

$$\begin{aligned}&\boldsymbol{\rho}^r(\boldsymbol{\Delta}^r,t)=\rho(\Delta_1,t),\cdots,\rho(\Delta_r,t),\\&\boldsymbol{\Delta}^r=\Delta_1\times\cdots\times\Delta_r,\\&\mathbf{p}^r=(p,\cdots,p)\ (r\text{ times}),\\&\mathbf{B}^r=B_1\times\cdots\times B_r.\end{aligned} \tag{A22}$$

Then after suppressing t which is common to all fields, we write

$$\begin{aligned}&N_\Psi(\boldsymbol{\rho}^r(\boldsymbol{\Delta}^r),\boldsymbol{\rho}^s(\boldsymbol{\Delta}^s);\mathbf{p}^{r+s}\,|\,\{\mathbf{1}\}^r\times\{-\mathbf{1}\}^s\times\mathbf{B}^r\times\mathbf{B}^s)\\&=\sum_{m,n=0}^{\infty}\binom{m+r}{r}\binom{n+s}{s}\\&\quad\times\langle\Psi|\,E^{(m+r,n+s)}(\boldsymbol{\Delta}^r\times\mathbb{R}^{3m}\times\boldsymbol{\Delta}^s\times\mathbb{R}^{3n})\\&\quad\times F^{(m+r,n+s)}(\mathbf{B}^r\times\mathbb{R}^{3m}\times\mathbf{B}^s\times\mathbb{R}^{3n})\,|\Psi\rangle.\end{aligned} \tag{A23}$$

By the same argument as the one applied earlier on (A20), we conclude that for $\Psi\in\mathfrak{L}$ we can recover all the set functions

$$\langle\Psi|\,E^{(m,n)}(\mathbf{R})F^{(m,n)}(\mathbf{S})\,|\Psi\rangle,\quad \mathbf{R},\mathbf{S}\in(\mathfrak{B}^3)^{\otimes(m+n)},$$
$$m,n=0,1,2,\cdots. \tag{A24}$$

In general, we define

$$\begin{aligned}&N_\Psi(\boldsymbol{\rho}^{r_1}(\boldsymbol{\Delta}^{r_1}),\boldsymbol{\rho}^{s_1}(\boldsymbol{\Delta}^{s_1});\mathbf{p}^{r_1+s_1};\cdots;\boldsymbol{\rho}^{r_k}(\boldsymbol{\Delta}^{r_k}),\boldsymbol{\rho}^{s_k}(\boldsymbol{\Delta}^{s_k});\\&\qquad\mathbf{p}^{r_k+s_k}\,|\,\{\mathbf{1}\}^{r_1}\times\{-\mathbf{1}\}^{s_1}\times\mathbf{B}^{r_1}\times\mathbf{B}^{s_1}\\&\qquad\quad\times\cdots\times\{\mathbf{1}\}^{r_k}\times\{-\mathbf{1}\}^{s_k}\times\mathbf{B}^{r_k}\times\mathbf{B}^{s_k})\\&=\sum_{m_1,n_1=0}^{\infty}\cdots\sum_{m_k,n_k=0}^{\infty}\binom{m_1+r_1}{r_1}\binom{n_1+s_1}{s_1}\cdots\\&\quad\times\langle\Psi|\,E^{(m_1+r_1,n_1+s_1)}(\boldsymbol{\Delta}^{r_1}\times\mathbb{R}^{3m_1}\times\boldsymbol{\Delta}^{s_1}\times\mathbb{R}^{3n_1})\\&\quad\times F^{(m_1+r_1,n_1+s_1)}(\mathbf{B}_1^{r_1}\times\mathbb{R}^{3m_1}\times\mathbf{B}^{s_1}\times\mathbb{R}^{3n_1})\cdots|\Psi\rangle\end{aligned} \tag{A25}$$

and deduce that from the above set functions we can recover any of the set functions (3.28). Consequently, the family of all set functions (A25) determines uniquely the physical state represented by the normalized vector $\Psi\in\mathfrak{L}$.

In order to establish the connection with the conventional treatment of free fields, consider

$$\begin{aligned}&N_\Psi(\cdots\mathbf{p}^{r_k+s_k};\boldsymbol{\rho}^r(\boldsymbol{\Delta}^r),\boldsymbol{\rho}^s(\boldsymbol{\Delta}^s);\mathbf{p}^{r+s};\\&\qquad\boldsymbol{\rho}^{r_{k+1}}(\boldsymbol{\Delta}^{r_{k+1}});\cdots\,|\cdots\times\mathbf{B}^{r_k}\times\mathbf{B}^{s_k}\times\{\mathbf{1}\}^r\\&\qquad\quad\times\{-\mathbf{1}\}^s\times\mathbf{B}^r\times\mathbf{B}^s\times\{\mathbf{1}\}^{r_{k+1}}\times\cdots)\\&=\sum_{m,n=0}^{\infty}\cdots\binom{m+r}{r}\binom{n+s}{s}\cdots\langle\Psi|\cdots F^{(m+r,n+s)}\\&\quad\times(\mathbf{B}^{r_k}\times\mathbb{R}^{3(m+r-r_k)}\times\mathbf{B}^{s_k}\times\mathbb{R}^{(n+s-s_k)})\\&\quad\times E^{(m+r,n+s)}(\boldsymbol{\Delta}^r\times\mathbb{R}^{3m}\times\boldsymbol{\Delta}^s\times\mathbb{R}^{3n})\\&\quad\times F^{(m+r,n+s)}(\mathbf{B}^r\times\mathbb{R}^{3m}\times\mathbf{B}^s\times\mathbb{R}^{3n})\\&\quad\times E^{(m+r,n+s)}(\boldsymbol{\Delta}^{r_{k+1}}\times\mathbb{R}^{3(m+r-r_{k+1})}\\&\quad\times\boldsymbol{\Delta}^{s_{k+1}}\times\mathbb{R}^{3(n+s-s_{k+1})})\cdots|\Psi\rangle.\end{aligned} \tag{A26}$$

Note that in writing (A25) in the form of (A26) we have taken into account that, for fixed r and s, all the terms under the summation sign in (A26) for which

$$m_i+r_i\neq m+n,\quad n_i+s_i\neq n+s,\quad i=1,\cdots,k \tag{A27}$$

are zero due to (3.27), and in addition the fact that $H^{(j,l)}$ are mutually orthogonal for different j's and l's. If we write

$$\Psi_1^{(j,l)} = \cdots \binom{j}{r_k}\binom{l}{s_k} F^{(j,l)}(\mathbf{B}^{r_k} \times \mathbb{R}^{3(j-r_k)} \times \mathbf{B}^{s_k} \times \mathbb{R}^{3(l-s_k)}) \cdots \Psi,$$

$$\Psi_2^{(j,l)} = \binom{j}{r_{k+1}}\binom{l}{s_{k+1}} \cdots E^{(j,l)}(\mathbf{\Delta}^{r_{k+1}} \times \mathbb{R}^{3(j-r_{k+1})} \times \mathbf{\Delta}^{s_{k+1}} \times \mathbb{R}^{3(l-s_{k+1})}) \cdots \Psi, \quad \text{(A28)}$$

then (A26) becomes

$$\sum_{m,n=0}^{\infty} \binom{m+r}{r}\binom{n+s}{s} \times \langle \Psi_1 | E^{(m+r,n+s)}(\mathbf{\Delta}^r \times \mathbb{R}^{3m} \times \mathbf{\Delta}^s \times \mathbb{R}^{3n}) \times F^{(m+r,n+s)}(\mathbf{B}^r \times \mathbb{R}^{3m} \times \mathbf{B}^s \times \mathbb{R}^{3n}) | \Psi_2 \rangle, \quad \text{(A29)}$$

where Ψ_1 and Ψ_2 are the vectors with the projections in $H^{(j,l)}$ equal to, respectively, $\Psi_1^{(j,l)}$ and $\Psi_2^{(j,l)}$ which are given by (A28). In order to simplify the notation, take $s = 0$. If

$$\Phi_\kappa^{(j,0)} = M_t^{-1}\Psi_\kappa^{(j,0)} \in K^{(j,0)}, \quad \kappa = 1, 2, \quad \text{(A29)}$$

then

$$\lim_{\mathbf{\Delta}^r \to \mathbf{x}^r} \frac{1}{|\mathbf{\Delta}^r|} \langle \Psi_1 | E^{(j,0)}(\mathbf{\Delta}^r \times \mathbb{R}^{3m}) | \Psi_2 \rangle \quad \text{(A30)}$$

exists, where the limit is taken for $\mathbf{\Delta}^r$ shrinking to $\mathbf{x}^r$, and

$$|\mathbf{\Delta}^r| = |\Delta_1| \cdots |\Delta_r| \quad \text{(A31)}$$

with $|\Delta_1|, \cdots, |\Delta_r|$ denoting the volumes of the three-dimensional regions $\Delta_1, \cdots, \Delta_r$; namely, we get from (3.36), (3.38), and (A29) that (A30) is equal to

$$i^j \int_t d\mathbf{y}^m \Phi_1^{(j,0)}(\mathbf{x}^r, \mathbf{y}^m) \frac{\overleftrightarrow{\partial}}{\partial \mathbf{x}_0^r} \frac{\overleftrightarrow{\partial}}{\partial \mathbf{y}_0^m} \Phi_2^{(j,0)}(\mathbf{x}^r, \mathbf{y}^m), \quad \text{(A32)}$$

where

$$\frac{\overleftrightarrow{\partial}}{\partial \mathbf{x}_0^r} = \frac{\overleftrightarrow{\partial}}{\partial x_1^0} \cdots \frac{\overleftrightarrow{\partial}}{\partial x_r^0}, \quad \frac{\overleftrightarrow{\partial}}{\partial y_0^m} = \frac{\overleftrightarrow{\partial}}{\partial y_1^0} \cdots \frac{\overleftrightarrow{\partial}}{\partial y_m^0}. \quad \text{(A33)}$$

Hence by combining (A26) and (A28) (in which we have taken $s = 0$) with (A32), we get

$$N_\Phi^{(t)}(\cdots; \boldsymbol{\rho}^r(\mathbf{x}^r); \cdots | \cdots \times \{1\}^r \times \cdots) = \lim_{\mathbf{\Delta}^r \to \mathbf{x}^r} \frac{1}{|\mathbf{\Delta}^r|} N_\Phi^{(t)}(\cdots; \boldsymbol{\rho}^r(\mathbf{\Delta}^r); \cdots | \cdots \times \{1\}^r \times \cdots) = \sum_{m=0}^{\infty} \binom{m+r}{r} i^{m+r} \int_t d\mathbf{y}^m \times \Phi_1^{(r+m,0)}(\mathbf{x}^r, \mathbf{y}^m) \frac{\overleftrightarrow{\partial}}{\partial x_0^r} \frac{\overleftrightarrow{\partial}}{\partial y_0^m} \Phi_2^{(r+m,0)}(\mathbf{x}^r, \mathbf{y}^m) \quad \text{(A34)}$$

where we have introduced the symbol

$$\boldsymbol{\rho}(\mathbf{x}^r) = ;\rho(\mathbf{x}_1) \cdots \rho(\mathbf{x}_r); \quad \text{(A35)}$$

We call $\rho(\mathbf{x})$ a field at a point. Comparison of (A34) and (A8) shows that we can formally write

$$N_\Phi^{(t)}(\cdots; \rho(\mathbf{x}_1) \cdots \rho(\mathbf{x}_r); \cdots | \cdots \times \{1\}^r \times \cdots) = i^r \langle \Phi_1 | : \left(\varphi_p^*(x_1) \frac{\overleftrightarrow{\partial}}{\partial x_1^0} \varphi_p(x_1)\right) \cdots \left(\varphi_p^*(x_r) \frac{\partial}{\partial x_r^0} \varphi_p(x^r)\right) : | \Phi_2 \rangle. \quad \text{(A36)}$$

For antiparticles we get a similar formal relation by replacing $\{1\}^r$ in (A30) by $\{-1\}^r$ and φ_p by φ_a.

Orear Behavior in Potential Scattering. II*

PORTER W. JOHNSON
Department of Physics, Case Western Reserve University, Cleveland, Ohio

(Received 24 May 1968)

Here we will extend certain results obtained in a previous paper [J. Math. Phys. 9, 712 (1968)] concerning high-energy large-angle scattering upon nonsingular potentials $V(r)$ which are even in r and analytic in a finite strip about the real r axis. (In the previous paper, it was required that the potentials have a larger domain of analyticity than is required here.) The scattering amplitude is compared with the first Born approximation for this wider class of potentials.

1. INTRODUCTION

In a previous paper[1] it was established that there is a class of potentials for which the corresponding differential cross section decreases very rapidly with respect to energy at fixed nonforward scattering angles, thus bearing some formal similarity to the properties of elastic pp scattering which were first pointed out by Orear.[2] The class of central potentials $V(r)$ considered in I was even in r and analytic in the region $(\mathrm{Im}\, r)^2 - (\mathrm{Re}\, r)^2 < r_0^2$, where $r_0 > 0$. It was shown that for a subclass of these potentials the scattering amplitude $f(k, \theta)$ obeys the following relation:

For all $b < r_0$, in the limit as $k \to \infty$ for fixed θ,

$$f(k, \theta) = O(e^{-2kb \sin \frac{1}{2}\theta}). \tag{1.1}$$

For the purposes of this paper we say that f has Orear behavior if it satisfies this order relation. In establishing this order relation in I, it was necessary to place conditions upon V which were such that the first Born approximation exhibited Orear behavior as well.

When V was further restricted, the following limit involving f and the first Born approximation f_1 was obtained:

$$\lim_{\substack{k\to\infty \\ \theta \neq s \text{ fixed}}} \frac{f(k, \theta)}{f_1(k, \theta)} = 1. \tag{1.2}$$

[Relation (1.2) is more restrictive than (1.1), because (1.1) gives an upper bound upon the asymptotic behavior of f, whereas (1.2) precisely indicates this asymptotic behavior.]

As will be seen below in Lemma 2, if the potential $V(r)$ is even in r, is analytic in the strip $|\mathrm{Im}\, r| < b$, for some number $b > 0$, and satisfies certain growth conditions in the strip, the corresponding first Born approximation satisfies an order relation of the form (1.1). This brings up the question of whether the additional analyticity assumed in I is *necessary* for relation (1.1) to apply to the entire amplitude. For example, the potential

$$V(r) = \lambda/(r^4 + b^4)$$

is not analytic in the region considered in I for any positive number r_0, whereas it is analytic in the region $|\mathrm{Im}\, r| < b$. In this paper we will show that the scattering amplitude corresponding to this potential, as well as other potentials which are even in r and analytic in a finite strip about the real r-axis, exhibits Orear behavior.

Within the class of potentials under consideration here the relation (1.2) is more restrictive than (1.1). The first Born approximation can easily be shown to be the high-energy, fixed-momentum transfer limit of the scattering amplitude. On the other hand, the large-momentum transfer, fixed-energy limit of f is in fact not f_1—this latter limit can most easily be examined through Regge poles. The limit in (1.2) is intermediate to these two well-known asymptotic limits and is not generally valid in potential scattering. In fact, as has long been known, (1.2) is not correct for the Gaussian potential, $V(r) = \lambda \exp(-ar^2)$. Precise necessary and sufficient conditions upon the potential for the validity of (1.2) are, however, not known. In this paper, we will find conditions upon V in our class which are sufficient for the validity of (1.2).

Let us make the initial restriction that $V(r)$ is square-integrable over $\mathbf{r}$, so that its three-dimensional Fourier transform $V(q)$ exists.[3] For definiteness let us further restrict the potential such that $V(q)$ decreases exponentially in q for large q; that is, there must exist a number $d > 0$ such that

$$\lim_{q\to\infty} |V(q)|\, e^{qd} = 0. \tag{1.3}$$

* Supported in part by the U.S. Atomic Energy Commission.

[1] P. Johnson, J. Math. Phys. 9, 712 (1968). It will be referred to as I hereafter.

[2] J. Orear, Phys. Rev. Letters 12, 112 (1964).

[3] Our convention for $V(q)$ is $V(q) = (2\pi)^{-3} \int d\mathbf{x} V(x) \exp(-i\mathbf{q} \cdot \mathbf{x})$. We will adhere to this convention for momentum-space representations of other functions. The results of this paper can be extended trivially to complex-valued, energy-independent potentials as well. For purposes of notation, we write the equations under the assumption that the potential is real for physical $\mathbf{r}$.

In nonpathological cases, this condition implies that $V(r)$ is even in r and analytic in the strip $|\mathrm{Im}\, r| < d$, as is seen in the following lemmas:

Lemma 1: If the integral $\int_0^\infty dqq\,|V(q)|\,e^{\lambda q}$ converges for $\lambda < b$, the Fourier transform $V(r)$ exists, is analytic in the region $|\mathrm{Im}\, r| < b$, and is bounded as $|r| \to \infty$ within the strip. Within this region of analyticity, the relation $V(r) = V(-r)$ is valid.

Lemma 2: If $V(r)$ is even in r and analytic in the strip $|\mathrm{Im}\, r| < b$ and has the property that

$$\int_{-\infty}^{\infty} ds\,|s + id|\,|V(s + id)|$$

converges for $d < b$, then relation (1.3) is valid for $d < b$.

It happens that one can state the conditions sufficient for the validity of (1.2) more generally and more naturally in terms of $V(q)$; in this way we will not be limited to any specific representation of $V(r)$, even though the conditions placed upon $V(q)$ certainly do imply that $V(r)$ be even in r and analytic in a strip of finite width about the real axis.[4]

For our purposes, it is advantageous to factor the potential; that is, to write $V(r) = v_1(r)v_2(r)$. With this factorization, one can write the Born series formally as

$$f = f_1 + v_1 W v_2 - v_1 W^2 v_2 \cdots, \tag{1.4}$$

where

$$W = v_2 G_k v_1. \tag{1.5}$$

The kernel W is square-integrable, and it will be seen that simple upper bounds upon $|W|$ lead to rather economical bounds upon $|f|$. The impetus of this paper is to obtain a bound upon $|W|$ which leads to bounds upon $|f|$ and $|f - f_1|$ which are sufficient to imply (1.1) and (1.2).

It is convenient to place conditions upon v_1 and v_2 which are sufficient for the validity of (1.1) and (1.2). Then (1.1) and (1.2), respectively, will be valid for the class of potentials V for which these conditions can be met by a suitably chosen factorization.

In Sec. 2 it is shown that the following conditions upon the Fourier transforms $v_1(q)$ and $v_2(q)$ are sufficient for (1.1) with $b < r_0$.

There exists a number $\lambda > -2$ and numbers K_1, K_2, and $r_0 > b$ such that

$$|v_1(q)| \leq K_1 q^\lambda \exp(-qr_0), \tag{1.6a}$$

$$\left|\frac{d}{dq} v_1(q)\right| \leq K_1 r_0 q^\lambda \exp(-qr_0). \tag{1.6b}$$

In Sec. 3 it is shown that the following conditions are sufficient for the validity of (1.2):

(1) There exists a number $b > 0$ and a number $p > 2$ such that[5]

$$\lim_{q\to\infty} |V(q)|\,e^{qb} \cdot q^n = 0 \quad \text{for} \quad n < p, \tag{1.7a}$$

$$\lim_{q\to\infty} \frac{1}{|V(q)|} e^{-qb} \frac{1}{q^n} = 0 \quad \text{for} \quad n > p. \tag{1.7b}$$

(2) For a suitable factorization of V, there must exist positive numbers C_1, C_2, and b and a number $s > 1$ such that

$$|v_1(q)| \leq C_1 \frac{1}{(1 + q^2b^2)^s} e^{-qb}, \tag{1.8a}$$

$$\left|\frac{d}{dq} v_1(q)\right| \leq C_1 b \frac{1}{(1 + q^2b^2)^s} e^{-qb}. \tag{1.8b}$$

(3) It is necessary that $2s + \frac{1}{2} > p$, where p is given in (1) and s in (2) above.

In Sec. 4 we discuss the classes of potentials for which the above results are proved.

2. OREAR BEHAVIOR

Here we will establish that conditions (1.6) upon the Fourier transforms of the potential factors of V are sufficient to imply the bound (1.1) for the entire scattering amplitude. We will consider only the case $\lambda = -1$ explicitly. The generalization to the case $\lambda > -1$ follows trivially, and the case $-2 < \lambda < -1$ can be obtained with additional effort.

The kernel W, which was written formally in Eq. (1.5), has the following explicit representations:

$$W(q_1, q_2; k) = \int d\mathbf{p} \frac{1}{p^2 - k^2 - i\epsilon} v_1(q_1 - p)v_2(p - q_2). \tag{2.1}$$

[4] One can construct the following representations for the class of potentials which is even in r and analytic in the region $|\mathrm{Im}\, r| < b$:

$$V(r) = \int_0^\infty d\alpha\sigma(\alpha) \exp(-\alpha b) \cos \alpha r,$$

$$V(r) = \int_0^\infty d\alpha\tau(\alpha) \frac{1}{r^2 + (b + i\alpha)^2} \frac{1}{r^2 + (b - i\alpha)^2}.$$

We could obtain conditions upon σ or τ, which would be sufficient for (1.1) or (1.2). We have avoided use of these rather cumbersome representations by considering $V(q)$ directly.

[5] The limit (1.7b) may not exist if the function $V(q)$ oscillates too rapidly as $q \to \infty$, as in the case $V(q) = \exp(-qb)(\sin qb/qb)^2$ for $n > 2$. In such cases one can define an alternative limit, namely

$$\lim_{q\to\infty} \int_q^\infty (dp/p^n) \exp(-pb) / \int_q^\infty dp\,|V(q)|,$$

which clearly reduces to (1.7b) if the latter exists. If one uses such limits here, he can prove that Eq. (1.2) is valid when the limit there is also taken in this manner.

The object here is to obtain a bound upon W which depends only upon $\Delta = q_2 - q_1$ and k. For this purpose it is convenient to consider the function

$$I(q_1, q_2; p) = \int d\Omega_p g(q_1 - p)g(p - q_2), \quad (2.2)$$

where

$$g(k) = k^{-1}e^{-kb}.$$

In Appendix A the following bound upon I is obtained[6]:

$$I(q_1, q_2; p) \leq (8\pi/p^2)K_0(\Delta b). \quad (2.3)$$

One may thus use the bound (2.3) along with Eq. (1.6a) to bound the imaginary part of W as follows:

$$|\text{Im } W| \leq (4\pi^2/k)K_0(\Delta b) \cdot K_1K_2. \quad (2.4)$$

The real part of the kernel W can be written as a principal-value integral over p. In order to bound Re W it is necessary to rewrite it in terms of nonsingular integrals. One can express Re W in terms of the following nonsingular integral:

$$\text{Re } W = k\int_0^1 \frac{du}{u^2 - 1}\left\{u^2h(u) - h\left(\frac{1}{u}\right)\right\}, \quad (2.5)$$

where

$$h(u) = \int d\Omega_l v_1(\mathbf{q}_1 - uk\mathbf{l})v_2(\mathbf{q}_1 - uk\mathbf{l}).$$

We use this nonsingular integral in Appendix B to obtain a bound upon Re W.[7] Here we state the result:

If we make the restriction $kb > 1$, for every number ϵ such that $0 \leq \epsilon < \frac{1}{2}$ there exists a number $M(\epsilon)$ such that

$$|W(q_1, q_2; k)| \leq [M(\epsilon)b/(kb)^\epsilon]K_0(\Delta b). \quad (2.6)$$

For convenience let $\lambda = M(\epsilon)b/(kb)^\epsilon$. We can now bound the scattering amplitude by the following series:

$$|f(k, \Delta)| \leq |f_1(\Delta)| + \sum_{\lambda=1}^{\infty} \lambda^n G_n(\Delta),$$

where

$$G_n(\Delta) = \int d\mathbf{p}_1 \cdots d\mathbf{p}_n g(|q_1 - p_1|)K_0(|p_1 - p_2|\, b) \cdots K_0(|p_{n-1} - p_n|\, b)g(|p_n - q_2|)$$

$$= \frac{1}{(2\pi)^3}\int d\mathbf{x}[\bar{g}(x)]^2[P(x)]^n e^{-i\Delta\cdot\mathbf{x}},$$

where

$$P(x) = \int d\mathbf{\Delta} e^{i\Delta\cdot\mathbf{x}}K_0(\Delta b) = \frac{\pi^2}{2}\frac{1}{(x^2 + b^2)^{\frac{3}{2}}},$$

$$\bar{g}(x) = \frac{4\pi}{b^2 + x^2}.$$

We can thus write

$$|f(k, \Delta)| \leq (4\pi C)\int d\mathbf{x}e^{-i\Delta\cdot\mathbf{x}}\frac{1}{(x^2 + b^2)^2}\frac{1}{1 - \lambda P(x)}. \quad (2.7)$$

We may integrate over the angular part in Eq. (2.7) to obtain

$$|f(k, \Delta)| \leq \frac{C'}{\Delta}\int_0^\infty dx x \sin \Delta x \frac{1}{(x^2 + b^2)^2}\frac{1}{1 - \lambda P(x)}.$$

Since $\bar{g}(x)$ and $P(x)$ are even functions of x we may write

$$|f(k, \Delta)| = \frac{C'}{2i\Delta}\int_{-\infty}^{\infty} dx x e^{i\Delta x}\frac{1}{(x^2 + b^2)^2}\frac{1}{1 - \lambda P(x)}.$$

Now for sufficiently small values of λ (i.e., for k sufficiently large) we can distort the contour of x integration arbitrarily closely to the line Im $x = b$. As a result, for every number $d < b$ there exists a sufficiently large value of k such that

$$|1 - \lambda P(x)|^{-1} \leq 2$$

for Im $x = d$. Then one has the bound

$$|f(k, \Delta)| \leq \frac{C'}{\Delta}e^{-\Delta b}\int_{-\infty+id}^{\infty+id} dx \frac{x}{(x^2 + b^2)^2}.$$

Since this latter integral is bounded, we have thus established that for k sufficiently large the scattering amplitude is bounded as follows:

$$|f(k, \Delta)| \leq Ce^{-2kd\sin\frac{1}{2}\theta}$$

for every number $d < r_0$. We have thus shown that condition (1.6) is sufficient to establish that the entire scattering amplitude exhibits Orear behavior [Eq. (1.1)]. We have used the bound (2.6) upon the kernel W to show that the nth term of the Born series is bounded by $C_ne^{-\Delta d}$ and that the constants C_n are such that the sum of the series is bounded by a term $Ce^{-\Delta d}$ as well.

3. THE ASYMPTOTIC LIMIT

In this section we will show that conditions (1.7) and (1.8) upon the Fourier transforms of the potential factors of V are sufficient to guarantee that the high-energy fixed-angle limit of f is f_1, as described in Eq. (1.2). The procedure is to obtain a bound upon the modulus of the kernel W which was defined formally by Eq. (1.5) and written explicitly in Eq. (2.1). Naturally, the bound upon $|W|$ must be more restrictive than (2.6), because one wishes to establish the more restrictive condition (1.2) here.

It is convenient to define the following functions:

$$I(q_1, q_2; p) = \int d\Omega_p f_s(q_1 - p)f_s(p - q_2), \quad (3.1)$$

[6] For the modified Bessel function $K_0(x)$, we follow the convention of I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press Inc., New York, 1965), p. 958ff.

[7] Techniques entirely analogous to those necessary to establish this result are exhibited in P. Johnson, J. Math. Phys. 9, 1633 (1968).

where

$$f_s(q) = \frac{1}{(1 + q^2b^2)^s} e^{-qb}.$$

A representation of $f_s(q)$ is considered in Appendix C, Eq. (C2). The bound (C4) is established there also. It is shown in Appendix D using the bound (C4) that for $s > 1$ there exists a number $D(s)$ such that

$$I(q_1, q_2; p) \leq [D(s)/(pb)^2] f_s(\Delta). \tag{3.2}$$

One can use the bound along with condition (1.8a) to bound the imaginary part of W as follows:

$$|\text{Im } W(q_1, q_2; k)| \leq [\pi D(s)/2kb^2] f_s(\Delta) \cdot C_1 C_2.$$

It is again possible to bound the real part of W by writing the principal-value integral as a nonsingular integral and using the inequality (3.2) along with conditions (1.8). We state the consequence here:

For every number ϵ such that $0 \leq \epsilon < \frac{1}{2}$ there exists a number $M(\epsilon)$ such that, for $kb > 1$ and $s > 1$,

$$|W(q_1, q_2; k)| \leq [M(\epsilon) D(s)/(kb)^\epsilon b] f_s(\Delta). \tag{3.3}$$

Let $\lambda = M(\epsilon)D(s)/(kb)^\epsilon b$. We can use this bound upon $|W|$ to bound the quantity $|f(k, \theta) - f_1(k, \theta)|$ as follows:

$$|f(k, \theta) - f_1(k, \theta)| \leq \sum_{n=1}^{\infty} \lambda^n g_n(\Delta),$$

where

$$g_n(\Delta) = \int d\mathbf{p}_1 \cdots d\mathbf{p}_n f_s(k' - p_1) \times f_s(p_1 - p_2) \cdots f_s(p_{n-1} - p_n) f_s(p_n - k).$$

The following lemma is established in Appendix E:

For every number $s > 1$ there exists a number $N(s)$ such that

$$\int dp f_s(q - p) f_s(p - r) \leq N(s) f_s(q - r). \tag{3.4}$$

Thus, we may bound $|f - f_1|$ as follows:

$$|f(k, \theta) - f_1(k, \theta)| \leq f_s(\Delta) N(s) \sum_{n=1}^{\infty} [\lambda N(s)]^n$$

If we choose k to be so large that $\lambda N(s) < 1$, one can then write

$$|f(k, \theta) - f_1(k, \theta)| \leq f_s(\Delta) \lambda [N(s)]^2 [1 - \lambda N(s)]^{-1}. \tag{3.5}$$

As a consequence we have established the following order relation valid in the limit as $k \to \infty$ with $\theta \neq 0$ fixed:

$$f(k, \theta) - f_1(k, \theta) = O\left(\frac{1}{k^{2s+\epsilon}} e^{-2kr_0 \sin \frac{1}{2}\theta}\right)$$

or $0 \leq \epsilon < \frac{1}{2}$.

One can now compare this with the restriction (1.7) upon $V(q)$ or, equivalently, upon $f_1(k, \theta)$. One can then conclude that

$$f(k, \theta) - f_1(k, \theta) = o[f_1(k, \theta)].$$

This relation is equivalent to (1.2), so that the anticipated results are obtained.

4. DISCUSSION

It is shown in Sec. 2 that conditions (1.6) upon the Fourier transform of the potential factors of $V(r)$ are sufficient to guarantee that the corresponding scattering amplitude exhibits Orear behavior [Eq. (1.1)]. As a result of Lemmas 1 and 2 being applied to the potential factors, one can see that the potentials being considered are even in r and analytic in the strip $|\text{Im } r| < b$. Let us consider a slightly pathological example, namely

$$v_1 = v_2 = v(q) = \mu(a - q), \quad \text{for } q < a,$$
$$= 0, \quad \text{for } q \geq a,$$

where μ and a are positive numbers. These functions satisfy (1.6) for $\lambda = 0$ for any value of r_0. The corresponding potential

$$V(r) = (4\pi\mu)^2 \left(\frac{2(1 - \cos ar) - ar \sin ar}{r^4}\right)^2$$

satisfies (1.1) for any value of r_0.

In Sec. 3 the more restrictive conditions (1.7) upon $V(q)$ and (1.8) upon the Fourier transform of the potential factors are shown to imply that (1.2) is valid for the scattering amplitude and its first Born term. The conditions (1.8) automatically imply that $v_1(r)$ and $v_2(r)$ and thus $V(r)$ are bounded everywhere in the region $|\text{Im } r| \leq r_0$, as is seen from the following lemma.[8]

Lemma 3: If a function $v(|\mathbf{q}|) = v(q)$ satisfies the condition that

$$\int_0^\infty dq q e^{qr_0} |v(q)|$$

converges for some value of $r_0 > 0$, then the three-dimensional Fourier transform $V(r)$ exists and has the following property: There exists a constant M such that $|v(r)| < M$ for all r such that $|\text{Im } r| \leq r_0$. A specific potential which is not so bounded and, thus, for which we have not established (1.2) is the example

$$V(r) = k(r^2 + r_0^2)^{-\frac{1}{2}} e^{-\mu(r^2 + r_0^2)^{\frac{1}{2}}}.$$

[8] Note that the above example would not be expected to have this property, since the corresponding first Born approximation is zero for $q > 2a$, whereas this property would not be expected to be true for the entire scattering amplitude.

Let us also note that conditions (1.8) contain the implicit requirement that $r^6V(r) \to 0$ as $v \to \infty$ through positive real values. In I it was necessary to require only that $r^3V(r) \to 0$ as $r \to \infty$ for the special class of potentials.

One can show that the following relatively more general conditions upon

$$v_1(q),$$
$$2$$

along with condition (1.7) upon $V(q)$, are sufficient for the validity of (1.2).

There exist numbers C, b, λ, and s such that

$$|v_1(q)| \leq \frac{C}{q2\lambda} \frac{1}{(1 + q^2b^2)^s} e^{-qb},$$
$$2$$
$$\left| \frac{d}{dq} v_1(q) \right| \leq \frac{Cb}{q2\lambda} \frac{1}{(1 + q^2b^2)^s} e^{-qb}, \qquad (4.1)$$
$$2$$

where $\lambda < 1$, $\lambda + s > 1$, and $p < 2s + \frac{1}{2}$. Details will not be given here. These latter conditions contain the implicit requirement that $x^4V(x) \to 0$ as $x \to \infty$ and are also sufficient to imply that $V(x)$ is bounded in the region $|\text{Im } r| \leq b$.

Finally, we recall that conditions (1.8) in conjunction with (1.7) are sufficient to imply (1.2) with $s > 1$. We will show here that they cannot possibly be sufficient to imply this for $s = 0$. Indeed, let $v_1(q) = v(q) = \lambda e^{-qb}$. This implies that $V(r) = C/(r^2 + b^2)^4$. One can show that for this potential[9]

$$f_1(k, \theta) \sim k^2 e^{-2kb \sin \frac{1}{2}\theta},$$
$$\text{as} \quad k \to \infty \quad \text{for} \quad \theta \neq 0 \quad \text{fixed},$$
$$f_2(k, \theta) \sim k^4 e^{-2kb \sin \frac{1}{2}\theta},$$
$$\text{as} \quad k \to \infty \quad \text{for} \quad \theta \neq 0 \quad \text{fixed},$$

so that the limit (1.2) cannot be valid for this potential for *every* value of the coupling strength C.

We note that conditions (1.7) and (1.8) are merely sufficient to establish convergence of the amplitude to the first Born approximation at high energies for fixed angles of scattering. Let us also recall that (1.2) is correct if (1.7) and (1.8) are met for $s > 1$ with $2s + \frac{1}{2} > p$ in addition. The fact that the above example satisfies Eq. (1.7) with $p = 2$ and Eq. (1.8) with $s = 0$ and yet fails to obey the inequality (1.2) indicates that the above procedure overestimates the scattering amplitude by only a few powers of k in a region in which the amplitudes are decreasing exponentially with k.

It has been shown here that for certain potentials which are even in r and analytic in a strip of finite width about the real axis, the corresponding scattering amplitude exhibits Orear behavior. We have further made it plausible to believe that *only* potentials which are even in r and analytic in a strip as considered above can exhibit such behavior, if one restricts considerations to energy-independent nonsingular potentials. We have also shown that the first Born approximation is the high-energy fixed-angle limit of the scattering amplitude for a more restrictive class. The question of asymptotic behavior for potentials for which $f \not\to f_1$ for fixed angles at high energies will be examined in another paper.

APPENDIX A

We want to prove the bound (2.3) of I defined by (2.2). We choose to write

$$I = \int d\Omega_p g(\mathbf{q}_1 - \mathbf{p}) g(\mathbf{p} - \mathbf{q}_2),$$

where

$$g(k) = k^{-1}e^{-kb}.$$

We rewrite I as an integral over the Fourier transform of $g(k)$ which is

$$\bar{g}(x) = 4\pi/(b^2 + x^2).$$

It is straightforward to write I in the form[10]

$$I = \frac{2}{p} \int_{-1}^{+1} du \int_0^\infty dr \frac{\sin Qr}{Q} \sin pr \left[r_0^2 + \frac{r^2}{4}(1 - u^2) \right]^{-\frac{1}{2}}$$
$$\times \exp -\Delta \left[r_0^2 + \frac{r^2}{4}(1 - u^2) \right]^{\frac{1}{2}}.$$

We now integrate the r integration by parts and obtain the bound

$$|I| \leq \frac{4}{p^2} \int_{-1}^{+1} du \int_0^\infty dr \frac{\exp \{-\Delta[r_0^2 + \frac{1}{4}r^2(1 - u^2)]^{\frac{1}{2}}\}}{[r_0^2 + \frac{1}{4}r^2(1 - u^2)]^{\frac{1}{2}}}$$
$$= \frac{8\pi}{p^2} K_0(\Delta r_0).$$

The bound upon I is thus obtained.

APPENDIX B

Let us write Eq. (3.5) in the form

$$\text{Re } W = k \int_0^1 \frac{du u^2}{u^2 - 1} \left[h(u) - h\left(\frac{1}{u}\right) \right] + k \int_0^1 du h\left(\frac{1}{u}\right). \qquad \text{(B1)}$$

The second term in (B1) may be bounded using the conditions (1.6) (for $\lambda = -1$) and the result of

[9] P. Johnson, to be published.

[10] $Q = \frac{u}{2}(\mathbf{q}_2 - \mathbf{q}_1) + \frac{1}{2}(\mathbf{q}_1 + \mathbf{q}_2)$.

Appendix A, to obtain

$$k\left|\int_0^1 duh\left(\frac{1}{u}\right)\right| \leq \frac{8\pi}{k} K_0(\Delta r_0)\int_0^1 duu^2, \quad \text{(B2)}$$

so that this part of Re W clearly satisfies the desired bound.

Now we wish to bound

$$B = k\int_0^1 du\,\frac{u^2}{u^2-1}\,g(u), \quad \text{(B3)}$$

where $g(u) = h(u) - h(1/u)$.

Using techniques analogous to those described in a previous paper[7] we use Holder's inequality to bound B as follows:

$$|B| \leq k\left[\int_0^1 duu^2\,|g(u)|\right]^{1/r}\left[\int_0^1 \frac{duu^2}{(1-u^2)^s}\,|g(u)|\right]^{1/s}, \quad \text{(B4)}$$

where $1/r + 1/s = 1$.

The first term in (B4) can easily be bounded in a manner similar to (B2) above. To bound the second term we use the inequalities

$$|g(u)| \leq \int_u^1 dV\,|g'(V)|$$

and

$$|g'(V)| \leq \frac{8\pi}{kb} K_0(\Delta b)\cdot\frac{1}{V^3}$$

The latter inequality is obtained from conditions (1.6a) and (1.6b).

We thus have a bound of the form

$$|B| \leq \frac{C'}{(kb)^{1/s}} K_0(\Delta b)\int_0^1 \frac{du}{(1-u^2)^s}\int_u^1 \frac{dV}{V^3}. \quad \text{(B5)}$$

The integral on the left side of (B5) converges for $1 \leq s < z$; thus for every number ϵ such that $0 \leq \epsilon < \frac{1}{2}$ there exists a number $N(\epsilon)$ such that

$$|B| \leq \frac{N(\epsilon)}{(kb)^{\epsilon}} K_0(\Delta b).$$

The result is thus proved.

APPENDIX C

Let us define the function $h_s(q)$ as follows:

$$h_s(q) = \frac{1}{(1+q^2b^2)^s}\,e^{-qb}. \quad \text{(C1)}$$

We construct the integral representation

$$h_s(q) = \int_0^\infty dx\rho_s(\alpha)e^{-\alpha b^2 - q^2/4\alpha}, \quad \text{(C2)}$$

where[11]

$$\rho_s(\alpha) = \frac{b}{\sqrt{\pi}\,\Gamma(s)}\frac{1}{(4b^2)^s}\frac{1}{\alpha^2} \times\int_0^\infty \frac{du}{(u+\alpha)^{s-\frac{1}{2}}}\left(\frac{u}{\alpha}\right)^{-1} e^{-b^2u-(1/4b^2)u/\alpha(u+\alpha)}. \quad \text{(C3)}$$

We will show here that for all $s \geq 0$ there exists a number $C(s)$ such that

$$\rho_s(\alpha) \leq \frac{b}{(\pi\alpha)^{\frac{1}{2}}}\frac{c(s)}{(1+b^2\alpha)^{2s}}. \quad \text{(C4)}$$

From (C3) we can bound $\rho_s(\alpha)$ as follows:

$$\rho_s(\alpha) \leq \frac{b}{(\pi\alpha)^{\frac{1}{2}}}\frac{1}{(4b^2)^s}\frac{1}{\alpha^s}\int_0^\infty \frac{du}{\Gamma(s)}\,u^{s-1}e^{-b^2u} = \frac{b}{(\pi\alpha)^{\frac{1}{2}}}\frac{1}{(2b^2\alpha)^s}. \quad \text{(C5)}$$

We now pick a number $\lambda > 0$ and split $\rho_s(\alpha)$ into $\rho_s(\alpha) = g_1(\alpha) + g_2(\alpha)$, where

$$g_1(\alpha) = \frac{b}{\sqrt{\pi}\,\Gamma(s)}\frac{1}{(4b^2)^s}\frac{1}{\alpha^2} \times\int_0^{\lambda\alpha}\frac{du}{(u+\alpha)^{s-\frac{1}{2}}}\left(\frac{u}{\alpha}\right)^{s-1} e^{-b^2u-(1/4b^2)[u/\alpha(u+\alpha)]} \leq \frac{1}{(\pi\alpha)^{\frac{1}{2}}\Gamma(s)}\frac{1}{(4b^2\alpha^2)^s}\int_0^{\lambda\alpha} duu^{s-1}e^{-u/4b^2(\lambda+1)\alpha^2}$$

or

$$g_1(\alpha) \leq \frac{b}{(\pi\alpha)^{\frac{1}{2}}}(\lambda+1)^s \quad \text{(C6)}$$

As a result, $g_2(\alpha)$ is given by

$$g_2(\alpha) = \frac{b}{\sqrt{\pi}\,\Gamma(s)}\frac{1}{(4b^2)^s}\frac{1}{\alpha^2} \times\int_{\lambda\alpha}^\infty\frac{du}{(u+\alpha)^{s-\frac{1}{2}}}\left(\frac{u}{\alpha}\right)^{s-1} e^{-b^2u-[u/4b^2\alpha(u+\alpha)]} \leq \frac{b}{(\pi\alpha)^{\frac{1}{2}}}\frac{1}{(4b^2)^s}\frac{1}{(\alpha)^{s-1}}e^{-(1/4\alpha b^2)(\lambda/1+\lambda)} \times\int_{\lambda\alpha}^\infty du\,\frac{u^{s-1}}{\Gamma(u)}\,e^{-b^2u} \leq \frac{4b}{(\pi\alpha)^{\frac{1}{2}}}\left(\frac{1+\lambda}{\lambda}\right)^{s+1}\mu^{s+1}e^{-\mu},$$

where

$$\mu = \frac{\lambda}{1+\lambda}\frac{1}{4\alpha b^2}.$$

Now for $\mu > 0$,

$$\mu^{s+1}e^{-\mu} \leq \left(\frac{s+1}{e}\right)^{s+1} = B(s).$$

[11] Note that $\rho_s(\alpha) \geq 0$ for $\alpha \geq 0$.

Thus,

$$g_2(\alpha) \leq \frac{4b}{(\pi\alpha)^{\frac{1}{2}}} B(s)\left(\frac{1+\lambda}{\lambda}\right)^s. \tag{C7}$$

As a result of (C5), (C6), and (C7), the inequality (C4) is valid for sufficiently large choice of $C(s)$.

APPENDIX D

We want to consider the function $I(q_1, q_2; p)$ which is defined as follows:

$$I(q_1, q_2; p) = \int d\Omega_p h_s(g_1 - p) h_s(p - q_2),$$

where $h_s(q)$ is defined in Appendix C. We will show that for $s > 1$, there exists a number $D(s)$ such that

$$I(q_1, q_2; p) \leq \frac{D(s)}{p^2 b^2} \frac{1}{(1+\Delta^2 b)^s} e^{-\Delta b}.$$

We use the integral representation for $h_s(q)$ shown in Appendix C. In terms of this representation, we are allowed to write the Fourier transform of $h_s(q)$, which we call $\bar{h}_s(x)$, as follows:

$$\bar{h}_s(x) = \int_0^\infty d\alpha (4\pi\alpha)^{\frac{3}{2}} \rho_s(\alpha) \exp[-\alpha(x^2 + b^2)].$$

Thus we may write our function I as follows:

$$I(q_1, q_2; p) = \frac{4}{\pi^2 p} \int_0^\infty\int d\alpha\, d\beta (\alpha\beta)^{\frac{3}{2}} \rho_s(\alpha)\rho_s(\beta) \times \exp[-(\alpha+\beta)b^2] J(q_1, q_2; k; \alpha, \beta),$$

where

$$J(q_1, q_2; p, \alpha, \beta) = \int\int dx\, dy \frac{\sin p(x-y)}{|x-y|} \times \exp[-(\alpha x^2 + \beta y^2)] \exp[-i(q_2 x - q_1 y)].$$

The following inequality can be derived:

$$|J| \leq \frac{4\pi^3}{p} \frac{1}{\alpha+\beta} \frac{1}{(\alpha\beta)^{\frac{1}{2}}} \exp[-\Delta^2/4(\alpha+\beta)],$$

so that

$$I \leq \frac{16\pi}{p^2} \int_0^\infty\int d\alpha\, d\beta \rho_s(\alpha)\rho_s(\beta) \frac{\alpha\beta}{\alpha+\beta} \times \exp[-(\alpha+\beta)b^2 - \Delta^2/4(\alpha+\beta)].$$

Define the new variables $u = \alpha + \beta$, $v = \beta$ and rewrite the above inequality as follows:

$$I \leq \frac{16\pi}{p^2} \int_0^\infty \frac{du}{u} \exp\left(-ub^2 - \frac{\Delta^2}{4u}\right) \times \int_0^u dv v(u-v)\rho_s(u-v)\rho_s(v)$$

$$= \frac{32\pi}{p^2} \int_0^\infty \frac{du}{u} \exp\left(-ub^2 - \frac{\Delta^2}{4a}\right) \times \int_0^{u/2} dv v(v-v)\rho_s(u-v)\rho_s(v).$$

One may now use the bound upon $\rho_s(\alpha)$ obtained in Appendix C to bound I sequentially as follows:

$$I \leq \frac{32[C(s)]^2 b^2}{p^2} \int_0^\infty \frac{du}{u} \exp\left(-ub^2 - \frac{\Delta^2}{4u}\right) \times \int_0^{u/2} \frac{dv[v(u-v)]^{\frac{1}{2}}}{(1+b^2 v)^{2s}[1+b^2(u-v)^{2s}]}$$

$$\leq \frac{32C(s)^2 b^2}{p^2} \int_0^\infty \frac{du}{(u)^{\frac{1}{2}}} \times \exp\left(-ub^2 - \frac{\Delta^2}{4u}\right) \frac{1}{(1+b^2 u/2)^{2s}} \times \int_0^{u/2} \frac{(v)^{\frac{1}{2}} dv}{(1+b^2 v)^{2s}}.$$

So that

$$I \leq 32\pi \left(\frac{C(s)b}{p}\right)^2 B(\tfrac{3}{2}, 2s - \tfrac{3}{2}) \int_0^\infty \frac{du}{(u)^{\frac{1}{2}}} \frac{1}{(1+b^2 u/2)^s} \times \exp\left(-ub^2 - \frac{\Delta^2}{4u}\right). \tag{D1}$$

One may use relation (1) above to write the following bound upon I:

$$I \leq H \frac{\exp(-\Delta b) b^2}{p^2} B(\tfrac{1}{2}, 2s - \tfrac{1}{2}), \tag{D2}$$

where

$$H = 32\pi C(s)^2 B(\tfrac{3}{2}, 2s - \tfrac{3}{2}).$$

Also, one may obtain the following bound upon I from (D1):

$$I \leq H \cdot 4^s \int_0^\infty \frac{du}{u^{2s+\frac{1}{2}}} \exp\left(-ub^2 - \frac{\Delta^2}{4u}\right)$$

$$= 2(2K)^{\frac{1}{2}} \left(\frac{2b}{\Delta}\right)^{2s-\frac{1}{2}} K_{2s-\frac{1}{2}}(\Delta b). \tag{D3}$$

Now $K_\nu(x) \to [\pi^{\frac{1}{2}}/(2x)^{\frac{1}{2}}]e^{-x}$ as $x \to \infty$, so that for $\Delta b > 2s - \frac{1}{2}$ there exists a number $c(\nu)$ such that $K_\nu(\Delta b) \leq [c(\nu)/(\Delta b)^{\frac{1}{2}}]e^{-\Delta b}$. As a result of (D2) and (D3), one can establish that there exists a number $D(s)$ for $s > 1$ such that

$$I \leq \frac{D(s)}{p^2 b^2} h_s(\Delta).$$

APPENDIX E

We wish to consider the function

$$Q_s(\Delta) = \int d\mathbf{p} h_s(q_1 - p) h_s(p - q_2),$$

where $h_s(q)$ is defined in Appendix C we use the representation for $h_s(q)$ in Appendix C to write

$$Q_s(\Delta) = \frac{\pi^{\frac{3}{2}}}{8}\int_0^\infty\int d\alpha\, d\beta \rho_s(\alpha)\rho_s(\beta)\left(\frac{\alpha\beta}{\alpha+\beta}\right)^{\frac{3}{2}} \times \exp\left[-\frac{\Delta^2}{4(\alpha+\beta)} - (\alpha+\beta)b^2\right].$$

We use the bound upon ρ_s obtained in Appendix C to write

$$Q_s(\Delta) \leq b^2\frac{[C(s)]^2\pi^{\frac{1}{2}}}{8} \times \int_0^\infty\int \frac{d\alpha\, d\beta}{(\alpha+\beta)^{\frac{3}{2}}}\frac{\alpha\beta}{(1+b^2\beta)^{2s}(1+b^2\beta)^{2s}} \times \exp\left[-\frac{\Delta^2}{4(\alpha+\beta)} - (\alpha+\beta)b^2\right].$$

One may make a transformation of variables to write

$$Q_s(\Delta) \leq \frac{(\pi)^{\frac{1}{2}}b^2}{4}c^2(s)\int_0^\infty \frac{du}{u^{\frac{3}{2}}}\exp\left(-ub^2 - \frac{\Delta^2}{4u}\right) \times \int_0^{u/2} d\beta \frac{\beta}{(1+b^2\beta)2^s}\frac{(u-\beta)}{[1+b^2(u-\beta)]^{2s}} \leq \frac{[C(s)b]^2(\pi)^{\frac{1}{2}}}{4}\int_0^\infty \frac{du}{(u)^{\frac{1}{2}}}\frac{1}{(1+b^2u/2)^{2s}} \times \exp(-ub^2 - \Delta^2/4u)\int_0^{u/2}\frac{d\beta\beta}{(1+\beta b^2)^{2s}}.$$

If we make the restriction $s > 1$, we can write

$$Q_s(\Delta) \leq \frac{[C(s)]^2\pi^{\frac{3}{2}}}{4b^2}B(2, 2s-2)\int_0^\infty \frac{du}{(u)^{\frac{1}{2}}}\frac{1}{(1+b^2u/2)^{2s}} \times \exp(-ub^2 - \Delta^2/4u).$$

We may bound this integral using the same techniques employed in Appendix D to show that for $s > 1$ there exists a number $N(s)$ such that

$$Q_s(\Delta) \leq \frac{N(s)}{(1+\Delta^2b^2)^s}e^{-\Delta b}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Stationary "Noncanonical" Solutions of the Einstein Vacuum Field Equations*

RICHARD B. HOFFMAN†
Department of Physics, Temple University, Philadelphia, Pennsylvania

(Received 18 September 1968)

The complete set of nonflat normal-hyperbolic solutions of the Einstein vacuum field equations is obtained for the metric tensor defined by the quadratic differential form $ds^2 = \alpha\, du^2 - 2\gamma\, du\, dv - \beta\, dv^2 - e^{\varphi}(dx^2 + dz^2)$ subject to the condition that $\alpha\beta + \gamma^2$ is constant. These solutions are characterized by the existence of a null hypersurface-orthogonal Killing vector, which is also a four-fold degenerate Debever vector with vanishing covariant derivative, and therefore are a special case of the class of plane-fronted gravitational waves.

1. INTRODUCTION

It is well known that the quadratic differential form (QDF)

$$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = \alpha\, du^2 - 2\gamma\, du\, dv - \beta\, dv^2 - e^{\varphi}(dx^2 + dz^2),$$
$$g_{\mu\nu} = g_{\mu\nu}(x, z), \quad \mu, \nu = 0, 1, 2, 3, \tag{1.1}$$

is form-invariant under the coordinate transformation

$$\bar{x} = \bar{x}(x, z), \quad \bar{z} = \bar{z}(x, z)$$

if $\bar{x}$ and $\bar{z}$ are restricted by the equations

$$\ddot{\bar{x}} + \bar{x}'' = 0, \tag{1.2}$$

$$\dot{\bar{x}} = \bar{z}', \quad \bar{x}' = -\dot{\bar{z}}, \tag{1.3}$$

where

$$\frac{\partial}{\partial x}f(x, z) \equiv \dot{f}, \quad \frac{\partial}{\partial z}f(x, z) \equiv f'.$$

It is a consequence of the field equations that ζ, defined by

$$\zeta^2 = \alpha\beta + \gamma^2, \tag{1.4}$$

satisfies Eq. (1.2), so it is customary to choose $\bar{x} = \zeta$, thereby reducing the number of dependent variables to three and simplifying the field equations.

* Supported by the Aerospace Laboratories of the Office of Aerospace Research, U.S. Air Force.

† Present address: Department of Physics, Franklin and Marshall College, Lancaster, Pa.

where $h_s(q)$ is defined in Appendix C we use the representation for $h_s(q)$ in Appendix C to write

$$Q_s(\Delta) = \frac{\pi^{\frac{3}{2}}}{8}\int_0^\infty\int d\alpha\, d\beta \rho_s(\alpha)\rho_s(\beta)\left(\frac{\alpha\beta}{\alpha+\beta}\right)^{\frac{3}{2}} \times \exp\left[-\frac{\Delta^2}{4(\alpha+\beta)} - (\alpha+\beta)b^2\right].$$

We use the bound upon ρ_s obtained in Appendix C to write

$$Q_s(\Delta) \leq b^2\frac{[C(s)]^2\pi^{\frac{1}{2}}}{8} \times \int_0^\infty\int \frac{d\alpha\, d\beta}{(\alpha+\beta)^{\frac{3}{2}}}\frac{\alpha\beta}{(1+b^2\beta)^{2s}(1+b^2\beta)^{2s}} \times \exp\left[-\frac{\Delta^2}{4(\alpha+\beta)} - (\alpha+\beta)b^2\right].$$

One may make a transformation of variables to write

$$Q_s(\Delta) \leq \frac{(\pi)^{\frac{1}{2}}b^2}{4}c^2(s)\int_0^\infty \frac{du}{u^{\frac{3}{2}}}\exp\left(-ub^2 - \frac{\Delta^2}{4u}\right) \times \int_0^{u/2} d\beta \frac{\beta}{(1+b^2\beta)2^s}\frac{(u-\beta)}{[1+b^2(u-\beta)]^{2s}} \leq \frac{[C(s)b]^2(\pi)^{\frac{1}{2}}}{4}\int_0^\infty \frac{du}{(u)^{\frac{1}{2}}}\frac{1}{(1+b^2u/2)^{2s}} \times \exp(-ub^2 - \Delta^2/4u)\int_0^{u/2}\frac{d\beta\beta}{(1+\beta b^2)^{2s}}.$$

If we make the restriction $s > 1$, we can write

$$Q_s(\Delta) \leq \frac{[C(s)]^2\pi^{\frac{3}{2}}}{4b^2}B(2, 2s-2)\int_0^\infty \frac{du}{(u)^{\frac{1}{2}}}\frac{1}{(1+b^2u/2)^{2s}} \times \exp(-ub^2 - \Delta^2/4u).$$

We may bound this integral using the same techniques employed in Appendix D to show that for $s > 1$ there exists a number $N(s)$ such that

$$Q_s(\Delta) \leq \frac{N(s)}{(1+\Delta^2b^2)^s}e^{-\Delta b}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 5 MAY 1969

Stationary "Noncanonical" Solutions of the Einstein Vacuum Field Equations*

RICHARD B. HOFFMAN†
Department of Physics, Temple University, Philadelphia, Pennsylvania

(Received 18 September 1968)

The complete set of nonflat normal-hyperbolic solutions of the Einstein vacuum field equations is obtained for the metric tensor defined by the quadratic differential form $ds^2 = \alpha\, du^2 - 2\gamma\, du\, dv - \beta\, dv^2 - e^{\varphi}(dx^2 + dz^2)$ subject to the condition that $\alpha\beta + \gamma^2$ is constant. These solutions are characterized by the existence of a null hypersurface-orthogonal Killing vector, which is also a four-fold degenerate Debever vector with vanishing covariant derivative, and therefore are a special case of the class of plane-fronted gravitational waves.

1. INTRODUCTION

It is well known that the quadratic differential form (QDF)

$$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = \alpha\, du^2 - 2\gamma\, du\, dv - \beta\, dv^2 - e^{\varphi}(dx^2 + dz^2),$$
$$g_{\mu\nu} = g_{\mu\nu}(x, z), \quad \mu, \nu = 0, 1, 2, 3, \tag{1.1}$$

is form-invariant under the coordinate transformation

$$\bar{x} = \bar{x}(x, z), \quad \bar{z} = \bar{z}(x, z)$$

if $\bar{x}$ and $\bar{z}$ are restricted by the equations

$$\ddot{\bar{x}} + \bar{x}'' = 0, \tag{1.2}$$

$$\dot{\bar{x}} = \bar{z}', \quad \bar{x}' = -\dot{\bar{z}}, \tag{1.3}$$

where

$$\frac{\partial}{\partial x}f(x, z) \equiv \dot{f}, \quad \frac{\partial}{\partial z}f(x, z) \equiv f'.$$

It is a consequence of the field equations that ζ, defined by

$$\zeta^2 = \alpha\beta + \gamma^2, \tag{1.4}$$

satisfies Eq. (1.2), so it is customary to choose $\bar{x} = \zeta$, thereby reducing the number of dependent variables to three and simplifying the field equations.

* Supported by the Aerospace Laboratories of the Office of Aerospace Research, U.S. Air Force.

† Present address: Department of Physics, Franklin and Marshall College, Lancaster, Pa.

Of course, ζ can only be chosen as one of the variables if the expression $\alpha\beta + \gamma^2$ is not constant. In this paper we find and discuss the mathematical properties of the most general solution of the Einstein vacuum field equations for the QDF (1.1) if $\alpha\beta + \gamma^2$ is constant.

2. SOLUTION OF THE FIELD EQUATIONS

Assume that $\alpha\beta + \gamma^2$ is constant. Since

$$g \equiv \|g_{\mu\nu}\| = -\zeta^2 e^{2\varphi}, \tag{2.1}$$

we must have $\zeta \neq 0$. In addition we see that, if we set

$$\alpha = |\zeta|\,\bar{\alpha}, \quad \beta = |\zeta|\,\bar{\beta}, \quad \gamma = |\zeta|\,\bar{\gamma},$$
$$u = |\zeta|^{-\frac{1}{2}}\,\bar{u}, \quad v = |\zeta|^{-\frac{1}{2}}\,\bar{v},$$

in the QDF (1.1) we obtain

$$ds^2 = \bar{\alpha}\,d\bar{u}^2 - 2\bar{\gamma}\,d\bar{u}\,d\bar{v} - \beta\,d\bar{v}^2 - e^{\varphi}(dx^2 + dz^2)$$

and

$$\bar{\alpha}\bar{\beta} + \bar{\gamma}^2 = \zeta^{-2}(\alpha\beta + \gamma^2) = 1;$$

thus, with no loss in generality, we set $\zeta^2 = 1$. The field equations for the QDF (1.1) with $\zeta^2 = 1$ are[1]

$$\alpha^{-1}(\ddot{\alpha} + \alpha'') + \dot{\alpha}\dot{\beta} + \dot{\gamma}^2 + \alpha'\beta' + \gamma'^2 = 0, \tag{2.2}$$

$$\beta^{-1}(\ddot{\beta} + \beta'') + \dot{\alpha}\dot{\beta} + \dot{\gamma}^2 + \alpha'\beta' + \gamma'^2 = 0, \tag{2.3}$$

$$\gamma^{-1}(\ddot{\gamma} + \gamma'') + \dot{\alpha}\dot{\beta} + \dot{\gamma}^2 + \alpha'\beta' + \gamma'^2 = 0, \tag{2.4}$$

$$\ddot{\varphi} + \varphi'' = \dot{\alpha}\dot{\beta} + \dot{\gamma}^2, \tag{2.5}$$

$$\ddot{\varphi} + \varphi'' = \alpha'\beta' + \gamma'^2, \tag{2.6}$$

$$\dot{\alpha}\beta' + \alpha'\dot{\beta} + 2\dot{\gamma}\gamma' = 0. \tag{2.7}$$

Comparing Eqs. (2.5) and (2.6) we see that

$$\dot{\alpha}\dot{\beta} + \dot{\gamma}^2 - \alpha'\beta' - \gamma'^2 = 0. \tag{2.8}$$

Since $\zeta^2 = 1$ we have, assuming $\alpha \neq 0$,

$$\beta = \alpha^{-1}(1 - \gamma^2); \tag{2.9}$$

substituting Eq. (2.9) in Eqs. (2.7) and (2.8) we obtain

$$(\dot{\alpha}\gamma' + \alpha'\dot{\gamma})\frac{\gamma}{\alpha} + \dot{\alpha}\alpha'\frac{1 - \gamma^2}{\alpha^2} = \dot{\gamma}\gamma' \tag{2.10}$$

and

$$2(\dot{\alpha}\dot{\gamma} - \alpha'\gamma')\frac{\gamma}{\alpha} + (\dot{\alpha}^2 - \alpha'^2)\frac{1 - \gamma^2}{\alpha^2} = \dot{\gamma}^2 - \gamma'^2, \tag{2.11}$$

respectively. Solving Eqs. (2.10) and (2.11) for $\gamma\alpha^{-1}$ and $(1 - \gamma^2)\alpha^{-2}$ we obtain

$$\frac{\gamma}{\alpha} = \frac{(\dot{\alpha}\dot{\gamma} + \alpha'\gamma')(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})}{(\dot{\alpha}^2 + \alpha'^2)(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})}, \tag{2.12}$$

$$\frac{1 - \gamma^2}{\alpha^2} = -\frac{(\dot{\gamma}^2 + \gamma'^2)(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})}{(\dot{\alpha}^2 + \alpha'^2)(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})}. \tag{2.13}$$

We now show that by assuming $(\dot{\alpha}\gamma' - \alpha'\dot{\gamma}) \neq 0$ we get a contradiction; in this case

$$\frac{\gamma}{\alpha} = \frac{\dot{\alpha}\dot{\gamma} + \alpha'\gamma'}{\dot{\alpha}^2 + \alpha'^2}, \quad \frac{1 - \gamma^2}{\alpha^2} = -\frac{\dot{\gamma}^2 + \gamma'^2}{\dot{\alpha}^2 + \alpha'^2}.$$

Substituting the first of these in the second we obtain

$$\frac{1}{\alpha^2} + \frac{(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})^2}{(\dot{\alpha}^2 + \alpha'^2)^2} = 0; \tag{2.14}$$

but this implies that

$$\frac{1}{\alpha^2} = \frac{(\dot{\alpha}\gamma' - \alpha'\dot{\gamma})^2}{(\dot{\alpha}^2 + \alpha'^2)^2} = 0,$$

or

$$\dot{\alpha}\gamma' - \alpha'\dot{\gamma} = 0, \tag{2.15}$$

which contradicts our initial assumption. We thus conclude that Eq. (2.15) must hold.

The most general solution of Eq. (2.15) is

$$\gamma = \gamma(\alpha); \tag{2.16}$$

it follows from Eqs. (2.9) and (2.16) that

$$\beta = \beta(\alpha). \tag{2.17}$$

Substituting Eqs. (2.16) and (2.17) in Eqs. (2.7) and (2.8) we find

$$\dot{\alpha}\alpha'\left[\frac{d\beta}{d\alpha} + \left(\frac{d\gamma}{d\alpha}\right)^2\right] = 0, \tag{2.18}$$

$$(\dot{\alpha}^2 - \alpha'^2)\left[\frac{d\beta}{d\alpha} + \left(\frac{d\gamma}{d\alpha}\right)^2\right] = 0. \tag{2.19}$$

If

$$\frac{d\beta}{d\alpha} + \left(\frac{d\gamma}{d\alpha}\right)^2 \neq 0,$$

then $\dot{\alpha} = \alpha' = 0$ and α, β, and γ are constants. Otherwise,

$$\frac{d\beta}{d\alpha} + \left(\frac{d\gamma}{d\alpha}\right)^2 = 0. \tag{2.20}$$

In view of Eq. (2.20) we see that

$$\dot{\alpha}\dot{\beta} + \dot{\gamma}^2 = \alpha'\beta' + \gamma'^2 = 0 \tag{2.21}$$

and the field equations (2.2), (2.3), and (2.4) take the form

$$\ddot{\alpha} + \alpha'' = \ddot{\beta} + \beta'' = \ddot{\gamma} + \gamma'' = 0. \tag{2.22}$$

[1] C. Møller, *The Theory of Relativity* (Oxford University Press, London, 1962), Sec. 118. We have set the cosmological constant equal to zero.

Substituting for γ in Eq. (2.22) from Eq. (2.16) we obtain

$$\frac{d^2\gamma}{d\alpha^2}(\dot{\alpha}^2 + \alpha'^2) + \frac{d\gamma}{d\alpha}(\ddot{\alpha} + \alpha'') = 0$$

or

$$\frac{d^2\gamma}{d\alpha^2} = 0 \tag{2.23}$$

with the solution

$$\gamma = C_1\alpha + C_2; \tag{2.24}$$

similarly, we obtain

$$\beta = C_3\alpha + C_4, \tag{2.25}$$

where $C_1, \cdots, C_4$ are constants. Since

$$\alpha\beta + \gamma^2 = 1,$$

we find that

$$\gamma = C_1\alpha \pm 1, \quad \beta = -C_1^2\alpha \mp 2C_1. \tag{2.26}$$

If we assume $C_1 > 0$ and choose the upper signs in Eqs. (2.26), upon setting

$$\alpha = (C_1)^{-1}\bar{\alpha}, \quad u = (C_1)^{\frac{1}{2}}(\bar{u} + \bar{v}),$$
$$v = (C_1)^{\frac{1}{2}}(2\bar{u} + \bar{v})$$

in the QDF (1.1), we find that

$$ds^2 = \bar{\alpha}\, d\bar{u}^2 + 2\, d\bar{u}\, d\bar{v} - e^{\varphi}(dx^2 + dz^2); \tag{2.27}$$

by following a similar procedure, one can reduce all other possibilities in Eqs. (2.26) to this form. If $\alpha = 0$, then $\gamma^2 = 1$ and the resulting QDF is seen to be that in Eq. (2.27) if we replace $\bar{\alpha}$ and u by $-\beta$ and v, respectively, since even in this case β satisfies Eq. (2.22).

From Eq. (2.22) we see that we may choose $\bar{\alpha}$ as our coordinate $\bar{x}$ and, if we choose $\bar{z}(x, z)$ to be the function conjugate to $\bar{\alpha}$ the QDF (2.27) reduces to

$$ds^2 = x\, du^2 + 2\, du\, dv - e^{\varphi}(dx^2 + dz^2) \tag{2.28}$$

in which we have dropped the bars for convenience. Of course this transformation cannot be made if $\bar{\alpha}$ is constant. In this case we see from the expressions for the affine connections in the Appendix that all $\Gamma^{\rho}_{\mu\nu}$ vanish except

$$\Gamma^1_{11} = \Gamma^3_{13} = -\Gamma^1_{33} = \tfrac{1}{2}\dot{\varphi}$$

and

$$-\Gamma^3_{11} = \Gamma^1_{13} = \Gamma^3_{33} = \tfrac{1}{2}\varphi'.$$

Thus the only component of the curvature tensor that does not vanish identically is

$$R_{1313} = \tfrac{1}{2}(\ddot{\varphi} + \varphi'')e^{\varphi};$$

but we see from Eqs. (2.21) and (2.5) that

$$\ddot{\varphi} + \varphi'' = 0. \tag{2.29}$$

Hence, R_{1313} is also zero and these solutions are flat. It follows that all nonflat vacuum solutions of the field equations for the QDF (1.1) with $\alpha\beta + \gamma^2$ constant may be put into the form in Eq. (2.28) in which φ is a solution of Eq. (2.29).

3. MATHEMATICAL PROPERTIES OF THE SOLUTIONS

A. Curvature Tensor

We find for the QDF (2.28) that the only nonvanishing components of the curvature tensor are

$$R_{0101} = -R_{0101} = \tfrac{1}{4}\ddot{\varphi},$$
$$R_{0103} = \tfrac{1}{4}\varphi', \tag{3.1}$$

modulo the symmetries of the curvature tensor, with

$$u \equiv x^0, \quad x \equiv x', \quad v \equiv x^2, \quad z \equiv x^3.$$

The QDF (2.28) is therefore seen to be flat if and only if φ is constant. The eigenvalues and, hence, the second-order differential invariants are all zero.[2]

B. Killing Vectors

In general, we have two and only two Killing vectors for the QDF (2.28):

$$l^{\mu} = (0, 0, 1, 0) = \delta^{\mu}_2 \tag{3.2}$$

and

$$l^{\mu} = (1, 0, 0, 0) = \delta^{\mu}_0, \tag{3.3}$$

where

$$\delta^{\mu}_{\nu} = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu. \end{cases}$$

If $\varphi = Ax + Bz$ we have the additional Killing vector

$$l^{\mu} = (0, B, 0, -A) \tag{3.4}$$

and if

$$\varphi = -\ln 4A^2[(x + B)^2 + (z + C)^2] + D \arctan\frac{x + B}{x + C}$$

we have the additional Killing vector

$$l^{\mu} = [u, -2(x + B), Bu - v, -2(z + C)]. \tag{3.5}$$

For all φ, the Killing vector in Eq. (3.2) is hypersurface-orthogonal; when $\varphi = Ax + Bz$, the Killing vector in Eq. (3.4) is also hypersurface-orthogonal. Since these vectors are null and spacelike, respectively, and since these are the only hypersurface-orthogonal Killing vectors, we conclude that all nonflat vacuum solutions for the QDF (2.28) are stationary and nonstatic.

C. Algebraic Classification

We find that for all solutions represented by the QDF (2.28) the null Killing vector $l^{\mu} = \delta^{\mu}_2$ satisfies

[2] A. S. Petrow, German transl.: *Einstein-Räume* (Akademie-Verlag, Berlin, 1964), Chap. III.

the equations

$$R_{\mu\nu\rho\sigma}l^{\sigma} = 0 \tag{3.6}$$

and

$$l_{\alpha;\beta} = 0. \tag{3.7}$$

Therefore, the solutions are type N with all optical parameters vanishing[3] and form a subclass of the general class of so-called *pp* waves.[4]

4. OTHER APPLICATIONS

We can apply a similar analysis to the QDF

$$ds^2 = e^{\varphi}(du^2 - dv^2) - \alpha\, dx^2 - 2\gamma\, dx\, dz - \beta\, dz^2, \tag{4.1}$$

where $g_{\mu\nu} = g_{\mu\nu}(u, v)$; in this case, the conformally flat two-space has signature zero compared to a signature of -2 for the QDF (1.1). For normal-hyperbolic noncanonical solutions of the field equations we set

$$\alpha\beta - \gamma^2 = 1 \tag{4.2}$$

and proceed as before. We again find that β and γ must be functions of α with

$$\ddot{\alpha} - \alpha'' = (\dot{\alpha}^2 - \alpha'^2)\frac{d^2\gamma}{d\alpha^2} = (\dot{\alpha}^2 - \alpha'^2)\frac{d^2\beta}{d\alpha^2} = 0. \tag{4.3}$$

If $\dot{\alpha}^2 - \alpha'^2 \neq 0$, β and γ have the forms in Eqs. (2.25) and (2.24); however, there are no choices for the constants $C_1, \cdots, C_4$ for which β, γ are real and satisfy Eq. (4.2) other than when α, β, and γ are all constant. If $\dot{\alpha}^2 - \alpha'^2 = 0$, then we are left with the equation

$$\frac{d\beta}{d\alpha} - \left(\frac{d\gamma}{d\alpha}\right)^2 = 0$$

or, with the use of Eq. (4.2),

$$\left(\frac{d\gamma}{d\alpha} - \frac{\gamma}{\alpha}\right)^2 + \frac{1}{\alpha^2} = 0, \tag{4.4}$$

which has no real solution. Thus, since, as for the QDF (1.1), the solutions are flat for α, β, γ constant, we conclude that there are no noncanonical solutions of the field equations for the QDF (4.1) other than those that are flat.

APPENDIX

For the metric tensor in the QDF (1.1) we find

$$g^{00} = \zeta^{-2}\beta, \quad g^{02} = -\zeta^{-2}\gamma,$$
$$g^{22} = -\zeta^{-2}\alpha, \quad g^{11} = g^{33} = -e^{-\varphi},$$

and all other $g_{\mu\nu} = 0$. The affine connections defined by

$$\Gamma^{\rho}_{\mu\nu} = \tfrac{1}{2}g^{\rho\sigma}(g_{\mu\sigma,\nu} + g_{\nu\sigma,\mu} - g_{\mu\nu,\sigma})$$

are given by

μ	ν \ ρ	0	1	2	3
0	0	0	$\frac{\dot{\alpha}}{2}e^{-\varphi}$	0	$\frac{\alpha'}{2}e^{-\varphi}$
0	1	$\frac{\beta\dot{\alpha} + \gamma\dot{\gamma}}{2\zeta^2}$	0	$\frac{\alpha\dot{\gamma} - \gamma\dot{\alpha}}{2\zeta^2}$	0
0	2	0	$-\frac{\dot{\gamma}}{2}e^{-\varphi}$	0	$-\frac{\gamma'}{2}e^{-\varphi}$
0	3	$\frac{\beta\alpha' + \gamma\gamma'}{2\zeta^2}$	0	$\frac{\alpha\gamma' - \gamma\alpha'}{2\zeta^2}$	0
1	1	0	$\frac{\dot{\varphi}}{2}$	0	$-\frac{\varphi'}{2}$
1	2	$\frac{\gamma\dot{\beta} - \beta\dot{\gamma}}{2\zeta^2}$	0	$\frac{\alpha\dot{\beta} + \gamma\dot{\gamma}}{2\zeta^2}$	0
1	3	0	$\frac{\varphi'}{2}$	0	$\frac{\dot{\varphi}}{2}$
2	2	0	$-\frac{\dot{\beta}}{2}e^{-\varphi}$	0	$-\frac{\beta'}{2}e^{-\varphi}$
2	3	$\frac{\gamma\beta' - \beta\gamma'}{2\zeta^2}$	0	$\frac{\alpha\beta' + \gamma\gamma'}{2\zeta^2}$	0
3	3	0	$-\frac{\dot{\varphi}}{2}$	0	$\frac{\varphi'}{2}$

[3] R. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).

[4] H. W. Brinkman, Proc. Nat. Acad. Sci. US **9**, 1 (1923). For a detailed discussion of these solutions, see J. Ehlers and W. Kundt, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), p. 49.