Weyl conform tensor for stationary gravitational fields*

Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 13 August 1973)

Our formulas for the Weyl conform tensor components generalize results published earlier by Z. Perjés for vacuum fields. We also offer an abstract version of these equations which may shed some light upon their structure. The expressions for the Weyl conform tensor are specialized to the case of small perturbations from a stationary axially symmetric background geometry. The resulting formulas supplement the expressions which Chandrasekhar and Friedman have developed for the components of the Ricci tensor. We anticipate that this will facilitate the comparison of the CF perturbation theory with the recent studies of perturbations of the Kerr metric by Press, Teukolsky, and Wald. In this connection we identify in terms of the CF field variables the fields which are involved in Teukolsky's separable field equations.

I. STATIONARY FIELDS

As in our earlier work[1] on stationary axially symmetric gravitational fields, we shall find it convenient to introduce a complex null tetrad system,

$$t=(e^1+ie^2)/2^{1/2},\quad t^*=(e^1-ie^2)/2^{1/2},\qquad k=(e^3-e^4)/2^{1/2},\ \text{and}\ m=(e^3+e^4)/2^{1/2}, \tag{1}$$

where

$$e^i=f^{-1/2}\epsilon^i\,(i=1,2,3)\ \text{and}\ e^4=f^{1/2}\,(dT-\bar{\omega}). \tag{2}$$

The five complex scalar fields C_i $(i=-2,\ldots,+2)$ associated with the Weyl conform tensor[2] are then given by the following expressions:

$$C_2-\tfrac{1}{2}R_{tt}=-\tfrac{1}{2}\tau\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}, \tag{3a}$$

$$C_1+\tfrac{1}{4}(R_{kt}+R_{mt})=(1/4\sqrt{2})[\kappa\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}+\tau\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\kappa}], \tag{3b}$$

$$C_0+R/12-\tfrac{1}{4}(R_{kk}+R_{mm})=\tfrac{1}{4}[\tau\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}^*+\tau^*\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}-f^{-1}\underline{G}\neg\underline{G}], \tag{3c}$$

$$C_{-1}-\tfrac{1}{4}(R_{kt^*}+R_{mt^*})=(-1/4\sqrt{2})[\kappa\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}^*+\tau^*\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\kappa}], \tag{3d}$$

$$C_{-2}-\tfrac{1}{2}R_{t^*t^*}=-\tfrac{1}{2}\tau^*\ulcorner(d\underline{G}+\tfrac{1}{2}f^{-1}G\underline{G})\neg\underline{\tau}^*, \tag{3e}$$

where the triad τ, τ^*, κ is defined by

$$\tau=(\epsilon^1+i\epsilon^2)/2^{1/2},\quad \tau^*=(\epsilon^1-i\epsilon^2)/2^{1/2},\quad \kappa=\epsilon^3. \tag{4}$$

The symbols $\ulcorner$ and $\neg$ denote the 3-dimensional Grassman inner products[3] for differential forms and tangent vectors, respectively. The latter are always distinguished from the former by underscoring.

It was shown in an earlier paper[4] that the complex 1-form

$$G=df-if^2*d\bar{\omega} \tag{5}$$

also occurs conspicuously in the Ricci tensor, the orthonormal components of which are given by the following expressions:

$$R^j_4=2^{-1/2}i(*\epsilon^j)\ulcorner dG, \tag{6a}$$

$$R^4_4=-\tfrac{1}{2}\epsilon\ulcorner(d\underline{G}-f^{-1}G\underline{G})\neg\underline{\epsilon}_j, \tag{6b}$$

$$R^j_k+\delta^j_kR^4_4=f^{(3)}R^j_k+\tfrac{1}{4}f^{-1}\epsilon^j\ulcorner(G^*\underline{G}+G\underline{G}^*)\neg\underline{\epsilon}_k, \tag{6c}$$

where the symbol $*$ denotes the 3-dimensional duality operator; *i.e.*, $*\epsilon^j=\epsilon^k\epsilon^l$ (j,k,l cyclic).

To render these equations into a form analogous to that to be found in the article of Perjés[5] we introduce an imaginary 1-form U and a complex 1-form V such that

$$U=d\underline{\tau}\neg\underline{\tau}^*\quad\text{and}\quad V=d\underline{\kappa}\neg\underline{\tau}. \tag{7}$$

Components of such forms relative to the triad τ, τ^*, κ are denoted by subscripts; e.g., $U_\kappa=\kappa\ulcorner U$. In particular, Eqs. (3a) through (3e) assume the following form:

$$C_2-\tfrac{1}{2}R_{tt}=-\tfrac{1}{2}[(d_\tau G_\tau+V_\tau G_\kappa-U_\tau G_\tau)+\tfrac{1}{2}f^{-1}G_\tau^2], \tag{8a}$$

$$C_1+\tfrac{1}{4}(R_{kt}+R_{mt})=(1/4\sqrt{2})[(d_\kappa G_\tau+V_\kappa G_\kappa-U_\kappa G_\tau)+(d_\tau G_\kappa-V_\tau G_{\tau^*}-V_{\tau^*}{}^*G_\tau)+f^{-1}G_\kappa G_\tau], \tag{8b}$$

$$C_0+R/12-\tfrac{1}{4}(R_{kk}+R_{mm})=\tfrac{1}{4}[(d_{\tau^*}G_\tau+V_{\tau^*}G_\kappa-U_{\tau^*}G_\tau)+(d_\tau G_{\tau^*}+V_{\tau^*}{}^*G_\kappa+U_\tau G_{\tau^*})-f^{-1}G_\tau G_{\tau^*}-f^{-1}G_\kappa^2], \tag{8c}$$

$$C_{-1}-\tfrac{1}{4}(R_{kt^*}+R_{mt^*})=(-1/4\sqrt{2})[(d_\kappa G_{\tau^*}+V_\kappa^*G_\kappa+U_\kappa G_{\tau^*})+(d_{\tau^*}G_\kappa-V_{\tau^*}G_{\tau^*}-V_\tau^*G_\tau)+f^{-1}G_\kappa G_{\tau^*}], \tag{8d}$$

$$C_{-2}-\tfrac{1}{2}R_{t^*t^*}=-\tfrac{1}{2}[(d_{\tau^*}G_{\tau^*}+V_\tau^*G_\kappa+U_{\tau^*}G_{\kappa^*})+\tfrac{1}{2}f^{-1}G_{\tau^*}^2]. \tag{8e}$$

In comparing these results with those of Perjés, one should note that an unfortunate factor $1/(2f)$ was included in the G field of Perjés, and our sign conventions are different. Aside from that our Eqs. (8a) through (8e) constitute a nonvacuum generalization of formulas given by Perjés.

The most significant components of the Ricci tensor will be written in a similar manner; namely,

$$R_{44}=\tfrac{1}{2}[(d_\kappa G_\kappa-V_\kappa G_{\tau^*}-V_\kappa^*G_\tau)+(d_\tau G_{\tau^*}+V_{\tau^*}{}^*G_\kappa+U_\tau G_{\tau^*})+(d_{\tau^*}G_\tau+V_{\tau^*}G_\kappa-U_{\tau^*}G_\tau)-f^{-1}(G_\kappa^2+2G_\tau G_{\tau^*})], \tag{9a}$$

$$R_{km}=\tfrac{1}{2}f[{}^{(3)}R_{\kappa\kappa}+\tfrac{1}{2}f^{-2}G_\kappa G_\kappa^*], \tag{9b}$$

$$R_{tt}=f[{}^{(3)}R_{\tau\tau}+\tfrac{1}{2}f^{-2}G_\tau G_{\tau^*}{}^*], \tag{9c}$$

$$R_{kt}+R_{mt}=2^{1/2}f[{}^{(3)}R_{\kappa\tau}+\tfrac{1}{4}f^{-2}(G_\kappa G_{\tau^*}{}^*+G_\tau G_\kappa^*)]. \tag{9d}$$

The components of the 3-dimensional Ricci tensor are given in turn by

$$\begin{aligned}{}^{(3)}R_{\kappa\kappa}&=2(\kappa\tau^*)\ulcorner(dV-UV)\\&=2[(d_\kappa V_{\tau^*}+V_\kappa^*V_\kappa+U_\kappa V_{\tau^*})\\&\quad-(d_{\tau^*}V_\kappa-V_{\tau^*}^2-V_\tau^*V_\tau)-U_\kappa V_{\tau^*}-V_{\tau^*}V_\kappa],\end{aligned} \tag{10a}$$

$$\begin{aligned}{}^{(3)}R_{\tau\tau}&=(\kappa\tau)\ulcorner(dV-UV)\\&=(d_\kappa V_\tau+V_\kappa^2-U_\kappa V_\tau)-(d_\tau V_\kappa-V_\tau V_{\tau^*}-V_{\tau^*}{}^*V_\tau)\\&\quad-U_\kappa V_\tau+U_\tau V_\kappa,\end{aligned} \tag{10b}$$

$$\begin{aligned}{}^{(3)}R_{\kappa\tau}&=(\tau\tau^*)\ulcorner(dV-UV)\\&=(d_\tau V_{\tau^*}+V_{\tau^*}{}^*V_\kappa+U_\tau V_{\tau^*})\\&\quad-(d_{\tau^*}V_\tau+V_{\tau^*}V_\kappa-U_{\tau^*}V_\tau)-U_\tau V_{\tau^*}+U_{\tau^*}V_\tau^*.\end{aligned} \tag{10c}$$

II. CHOICE OF TRIAD

In the case of stationary space—times with geodetic eigenrays, one may introduce a *Perjés* triad; i.e., a triad such that $G_\tau=V_\kappa=0$. Such space—times have been the subject of a recent investigation by Kota and Perjés.[6]

More generally it might be advantageous to choose a triad such that $G_\kappa=0$. However, because our principal objective in this paper is to facilitate a comparison of the Chandrasekhar—Friedman[7] and the Press—Teukolsky—Wald[8] treatments of perturbations of the Kerr metric, we shall direct our attention to an alternative triad choice for which $U_\kappa=\operatorname{Im}V_\tau=\operatorname{Im}V_{\tau^*}=0$.

From the definitions, Eqs. (7), of the fields U and V we may infer that

$$d\kappa=V^*\tau+V\tau^* \text{ and } d\tau=U\tau-V\kappa. \tag{11}$$

Under the assumption that $\operatorname{Im}V_{\tau^*}=0$ there are no $\tau\tau^*$ terms in the first of these equations. By the Frobenius theorem there must exist real fields R and ϕ such that $\kappa=R\,d\phi$. Under the further assumption that $\operatorname{Im}V_\tau=0$ we may define a real field α such that $V_\tau=-d_\kappa\alpha$. It is then possible to establish the existence of fields P and ζ such that

$$\tau\cosh\alpha+\tau^*\sinh\alpha=P^{-1}d\zeta.$$

Thus, we arrive at the *Chandrasekhar—Friedman* triad

$$\kappa=Rd\phi,\ \tau=P^{-1}[\cosh\alpha\, d\zeta-\sinh\alpha\, d\zeta^*], \tag{12}$$

for which

$$\begin{aligned}&U_\kappa=0,\ U_\tau=-d_\tau(\ln P)+d_{\tau^*}\alpha,\ U_{\tau^*}=d_{\tau^*}(\ln P)-d_\tau\alpha,\\&V_\kappa=-d_\tau(\ln R),\ V_\tau=-d_\kappa\alpha,\ V_{\tau^*}=-d_\kappa(\ln P),\\&d_\kappa=R^{-1}\partial/\partial\phi,\ d_\tau=2^{-1/2}P[\cosh\alpha\nabla+\sinh\alpha\nabla^*].\end{aligned} \tag{13}$$

By substituting Eqs. (13) into Eqs. (8), (9), and (10) one may derive the CF form of the equations governing stationary gravitational fields.

III. PERTURBATION THEORY

We shall restrict attention to stationary vacuum fields which are almost axially symmetric, for which we may write

$$d_\tau=2^{-1/2}P(\nabla+\delta\alpha\nabla^*) \text{ and } d_{\tau^*}=2^{-1/2}P(\nabla^*+\delta\overline{\alpha}\nabla). \tag{14}$$

The first-order perturbations of the vacuum field equations assume the following form:

$$\begin{aligned}0=\delta R_{44}&=\tfrac{1}{4}R^{-1}[\eth(R\eth^*\delta\epsilon)+\eth^*(R\eth\delta\epsilon)]\\&\quad+\tfrac{1}{4}[\eth\delta(\ln R)\eth^*\epsilon+\eth^*\delta(\ln R)\eth\epsilon]\\&\quad-\tfrac{1}{2}f^{-1}[\eth\epsilon\eth^*(\delta\epsilon)+\eth(\delta\epsilon)\eth^*\epsilon]\\&\quad+\tfrac{1}{2}f^{-1}\delta(\ln f)\eth\epsilon\eth^*\epsilon\\&\quad+\tfrac{1}{2}R^{-1}[\eth(R\delta\overline{\alpha}\eth\epsilon)+\eth^*(R\delta\alpha\eth^*\epsilon)]\\&\quad-\tfrac{1}{2}f^{-1}[\delta\overline{\alpha}(\eth\epsilon)^2+\delta\alpha(\eth^*\epsilon)^2]+\tfrac{1}{2}R^{-2}\frac{\partial^2}{\partial\phi^2}(\delta\epsilon),\end{aligned} \tag{15a}$$

$$\begin{aligned}0=\delta[f^{-1}R_{km}]&=\tfrac{1}{2}R^{-1}\eth^*\eth(\delta R)\\&\quad+\tfrac{1}{2}R^{-1}[\eth(\delta\overline{\alpha}\eth R)+\eth^*(\delta\alpha\eth^*R)]\\&\quad-R^{-2}\frac{\partial^2}{\partial\phi^2}\delta(\ln P),\end{aligned} \tag{15b}$$

$$\begin{aligned}0=\delta[f^{-1}R_{tt}]&=\tfrac{1}{2}\eth[\eth\delta(\ln R)]+\eth(\ln R)\eth\delta(\ln R+\ln P)\\&\quad+\tfrac{1}{4}f^{-2}[\eth\epsilon\eth(\delta\epsilon^*)+\eth(\delta\epsilon)\eth\epsilon^*]\\&\quad-\tfrac{1}{2}f^{-2}\delta(\ln f)\eth\epsilon\eth\epsilon^*\\&\quad-\tfrac{1}{2}\eth^*(\delta\alpha)\eth(\ln R)+\tfrac{1}{2}\eth(\delta\alpha)\eth^*(\ln R)\\&\quad+\tfrac{1}{4}f^{-2}\delta\alpha[\eth\epsilon\eth^*\epsilon^*+\eth^*\epsilon\eth\epsilon^*]\\&\quad-R^{-2}\frac{\partial^2}{\partial\phi^2}(\delta\alpha),\end{aligned} \tag{15c}$$

$$\begin{aligned}0&=\delta[f^{-1}(R_{kt}+R_{mt})]\\&=\frac{\partial}{\partial\phi}\{\eth^*(R^{-1}\delta\alpha)-\eth[R^{-1}\delta(\ln P)]\\&\quad+\tfrac{1}{4}R^{-1}f^{-2}(\eth\epsilon^*\delta\epsilon+\eth\epsilon\delta\epsilon^*)\}.\end{aligned} \tag{15d}$$

The last of these equations is designated by Chandrasekhar and Friedman as an "initial value" equation, while the others are "dynamical" equations. By treating the fields $\delta\alpha$ and $\delta\overline{\alpha}$ as fields of spin weights $+2$ and -2, respectively, we have been able to make use of the "thop" idea introduced in Ref. 1.

All of the foregoing equations may be expressed in terms of the notation of Chandrasekhar and Friedman by making the substitutions

$$\begin{aligned}&R=e^{n+p},\ f=e^{2n},\ P=e^{-n-\mu},\ \alpha=\tau,\\&\delta\epsilon=2e^{2n}\delta n+iQ,\ \nabla\epsilon=2e^{2n}\nabla n-e^{3n-p}\nabla\omega.\end{aligned} \tag{16}$$

For example, the dynamical equation for Q arises from the imaginary part of Eq. (15a):

$$\begin{aligned}&\tfrac{1}{2}\nabla(e^{-3n+p}\nabla^*Q)+\tfrac{1}{2}\nabla^*(e^{-3n+p}\nabla Q)\\&\quad+e^{-3n-p+2\mu}\frac{\partial^2}{\partial\phi^2}(Q)\\&\quad+\tfrac{1}{2}i\nabla[(3\delta n-\delta p)\nabla^*\omega+2\delta\tau\nabla\omega]\\&\quad-\tfrac{1}{2}i\nabla^*[(3\delta n-\delta p)\nabla\omega+2\delta\tau\nabla^*\omega]=0.\end{aligned} \tag{17}$$

On the other hand the initial value equation assumes the form

$$\frac{\partial}{\partial\phi}[\nabla(\delta n+\delta\mu)-\nabla p(\delta n+\delta\mu)+\nabla n(\delta n-\delta\mu)$$
$$+\tfrac{1}{2}ie^{-n-p}\nabla\omega Q+\nabla^*(\delta\tau)+\nabla^*(n-p+2\mu)\delta\tau]=0. \tag{18}$$

Equations (17) and (18) should be compared with CF Eq. (45) and CF Eq. (43), respectively.

We shall not write out the other dynamical equations, for they have been given by Chandrasekhar and Friedman, and the above examples should suffice to show the relation between their formalism and ours.

In CF perturbation theory it is impossible to assess results without a knowledge of the curvature tensor. Therefore, we offer the following expressions for the perturbations of the five complex scalars associated with the Weyl conform tensor:

$$\begin{aligned}\delta C_2 = {} & 2C_2\delta(\ln P)-\tfrac{1}{4}\eth\eth(\delta\epsilon)-\tfrac{1}{2}\eth\,\delta(\ln P)\eth\epsilon\\ & -\tfrac{1}{4}f^{-1}\eth\epsilon\eth(\delta\epsilon)+\tfrac{1}{8}f^{-1}\delta(\ln f)(\eth\epsilon)^2\\ & -\tfrac{1}{2}\delta\alpha\eth^*\eth\epsilon+\tfrac{1}{4}\eth^*(\delta\alpha)\eth\epsilon-\tfrac{1}{4}\eth(\delta\alpha)\eth^*\epsilon\\ & -\tfrac{1}{4}f^{-1}\delta\alpha\eth\epsilon\eth^*\epsilon,\end{aligned} \tag{19a}$$

$$\begin{aligned}\delta C_1 = {} & \tfrac{1}{4}\frac{\partial}{\partial\phi}\{\eth(R^{-1}\delta\epsilon)+R^{-1}\delta(\ln P)\eth\epsilon\\ & +R^{-1}\delta\alpha\eth^*\epsilon+\tfrac{1}{2}f^{-1}R^{-1}\delta\epsilon\eth\epsilon\},\end{aligned} \tag{19b}$$

$$\begin{aligned}\delta C_0 = {} & 2C_0\delta(\ln P)+\tfrac{1}{4}\eth^*\eth(\delta\epsilon)-\tfrac{1}{8}f^{-1}\eth\epsilon\eth^*(\delta\epsilon)\\ & -\tfrac{1}{8}f^{-1}\eth(\delta\epsilon)\eth^*\epsilon+\tfrac{1}{8}f^{-1}\delta(\ln f)\eth\epsilon\eth^*\epsilon\\ & +\tfrac{1}{4}\eth^*(\delta\alpha\eth^*\epsilon)+\tfrac{1}{4}\eth(\delta\bar{\alpha}\eth\epsilon),\end{aligned} \tag{19c}$$

$$\begin{aligned}\delta C_{-1} = {} & -\tfrac{1}{4}\frac{\partial}{\partial\phi}\{\eth^*(R^{-1}\delta\epsilon)+R^{-1}\delta(\ln P)\eth^*\epsilon\\ & +R^{-1}\delta\bar{\alpha}\eth\epsilon+\tfrac{1}{2}f^{-1}R^{-1}\delta\epsilon\eth^*\epsilon\},\end{aligned} \tag{19d}$$

$$\begin{aligned}\delta C_{-2} = {} & 2C_{-2}\delta(\ln P)-\tfrac{1}{4}\eth^*\eth^*(\delta\epsilon)-\tfrac{1}{2}\eth^*\delta(\ln P)\eth^*\epsilon\\ & -\tfrac{1}{4}f^{-1}\eth^*\epsilon\eth^*(\delta\epsilon)+\tfrac{1}{8}f^{-1}\delta(\ln f)(\eth^*\epsilon)^2\\ & -\tfrac{1}{2}\delta\bar{\alpha}\eth^*\eth\epsilon+\tfrac{1}{4}\eth(\delta\bar{\alpha})\eth^*\epsilon-\tfrac{1}{4}\eth^*(\delta\bar{\alpha})\eth\epsilon\\ & -\tfrac{1}{4}f^{-1}\delta\bar{\alpha}\eth\epsilon\eth^*\epsilon.\end{aligned} \tag{19e}$$

It should be noted that throughout this section we have taken advantage of the vacuum field equations in order to introduce a complex potential ϵ such that $G=d\epsilon$.[9]

IV. PERTURBATIONS OF THE KERR METRIC

In order to identify the fields which satisfy Teukolsky's separable wave equations, we turn our attention to the relation between the Kinnersley tetrad[10]

$$\begin{aligned} k' &= dT - a\sin^2\theta d\phi + (\rho\rho^*\Delta)^{-1}dr,\\ m' &= dr - \tfrac{1}{2}\rho\rho^*\Delta k',\\ t' &= 2^{-1/2}\rho^{-1}\{d\theta + i\rho\rho^*\sin\theta[a\,dT-(r^2+a^2)d\phi]\},\end{aligned} \tag{20}$$

and the tetrad k, m, t, t^* which we have been employing in this paper. A straightforward calculation results in the relations

$$\begin{aligned} k' &= 2^{1/2}(\rho\rho^*\Delta)^{-1/2}[\mathrm{Re}\,t-\tfrac{1}{2}(\Lambda k-\Lambda^{-1}m)],\\ m' &= 2^{-1/2}(\rho\rho^*\Delta)^{1/2}[\mathrm{Re}\,t+\tfrac{1}{2}(\Lambda k-\Lambda^{-1}m)],\\ t' &= \rho^{-1}(\rho\rho^*)^{1/2}[\mathrm{Im}\,t-\tfrac{1}{2}i(\Lambda k+\Lambda^{-1}m)],\end{aligned} \tag{21}$$

where $\Delta = r^2+a^2-2mr$, $\rho^{-1}=r-ia\cos\theta$ and

$$\Lambda^2 = (\Delta^{1/2}+a\sin\theta)/(\Delta^{1/2}-a\sin\theta).$$

The corresponding relations between the bivectors[11] are

$$\begin{aligned} (\rho\Delta^{1/2})B'_+ &= 2^{-1/2}i[\Lambda B_+ + B_0 + \Lambda^{-1}B_-],\\ B'_0 &= -\Lambda B_+ + \Lambda^{-1}B_-,\\ (\rho\Delta^{1/2})^{-1}B'_- &= \tfrac{1}{2}2^{-1/2}i[-\Lambda B_+ + B_0 - \Lambda^{-1}B_-].\end{aligned} \tag{22}$$

In particular, it should be noted that B'_+ and B'_- are the principal null bivectors of the type D Kerr field.

From Eqs. (22) we may immediately infer that the Weyl tensor components C'_i relative to the Kinnersley tetrad are related to the Weyl tensor components C_j relative to our tetrad as follows:

$$(\rho^2\Delta)C'_2 = -\tfrac{1}{2}\Lambda^2C_2+2\Lambda C_1-\tfrac{3}{2}C_0+2\Lambda^{-1}C_{-1}-\tfrac{1}{2}\Lambda^{-2}C_{-2}, \tag{23a}$$

$$4(\rho^2\Delta)^{-1}C'_{-2} = -\tfrac{1}{2}\Lambda^2C_2-2\Lambda C_1-\tfrac{3}{2}C_0-2\Lambda^{-1}C_{-1}-\tfrac{1}{2}\Lambda^{-2}C_{-2}. \tag{23b}$$

In the case of the Kerr metric one has

$$\epsilon = 1-2m\rho,\quad R=\Delta^{1/2}\sin\theta,$$

$$P=(\Delta-a^2\sin^2\theta)^{-1/2},\ \text{and}\ \nabla = \Delta^{1/2}\frac{\partial}{\partial r}+i\frac{\partial}{\partial\theta}. \tag{24}$$

Of course, both C'_2 and C'_{-2} vanish for the unperturbed field. The quantities which satisfy Teukolsky's separable wave equations are $\delta C'_2$ and $\rho^{-4}\delta C'_{-2}$. Fortunately, both of these quantities are invariant under infinitesimal null rotations. Therefore, it suffices to correlate infinitesimal perturbations of our tetrad with infinitesimal perturbations of the Kinnersley tetrad using Eqs. (21). Accordingly, Eqs. (23a) and (23b) provide the relation between the fields which satisfy Teukolsky's equations and the perturbed fields δC_i $(i=-2,\ldots,+2)$, which are given in turn by Eqs. (19).

In the near future it is anticipated that a joint effort will be undertaken with S. Chandrasekhar to clarify further the relation between CF perturbation theory and the recent investigations of Press, Teukolsky, and Wald, particularly as regards inferences which may be drawn concerning the stability of the Kerr metric. The formulas which we have displayed in this paper will play an important role in our subsequent efforts. In addition, it is hoped that Eqs. (3) for the Weyl tensor of stationary fields will be found useful by other researchers.

ACKNOWLEDGMENTS

The author is grateful to S. Chandrasekhar for frequent discussions of CF perturbation theory and for suggesting that an attempt be made to relate this approach to that of Press, Teukolsky, and Wald. He would also like to take the opportunity to thank I. Hauser for his continuing encouragement and helpful suggestions.

*Work supported in part by National Science Foundation grant GP-8553.

[1]F.J. Ernst, J. Math. Phys. 15, 1409 (1974).

[2]These fields, which are analogous to the fields $\psi_4, \ldots, \psi_0$ of Newman and Penrose, were defined in Ref. 1.

[3]We employ throughout this paper notation of *I*. Hauser, concerning which we intend to publish a monograph in due course. In terms of tensor notation Eq. (3a) might be written as follows:

$$C_2 - \tfrac{1}{2}R_{tt} = -\tfrac{1}{2}\tau^j(G^k_{;j} + \tfrac{1}{2}f^{-1}G_jG^k)\tau_k.$$

[4]F.J. Ernst, J. Math. Phys. **12**, 2395 (1971).

[5]Z. Perjés, J. Math. Phys. **11**, 3383 (1970).

[6]J. Kota and Z. Perjés, J. Math. Phys. **13**, 1695 (1972).

[7]S. Chandrasekhar and J. Friedman, Ap. J. **185**, 1 (1973).

[8]S. Teukolsky (to be published); W. Press and S. Teukolsky (to be published); R. Wald, J. Math. Phys. 14, 1453 (1973).

[9]The complex gravitational potential ϵ appeared for the first time in the article by F. Ernst, Phys. Rev. **167**, 1175 (1968).

[10]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).

[11]The bivectors $B_+ = kt$, $B_0 = km + tt^*$, $B_- = mt^*$ were introduced in Ref. 1.

On irreducible corepresentations of finite magnetic groups

P. Rudra

Department of Physics, University of Kalyani, Kalyani, West Bengal 741235, India
(Received 26 June 1973)

We have obtained a set of homogeneous linear equations in the Clebsch–Gordan coefficients for the Kronecker inner direct product of two irreducible corepresentations of a finite magnetic group. The solutions of these equations give the Clebsch–Gordan coefficients even when the group is not simply reducible. The nontrivial Clebsch–Gordan coefficients for the magnetic group $C_{4v}(C_{2v})$ have been evaluated. We have also investigated the criterion determining whether a particular irreducible corepresentation is equivalent to its complex conjugate representation. A projection operator has been constructed for obtaining the basis pertaining to a particular irreducible corepresentation.

1. INTRODUCTION

The physical properties of magnetic crystals are governed not only by groups of space rotations (proper and improper) and translations of space but also by groups which contain time reversal either singly or in conjunction with other rotation or translation operators. The group elements that contain time reversal are antilinear operators.[1,2] The symmetry group M of order $|M|$ for such systems contains an invariant subgroup G of index 2 containing unitary and linear operations on space and a coset of unitary antilinear operations containing time reversal.

$$M = G \cup a_0 G. \tag{1}$$

a_0 is an antilinear operator containing the time reversal operator θ. a_0 can be written as

$$a_0 = \theta v_0, \tag{2}$$

where v_0 is a fixed linear operator so that

$$v_0^2 \in G.$$

The representation theory of such groups has been worked out in detail by Wigner,[1] Dimmock and Wheeler,[3] and by Dimmock.[4] This has been summarized in Sec. 2. The application of this theory to magnetic space groups[5] is also extensive.[6–12] In Sec. 3 we have constructed the projection operator for obtaining the basis functions for an irreducible corepresentations. Their orthogonality properties have also been investigated here. Krocecker inner direct product of two irreducible corepresentations as also the frequency of a particular irreducible corepresentation in the Clebsch–Gordan series can be obtained[13,14] from the unitary linear subgroup G of M. However there is no simple relation for obtaining the actual Clebsch–Gordan coefficients as exists[1] for unitary linear groups. This is due to the peculiar form of orthogonality relations for irreducible corepresentations.[4] In Sec. 4 we have obtained a set of homogeneous linear equations in the Clebsch–Gordan coefficients, the solution of which gives these coefficients for each multiplicity. As an example we have worked out the nontrivial case for C_{4v} (C_{2v}). For linear groups Frobenius and Schur analysed[15] the condition for the equivalence of an irreducible representation with its complex conjugate. In Sec. 5 we have extended this procedure for irreducible corepresentations.

2. COREPRESENTATIONS OF MAGNETIC GROUPS[3,4,14]

The standard results of corepresentation theory have been summarized here for later references. We denote the elements of the linear subgroup G of M by u and elements of the coset a_0G by a. θ commutes with any linear operator. When we consider the representation matrix $D(\theta^2 u)$, we put

$$\theta^2 = \omega = \begin{cases} 1 & \text{for bosons or even number of fermions,} \\ -1 & \text{for odd number of fermions when spins are considered.} \end{cases} \tag{3}$$

The corepresentation matrices satisfy

$$D(\alpha) D^{\alpha}(\beta) = D(\alpha\beta), \ \forall\, \alpha, \beta \in M. \tag{4}$$

A group element $\alpha \in M$ used as the superscript of either a matrix or a complex number will mean

$$A^{\alpha} = \begin{cases} A, & \text{if } \alpha \in G, \\ A^*, & \text{if } \alpha \in M - G, \end{cases}$$

where A is either a matrix or a complex number. Two corepresentations D and D' are V equivalent (we write $D' \overset{V}{\equiv} D$) if there exists a nonsingular matrix V such that

$$D'(u) = V^{-1} D(u) V, \ \forall u \in G,$$
$$D'(a) = V^{-1} D(a) V^*, \ \forall a \in M - G. \tag{5}$$

A similar notation will be used for equivalent representations of linear groups. The bases $|\psi_i\rangle$ forming D and $|\psi_i'\rangle$ forming D' will have the properties,

$$O_\alpha |\psi_i\rangle = \sum_j D(\alpha)_{ji} |\psi_j\rangle,$$
$$O_\alpha |\psi_i'\rangle = \sum_j D'(\alpha)_{ji} |\psi_i'\rangle,$$
$$|\psi_i'\rangle = \sum_j V_{ji} |\psi_j\rangle, \tag{6}$$

where O_α's are the Wigner operators for $\alpha \in M$. The definition of reducibility is the same as for linear groups compatible with the transformation rule (5). The corepresentation matrices can be taken as unitary just as in the case of linear groups. Also Maschke's theorem of reducibility meaning full reducibility is valid.

Schur's lemma for the same irreducible corepresentation[16] D^{μ} would take the form:

If there exists a matrix N, having *at least one real eigenvalue*, satisfying

$$ND^{\mu}(u)=D^{\mu}(u)N, \quad ND^{\mu}(a)=D^{\mu}(a)N^{*} \tag{7}$$

for all $u \in G$ and for all $a \in M-G$
then $N=r\cdot E$, where r is a real number.

The restriction that N shall have at least one real eigenvalue is essential. If N is a Hermitian matrix it satisfies this restriction and we get the form given by Dimmock.[4] Because of this restriction on N, the orthognality relations for the irreducible corepresentations take the rather inconvenient forms[4]

$$\sum_u D^{\mu}(u)_{im}D^{\nu}(u)^{*}_{jn}=0,$$
$$\sum_a D^{\mu}(a)_{in}D^{\nu}(a)^{*}_{jm}=0, \tag{8}$$
$$\text{if } D^{\mu}\cdot\not\equiv\cdot D^{\nu},$$

and

$$\sum_u D^{\mu}(u)_{im}D^{\mu}(u)^{*}_{jn}+\sum_a D^{\mu}(a)_{in}D^{\mu}(a)^{*}_{jm}=\frac{|M|}{d_{\mu}}\delta_{ij}\delta_{mn}. \tag{9a}$$

Also

$$\sum_u |x^{\mu}(u)|^2+\sum_a x^{\mu}(a^2)=|M|, \tag{9b}$$

where

$$x^{\mu}(\alpha)=\text{Trace } D^{\mu}(\alpha).$$

Here d_{μ} is the dimension of the irreducible corepresentation. This inconvenient form of the orthogonality relations and the fact that the traces of the matrices for the antilinear operators are not invariant for equivalent corepresentations[cf. Eq. (5)] makes the powerful method of group algebra and characters inapplicable to magnetic groups.

Wigner's analysis[1] using the irreducible representations of G to classify the complete set of irreducible corepresentations D^{μ} of M in 3 types gives the following result.

(i) *Type a*

$$\Delta^{\bar{\mu}}(u)\equiv\Delta^{\mu}(a_0^{-1}ua_0)^{*}=P^{-1}\Delta^{\mu}(u)P, \quad \forall u\in G$$

with

$$PP^{*}=+\Delta^{\mu}(a_0^2).$$

In this case

$$D^{\mu}(u)=\Delta^{\mu}(u)$$

and

$$D^{\mu}(a)=\pm\Delta^{\mu}(aa_0^{-1})P$$

and

$$\sum_u |x^{\mu}(u)|^2=|M|/2, \sum_a x^{\mu}(a^2)=|M|/2. \tag{10a}$$

A necessary and sufficient condition for this case to happen is that the trace $\psi^{\mu}(u)$ of $\Delta^{\mu}(u)$ will satisfy

$$\sum_a \psi^{\mu}(a^2)=|G|=|M|/2. \tag{10b}$$

(ii) *Type b*

$$\Delta^{\bar{\mu}}(u)\equiv\Delta^{\mu}(a_0^{-1}ua_0)^{*}=P^{-1}\Delta^{\mu}(u)P, \quad \forall u\in G$$

with

$$PP^{*}=-\Delta^{\mu}(a_0^2).$$

In this case

$$D^{\mu}(u)=\begin{pmatrix}\Delta^{\mu}(u) & 0\\ 0 & \Delta^{\mu}(u)\end{pmatrix},$$

$$D^{\mu}(a)=\begin{pmatrix}0 & -\Delta^{\mu}(aa_0^{-1})P\\ \Delta^{\mu}(aa_0^{-1})P & 0\end{pmatrix},$$

and

$$\sum_u |x^{\mu}(u)|^2=2\cdot|M|$$

and

$$\sum_a x^{\mu}(a^2)=-|M|. \tag{11a}$$

The necessary and sufficient condition for this case is

$$\sum_a \psi^{\mu}(a^2)=-|G|=-|M|/2. \tag{11b}$$

(iii) *Type c*

$$\Delta^{\bar{\mu}}(u)\equiv\Delta^{\mu}(a_0^{-1}ua_0)^{*}\cdot\not\equiv\cdot\Delta^{\mu}(u),$$

i.e.,

$$\sum_u \psi^{\mu}(u)\psi^{\bar{\mu}}(u)^{*}=0,$$

where ψ^{μ} and $\psi^{\bar{\mu}}$ are the traces of Δ^{μ} and $\Delta^{\bar{\mu}}$, respectively.

In this case

$$D^{\mu}(u)=\begin{pmatrix}\Delta^{\mu}(u) & 0\\ 0 & \Delta^{\mu}(a_0^{-1}ua_0)^{*}\end{pmatrix},$$

$$D^{\mu}(a)=\begin{pmatrix}0 & \Delta^{\mu}(aa_0)\\ \Delta^{\mu}(a_0^{-1}a)^{*} & 0\end{pmatrix},$$

and

$$\sum_a x^{\mu}(a^2)=0. \tag{12a}$$

The necessary and sufficient condition is

$$\sum_a \psi^{\mu}(a^2)=\sum_a \psi^{\bar{\mu}}(a^2)=0. \tag{12b}$$

These notations will be retained in all subsequent sections.

3. PROJECTION OPERATORS

We first outline the procedure for obtaining the basis functions that transform according a particular irreducible corepresentation D^{μ}. In this section and later on a particular corepresentation will be assumed to be irreducible unless mentioned otherwise. Wigner has given[1] the prescription for obtaining the bases $|\psi_i^{\mu}\rangle$ of D^{μ} starting from $|\varphi_i^{\mu}\rangle$ the bases for Δ^{μ}. A convenient method will be defining a projection operator[1] P_i^{μ} whose action on any arbitrary $|\psi\rangle$ will be the basis function $|\psi_i^{\mu}\rangle$ other than a normalization constant. One such projection operator is

$$P_i^{\mu}=\sum_u D^{\mu}(u)^{*}_{ii_0}O_u+\sum_a D^{\mu}(a)^{*}_{ii_0}O_a \tag{13}$$

so that

$$P_i^\mu |\psi\rangle = A|\psi_i^\mu\rangle,$$

with

$$|A|^{-2} = \frac{|M|}{d_\mu}\mathrm{Re}[\sum_\mu D^\mu(u)^*_{i_0 i_0}\langle\psi|O_u|\psi\rangle + \sum_a D^\mu(a)^*_{i_0 i_0}\langle\psi|O_a|\psi\rangle], \quad (14)$$

where O_α's are the Wigner operators corresponding to $\alpha \in M$. The proof that $|\psi_i^\mu\rangle$ of (13) transform according to D^μ is straightforward. If D^μ's in Eq. (13) are replaced by the projective corepresentations[12,14,17] $D^{\omega,\mu}$ defined by

$$D^{\omega,\mu}(\alpha)D^{\omega,\mu;\alpha}(\beta) = \omega(\alpha,\beta)^{\alpha\beta}D^{\omega,\mu}(\alpha\beta), \ \forall\alpha,\beta \in M \quad (15)$$

with $|\omega(\alpha,\beta)| = 1$ and the $\omega(\alpha,\beta)$'s satisfying[14]

$$\omega(\alpha,\beta)^y\omega(\alpha\beta,y) = \omega(\alpha,\beta y)\omega(\beta,y) \ \forall\alpha,\beta,y \in M \quad (16)$$

then the resulting $P_i^{\omega,\mu}$ acting on $|\psi\rangle$ will give the bases $|\psi_i^{\omega,\mu}\rangle$ pertaining to $D^{\omega,\mu}$. The only restriction is that the starting function should be such that

$$O_\alpha O_\beta|\psi\rangle = \omega(\alpha,\beta)O_{\alpha\beta}|\psi\rangle. \quad (17)$$

For example if $D^{\omega,\mu}$ is a double group representation, $|\psi\rangle$ should be half-integral spin states. This proof is also straightforward manipulation with the $\omega(\alpha,\beta)$'s. We now investigate the orthogonality of the bases obtained according to Eqs. (13) and (14) for D^μ.

If D^μ is of type (a), the dimension d_μ of D^μ is the same as the dimension of Δ^μ and

$$|\mu,i\rangle = A[\sum_\mu \Delta^\mu(u)^*_{i i_0}O_u|\psi\rangle \pm \sum_a [\Delta^\mu(aa_0^{-1})P]^*_{i i_0}O_a|\psi\rangle]$$

and

$$\langle\mu', i'|\mu, i\rangle = \delta_{\mu'\mu}\delta_{i'i}. \quad (18)$$

If D^μ is of type (b), the dimension d_u of D^u is twice the dimension of Δ^μ. We can choose i_0 in Eq. (13) so that $i_0 \leq d_\mu/2$. We divide the bases of D^u in two groups

$$|\mu,i\rangle = A\sum_u \Delta^\mu(u)^*_{i i_0}O_u|\psi\rangle$$

and

$$|\mu,(d_u/2)+i\rangle = A\sum_a [\Delta^\mu(aa_0^{-1})P]^*_{i i_0}O_a|\psi\rangle, \quad (19)$$

with $i \leq d_u/2$.

We obtain

$$\langle\mu', i'|\mu, i\rangle = \langle\mu', (d_u/2)+i'|\mu, (d_\mu/2)+i\rangle = \delta_{\mu'\mu}\delta_{i'i},$$

$$\langle\mu', i'|\mu, (d_\mu/2)+i\rangle = \delta_{\mu'\mu}\delta_{i'i} \times\{\sum_a [\Delta^\mu(aa_0^{-1})P]^*_{i_0 i_0}\langle\psi|O_a|\psi\rangle\} \times[\sum_u \Delta^\mu(u)^*_{i_0 i_0}\langle\psi|O_u|\psi\rangle]^{-1}, \quad (20)$$

for $i', i \leq d_u/2$.

If D^u is of type (c) the dimensionality d_μ of D^u is again twice that of Δ^u and we can choose $i_0 \leq d_u/2$.

Here also,

$$|\mu,i\rangle = A\sum_u \Delta^\mu(u)^*_{i i_0}O_u|\psi\rangle,$$

$$|\mu,(d_u/2)+i\rangle = A\sum_a \Delta^\mu(a_0^{-1}a)_{i i_0}O_a|\psi\rangle, \ i \leq d_u/2,$$

$$\langle\mu', j'|\mu, j\rangle = \delta_{\mu'\mu}\delta_{j'j}, \ j', j \leq d_u. \quad (21)$$

In all the deductions we have used the following property of scalar product for linear and antilinear operators.

$$\langle O_a Z\psi|O_{a'}Z'\psi'\rangle = ZZ'^*\langle\psi'|O_{a'^{-1}a}|\psi\rangle,$$

$$\langle O_a Z\psi|O_u Z'\psi'\rangle = ZZ'\langle\psi|O_{a^{-1}u}|\psi'\rangle^*,$$

$$\langle O_u Z\psi|O_a Z'\psi'\rangle = Z^*Z'^*\langle\psi|O_{u^{-1}a}|\psi'\rangle,$$

and of course

$$\langle O_u Z\psi|O_{u'}Z'\psi'\rangle = Z^*Z'\langle\psi|O_{u^{-1}u'}|\psi'\rangle, \quad (22a)$$

where Z and Z' are arbitrary complex numbers. It also follows from the properties of antilinear operators that

$$\langle\psi|O_a|\psi\rangle = \langle\psi|O_{a^{-1}}|\psi\rangle \quad (22b)$$

Thus we see that only in the case of type (b) corepresentation the bases are not orthogonal. In case (b) if we try to orthogonalize the bases by Schmidt procedure the matrices cease to be unitary. This difficulty remains. In most of the problems it is more convenient to work with unitary matrices and hence we give up working with an orthogonal base.

4. CLEBSCH-GORDAN COEFFICIENTS

The Kronecker inner direct product of two corepresentations D^μ and D^ν is in general reducible and hence is a direct sum of irreducible components D^λ:

$$D^{\mu\otimes\nu} \equiv D^\mu \otimes D^\nu = \sum_\lambda{}^{\oplus} d_\lambda^{\mu\nu}D^\lambda. \quad (23)$$

The frequency of D^λ in $D^{\mu\otimes\nu}$, $d_\lambda^{\mu\nu}$ is determined by the linear part G of the full group M and is given by Karavaev[13]

$$d_\lambda^{\mu\nu} = \sum_u x^\mu(u)x^\nu(u)x^\lambda(u)^* / \sum_u |x^\lambda(u)|^2. \quad (24)$$

Bradley and Davies[14] have investigated the connection between the numbers $d_\lambda^{\mu\nu}$ and the corresponding numbers $C_\lambda^{\mu\nu}$ for the irreducible representations Δ^μ's of G

$$\Delta^{\mu\otimes\nu} \equiv \Delta^\mu \otimes \Delta^\nu = \sum_\lambda{}^{\oplus} C_\lambda^{\mu\nu}\Delta^\lambda. \quad (25)$$

The Clebsch–Gordan coefficients are required when we are interested in the basis of different corepresentations D^λ formed from the product space of the bases of D^μ and D^ν. Since $d_\lambda^{\mu\nu}$ may be greater than 1, we introduce the index τ_λ for a particular repetition of D^λ. The lth basis of the τ_λth repetition of D^λ in $D^{\mu\otimes\nu}$ is given in terms of the product bases $|\mu m;\nu n\rangle$ of $D^{\mu\otimes\nu}$ in terms of the Clebsch–Gordan coefficient $\langle\mu m;\nu n|\tau_\lambda\lambda l\rangle$:

$$|\tau_\lambda\lambda l\rangle = \sum_{m,n}\langle\mu m;\nu n|\tau_\lambda\lambda l\rangle|\mu m;\nu n\rangle. \quad (26)$$

Operating by O_u and O_a on both sides, summing over all $u \in G$ and $a \in M - G$, remembering the antilinearity of O_a's and the orthogonality relations (8) and (9), we obtain a set of homogeneous linear equations in the Clebsch–Gordan coefficients

$$\sum_{m,n}[\langle\mu m;\nu n|\tau_\lambda\lambda l\rangle\sum_u D^\mu(u)_{im}D^\nu(u)_{jn}D^\lambda(u)^*_{s's}$$
$$+\langle\mu m;\nu n|\tau_\lambda\lambda s\rangle^*\sum_a D^\mu(a)_{im}D^\nu(a)_{jn}D^\lambda(a)^*_{s'l}]$$
$$=\delta_{l,s}(|M|/d_\lambda)\cdot\langle\mu i;\nu j|\tau_\lambda\lambda s'\rangle. \tag{27}$$

The orthogonality of the Clebsch–Gordan coefficients will be

$$\sum_{\substack{m,n\\ m',n'}}\langle\mu m';\nu n'|\tau'_{\lambda'}\lambda' l'\rangle^*\langle\mu m;\nu n|\tau_\lambda\lambda l\rangle\langle\mu m'|\mu m\rangle\langle\nu n'|\nu n\rangle$$
$$=\delta_{\tau'_{\lambda'},\tau_\lambda}\delta_{\lambda',\lambda}\langle\lambda l'|\lambda l\rangle. \tag{28}$$

In Eq. (28) the expression $\langle\mu m'|\mu m\rangle$ and similar ones are Kronecker $\delta_{m',m}$ if D^u is of type (a) or type (c) but are not always $\delta_{m',m}$ for type (b) [cf. Eqs. (18), (20), (21)]. Moreover for type (b) representation $\langle\mu m'|\mu m\rangle$ depends on the particular choice of the bases. This is very important for groups which are not simply reducible. These factors may be different for different repetition of particular irreducible component, thus introducing a lack of essential uniqueness in the Clebsch–Gordan coefficients for this type of groups. Solving Eqs. (27) and (28), we shall get all the Clebsch–Gordan coefficients. We work this out for the magnetic group

$$C_{4v}(C_{2v})=C_{2v}\cup\theta C_4^{13}C_{2v}.$$

It has[14] three irreducible corepresentations, two of them D^1 and D^2 being one-dimensional and the remaining one, D^3, being two-dimensional. D^3 is of type (c). Table V of Bradley and Devies[14] shows that the nontrivial case is the Kronecker inner direct product $D^{3\otimes 3}=2D^1\oplus 2D^2$. We have here 8 unknown C–G coefficients $\langle D^3_m;D^3_n|^{\tau_\lambda}D^\lambda_1\rangle$ where m and n can take the values 1 and 2.

Solutions of Eqs. (27) and (28) give the nonvanishing Clebsch–Gordan efficients

$$\langle D_1^3;D_1^3|^1D_1^1\rangle=\frac{1}{\sqrt2}\exp[-i\alpha],\ \langle D_2^3;D_2^3|^1D_1^1\rangle=\frac{1}{\sqrt2}\exp[i\alpha],$$
$$\langle D_1^3;D_2^3|^1D_1^2\rangle=\frac{1}{\sqrt2}\exp[-i\beta],\ \langle D_2^3;D_1^3|^1D_1^2\rangle=-\frac{1}{\sqrt2}\times\exp[i\beta],$$
$$\langle D_1^3;D_1^3|^2D_1^1\rangle=\mp\frac{i}{\sqrt2}\exp[-i\alpha],\ \langle D_2^3;D_2^3|^2D_1^1\rangle=\pm\frac{i}{\sqrt2}\exp[i\alpha],$$
$$\langle D_1^3;D_2^3|^2D_1^2\rangle=\mp\frac{i}{\sqrt2}\exp[-i\beta],\ \langle D_2^3;D_1^3|^2D_1^2\rangle=\mp\frac{i}{\sqrt2}\exp[i\beta], \tag{29}$$

where α and β are arbitrary real numbers. We have written here $\langle D^\mu_m;D^\nu_n|^{\tau_\lambda}D^\lambda_l\rangle$ for $\langle\mu m;\nu n|\tau_\lambda\lambda l\rangle$.

5. REALITY OF COREPRESENTATIONS

In this section we investigate the criterion when an irreducible corepresentation $D^{\mu*}$ is equivalent to D^μ in the sense of Eqs. (5), i.e.,

$$D^{\mu*}\cdot\overset{C}{\equiv}\cdot D^\mu.$$

A simple calculation shows that CC^+ satisfies all the conditions of Schur's lemma for irreducible corepresentations, hence $CC^+=r\cdot E$, where r is a real number. Thus C can be chosen as a unitary matrix. The procedure for linear groups is not applicable because of the restriction on N having at least one real eigenvalue in proving the Schur's lemma for irreducible corepresentations. Hence we investigate the critertion for the 3 types of corepresentations separately. The proof of the following result is straightforward but laborious application of Schur's lemma and is omitted.

(i) D^μ is of type (a) or type (b)

$$D^{\mu*}\cdot\overset{C}{\equiv}\cdot D^\mu\Longleftrightarrow\Delta^{\mu*}\cdot\overset{K}{\equiv}\cdot\Delta^\mu \tag{30}$$

Now, if $\Delta^{u*}\cdot\overset{K}{\equiv}\cdot\Delta^u$, we have[15] $\tilde K=\alpha K$, with $\alpha=\pm1$.

Also, a necessary and sufficient condition for $\Delta^\mu\circ\overset{k}{\equiv}\cdot\Delta^\mu$ is

$$\sum_u\psi^\mu(u^2)=\alpha|G|=\alpha(|M|/2). \tag{31}$$

In both cases we can choose

$$PK^*=KP. \tag{32}$$

For type (a) corepresentations

$$C=K \tag{33a}$$

and for type (b) corepresentations

$$C=\begin{pmatrix}K&0\\0&K\end{pmatrix} \tag{33b}$$

so that for both the cases

$$\tilde C=\alpha C$$

with

$$\alpha=\pm1. \tag{33c}$$

(ii) D^μ is of type (c)

$$D^{\mu*}\circ\overset{C}{\equiv}\cdot D^\mu\Longleftrightarrow\text{either }\Delta^{\mu*}\circ\overset{K}{\equiv}\circ\Delta^\mu,$$
$$\text{or }\Delta^{\bar\mu*}\circ\overset{K}{\equiv}\circ\Delta^\mu. \tag{34}$$

If $\Delta^{\mu*}\circ\overset{K}{\equiv}\circ\Delta^\mu$, $\tilde K=\alpha K$, with $\alpha=\pm1$ and the necessary and sufficient conditions are

$$\sum_u\psi^\mu(u^2)=\sum_u\psi^{\bar\mu}(u^2)=\alpha(|M|/2)$$

and

$$\sum_u\psi^\mu(u)\psi^{\bar\mu}(u)=0. \tag{35}$$

In this case

$$C=\begin{pmatrix}K&0\\0&K^*\end{pmatrix} \tag{36}$$

with $\tilde C=\alpha C$, where $\alpha=\pm1$.

On the other hand, when $\Delta^{\bar\mu}*\circ\overset{K}{\equiv}\circ\Delta^\mu$ with $KK^*=E$, the necessary and sufficient conditions are

$$\sum_u\psi^\mu(u^2)=\sum_u\psi^{\bar\mu}(u^2)=0$$

and

$$\sum_u\psi^\mu(u)\psi^{\bar\mu}(u)=|M|/2. \tag{37}$$

In this case

$$C=\begin{pmatrix} 0 & K \\ \tilde{K} & 0 \end{pmatrix} \tag{38}$$

with $\tilde{C}=\alpha C$, where $\alpha=+1$. For all the 3 types of irreducible corepresentations we find on direct calculations

$$\sum_u x^\mu(u^2)+\alpha\sum_a \left|\mathrm{Tr}\ CD^\mu(a)^*\right|^2=\alpha\left|M\right|$$

$$\text{if } D^{\mu *}\cdot\overset{c}{\equiv}\cdot D^\mu . \tag{39}$$

[1] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), English translation.

[2] E. P. Wigner, J. Math. Phys. **1**, 409, 414 (1960).

[3] J. O. Dimmock and R. G. Wheeler, J. Phys. Chem. Solids **23**, 729 (1962).

[4] J. O. Dimmock, J. Math. Phys. **4**, 1307 (1963).

[5] W. Opechowski and R. Guccione, in *Magnetism*, edited by G. T. Rado and H. Suhl (Academic, New York, 1965), Vol. IIA.

[6] G. F. Karavaev, N. V. Kudryavsteva, and V. A. Chaldyshev, Sov. Phys.—Solid State **4**, 2540 (1963).

[7] A. P. Cracknell, Prog. Theor. Phys. **33**, 812 (1965).

[8] C. J. Bradley and A. P. Cracknell, Prog. Theor. Phys. **36**, 648 (1966).

[9] S. L. Altmann and A. P. Cracknell, Rev. Mod. Phys. **37**, 19 (1965).

[10] A. P. Cracknell, Prog. Theor. Phys. **35**, 196 (1966).

[11] N. V. Kudryavtseva, Sov. Phys.—Solid State **7**, 803 (1965).

[12] M. V. Murthy, J. Math. Phys. **7**, 853 (1966).

[13] G. F. Karavaev, Sov. Phys.—Solid State **6**, 2943 (1965).

[14] C. J. Bradley and B. L. Davies, Rev. Mod. Phys. **40**, 359 (1968).

[15] G. Frobenius and I. Schur, Sitzungsber. Deut. Akad. Wiss. **49**, 186 (1906).

[16] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1970), English translation.

[17] T. Janssen, J. Math. Phys. **13**, 342 (1972).

Dynamics of a multilevel Wigner–Weisskopf atom

E. B. Davies

Mathematical Institute, Oxford, England
(Received 14 March 1973)

We study the dynamics of an atom with a finite number of discrete energy levels weakly coupled to a continuum of energy levels, showing that any bound state undergoes a decay into the continuum which, in the limit as the coupling constant goes to zero, becomes rigorously exponential.

1. INTRODUCTION

We study a Hamiltonian H_λ on $\mathbf{C}^n \oplus L^2(0,\infty)$ given by

$$H_\lambda = H_0 + \lambda H_1, \tag{1.1}$$

where

$$H_0 = \left[\begin{array}{c|c} S & 0 \\ \hline 0 & Q \end{array}\right], \quad H_1 = \left[\begin{array}{c|c} 0 & a\otimes \bar{f} \\ \hline f\otimes \bar{a} & 0 \end{array}\right]. \tag{1.2}$$

In these equations $S: \mathbf{C}^n \to \mathbf{C}^n$ is defined by $Se_r = \omega_r e_r$ where $e_1, \ldots, e_n$ is the standard orthonormal basis of $\mathbf{C}^n$ and $\omega_r > 0$ are all different. The operator Q on $L^2(0,\infty)$ is defined by

$$(Q\varphi)(x) = x\varphi(x) \tag{1.3}$$

and is an unbounded, positive, self-adjoint operator. All the coefficients of the vector $a \in \mathbf{C}^n$ are supposed to be nonzero. The function $f \in L^2(0,\infty)$ is supposed to satisfy some regularity conditions which are discussed in the appendix, and also the more physically significant conditions

$$f(\omega_r) \neq 0 \text{ for } r = 1, \ldots, n. \tag{1.4}$$

The orthogonal projection P is defined by

$$P(u \oplus \varphi) = u \oplus 0 \tag{1.5}$$

for $u \in \mathbf{C}^n$ and $\varphi \in L^2(0,\infty)$. The problem of the paper is to calculate the asymptotic form of the matrix

$$P\, e^{iH_\lambda t} P \tag{1.6}$$

in the limit $\lambda \to 0$. For fixed t it is trivial that this limit exists and is equal to $P\, e^{iH_0 t} P$. To get a more interesting asymptotic expression we work in the interaction representation and write

$$T_\lambda(\tau) = P \exp(-iH_0 t) \exp(iH_\lambda t) P, \tag{1.7}$$

where $\tau = \lambda^2 t$. Then $T_\lambda(\tau): \mathbf{C}^n \to \mathbf{C}^n$ is a linear contraction and the limiting expression is

$$T(\tau) = \lim_{\lambda\to 0} T_\lambda(\tau); \tag{1.8}$$

this limit being taken for constant τ rather than constant t. We show that the limit exists *uniformly with respect to τ for* $0 \leq \tau < \infty$ and obtain an explicit expression for it. Two qualitative features of the limit may be observed. The first is that T_τ is a semigroup:

$$T_\sigma T_\tau = T_{\sigma+\tau} \tag{1.9}$$

for all $\sigma, \tau \geq 0$. This is closely related to the property of exponential decay which holds for the limiting system —note that strict exponential decay cannot hold before taking the weak coupling limit since, for finite λ, H_λ is semibounded.[1] The second feature is that

$$H_0 T_\tau = T_\tau H_0 \tag{1.10}$$

for all $\tau \geq 0$. As we have obtained the results both of these equations appear entirely coincidental, but we feel that there should be deeper reasons for expecting them to hold, in this and possibly more general systems.

In the case $n=1$ the form of the limit T_τ is well-known in the physical literature.[2–4] The point of this work is to investigate what types of estimate are necessary for a rigorous proof of the exponential decay law. For $n>1$ we also show that the different eigenstates of H_0 decay *independently* in the weak coupling limit, although the interaction term H_1 involves only one particular linear combination a of the bound states.

We make some final comments about the assumptions on the model. The continuum is represented by the operator Q, but this may be replaced by any operator with absolutely continuous spectrum. There is no need to assume that the spectrum of S is multiplicity free. For if S has eigenvalues $\omega_1, \ldots, \omega_n$ with corresponding spectral projections $P_1, \ldots, P_n$ and acts on the finite dimensional space V, we may define

$$e_r = P_r a / \|P_r a\| \tag{1.11}$$

to obtain an orthonormal set such that $Se_r = \omega_r e_r$ and $a = \Sigma\langle a, e_r\rangle e_r$. If W is the subspace spanned by $e_1, \ldots, e_n$ then the methods of the paper apply to H_λ restricted to $W \oplus L^2(0,\infty)$, while $H_\lambda = S$ on $V \cap W^\perp$.

2. ESTIMATES OF THE DECAY RATE FOR VERY LARGE TIMES

The method we use is an extension and generalization of that in Ref. 5, where the problem was solved for the case where H_0 has only one discrete energy level.

Lemma 2.1: *If λ is sufficiently small but not zero then H_λ has no pure point spectrum.*

Proof: Suppose

$$\left[\begin{array}{c|c} S & \lambda a\otimes\bar{f} \\ \hline \lambda f\otimes\bar{a} & Q \end{array}\right] \left[\begin{array}{c} u \\ \hline \psi \end{array}\right] = \alpha \left[\begin{array}{c} u \\ \hline \psi \end{array}\right], \tag{2.1}$$

or

$$Su + \lambda a\langle \psi, f\rangle = \alpha u, \tag{2.2}$$

$$\lambda\langle u, a\rangle f + Q\psi = \alpha\psi, \tag{2.3}$$

where either u or ψ is nonzero. If $\psi = 0$ then $Su = \alpha u$, so $\alpha = \omega_r$ for some r and $u = \beta e_r$; also $\langle u, a\rangle = 0$ which contradicts our assumption that all the coefficients of a are nonzero. On the other hand, if $\langle u, a\rangle = 0$ then $Q\psi = \alpha\psi$ and since Q has no point spectrum $\psi = 0$; the contradiction implies that $\langle u, a\rangle \neq 0$. Since

$$(\alpha - Q)\psi = \lambda\langle u, a\rangle f \tag{2.4}$$

a condition for solubility of the equations is that

$$(\alpha - Q)^{-1} f \in L^2(0, \infty). \tag{2.5}$$

Since f is continuous this implies $f(\alpha) = 0$. Substituting this value of ψ into Eq. (2.2) yields

$$Su + \lambda^2 a\langle u, a\rangle \langle (\alpha - Q)^{-1} f, f\rangle = \alpha u \tag{2.6}$$

which can only have a nontrivial solution if

$$\det[S - \alpha + \lambda^2 \langle (\alpha - Q)^{-1} f, f\rangle a \otimes \bar{a}] = 0. \tag{2.7}$$

Since $\langle (\alpha - Q)^{-1} f, f\rangle$ is a bounded function of α the only solutions of this for small λ are near one of $\omega_1, \ldots, \omega_n$. By condition (1.4) it is therefore impossible to satisfy the condition $f(\alpha) = 0$ as well as Eq. (2.7) for small enough λ.

Lemma 2.2: *For all* Im $z > 0$ *and* $u, v \in \mathbb{C}^n$

$$\langle R(z, H_\lambda) u, v\rangle = \frac{p_1(z) + \lambda^2 h(z) p_2(z)}{\prod_{r=1}^{n} (z - \omega_r) - \lambda^2 h(z) p_3(z)}, \tag{2.8}$$

where p_1, p_2, p_3 *are polynomials and*

$$h(z) = \langle R(z, Q) f, f\rangle. \tag{2.9}$$

Proof: We use the well-known perturbation series for the resolvent.[6] If

$$K = H_0 + \lambda\, a \otimes \bar{f} \tag{2.10}$$

then

$$\begin{aligned} R(z, K) &= R(z, H_0) + \lambda R(z, H_0)\, a \otimes \bar{f} R(z, K) \\ &= R(z, H_0) + \lambda R(z, H_0)\, a \otimes \bar{f}\, R(z, H_0) \\ &\quad + \lambda^2 R(z, H_0)\, a \otimes \bar{f} R(z, H_0)\, a \otimes \bar{f} R(z, K) \\ &= R(z, H_0) + \lambda R(z, H_0)\, a \otimes \bar{f}\, R(z, H_0). \end{aligned} \tag{2.11}$$

Since

$$\langle R(z, H_0) a, f\rangle = 0. \tag{2.12}$$

In matrix notation

$$R(z, K) = \left[\begin{array}{c|c} R(z, S) & \lambda R(z, S)\, a \otimes \bar{f} R(z, Q) \\ \hline 0 & R(z, Q) \end{array}\right]. \tag{2.13}$$

Now

$$H_\lambda = K + \lambda f \otimes \bar{a}$$

so

$$\begin{aligned} R(z, H_\lambda) &= R(z, K) + \lambda R(z, K) f \otimes \bar{a}\, R(z, H_\lambda) \\ &= R(z, K) + [1 - \lambda \langle R(z, K) f, a\rangle]^{-1} \\ &\quad \times \lambda R(z, K) f \otimes \bar{a} R(z, K) \end{aligned} \tag{2.14}$$

by repeated resubstitution and taking the limit. If $u, v \in \mathbb{C}^n$ then

$$\begin{aligned} &\langle R(z, H_\lambda) u, v\rangle = \langle R(z, K) u, v\rangle + \\ &\lambda [1 - \lambda \langle R(z, K) f, a\rangle]^{-1} \langle R(z, K) u, a\rangle \langle R(z, K) f, v\rangle \\ &= \langle R(z, S) u, v\rangle + \\ &\lambda^2 [1 - \lambda^2 \langle R(z, Q) f, f\rangle \langle R(z, S) a, a\rangle]^{-1} \\ &\times \langle R(z, S) u, a\rangle \langle R(z, Q) f, f\rangle \langle R(z, S) a, v\rangle \\ &= \{\langle R(z, S) u, v\rangle + \lambda^2 \langle R(z, Q) f, f\rangle \\ &\times [\langle R(z, S) a, v\rangle \langle R(z, S) u, a\rangle - \langle R(z, S) a, a\rangle \langle R(z, S) u, v\rangle]\} \\ &\times \{1 - \lambda^2 \langle R(z, Q) f, f\rangle \langle R(z, S) a, a\rangle\}^{-1} \\ &= \frac{p_1(z) + \lambda^2 h(z) p_2(z)}{\prod_{r=1}^{n} (z - \omega_r) - \lambda^2 h(z) p_3(z)}, \end{aligned}$$

where

$$h(z) = \langle R(z, Q) f, f\rangle, \tag{2.15}$$

$$\begin{aligned} p_1(z) &= \prod_{r=1}^{n} (z - \omega_r) \cdot \langle R(z, S) u, v\rangle, \\ &= \prod_{r=1}^{n} (z - \omega_r) \cdot \sum_{s=1}^{n} (z - \omega_s)^{-1} u_s \bar{v}_s \end{aligned} \tag{2.16}$$

is a polynomial of order $(n - 1)$;

$$\begin{aligned} p_2(z) &= \prod_{r=1}^{n} (z - w_r) [\langle R(z, S) a, v\rangle \langle R(z, S) u, a\rangle \\ &\quad - \langle R(z, S) a, a\rangle \langle R(z, S) u, v\rangle] \\ &= \prod_{r=1}^{n} (z - w_r) \cdot [\sum_{s,t=1}^{n} (z - \omega_s)^{-1} (z - \omega_t)^{-1} \{a_s \bar{v}_s u_t \bar{a}_t - a_s \bar{a}_s u_t \bar{v}_t\}] \\ &= \prod_{r=1}^{n} (z - \omega_r) \cdot [\sum_{s \neq t} (z - \omega_s)^{-1} (z - \omega_t)^{-1} \{a_s \bar{v}_s u_t \bar{a}_t - a_s \bar{a}_s u_t \bar{v}_t\}] \end{aligned} \tag{2.17}$$

is a polynomial of order $(n - 2)$; and

$$p_3(z) = \prod_{r=1}^{n} (z - \omega_r) \cdot \sum_{s=1}^{n} (z - \omega_s)^{-1} a_s \bar{a}_s \tag{2.18}$$

is a polynomial of order $(n - 1)$ with real coefficients such that for $r = 1, \ldots, n$

$$p_3(\omega_r) = \prod_{\{s: s \neq r\}} (\omega_r - \omega_s) |a_s|^2 \neq 0. \tag{2.19}$$

Theorem 2.3. *There are constants* A, $\lambda_0 > 0$ *such that if* $|\lambda| \leq \lambda_0$ *and* $t \geq 0$ *then*

$$\|P \exp(i H_\lambda t) P\| \leq \min\{1, A/\lambda^2 t\}. \tag{2.20}$$

Proof. We note that for Im $z > 0$

$$\begin{aligned} &\int_0^\infty \langle \exp(-i H_\lambda t)\, e_\alpha, e_\beta\rangle \exp(izt)\, dt \\ &= i \langle R(z, H_\lambda) e_\alpha, e_\beta\rangle = i\varphi(z), \text{ say.} \end{aligned} \tag{2.21}$$

Now φ, which we have already computed, has boundary values as Im $z \to 0$, so

$$\int_0^\infty \langle \exp(-i H_\lambda t)\, e_\alpha, e_\beta\rangle \exp(ixt) dt = i\varphi(x + i0), \tag{2.22}$$

where both sides are interpreted as tempered distributions.[7] Therefore

$$\begin{aligned} \langle \exp(-i H_\lambda t) e_\alpha, e_\beta\rangle &= \frac{i}{2\pi} \int_{-\infty}^{\infty} \varphi(x + i0) \exp(-ixt)\, dx, \\ &= \frac{1}{2\pi t} \int_{-\infty}^{\infty} \varphi'(x + i0) \exp(-ixt)\, dx, \end{aligned} \tag{2.23}$$

so

$$|\langle \exp(-i H_\lambda t) e_\alpha, e_\beta\rangle| \leq \frac{1}{2\pi t} \int_{-\infty}^{\infty} |\varphi'(x + i0)|\, dx. \tag{2.24}$$

By Lemma 2.2

$$\varphi(x + i0) = q(x)/\psi(x), \tag{2.25}$$

where

$$q(x) = p_1(x) + \lambda^2 h(x + i0) p_2(x) \tag{2.26}$$

satisfies inequalities of the form

$$|q(x)| \leq A_1 (1 + |x|)^{n-1}, \tag{2.27}$$

$$|q'(x)| \leq A_2 (1 + |x|)^{n-2}, \tag{2.28}$$

and

$$\psi(x) = \prod_{r=1}^{n} (x - \omega_r) - \lambda^2 h(x + i0) p_3(x) \tag{2.29}$$

satisfies inequalities of the form

$$|\psi(x)| \le A_3(1 + |x|)^n, \tag{2.30}$$

$$|\psi'(x)| \le A_4(1 + |x|)^{n-1}. \tag{2.31}$$

Therefore

$$|\varphi'(x + i0)| = |\psi(x) q'(x) - \psi'(x) q(x)| / |\psi(x)|^2$$

$$\le A_5(1 + |x|)^{2n-2} / |\psi(x)|^2. \tag{2.32}$$

To estimate this further we need a *lower* bound on $|\psi|$. We choose a constant $c > \omega_r$ for all $r = 1, \ldots, n$ and extimate ψ separately in the three regions:

$$\Omega_1 = \{x : x \ge c\}, \tag{2.33}$$

$$\Omega_2 = \{x : |x - \omega_r| \le B\lambda^2, \text{ for some } r = 1, \ldots, n\}, \tag{2.34}$$

$$\Omega_3 = \{x : 0 \le x \le c, \text{ but } |x - \omega_r| \ge B\lambda^2, \text{ for all } r\}, \tag{2.35}$$

where B is a constant to be determined.

In Ω_1,

$$\prod_{r=1}^{n} (x - \omega_r)$$

does not vanish; since it increases at infinity faster than $h(x)p_3(x)$ there is a constant A_6 with $0 < A_6 < 1$ and

$$|\psi(x)| \ge A_6(1 + |x|)^n. \tag{2.36}$$

In Ω_2 we let

$$A_7 = \tfrac{1}{2}\min\{|p_3(\omega_r)| : r = 1, \ldots, n\} \tag{2.37}$$

which is nonzero by Eq. (2.19). If λ is small enough then for $|x - \omega_r| \le B\lambda^2$

$$|p_3(x)| \ge A_7 \tag{2.38}$$

and so

$$|\psi(x)| \ge |\operatorname{Im} \psi(x)| = \lambda^2 |\operatorname{Im} h(x + i0)| |p_3(x)|,$$

$$\ge \lambda^2 A_7 |\operatorname{Im} h(x + i0)|,$$

$$= \lambda^2 \pi A_7 |f(x)|^2, \tag{2.39}$$

by Eq. (A12). By Eq. (1.4) and the continuity of f, if λ is small enough

$$|\psi(x)| \ge \lambda^2 A_8, \tag{2.40}$$

where $A_8 > 0$.

In Ω_3 we use the fact that

$$\prod_{r=1}^{n} (x - \omega_r)$$

is continuously differentiable with simple zeros at $\omega_1, \ldots, \omega_n$ alone, to find a constant B_1 such that

$$\left|\prod_{r=1}^{n} (x - \omega_r)\right| \ge B_1 \min\{|x - \omega_r| : r = 1, \ldots, n\}. \tag{2.41}$$

We also let

$$B_2 = \max\{|h(x)p_3(x)| : 0 \le x \le c\}. \tag{2.42}$$

Then if $|x - \omega_r| \ge B\lambda^2$ for all $r = 1, \ldots, n$

$$|\psi(x)| \ge B_1 \min_r |x - \omega_r| - \lambda^2 B_2,$$

$$\ge (B_1 - B_2 B^{-1}) \min_r |x - \omega_r|,$$

$$= \tfrac{1}{2} B_1 \min_r |x - \omega_r|, \tag{2.43}$$

if we define $B = 2B_2 B_1^{-1}$.

Returning to Eq. (2.24), we now obtain

$$|\langle \exp(-iH_\lambda t) e_\alpha, e_\beta \rangle| \le \frac{1}{2\pi t} \sum_{r=1}^{3} \int_{\Omega_r} |\varphi'(x + i0)| \, dx,$$

$$\le \frac{1}{2\pi t} \left[\int_c^\infty A_5 A_6^{-2} (1 + |x|)^{-2} \, dx \right.$$

$$+ n \int_{-B\lambda^2}^{B\lambda^2} A_5 (1 + c)^{2n-2} A_8^{-2} \lambda^{-4} \, dx$$

$$\left. + 2n \int_{B\lambda^2}^{\infty} A_5 (1 + c)^{2n-2} \, 4B_1^{-2} x^{-2} \, dx \right],$$

$$= \frac{1}{2\pi t} (B_3 + B_4 \lambda^{-2}),$$

$$\le A / \lambda^2 t, \tag{2.44}$$

provided λ is small enough.

Before going on to the next section we comment that this calculation gives the initial reason for believing that the rescaled time $\tau = \lambda^2 t$ should be relevant for the description of the decay of the system.

3. THE EXACT WEAK COUPLING LIMIT

Throughout this section we let $\tau = \lambda^2 t$ and take λ, τ to be the independent variables. We study the asymptotic form as $\lambda \to 0$ of

$$T_\lambda(\tau) = P \exp(-iH_0 t) \exp(iH_\lambda t) P. \tag{3.1}$$

The method is to expand $\exp(iH_\lambda t)$ as a perturbation series in λ, which is known to converge for all t and λ since H_1 is a bounded perturbation. [6] The series is

$$\exp(iH_\lambda t) = \exp(iH_0 t) + i\lambda \int_{t_1=0}^{t} \exp[iH_0(t - t_1)] H_1 \exp(iH_0 t_1) dt_1$$

$$+ (i\lambda)^2 \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} \exp(iH_0(t - t_1) H_1 \exp[iH_0(t_1 - t_2) H_1 \exp(iH_0 t_2)$$

$$dt_2 dt_1 + \ldots \tag{3.2}$$

so if $u, v \in \mathbb{C}^n$

$$\langle T_\lambda(\tau) u, v \rangle = \langle u, v \rangle + i\lambda \int_{t_1=0}^{t} \langle \exp(-iH_0 t_1) H_1 \exp(iH_0 t_1) u, v \rangle dt_1$$

$$+ (i\lambda)^2 \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} \langle \exp(-iH_0 t_1) H_1 \exp[iH_0(t_1 - t_2)] H_1 \exp(iH_0 t_2)$$

$$u, v \rangle \, dt_2 dt_1 + \ldots . \tag{3.3}$$

Since H_1 interchanges the spaces $\mathbb{C}^n$ and $L^2(0, \infty)$, the even terms of this series vanish and we get

$$\langle T_\lambda(\tau) u v \rangle = \langle u, v \rangle$$

$$+ (i\lambda)^2 \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} \langle \exp(-iSt_1) a, v \rangle \langle \exp[iQ(t_1 - t_2)] f, f \rangle$$

$$\langle \exp(iSt_2) u, a \rangle dt_2 dt_1$$

$$+ (i\lambda)^4 \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} \int_{t_3=0}^{t_2} \int_{t_4=0}^{t_3} \langle \exp(-iSt_1) a, v \rangle \langle iQ(t_1 - t_2) f, f \rangle .$$

$$\langle \exp[iS(t_2 - t_3)] a, a \rangle \langle \exp[iQ(t_3 - t_4)] f, f \rangle \langle \exp(iSt_4) u, a \rangle$$

$$dt_4 \ldots dt_1 + \ldots$$

$$= \sum_{n=0}^{\infty} I_n(\lambda, \tau), \tag{3.4}$$

where

$$I_n(\lambda,\tau)=(i\lambda)^{2n}\int_{\Lambda_n} q_1(t_1)g(t_1-t_2)q_2(t_2-t_3)g(t_3-t_4)\ldots q_2(t_{2n-2}-t_{2n-1})g(t_{2n-1}-t_{2n})q_3(t_{2n})dt_{2n}\ldots dt_1, \tag{3.5}$$

and

$$g(s)=\langle\exp(iQs)f,f\rangle, \tag{3.6}$$

$$q_1(x)=\langle\exp(-iSx)a,v\rangle, \tag{3.7}$$

$$q_2(x)=\langle\exp(iSx)a,a\rangle, \tag{3.8}$$

$$q_3(x)=\langle\exp(iSx)u,a\rangle, \tag{3.9}$$

and

$$\Lambda_n=\{(t_1\ldots t_{2n}):0\le t_{2n}\le t_{2n-1}\le\ldots\le t_1\le t\}. \tag{3.10}$$

Writing

$$s_r=t_{2r-1}-t_{2r}, \tag{3.11}$$

$$x_r=t_{2r}-t_{2r+1}, \tag{3.12}$$

so that

$$\Lambda_n=\{(s_1,\ldots,s_n,x_1,\ldots,x_n):s_r\ge0,\ x_r\ge0,\ \sum_{r=1}^n(s_r+x_r)\equiv t_1\le t\} \tag{3.13}$$

we obtain

$$I_n(\lambda,\tau)=(i\lambda)^{2n}\int_{\Lambda_n}q_1\{\sum_{r=1}^n(s_r+x_r)\}q_3(x_n)\prod_{r=1}^{n-1}q_2(x_r)\times\prod_{r=1}^n g(s_r)ds_r dx_r$$

$$=\int_{s_1=0}^\infty\int_{s_n=0}^\infty K_n(t-\sum_{r=1}^n s_r,\sum_{r=1}^n s_r)g(s_1)\ldots g(s_n)ds_1\ldots ds_n, \tag{3.14}$$

where $K_n(x,y)=0$ unless $x,y\ge0$, and if this is the case

$$K_n(x,y)=(i\lambda)^{2n}\int_{\Delta_n}q_1\{y+\sum_{r=1}^n x_r\}q_2(x_1)\ldots q_2(x_{n-1})q_3(x_n)dx_1\ldots dx_n, \tag{3.15}$$

where

$$\Delta_n=\{(x_1\ldots x_n):x_r\ge0 \text{ and } \sum x_r\le x\}. \tag{3.16}$$

Further progress depends on estimates of the kernel K.

Lemma 3.1. *If one of* $\nu_1,\ldots,\nu_n$ *is non-zero then there is a constant* A *such that for all* $x\ge0$

$$J_x=|\int_{\Delta_n}\exp[i\sum_{r=1}^n\nu_r x_r]dx_1\ldots dx_n|\le Ax^{n-1}. \tag{3.17}$$

Proof: By permuting indices we may suppose that $\nu_n\ne0$. Writing

$$\sum_{i=1}^r x_i=y_r,\quad \nu_r=\sum_{i=r}^n\mu_i. \tag{3.18}$$

so that

$$\sum_{r=1}^n\nu_r x_r=\sum_{r=1}^n\mu_r y_r, \tag{3.19}$$

$$J_x=|\int_{y_1=0}^x\int_{y_2=y_1}^x\int_{y_n=y_{n-1}}^x\exp\left(i\sum_{r=1}^n\mu_r y_r\right)dy_n\ldots dy_1$$

$$=|\int_{y_1=0}^x\int_{y_{n-1}=y_{n-2}}^x\exp\left(i\sum_{r=1}^{n-1}\mu_r y_r\right)(i\nu_n)^{-1}\{\exp(i\nu_n x)-\exp(i\nu_n y_{n-1})\}dy_{n-1}\ldots dy_1|\le2|\nu_n|^{-1}x^{n-1}/(n-1)!. \tag{3.20}$$

For the remainder of the calculations we let u and v be, respectively, the elements e_p and e_q of the standard orthonormal basis of C^n. Since C^n is finite dimensional and $T_\lambda(\tau)$ is linear the general case can be immediately obtained from the special one.

Lemma 3.2: *There are constants* A_n *such that for all* $x\ge0$ *and* $y\ge0$

$$|K_n(x,y)-(i\lambda)^{2n}\delta_{pq}|a_q|^{2n}\exp(-i\omega_q y)x^n/n!|\le A_n\lambda^{2n}x^{n-1}. \tag{3.21}$$

Proof: We note that

$$q_1(x)=\exp(-i\omega_q x)\,a_q, \tag{3.22}$$

$$q_2(x)=\sum_{r=1}^n\exp(i\omega_r x|a_r|^2, \tag{3.23}$$

$$q_3(x)=\exp(i\omega_p x)\,\bar a_p, \tag{3.24}$$

so

$$K_n(x,y)=(i\lambda)^{2n}a_q\bar a_p\exp(-i\omega_q y)\int_{\Delta_n}\{\prod_{r=1}^{n-1}\exp(-i\omega_q x_r)q_2(x_r)\}\exp[i(\omega_p-\omega_q)x_n\,dx_1\ldots dx_n \tag{3.25}$$

and

$$\exp(-i\omega_q r)q_2(x)=|a_q|^2+\sum_{r\ne q}|a_r|^2\exp[i(\omega_r-\omega_q)x]. \tag{3.26}$$

Expanding the integrand as a sum of exponentials the main contribution comes from the constant term, which is $\delta_{pq}|a_q|^{2n-2}$, while the other integrals can be bounded as required by Lemma 3.1.

We now define

$$I_n(\tau)=[(-\tau)^n/n!]\,\delta_{pq}\,\alpha_q^n, \tag{3.27}$$

where

$$\alpha_q=|a_q|^2\int_0^\infty g(s)\exp(-i\omega_q s)\,ds \tag{3.28}$$

and

$$\operatorname{Re}\alpha_q=\pi|a_q|^2\ |f(\omega_q)|^2>0, \tag{3.29}$$

by the Appendix and Eq. (1.4).

Lemma 3.3: $\lim_{\lambda\to0}I_n(\lambda,\tau)=I_n(\tau)$

uniformly for $0\le\tau\le\tau_0$.

Proof: We first observe that for all x,y

$$|K_n(x,y)|\le\lambda^{2n}x^n\,\|q_1\|_\infty\|q_2\|_\infty^{n-1}\|q_3\|_\infty/n! \tag{3.30}$$

so

$$|I_n(\lambda,\tau)|\le\lambda^{2n}t^n\|q_2\|_\infty^{n-1}\|q_3\|_\infty\|g\|_1^n/n! \tag{3.31}$$

Given $\epsilon>0$ there exists $\delta>0$ such that if $0\le\tau\le\delta$ then

$$|I_n(\lambda,\tau)|<\epsilon/2,\quad |I_n(\tau)|<\epsilon/2, \tag{3.32}$$

so

$$|I_n(\lambda,\tau)-I_n(\tau)|<\epsilon. \tag{3.33}$$

We now estimate the difference for $\delta\le\tau\le\tau_0$.

$$I_n(\lambda,\tau)-I_n(\tau)=R_1+R_2+R_3, \tag{3.34}$$

where

$$R_1=-\int_{s_r\ge0,\,\Sigma s_r\ge t}\frac{1}{n!}(-\tau)^n\delta_{pq}|a_q|^{2n}\exp(i\omega_q\Sigma s_r)\,g(s_1)\ldots\times g(s_n)\,ds_1\ldots ds_n, \tag{3.35}$$

$$R_2=\int_{s_r\geq 0,\Sigma s_r\leq t}[K_n(t-\Sigma s_r,\Sigma s_r)$$
$$-\frac{1}{n!}(i\lambda)^{2n}\delta_{pq}|a_q|^{2n}\exp(-i\omega_q\Sigma s_r)(t-\Sigma s_r)^n]g(s_1)\ldots g(s_n)$$
$$\times ds_1\ldots ds_n, \quad (3.36)$$

$$R_3=\int_{s_r\geq 0,\Sigma s_r\leq t}(i\lambda)^{2n}\delta_{pq}|a_q|^{2n}\exp(-i\omega_q\Sigma s_r)$$
$$\times[(t-\Sigma s_r)^n-t^n]g(s_1)\ldots g(s_n)ds_1\ldots ds_n. \quad (3.37)$$

We estimate these terms separately.

$$|R_1|\leq\frac{1}{n!}\tau_0^n\delta_{pq}|a_q|^{2n}\int_{s_r\geq 0,\Sigma s_r\geq\delta\lambda^{-2}}|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n$$
$$<\epsilon/4 \quad (3.38)$$

if λ is small enough. By Lemma 3.2

$$|R_2|\leq\int_{s_r\geq 0,\Sigma s_r\leq t}A_n\lambda^{2n}(t-\Sigma s_r)^{n-1}|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n,$$
$$\leq A_n\lambda^2\tau_0^{n-1}\int_{s_r\geq 0,\Sigma s_r\leq t}|g(s_1)\ldots g(s_n)|ds_1\ldots ds_1,$$
$$\leq A_n\lambda^2\tau_0^{n-1}\|g\|_1^n,$$
$$<\epsilon/4, \quad (3.39)$$

if λ is small enough.

To estimate R_3 we let a be a constant large enough so that

$$\int_{s_r\geq 0,\Sigma s_r\geq a}\frac{1}{n!}\tau_0^n\delta_{pq}|a_q|^{2n}|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n<\epsilon/4. \quad (3.40)$$

If $\lambda^2\leq\delta/a$ and $\delta\leq\tau\leq\tau_0$ then $t\geq\delta\lambda^{-2}\geq a$ and

$$|R_3|<\int_{s_r\geq 0,\, t\geq\Sigma s_r\geq a}\frac{1}{n!}\lambda^{2n}\delta_{pq}|a_q|^{2n}[t^n-(t-\Sigma s_r)^n]$$
$$\times|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n$$
$$+\int_{s_r\geq 0,\Sigma s_r\leq a}\frac{1}{n!}\lambda^{2n}\delta_{pq}|a_q|^{2n}[t^n-(t-\Sigma s_r)^n]$$
$$\times|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n,$$
$$<\epsilon/4+\frac{1}{n!}\tau_0^n\delta_{pq}|a_q|^{2n}\int_{s_r\geq 0,\Sigma s_r\leq a}\left[1-\left(1-\frac{\Sigma s_r}{t}\right)^n\right]$$
$$\times|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n,$$
$$\leq\epsilon/4+\frac{1}{n!}\tau_0^n\delta_{pq}|a_q|^{2n}\int_{s_r\geq 0,\Sigma s_r\leq a}[1-(1-\lambda^2a\delta^{-1})^n]$$
$$\times|g(s_1)\ldots g(s_n)|ds_1\ldots ds_n,$$
$$\leq\epsilon/4+\frac{1}{n!}\tau_0^n\delta_{pq}|a_q|^{2n}\|g\|_1^n[1-(1-\lambda^2a\delta^{-1})^n],$$
$$<\epsilon/2, \quad (3.41)$$

if λ is sufficiently small. Putting these estimates together proves that if $\delta\leq\tau\leq\tau_0$ and λ is small enough then

$$|I_n(\lambda,\tau)-I_n(\tau)|<\epsilon \quad (3.42)$$

which, with Eq. (3.33), proves the lemma.

After these preliminaries we are now able to prove the main theorem of the paper.

Theorem 3.4:

$$\lim_{\lambda\to 0}\langle T_\lambda(\tau)e_p,e_q\rangle=\delta_{pq}\exp(-\alpha_q\tau) \quad (3.43)$$

uniformly for $0\leq\tau<\infty$, *where* α_q *is given by* Eq. (3.28).

Proof: Given $\epsilon>0$ we can by Theorem 2.3 choose τ_0 large enough so that if $\tau\geq\tau_0$ and $|\lambda|\leq\lambda_0$ then

$$|\exp(-\alpha_q\tau)|<\epsilon/2,\quad|\langle T_\lambda(\tau)e_p,e_q\rangle|<\epsilon/2. \quad (3.44)$$

Therefore

$$|\langle T_\lambda(\tau)e_p,e_q\rangle-\delta_{pq}\exp(-\alpha_q\tau)|<\epsilon. \quad (3.45)$$

For $0\leq\tau\leq\tau_0$ we use the expansion

$$\langle T_\lambda(\tau)e_p,e_q\rangle=\sum_{n=0}^{\infty}I_n(\lambda,\tau). \quad (3.46)$$

Each term of the series is uniformly convergent by Lemma 3.3 and the series is bounded by

$$\sum_{n=0}^{\infty}|I_n(\lambda,\tau)|\leq\sum_{n=0}^{\infty}\tau_0^n\|q_1\|_\infty\|q_2\|_\infty^{n-1}\|q_3\|_\infty\|g\|_1^n/n!$$
$$<\infty. \quad (3.47)$$

Therefore the series converges absolutely uniformly and the limit is

$$\sum_{n=0}^{\infty}I_n(\tau)=\sum_{n=0}^{\infty}(-\tau)^n\delta_{pq}\alpha_q^n/n!$$
$$=\delta_{pq}\exp(-\alpha q\tau). \quad (3.48)$$

If $0\leq\tau\leq\tau_0$ and λ is small enough then

$$|\langle T_\lambda(\tau)e_p,e_q\rangle-\delta_{pq}\exp(-\alpha_q\tau)|<\epsilon \quad (3.49)$$

which, with Eq. (3.45), proves the theorem.

The limiting map T_τ: $C^n\to C^n$ is given by

$$\langle T_\tau u,v\rangle=\sum_{r=1}^{n}\exp(-\alpha_r\tau)u_r\bar{v}_r. \quad (3.50)$$

The fact, stated in the introduction, that T_τ is a semigroup which commutes with H_0, is immediately apparent from the solution.

APPENDIX

The function $f\in L^2(0,\infty)$ is supposed to satisfy some regularity conditions. It is certainly sufficient that f be a C^∞ function of compact support, but in fact we need very much less. Let

$$g(s)=\langle\exp(iQs)f,f\rangle \quad (A1)$$

so that g is a bounded continuous function. *We suppose that*

$$\int_0^\infty(1+s)|g(s)|ds<\infty \quad (A2)$$

so that $g\in L^1(R)$ and $|f|^2$ is a continuous bounded function with

$$|f(x)|^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}g(s)\exp(-isx)ds \quad (A3)$$

for all $-\infty<x<\infty$. Now for all Re $z>0$

$$\int_0^\infty\exp(-iQs)\exp(izs)\,ds=iR(z,Q) \quad (A4)$$

so if

$$h(z)=\langle R(z,Q)f,f\rangle \quad (A5)$$

then

$$h(z)=i\int_0^\infty g(-s)\exp(isz)\,ds. \tag{A6}$$

This function is analytic and bounded in the upper half-plane and can be continuously extended to the real axis with

$$h(x+i0)=i\int_0^\infty g(-s)\exp(isx)ds. \tag{A7}$$

Similarly for Re $z<0$

$$\int_0^\infty \exp(iQs)\exp(-izs)\,ds=-iR(z,Q) \tag{A8}$$

so

$$h(z)=-i\int_0^\infty g(s)\exp(-isz)\,ds \tag{A9}$$

and

$$h(x-i0)=-\int_0^\infty g(s)\exp(-isx)\,ds. \tag{A10}$$

Since $g(-s)=\overline{g(s)}$ for all s it follows that

$$h(x-i0)=\overline{h(x+i0)}. \tag{A11}$$

Also

$$\begin{aligned}\operatorname{Im} h(x+i0)&=(1/2i)\,[h(x+i0)-h(x-i0)],\\ &=\tfrac{1}{2}\int_{-\infty}^\infty g(s)\exp(-isx)\,ds,\\ &=\pi\,|f(x)|^2.\end{aligned} \tag{A12}$$

The general behavior of h is that it is bounded and analytic in the entire complex plane with a cut along that part of the real axis where $f(x)\neq 0$. Finally

$$\begin{aligned}h'(x+i0)&=i\frac{d}{dx}\int_0^\infty g(-s)\exp(isx)\,ds\\ &=-\int_0^\infty sg(-s)\exp(isx)\,ds,\end{aligned} \tag{A13}$$

which is continuous and bounded by our hypothesis on g.

[1]D. N. Williams, Commun. Math. Phys. 314 (1971).
[2]W. Heitler, *The Quantum Theory of Radiation* (Oxford U. P., London, 1954), 3rd ed., p. 163.
[3]M. Levy, Nuovo Cimento 13, 115 (1959).
[4]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1966), Vol. 2, Chap. 21, Sec. 13.
[5]E. B. Davies, "The Harmonic Oscillator in a Heat Bath" (to be published).
[6]T. Kato, *Perturbation Theory of Linear Operators* (Springer, New York, 1966).
[7]H. Bremermann, *Distributions, Complex Variables and Fourier Transforms* (Addison—Wesley, Reading, Mass., 1965).

Clebsch–Gordan coefficients and special functions related to the Euclidean group in three-space*

Jung Sik Rno†

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455†
(Received 16 July 1973)

In this paper the Clebsch–Gordan coefficients of the Euclidean group in 3-space are explicitly and rigorously determined. The results are used to give elegant derivations of identities involving Wigner D functions and spinor functions.

1. INTRODUCTION

Several authors[1–3] studied special functions associated with the representation theory of the Euclidean group in 3-space [$E(3)$]. In 1969 Holman[4] obtained the Clebsch–Gordan (C–G) coefficients of $E(3)$ from the study of the corresponding structure $SO(4)$ by an Inonu–Wigner contraction. His work is, however, incomplete in that full mathematical justification still remains to be made regarding use of contractions for the derivation of C–G coefficients, and that the method of contraction, even if justified, yields no information on the coupling scheme of helicities in the final results. Furthermore, no author, to the best of our knowledge, has explicitly linked the C–G coefficients of $E(3)$ with special functions. This paper fills these gaps.

The twin purposes of the present work are to determine the C–G coefficients in a direct and rigorous manner and to apply them to special functions related to $E(3)$. The latter process not only leads to new generalized identities involving Wigner D functions and spinor functions, but also unveils the hidden group-theoretic structure of known results. The identities of Jackson and Maximon,[5] for instance, turn out to be expressible as a product of two C–G coefficients of $E(3)$.

Our investigations are carried out within the framework of theory of induced representations. Section 2 is devoted to an explicit construction of the unitary irreducible representations of the simply connected covering group of the Euclidean group in 3-space [$E(3)$]. (In this paper this covering group is often named implicitly the Euclidean group in 3-space.) The problem of decomposing the tensor product representation as a direct integral is solved in Sec. 3 together with computation of C–G coefficients. In Sec. 4 we derive an addition-product theorem and integral formula for Wigner D functions. We also determine the matrix elements of tensor product representations and an integral containing a product of three matrix elements. All these lead to various identities involving Wigner D functions and spinor functions in terms of the C–G coefficients. Most of these results are new in such generalized forms. Wigner D functions comprise such special functions as Jacobi polynomials, ultraspherical polynomials, associated Legendre polynomials, and spherical harmonics, while spinor functions become generalized and ordinary spherical Bessel functions as special cases.

2. THE UIR OF $\widetilde{E(3)}$

A. The Euclidean group in 3-space

In this paper we are concerned with the simply connected covering group $\widetilde{E(3)}$ of the proper Euclidean group in 3-space $E(3)$. It is the semidirect product $R^3 \times_\eta SU(2)$ relative to the homomorphism η of $SU(2)$ into the group of automorphisms of R^3. The matrices $\pm A \in SU(2)$ determine the same rotation $\eta(A)$ given by

$$A(r\cdot\sigma)A^{-1} = (\eta(A)r)\cdot\sigma, \tag{1}$$

where σ stands for the Pauli matrices

$$\sigma^1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma^2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma^3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}. \tag{2}$$

We usually write Ar instead of $\eta(A)r$. If

$$A=\begin{pmatrix}a&b\\-\bar b&\bar a\end{pmatrix}$$

with $a\bar a+b\bar b=1$, then $\eta(A)$ has the explicit expression[6]

$$\eta(A)=\begin{pmatrix}\dfrac{a^2-b^2+\bar a^2-\bar b^2}{2} & \dfrac{i(\bar a^2+\bar b^2-a^2-b^2)}{2} & \bar a\bar b+ab\\ \dfrac{i(a^2-b^2-\bar a^2+\bar b^2)}{2} & \dfrac{(\bar a^2+\bar b^2+a^2+b^2)}{2} & i(-\bar a\bar b+ab)\\ -(\bar a b+a\bar b) & i(-\bar a b+a\bar b) & a\bar a-b\bar b\end{pmatrix}. \tag{3}$$

B. The construction of UIRs of $\widetilde{E(3)}$

The dual group $\hat R^3$ of R^3 consists of the unitary characters $\chi^p : a \mapsto e^{ip\cdot a}$ for $a\in R^3$. We identify $\hat R^3$ with the momentum space P^3. Then the group $SU(2)$ acts on P^3 as well as on R^3. The $SU(2)$ orbit of a given $p\in P^3$ consists of all p', where $\|p'\|=\|p\|$ and hence the $SU(2)$ orbits in P^3 are spheres Ω_ρ, $\Omega_\rho=\{p\in P^3 : \|p\|=\rho\geq 0\}$. Thus we can characterize the partition of P^3 into orbits by choosing the following set K of representing the standard momentum $\mathring{p}$:

$$P^3=\bigcup_{\mathring{p}\in K}\Omega(\mathring{p})\equiv\bigcup_{\rho\geq 0}\Omega_\rho, \tag{4}$$

where

$$K=\{\mathring{p}=(0,0,\rho):\rho\geq 0\}.$$

Hence there exist only two different stability groups (little groups),

$$G_{\mathring{p}}=SU(2)\quad\text{for }\mathring{p}\in\Omega_0,$$
$$G_{\mathring{p}}=\widetilde{SO(2)}\quad\text{for }\mathring{p}\in\Omega_\rho\ \ (\rho>0), \tag{5}$$

where $\widetilde{SO(2)}$ is the twofold covering group of $SO(2)$, the group of rotations around the z axis, and it is isomorphic to the multiplicative group of the complex numbers $e^{i\psi/2}$, $0\leq\psi<4\pi$. Thus its UIRs are one-dimensional and of the form

$$\Gamma^s\left(\begin{bmatrix}e^{i\psi/2}&0\\0&e^{-i\psi/2}\end{bmatrix}\right)=e^{is\psi}, \tag{6}$$

where $2s=0,\pm1,\pm2,\ldots$.

The UIRs associated with the trivial orbit Ω_0 are those which act trivially on the translation subgroup and are uniquely determined by a UIR of $SU(2)$. They are of little interest in the present work. The UIRs (ρ, s) associated with an orbit $\Omega_\rho(\rho>0)$ are given by

$$[U^{\rho,S}(a,A)f](p)=e^{ip\cdot a}(\Gamma^S \uparrow SU(2))(p,A)f(A^{-1}p), \tag{7}$$

where $\uparrow$ denotes "induced."
The carrier space of (ρ, s) is $H(\rho, s)$, the Hilbert space of Lebesgue square integrable functions on the manifold Ω_ρ with inner product

$$\langle f,g\rangle=\int_{\Omega_\rho}\overline{f(p)}\cdot g(p)dw(p), \quad f,g\in H(\rho,s), \tag{8}$$

where $dw(p)=\sin\theta d\theta d\varphi$ for $p=(\rho\sin\theta\cos\varphi,\ \rho\sin\theta\sin\varphi,\ \rho\cos\theta)\in\Omega_\rho$. We recall the set K in (4) meets each orbit just once and it is certainly a Borel set in P^3. Thus $\widetilde{E(3)}$ is a regular[7] semidirect product. Therefore, one can conclude that (i) every UIR of $\widetilde{E(3)}$ which acts nontrivially on the translation subgroup is unitarily equivalent to a representation of the form (7) for some choice of constants ρ, s and (ii) two such representations U_1 and U_2 are unitarily equivalent if and only if $\rho_1=\rho_2$ and $s_1=s_2$.

In (7) we set

$$Q(p,A)=(\Gamma^S\uparrow SU(2))(p,A) \tag{9}$$

which is called a multiplier and satisfies

$$Q(p,A_1)Q(A_1^{-1}p,A_2)=Q(p,A_1A_2). \tag{10}$$

Hence we can get[8]

$$Q(p,A)=Q(\mathring{p},A^{-1}_{p\text{-}\mathring{p}})^{-1}Q(\mathring{p},A^{-1}_{p\text{-}\mathring{p}}AA_{A^{-1}p\text{-}\mathring{p}})Q(\mathring{p},A^{-1}_{A^{-1}p\text{-}\mathring{p}}), \tag{11}$$

where the rotation $A_{p\text{-}\mathring{p}}$, for instance, denotes $\eta(A_{p\text{-}\mathring{p}})\mathring{p}=p$ and $R(p,A)\equiv A^{-1}_{p\text{-}\mathring{p}}AA_{A^{-1}p\text{-}\mathring{p}}$ is called the Wigner's rotation with a property

$$\eta(A^{-1}_{p\text{-}\mathring{p}}AA_{A^{-1}p\text{-}\mathring{p}})\mathring{p}=\eta(A^{-1}_{p\text{-}\mathring{p}}A)A^{-1}p=\eta(A^{-1}_{p\text{-}\mathring{p}})p=\mathring{p}. \tag{12}$$

We can see that (11) implies[8] the unitary equivalence between UIRs (ρ, s) corresponding to $Q(p,A)$ and $Q(\mathring{p},R(p,A))$. Thus we often write $Q(p,A)=Q(\mathring{p},R(p,A))$. Making use of (3) and parametrizing $SU(2)$ in Eulerian angles, we can compute $Q(\mathring{p},R(p,A))$. From (7) we can write

$$[U^{\rho,S}(a,A)f](p)=e^{ip\cdot a}Q(\mathring{p},R(p,A))f(A^{-1}p), \tag{13}$$

where if $p=(\rho\sin\theta\cos\varphi,\ \rho\sin\theta\sin\varphi,\ \rho\cos\theta)$ and $A=\begin{pmatrix} a & b\\ -\bar{b} & \bar{a}\end{pmatrix}$, then

$$Q(\mathring{p},R(p,A))$$
$$=\left(\frac{(1-2b\bar{b})\sin\theta-abe^{i\varphi}(1+\cos\theta)+\bar{a}\bar{b}e^{-i\varphi}(1-\cos\theta)}{(1-2b\bar{b})\sin\theta+abe^{i\varphi}(1-\cos\theta)-\bar{a}\bar{b}e^{-i\varphi}(1+\cos\theta)}\right)^{S/2}$$

and if $A^{-1}p=(\rho\sin\theta'\cos\varphi',\ \rho\sin\theta'\sin\varphi',\ \rho\cos\theta')$, then

$$\cos\theta'=(1-2b\bar{b})\cos\theta+(abe^{i\varphi}+\bar{a}\bar{b}e^{-i\varphi})\sin\theta,$$

$$e^{i\varphi'}=\left(\frac{a^2\sin\theta\cdot e^{i\varphi}-\bar{b}^2\sin\theta\cdot e^{-i\varphi}-2a\bar{b}\cos\theta}{\bar{a}^2\sin\theta\cdot e^{-i\varphi}-b^2\sin\theta\cdot e^{i\varphi}-2\bar{a}b\cos\theta}\right)^{1/2}.$$

An orthonormal basis for $H(\rho, s)$ is known[6] as

$$h^u_m(\theta,\varphi)=(-1)^m\cdot(\sqrt{2\pi})^{-1}D^u_{s,m}(\cos\theta)e^{im\varphi},$$

$$u=|s|,\ |s|+1,\cdots,\quad m=-u,-u+1,\cdots,u, \tag{14}$$

where the $D^u_{s,m}(\cos\theta)$ are Wigner D functions [see (A1)].

3. THE TENSOR PRODUCT OF TWO UIRs OF $\widetilde{E(3)}$

A. Clebsch-Gordan series of $\widetilde{E(3)}$

The tensor product $(\rho_1, s_1)\otimes(\rho_2, s_2)$ of two UIRs of $\widetilde{E(3)}$ is defined as a representation,

$$[(U^{\rho_1,S_1}\otimes U^{\rho_2,S_2})(a,A)f](p_1,p_2)\equiv[U^{1,2}(a,A)f](p_1,p_2)$$
$$=\exp[i(p_1+p_2)\cdot a]Q(p_1,A)Q(p_2,A)f(A^{-1}p_1,A^{-1}p_2), \tag{15}$$

on the Hilbert space $H^{1,2}=H(\rho_1,s_1)\otimes H(\rho_2,s_2)$ of Lebesque square integrable functions on $\Omega_{\rho_1}\otimes\Omega_{\rho_2}$ with scalar product

$$\langle f,g\rangle^{1,2}=\int_{\Omega_{\rho_1}\otimes\Omega_{\rho_2}}dw(p_1)dw(p_2)\overline{f(p_1,p_2)}g(p_1,p_2). \tag{16}$$

To the representation $U^{1,2}$ according to (15) belongs the character $\chi^p(a)=e^{ip\cdot a}=\exp[i(p_1+p_2)]\cdot a$ on the momentum space P^3 of $E(3)$. In $U^{1,2}$, therefore, occur only once those representations $U^{\rho,s}$ for which $p_1\in\Omega_{\rho_1}$ and $p_2\in\Omega_{\rho_2}$ exist with the total momentum

$$p=p_1+p_2 \tag{17}$$

from Ω_ρ where $0\le|\rho_1-\rho_2|<\rho<\rho_1+\rho_2$.

We introduce a unit vector q perpendicular to p,

$$q=[(\rho_1^2-\rho_2^2-\rho^2)|p_1+(\rho^2+\rho_1^2-\rho_2^2)p_2]/\rho\sqrt{\lambda(\rho^2,\rho_1^2,\rho_2^2)} \tag{18}$$

where $\lambda(a,b,c)=2(ab+bc+ac)-a^2-b^2-c^2$.

Let $M(\mathring{p})=\{q\in P^3\,|\,q\perp\mathring{p},\ ||q||=1\}$. Because $G_{\mathring{p}}=\widetilde{SO(2)}$ acts on $M(\mathring{p})$ transitively, $M(\mathring{p})$ can be characterized by a standard element, $q_\alpha=(\cos\alpha,\ \sin\alpha,\ 0)$ such that $M(\mathring{p})=G_{\mathring{p}}\cdot q_\alpha$. Introducing a linear transformation that carries simultaneously $\mathring{p}$ to p and q_α to q, we can express q in terms of p and q_α. If $p=(\rho\sin\theta\cos\varphi,\rho\sin\theta\sin\varphi,\rho\cos\theta)$, then

$$q=(-\cos\alpha\sin\varphi-\sin\alpha\cos\varphi\cos\theta,\ \cos\alpha\cos\varphi$$
$$-\sin\alpha\sin\varphi\cos\theta,\ \sin\alpha\sin\theta). \tag{19}$$

Thus we can pass from (p_1,p_2) space to (p,q) space, i.e., from $(\theta_1,\varphi_1,\theta_2,\varphi_2)$ coordinates to $(\rho,\theta,\varphi,\alpha)$ coordinates. By a direct computation we obtain the Jacobian,

$$\rho_1dw(p_1)\cdot\rho_2dw(p_2)=\rho d\rho d\alpha dw(p). \tag{20}$$

One can easily see that the tensor product $(\rho_1,s_1)\otimes(\rho_2,s_2)$ itself is induced by the stability group $G_{\mathring{p}}$. Hence it is sufficient to decompose $(U^{\rho_1,s_1}\otimes U^{\rho_2,s_2})(a,A)$ for $A\in G_{\mathring{p}}$. We obtain

$$[(U^{\rho_1,s_1}\otimes U^{\rho_2,s_2})(a,A)f](\mathring{p},q_\alpha)=\sum_{n=-\infty}^{\infty}\exp\{i(s_1+s_2$$
$$+n)\alpha\}[U^{\rho,s_1+s_2+n}(a,A)f_n](\mathring{p}), \tag{21}$$

where $\sum_n$ is due to the Fourier series expansion with respect to $A^{-1}q_\alpha$, and we made use of

$$Q(\mathring{p},R(p_i(\mathring{p},q_\alpha),A))=e^{is_i(\psi+\alpha)} \tag{22}$$

where

$$A=\begin{pmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2}\end{pmatrix}\in G_{\mathring{p}}^{\circ},\quad 0\leq\psi<4\pi.$$

Therefore, we showed symbolically

Theorem 1: $(\rho_1, s_1)\oplus(\rho_2, s_2)\approx\int_{|\rho_1-\rho_2|}^{\oplus\rho_1+\rho_2} d\rho \sum_{n=-\infty}^{\infty}\oplus(\rho, s_1+s_2+n)$ (23)

The immediate consequences of Theorem 1 are contained in

Corollary 1:

(i) $(\rho_1, s_1)\otimes(\rho_2, s_2)\approx(\rho_1, s_2)\otimes(\rho_2, s_1),$ (24)

(ii) $(\rho_1, s_1)\otimes(\rho_2, s_2)\approx(\rho_1, s_3)\otimes(\rho_2, s_4),$ (25)

if and only if $s_1+s_2=s_3+s_4$ (mod 1).

B. The Clebsch-Gordan coefficients of $\widetilde{E(3)}$

In the decomposition (23) the representations on each side act on functions of different variables and, hence, care is needed in handling the multipliers and basis. Choosing the standard multiplier $Q(\mathring{p}, R(p, A))$ for (ρ, s_1+s_2+n), we must take the "consistent" multipliers $Q(p_i(\mathring{p}, q_\alpha), R(p_i(p, q_\alpha), A))$ rather than the standard multipliers for (ρ_i, s_i), $i=1, 2$. Using the property of multipliers, we obtain[9] the intertwining multiplication functions $Q(\mathring{p}, R(p_i(\mathring{p}, q_\alpha), A_{p\text{-}\mathring{p}}^{-1}))$, $i=1, 2$. Thus the basis of (ρ, s_1+s_2+n) is

$$h_m^u(\theta, \varphi)=[(-1)^m/\sqrt{2\pi}]\cdot D^u_{s_1+s_2+n, m}(\cos\theta)\cdot e^{im\varphi}, \tag{26}$$

$$u=|s|, |s|+1,\cdots,\quad m=-u, -u+1, \ldots, u$$

while the bases of (ρ_1, s_1) and (ρ_2, s_2) are, respectively,

$$Q(\mathring{p}, R(p_1(\mathring{p}, q_\alpha), A_{p\text{-}\mathring{p}}^{-1}))\cdot h^{u_1}_{m_1}(\theta_1, \varphi_1)$$
$$=\frac{(-1)^{m_1}}{\sqrt{2\pi}}\cdot\left(\frac{\cos\theta\sin\chi_1+\sin\theta(\sin\alpha\cos\chi_1-i\cos\alpha)}{\cos\theta\sin\chi_1+\sin\theta(\sin\alpha\cos\chi_1+i\cos\alpha)}\right)^{s_1/2}$$
$$\times D^u_{s_1, m_1}(\cos\theta_1)\cdot e^{im_1\varphi_1}, \tag{27}$$

and

$$Q(\mathring{p}, R(p_2(\mathring{p}, q_\alpha), A_{p\text{-}\mathring{p}}^{-1}))\cdot h^{u_2}_{m_2}(\theta_2, \varphi_2)$$
$$=\frac{(-1)^{m_2}}{\sqrt{2\pi}}\cdot\left(\frac{\cos\theta\sin\chi_2-\sin\theta(\sin\alpha\cos\chi_2-i\cos\alpha)}{\cos\theta\sin\chi_2-\sin\theta(\sin\alpha\cos\chi_2+i\cos\alpha)}\right)^{s_2/2}$$
$$\times D^{u_2}_{s_2, m_2}(\cos\theta_2)\cdot e^{im_2\varphi_2},$$

$$u_i=|s_i|, |s_i|+1,\cdots,\ m_i=-u_i, -u_i+1\cdots, u_i,\ i=1,2,$$

where

$$\sin\chi_i=\sqrt{\lambda}/2\rho_i\rho,\quad i=1,2,$$
$$\cos\chi_1=(\rho^2+\rho_1^2-\rho_2^2)/2\rho_1\rho,\quad \cos\chi_2=(\rho^2+\rho_2^2-\rho_1^2)/2\rho_2\rho. \tag{28}$$

Equipped with these we can define the C—G coefficients of $\widetilde{E(3)}$,

$$\prod_{i=1}^{2} Q(\mathring{p}, R(p_i(\mathring{p}, q_\alpha), A_{p\text{-}\mathring{p}}^{-1}))h^{u_i}_{m_i}(\theta_i, \varphi_i)=\sum_{u=|u_1-u_2|}^{u_1+u_2}\sum_{m=-u}^{u}\sum_{n=-\infty}^{\infty}E(\rho, u, m, s_1+s_2+n|\rho_1, u, m_1, s_1;\rho_2, u_2, m_2, s_2)$$
$$\times h^u_m(\theta, \varphi)[\exp\{i(s_1+s_2+n)\alpha\}]\sqrt{2\pi}. \tag{29}$$

Using orthonormality of basis and substituting multipliers and variables we can express the C—G coefficients in an integral form,

$$E(\rho, u, m, s_1+s_2+n|\rho_1, u_1, m_1, s_1, ;\rho_2, u_2, m_2, s_2)$$
$$=\frac{(-1)^{m_1+m_2-m}}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\exp[-i(s_1+s_2+n)\alpha]$$
$$\times\exp[i(m_1+m_2-m)\varphi]\,\overline{D^u_{s_1+s_2+n, m}(\cos\theta)}$$
$$\times\left(\frac{\cos\theta\sin\chi_1+\sin\theta(\sin\alpha\cos\chi_1-i\cos\alpha)}{\cos\theta\sin\chi_1+\sin\theta(\sin\alpha\cos\chi_1+i\cos\alpha)}\right)^{s_1/2}$$
$$\times\left(\frac{\cos\chi_1\sin\theta+\sin\chi_1(\cos\theta\sin\alpha-i\cos\alpha)}{\cos\chi_1\sin\theta+\sin\chi_1(\cos\theta\sin\alpha+i\cos\alpha)}\right)^{m_1/2}$$
$$\times D^u_{s_1, m_1}(\cos\chi_1\cos\theta-\sin\chi_1\sin\theta\sin\alpha)$$
$$\times\left(\frac{\cos\theta\sin\chi_2-\sin\theta(\sin\alpha\cos\chi_2-i\cos\alpha)}{\cos\theta\sin\chi_2-\sin\theta(\sin\alpha\cos\chi_2+i\cos\alpha)}\right)^{s_2/2}$$
$$\times\left(\frac{\cos\chi_2\sin\theta-\sin\chi_2(\cos\theta\sin\alpha-i\cos\alpha)}{\cos\chi_2\sin\theta-\sin\chi_2(\cos\theta\sin\alpha+i\cos\alpha)}\right)^{m_2/2}$$
$$\times D^{u_2}_{s_2, m_2}(\cos\chi_2\cos\theta+\sin\chi_2\sin\theta\sin\alpha)$$
$$\times\sin\theta\, d\theta\, d\varphi\, d\alpha. \tag{30}$$

Making use of (A11) and (A3) and integrating with respect to θ, φ, and α we obtain[9]

$$E(\rho, u, m_1+m_2, s_1+s_2+n|\rho_1, u_1, m_1, s_1;\rho_2, u_2, m_2, s_2)$$
$$=i^{-s_1-s_2-n}[2/(2u+1)]^{1/2}C(u_1, m_1;u_2, m_2|u, m_1+m_2)$$
$$\times\sum_{k=-u_1}^{u_1}(-1)^k C(u_1, k;u_2, -k-s_1-s_2-n|u, -s_1-s_2-n)$$
$$\times D^{u_1}_{s_1, k}(\cos\chi_1)D^{u_2}_{s_2, -k-s_1-s_2-n}(\cos\chi_2), \tag{31}$$

where $C(\cdot\,;\cdot|\cdot)$ are the C—G coefficients of $SU(2)$.

It is similar to the result obtained by Holman[4] in that two C—G coefficients of $SU(2)$ and two Wigner D functions are present in each expression of the C—G coefficient. Holman derived it indirectly by an Inonu—Wigner contraction from the study of $SO(4)$. His work is, however, incomplete in that full mathematical justification still remains to be made regarding use of contractions for the derivation of C—G coefficients, and his result has no information on the coupling scheme of helicities. The present work is explicit and complete using direct and rigorous method.

In (31) we normalized the $E(\cdot|\cdot\,;\cdot)$ such that the following orthonormal relations hold:

$$\int_{|\rho_1-\rho_2|}^{\rho_1+\rho_2}\rho\, d\rho\sum_{u=|u_1-u_2|}^{u_1+u_2}\sum_{n=-\infty}^{\infty}$$
$$\overline{E(\rho, u, m_1'+m_2', s_1'+s_2'+n|\rho_1, u_1', m_1', s_1';\rho_2, u_2', m_2', s_2')}$$
$$\times E(\rho, u, m_1+m_2, s_1+s_2+n|\rho_1, u_1, m_1, s_1;\rho_2, u_2, m_2, s_2)$$
$$=\rho_1\rho_2\delta_{u_1', u_1}\delta_{m_1', m_1}\delta_{s_1', s_1}\delta_{u_2', u_2}\delta_{m_2', m_2}\delta_{s_2', s_2}. \tag{32}$$

4. The C-G COEFFICIENTS OF $\widetilde{E(3)}$ AND SPECIAL FUNCTIONS

A. C-G coefficients and Wigner D-functions

From (29) we can derive the addition—product theorem for Wigner D functions

$$D^{u_1}_{s_1,m_1}(\cos\chi_1\cos\theta-\sin\chi_1\sin\theta\sin\alpha)\,D^{u_2}_{s_2,m_2}(\cos\chi_2\cos\theta+\sin\chi_2\sin\theta\sin\alpha)$$
$$=\sum_{u=|u_1-u_2|}^{u_1+u_2}\sum_{n=-\infty}^{\infty}\times E(\rho,u,m_1+m_2,s_1+s_2+n\,|\,\rho_1,u_1,m_1,s_1;\rho_2,u_2,m_2,s_2)$$
$$\times\left(\frac{\sin\chi_1\cos\theta+(\sin\alpha\cos\chi_1+i\cos\alpha)\sin\theta}{\sin\chi_1\cos\theta+(\sin\alpha\cos\chi_1-i\cos\alpha)\sin\theta}\right)^{s_1/2}$$
$$\times\left(\frac{\sin\chi_2\cos\theta-(\sin\alpha\cos\chi_2+i\cos\alpha)\sin\theta}{\sin\chi_2\cos\theta-(\sin\alpha\cos\chi_2-i\cos\alpha)\sin\theta}\right)^{s_2/2}$$
$$\times\left(\frac{\cos\chi_1\sin\theta+(\sin\alpha\cos\theta+i\cos\alpha)\sin\chi_1}{\cos\chi_1\sin\theta+(\sin\alpha\cos\theta-i\cos\alpha)\sin\chi_1}\right)^{m_1/2}$$
$$\times\left(\frac{\cos\chi_2\sin\theta-(\sin\alpha\cos\theta+i\cos\alpha)\sin\chi_2}{\cos\chi_2\sin\theta-(\sin\alpha\cos\theta-i\cos\alpha)\sin\chi_2}\right)^{m_2/2}$$
$$\times D^{u}_{s_1+s_2+n,m_1+m_2}(\cos\theta)\exp[i(s_1+s_2+n)\alpha] \tag{33}$$

Expressing (29) in the other set of variables we are given a product theorem for Wigner D functions. We can also get an integral formula for Wigner D functions from (30).

B. The C-G coefficients and matrix elements of (ρ, s)

In the following we write the operator U instead of $U^{\rho,s}$, and we consider the matrix element with respect to the orthonormal basis $\{h^u_m(\theta,\varphi)\}$ given in (14). If $A\in SU(2)$ has Eulerian coordinates $(\varphi,\alpha,\varphi_2)$, we know[6]

$$\langle h^v_n, U(0,A)h^u_m\rangle=T^u_{nm}(A)\delta_{u,v} \tag{34}$$

where

$$T^u_{nm}(A)=(-i)^{n-m}\exp[i(n\varphi_1+m\varphi_2)][2/(2u+1)]^{1/2}D^u_{-n,m}(\cos\alpha)$$

and

$$\langle h^v_n,U(\mathbf{r},I)h^u_m\rangle=[v,n\,|\,\rho,s\,|\,u,m](\mathbf{r})$$
$$=\sqrt{4\pi}\sum_{l=|u-v|}^{u+v}\left[\frac{(2u+1)(2l+1)}{(2v+1)}\right]^{1/2}i^{-l}j_l(\rho r)$$
$$\times\overline{Y^{n-m}_l(\theta_r,\varphi_r)}C(l,0;u,s\,|\,v,s)C(l,n-m;u,m\,|\,v,n), \tag{35}$$

where $C(\cdot;\cdot|\cdot)$ are the C—G coefficients of $SU(2)$ and the $j_l(\rho r)$ are spherical Bessel functions. The $[v,n\,|\,\rho,s\,|\,u,m](\mathbf{r})$ for fixed v are called spinor functions. The functions $j^{v,u}_{s,n}(\rho r)\equiv i^{u-v}[v,n\,|\,\rho,s\,|\,u,n]((0,0,r))$ are called generalized spherical Bessel functions.[3] In particular, $j^{l,0}_{0,0}(\rho r)=j_l(\rho r)$.

By the group property $U(\mathbf{r},A)=U(\mathbf{r},I)U(0,A)=U(0,A)U(A^{-1}\mathbf{r},I)$, we obtain the matrix elements of $E(3)$,

$$\{v,n\,|\,\rho,s\,|\,u,m\}(\mathbf{r},A)\equiv\langle h^v_n,U(\mathbf{r},A)h^u_m\rangle$$
$$=\sum_{m'=-u}^{u}[v,n\,|\,\rho,s\,|\,u,m'](\mathbf{r})\cdot T^u_{m',m}(A)$$
$$=\sum_{n'=-v}^{v}T^v_{n,n'}(A)\cdot[v,n'\,|\,\rho,s\,|\,u,m](A^{-1}\mathbf{r}). \tag{36}$$

The matrix elements $\{v,n\,|\,\rho,s\,|\,u,m\}(\mathbf{r},A)$ satisfy the orthogonality relations,[2]

$$\int_{R^3}d^3\mathbf{r}\int_{SU(2)}dA\,\overline{\{v_1,n_1\,|\,\rho_1,s_1\,|\,u_1,m_1\}(\mathbf{r},A)}\{v_2,n_2\,|\,\rho_2,s_2\,|\,u_2,m_2\}(\mathbf{r},A)=(4\pi^2/\rho_1^2)\delta(\rho_1-\rho_2)\delta_{s_1,s_2}\delta_{u_1,u_2}\delta_{v_1,v_2}\delta_{m_1,m_2}\delta_{n_1,n_2}. \tag{37}$$

The matrix elements of $(U^{\rho_1,s_1}\otimes U^{\rho_2,s_2})(\mathbf{r},A)$ give rise to the product theorem:

$$\{v_1,n_1\,|\,\rho_1,s_1\,|\,u_1,m_1\}(\mathbf{r},A)\{v_2,n_2\,|\,\rho_2,s_2\,|\,u_2,m_2\}(\mathbf{r},A)$$
$$=\frac{1}{\rho_1\rho_2}\sum_{u=|u_1-u_2|}^{u_1+u_2}\sum_{v=|v_1-v_2|}^{v_1+v_2}\sum_{n=-\infty}^{\infty}\int_{|\rho_1-\rho_2|}^{\rho_1+\rho_2}\rho\,d\rho$$
$$\times\overline{E(\rho,v,n_1+n_2,s_1+s_2+n\,|\,\rho_1,v_1,n_1,s_1;\rho_2,v_2,m_2,s_2)}$$
$$\times E(\rho,u,m_1+m_2,s_1+s_2+n\,|\,\rho_1,u_1,m_1,s_1;\rho_2,u_2,m_2,s_2)$$
$$\times\{v,n_1+n_2\,|\,\rho,s_1+s_2+n\,|\,u,m_1+m_2\}(\mathbf{r},A). \tag{38}$$

The above, among others, reduces to a product theorem for generalized and ordinary spherical Bessel functions, respectively.

We now turn to an integral containing a triple product of matrix elements. Making use of (37) and (38) we obtain, for $|\rho_1-\rho_2|<\rho_3<\rho_1+\rho_2$,

$$\int_{R^3}d^3\mathbf{r}\int_{SU(2)}dA\{l_1,m_1,|\rho_1,s_1\,|\,l_1',m_1'\}(\mathbf{r},A)$$
$$\{l_2,m_2\,|\,\rho_2,s_2\,|\,l_2',m_2'\}(\mathbf{r},A)\ \overline{\{l_3,m_3\,|\,\rho_3,s_3\,|\,l_3',m_3'\}(\mathbf{r},A)}$$
$$=\frac{4\pi^2}{\rho_1\rho_2\rho_3}\overline{E(\rho_3,l_3,m_3,s_3\,|\,\rho_1,l_1,m_1,s_1;\rho_2,l_2,m_2,s_2)}$$
$$\times E(\rho_3,l_3',m_3',s_3\,|\,\rho_1,l_1',m_1',s_1;\rho_2,l_2',m_2',s_2)$$
$$\times\delta_{m_3,m_1+m_2}\delta_{m_3',m_1'+m_2'}\delta_{s_3,s_1+s_2+n} \tag{39}$$

for some integer n.

Setting $l_i'=0$, $i=1,2,3$ and using (35), (36), and (A4), we get

$$(4\pi)^{3/2}i^{-l_1-l_2+l_3}\int_0^\infty j_{l_1}(\rho_1 r)j_{l_2}(\rho_2 r)j_{l_3}(\rho_3 r)r^2\,dr$$
$$\times\int_0^{2\pi}\int_0^{\pi}\overline{Y^{m_1}_{l_1}(\theta_r,\varphi_r)}\cdot\overline{Y^{m_2}_{l_2}(\theta_r,\varphi_r)}\,Y^{m_1+m_2}_{l_3}(\theta_r,\varphi_r)$$
$$\times\sin\theta_r d\theta_r d\varphi_r$$
$$=\frac{4\pi^2}{\rho_1\rho_2\rho_3}E(\rho_3,0,0,0\,|\,\rho_1,0,0,0;\rho_2,0,0,0)$$
$$\times\overline{E(\rho_3,l_3,m_1+m_2,0\,|\,\rho_1,l_1,m_1,0;\rho_2,l_2,m_2,0)}, \tag{40}$$

where l_1, l_2, l_3 are nonnegative integers.

Using (31), (A5) and the known integral in terms of 3-j coefficients,[10]

$$\int_0^{2\pi}\int_0^{\pi}Y^{m_1}_{l_1}(\theta,\varphi)Y^{m_2}_{l_2}(\theta,\varphi)Y^{m_3}_{l_3}(\theta,\varphi)\sin\theta\,d\theta\,d\varphi$$

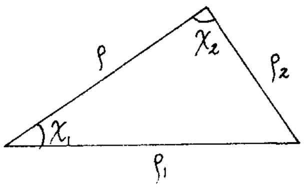

FIG. 1. Angles χ_1 and χ_2.

$$= \left(\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}\right)^{1/2} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{41}$$

we can easily obtain

$$\int_0^\infty j_{l_1}(\rho_1 r) j_{l_2}(\rho_2 r) j_{l_3}(\rho_3 r) r^2\, dr = \frac{\pi}{\rho_1\rho_2\rho_3} i^{l_1+l_2-l_3}[(2l_1+1)(2l_2+1)(2l_3+1)]^{-1/2}$$
$$\times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^{-2} E(\rho_3, 0, 0, 0 \,|\, \rho_1, 0, 0, 0; \rho_2, 0, 0, 0)$$
$$\times E(\rho_3, l_3, 0, 0 \,|\, \rho_1, l_1, 0, 0; \rho_2 l_2, 0, 0)$$
$$= \frac{\pi}{4\rho_1\rho_2\rho_3} i^{l_1+l_2-l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^{-1}$$
$$\times \sum_{m=-\min(l_1,l_2)}^{\min(l_1,l_2)} (-1)^m \begin{pmatrix} l_1 & l_2 & l_3 \\ m & -m & 0 \end{pmatrix} \left(\frac{(l_1-m)!\,(l_2+m)!}{(l_1+m)!\,(l_2-m)!}\right)^{1/2}$$
$$\times P_{l_1}^m(\cos\chi_1) P_{l_2}^{-m}(\cos\chi_2), \tag{42}$$

where the P_l^m are associated Legendre polynomials. Thus we need the restriction that $l_1+l_2+l_3$ is even together with $|l_1-l_2| \le l_3 \le l_1+l_2$ for (42) not to vanish from the properties of the C–G coefficients of $SU(2)$. The above (42) agrees with the known results[5,11] up to an arbitrariness of multiplicative factor of unit modulus, which is applicable to $E(\cdot|\cdot;\cdot)$.

APPENDIX A: WIGNER D-FUNCTIONS

We define Wigner D functions as

$$D^u_{s,m}(\cos\theta) = \left(\frac{(2u+1)(u+s)!\,(u+m)!}{2(u-s)!\,(u-m)!}\right)^{1/2}$$
$$\times \frac{(\sin\theta)^{m+s}(1+\cos\theta)^{u-s-m}}{2^u\Gamma(m+s+1)}$$
$$\times {}_2F_1\left(-u+s, m-u; m+s+1; \frac{\cos\theta-1}{\cos\theta+1}\right), \tag{A1}$$
$$s, m = -u, -u+1, \ldots, u,$$

where the ${}_2F_1$ denote hypergoemetric functions. The following orthogonality relations hold:

$$\int_0^\pi D^u_{n,m}(\cos\theta) D^v_{n,m}(\cos\theta) \sin\theta\, d\theta = \delta_{u,v}. \tag{A2}$$

The Wigner D functions satisfy the relations

$$D^u_{n,m}(\cos\theta) D^{u'}_{n',m'}(\cos\theta) = \sum_{l=|u-u'|}^{u+u'} \left(\frac{(2u+1)(2u'+1)}{2(2l+1)}\right)^{1/2}$$
$$\times C(u, -n; u', -n' \,|\, l, -n-n')$$
$$\times C(u, m; u', m' \,|\, l, m+m') D^l_{n+n', m+m'}(\cos\theta) \tag{A3}$$

and

$$\int_0^\pi D^u_{n,m}(\cos\theta) D^{u'}_{n',m'}(\cos\theta) D^l_{n+n',m+m'}(\cos\theta) \sin\theta\, d\theta$$
$$= \left(\frac{(2u+1)(2u'+1)}{2(2l+1)}\right)^{1/2} C(u, -n; u', -n' \,|\, l, -n-n')$$
$$\times C(u, m; u', m' \,|\, l, m+m'). \tag{A4}$$

The Wigner D functions are related to associated Legendre polynomials as

$$D^l_{m,0}(z) = D^l_{0,m}(z) = \frac{(2l+1)^{1/2}}{2} \left(\frac{(l-m)!}{(l+m)!}\right)^{1/2} P^m_l(z). \tag{A5}$$

APPENDIX B: ADDITION THEOREM FOR WIGNER D FUNCTIONS

We shall derive the addition theorem for Wigner D functions indispensable to computation of the C–G coefficients of $E(3)$. From (13) we can write, restricting $U^{\rho,s}(r, A)$ on $SU(2)$,

$$[U(A)f](\cos\theta, e^{i\varphi})$$
$$= \left(\frac{(1-2b\bar b)\sin\theta - abe^{i\varphi}(1+\cos\theta) + \bar a\bar b e^{-i\varphi}(1-\cos\theta)}{(1-2b\bar b)\sin\theta + abe^{i\varphi}(1-\cos\theta) - \bar a\bar b e^{-i\varphi}(1+\cos\theta)}\right)^{s/2}$$
$$\times f\left[(1-2b\bar b)\cos\theta + (abe^{i\varphi} + \bar a\bar b e^{-i\varphi})\sin\theta, \left(\frac{a^2 e^{i\varphi} - \bar b^2 e^{-i\varphi})\sin\theta - 2a\bar b\cos\theta}{\bar a^2 e^{-i\varphi} - b^2 e^{i\varphi})\sin\theta - 2\bar a b\cos\theta}\right)\right],$$
$$A = \begin{pmatrix} a & b \\ -\bar b & \bar a \end{pmatrix} \in SU(2). \tag{B1}$$

As is well known,[2,9,12] the $U(A)$ in (B1) is unitarily equivalent to $D(u)$, a UIR of $SU(2)$ where $u \ge |s|$ and $2u$ is a nonnegative integer. As in (14), we can choose the basis vectors[6]

$$h_m(\cos\theta, e^{i\varphi}) = (-1)^m(\sqrt{2}\pi)^{-1} D^u_{s,m}(\cos\theta)\cdot e^{im\varphi} \tag{B2}$$

and the matrix elements of $D(u)$ with respect to these basis vectors are, from (34)

$$T^u_{l-u,k-u}(A) = [2/(2u+1)]^{1/2} D^u_{u-l,k-u}(\cos\alpha), \tag{B3}$$

for

$$A = \begin{pmatrix} \cos(\alpha/2) & \sin(\alpha/2) \\ -\sin(\alpha/2) & \cos(\alpha/2) \end{pmatrix}, \quad l, k = 0, 1, \ldots, 2u.$$

Consequently, we write

$$[U(A)h_{-u+k}](\cos\theta, e^{i\varphi}) = \sum_{l=0}^{2u} T^u_{l-u,k-u}(A) h_{-u+l}(\cos\theta, e^{i\varphi}). \tag{B4}$$

Substituting (B1), (B2), and (B3) into (B4) we obtain the addition theorem for Wigner D functions,

$$\left(\frac{\cos\alpha\sin\theta - \sin\alpha(\cos\theta\cos\varphi + i\sin\varphi)}{\cos\alpha\sin\theta - \sin\alpha(\cos\theta\cos\varphi - i\sin\varphi)}\right)^{s/2}$$
$$\times \left(\frac{\cos\theta\sin\alpha - \sin\theta(\cos\alpha\cos\varphi + i\sin\varphi)}{\cos\theta\sin\alpha - \sin\theta(\cos\alpha\cos\varphi - i\sin\varphi)}\right)^{m/2}$$
$$\times D^u_{s,m}(\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos\varphi)$$
$$= \left[\frac{2}{2u+1}\right]^{1/2} \sum_{l=0}^{2u} D^u_{u-l,m}(\cos\alpha) D^u_{s,l-u}(\cos\theta) e^{i(l-u)\varphi}. \tag{B5}$$

Replacing φ by $(\pi/2)-\varphi$ and using $D^u_{n,m}(z) = (-1)^{n+m} D^u_{-n,-m}(z)$, we can easily express (B5) in the form

$$\left(\frac{\cos\alpha\sin\theta-\sin\alpha(\cos\theta\sin\varphi-i\cos\varphi)}{\cos\alpha\sin\theta-\sin\alpha(\cos\theta\sin\varphi+i\cos\varphi)}\right)^{s/2}$$

$$\times\left(\frac{\cos\theta\sin\alpha-\sin\theta(\cos\alpha\sin\varphi-i\cos\varphi)}{\cos\theta\sin\alpha-\sin\theta(\cos\alpha\sin\varphi+i\cos\varphi)}\right)^{m/2}$$

$$\times D^u_{s,m}(\cos\alpha\cos\theta+\sin\alpha\sin\theta\sin\varphi)$$

$$=\left(\frac{2}{2u+1}\right)^{1/2}\sum_{l=0}^{2u} i^{u-l} D^u_{l-u,m}(\cos\alpha) D^u_{s,u-l}(\cos\theta)\, e^{i(l-u)\varphi}. \qquad \text{(B6)}$$

ACKNOWLEDGMENT

The author wishes to thank Professor Willard Miller, Jr. for his enlightening suggestions and ceaseless encouragements.

*A part of the author's Ph.D. thesis (see Ref. 9).

†Present address: Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742.

[1] N.Y. Vilenkin, E.L. Akim, and A.A. Levin, Dokl. Akad. Nauk. SSSR 112, 987 (1957).

[2] W. Miller, Comm. Pure Appl. Math. 17, 527 (1964).

[3] H. Ogura, J. Phys. Soc. Japan 25, No. 2, 586 (August 1968).

[4] W.J. Holman, Ann. Phys. 52, 176 (1969).

[5] A.D. Jackson and L.C. Maximon, SIAM J. Math. Anal. 3, No. 3, 446 (August 1972).

[6] W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).

[7] G.W. Mackey, in *Functional Analysis and Related Fields* edited by F.E. Browder (Springer, Berlin, 1970).

[8] A.S. Wightman, in *Relations de Dispersion et Particules Élementaries*, edited by C. de Witt and R. Omnes (Hermann, Paris, 1960), pp. 161–226.

[9] J.S. Rno, Ph.D. thesis, University of Minnesota (1973).

[10] A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J. 1960).

[11] E.G. Kalnins, J. Patera, R.T. Sharp, and P. Winternitz, "Two Variable Galilei Group Expansions of Non-Relativistic Scattering Amplitudes" (in preparation).

[12] J.D. Talman, *Special Functions* (Benjamin, New York, 1968).

Scattering theory for Schrödinger operators with L^∞ potentials and distorted Bloch waves

Giovanni M. Troianiello

Istituto per le Applicazioni del Calcolo (IAC) del Consiglio Nazionale delle Ricerche, Rome, Italy
(Received 19 July 1973)

We prove that, if $q_1 \in C^0(R^3) \cap L^\infty(R^3)$ and $q_2 \in L^1(R^3) \cap L^2(R^3)$ are real-valued functions, the wave operators associated with the self-adjoint operators $H_1 = -\Delta + q_1$ and $H_2 = -\Delta + q_1 + q_2$ in $L^2(R^3)$ exist and are complete. We also prove that, if q_1 is periodic and q_2 is in a certain weighted L^2 space $\mathcal{X}$, the absolutely continuous part of H_2 possesses two sets of generalized eigenfunctions which belong to the dual space $\mathcal{X}^*$ of $\mathcal{X}$ and are solutions of linear equations involving the generalized eigenfunctions of H_1.

INTRODUCTION

The present paper deals with some problems of spectral theory relative to the Schrödinger operator $-\Delta + q_1 + q_2$ in $L^2(R^3)$, q_1 and q_2 being real-valued potential functions such that $q_1 \in C^0(R^3) \cap L^\infty(R^3)$, $q_2 \in L^1(R^3) \cap L^2(R^3)$.

Our approach is based on scattering theory, with reference to $-\Delta + q_1$ as the unperturbed operator H_1 and to $-\Delta + q_1 + q_2$ as the perturbed operator H_2.

We proceed as follows. In Sec. 1 we assume q_1 and q_2 as above, and prove (Theorem 1) that the wave operators associated with the pair H_1, H_2 exist and are complete. In Sec. 2 we make the additional assumptions that q_1 is periodic, that H_1 is spectrally absolutely continuous, and that q_2 is in a certain weighted L^2 space $\mathcal{X}$. Because of the periodicity, H_1 possesses a "complete" set of generalized eigenfunctions (the so-called Bloch waves) which belong to the dual space $\mathcal{X}^*$ of $\mathcal{X}$. We prove (Theorem 2) that the absolutely continuous part of H_2 possesses two "complete" sets of generalized eigenfunctions which belong to $\mathcal{X}^*$ and are solutions of certain linear equations (the Lippmann–Schwinger equations in an abstract form) involving the generalized eigenfunctions of H_1.

Our method consists in passing from H_1 and H_2 to the resolvents $(H_1 - a)^{-1}$ and $(H_2 - a)^{-1}$, a being a suitable negative real number. This enables us, by means of a convenient estimate of $(H_1 - a)^{-1}$, to apply the trace class method of scattering theory, as well as a perturbation method for eigenfunction expansions based on results of Kato and Kuroda.[1,2]

Let us recall that Kuroda[3] has given a somewhat different treatment of distorted Bloch waves, by directly investigating the existence of strong boundary values for $(H_1 - \lambda \mp i\epsilon)^{-1}$ as $\epsilon \downarrow 0$.

1. EXISTENCE AND COMPLETENESS OF WAVE OPERATORS IN THE GENERAL CASE

We denote by H_0 the self-adjoint realization in $L^2(R^3)$ of the operator

$$-\Delta = -\sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2},$$

and by Q_1 and Q_2 the maximal multiplication operators corresponding to real-valued functions $q_1 \in L^\infty(R^3)$ and $q_2 \in L^2(R^3)$, respectively.

It is well known (see for instance Chap. V of Kato's book[4]) that the self-adjoint operators H_0, $H_1 = H_0 + Q_1$ and $H_2 = H_0 + Q_1 + Q_2$ have the same domain: moreover, they are all bounded from below, that is,

$$H_j - r \geq 0, \quad j = 0, 1, 2,$$

for a suitable real constant r. We have the spectral representation

$$H_j = \int_{-\infty}^{+\infty} \lambda \, dE_j(\lambda),$$

$\{E_j(\lambda)\}$ being the right-continuous spectral family associated with H_j; hence, setting $R_j(\zeta) = (H_j - \zeta)^{-1}$ for ζ in the resolvent set of H_j, we also have

$$R_j(\zeta) = \int_{-\infty}^{+\infty} (\lambda - \zeta)^{-1} \, dE_j(\lambda).$$

If, specifically, ζ equals a real number $a < r$, the spectral family $\{F_j(\lambda)\}$ associated with the self-adjoint operator $-R_j(a)$ satisfies[5]

$$F_j(\Gamma) = E_j(\rho^{-1}(\Gamma)), \tag{1.1}$$

for every Borel set $\Gamma \subset R^1$, with

$$\rho(t) = \begin{cases} -(t-a)^{-1} & \text{for } t \geq r, \\ -(r-a)^{-1} & \text{for } t < r. \end{cases}$$

Throughout this paper we shall deal with real-valued functions q_1 and q_2 such that

$$q_1 \in C^0(R^3) \cap L^\infty(R^3)$$

and

$$q_2 \in L^1(R^3) \cap L^2(R^3).$$

Let us call K the maximum of $|q_1(x)|$ for $x \in R^3$. Let us also factor $q_2(x)$ as the product of the two square-integrable functions $q_{21}(x) = |q_2(x)|^{1/2}$ and $q_{22}(x) = (\text{sign} q_2(x)) |q_2(x)|^{1/2}$; Q_{2j} is then the maximal multiplication operator by $q_{2j}(x)$, and $Q_2 = Q_{21} Q_{22}$.

In the next lemma we present an estimate of $R_1(a)$ which follows from theorems about Wiener integrals; the negative real number a is fixed as above, so that, in particular, $a < -K$.

Lemma 1.1: $R_1(a)$ is an integral operator whose kernel $g_1(x, y; a)$ satisfies the inequalities

$$0 \leq g_1(x, y; a) \leq g_0(x, y; a + K), \text{ a.e. } (x, y) \in R^3 \times R^3, \tag{1.2}$$

where $g_0(x, y; a + K)$ is the kernel of the integral operator

$R_0(a+K)$, *that is*,

$$g_0(x,y;a+K)=\exp(-|a+K|^{1/2}|x-y|)/4\pi|x-y|.$$

Proof: The bounded operators e^{-tH_0} and e^{-tH_1}, $0<t<+\infty$, are integral operators whose respective kernels

$p_0(x,y;t)=\exp(-|x-y|^2/4t)/(4\pi t)^{3/2}$ and $p_1(x,y;t)$

satisfy[6,7]

$$0\leq p_1(x,y;t)\leq e^{tK}p_0(x,y;t),\ (x,y)\in R^3\times R^3. \quad (1.2')$$

For $x\neq y$ the integral

$$\int_0^{+\infty} e^{ta}e^{tK}p_0(x,y;t)dt$$

exists and equals $g_0(x,y;a+K)$,[8] so that, by (1.2′), the integral

$$g_1(x,y;a)=\int_0^{+\infty} e^{ta}p_1(x,y;t)dt$$

also exists, and satisfies

$$0\leq g_1(x,y;a)\leq g_0(x,y;a+K).$$

By applying the formula

$$R_1(a)=\int_0^{+\infty} e^{ta}e^{-tH_1}dt$$

the lemma is proved.

We are now in a position to investigate the perturbation problem relative to the operators $-R_1(a)$ and $-R_2(a)$.

Lemma 1.2: $V=-R_2(a)-(-R_1(a))$ *belongs to the trace class of operators in* $L^2(R^3)$.

Proof: The second of the resolvent equations

$$R_1(a)-R_2(a)=R_1(a)Q_2R_2(a),\quad R_1(a)-R_2(a)=R_2(a)Q_2R_1(a) \quad (1.3)$$

yields

$$V=R_1(a)-R_2(a)=[R_2(a)(H_1-a)][R_1(a)Q_2R_1(a)].$$

Since $R_2(a)(H_1-a)$ is bounded, it suffices to show that $R_1(a)Q_2R_1(a)=R_1(a)Q_{21}Q_{22}R_1(a)$ belongs to the trace class. Now, it follows from Lemma 1.1 that $R_1(a)Q_{21}$ is an integral operator whose kernel $k(x,y)=g_1(x,y;a)q_{21}(y)$ satisfies

$$|k(x,y)|\leq[\exp(-|a+K|^{1/2}|x-y|)/4\pi|x-y|]|q_{21}(y)|,$$

$$\text{a.e. } (x,y)\in R^3\times R^3,$$

that is,

$$k(x,y)\in L^2(R^3\times R^3).$$

Hence, $R_1(a)\,Q_{21}$ is in the Banach space $\mathcal{B}_2(\mathcal{H})$ of Hilbert–Schmidt operators in $\mathcal{H}=L^2(R^3)$. The same is true for $Q_{22}R_1(a)$, and the lemma is proved.

We shall now recall some basic notions of scattering theory; for a detailed exposition, see Kato's book,[4] Chap. X.

Let T_1 and T_2 be self-adjoint operators in a Hilbert space $\mathcal{H}$. If $\mathcal{H}_{j,ac}$ denotes the subspace of absolute continuity with respect to T_j, and $P_{j,ac}$ the projection of $\mathcal{H}$ onto $\mathcal{H}_{j,ac}$, $j=1,\ 2$, the limits

$$W^{(\pm)}(T_2,T_1)=\underset{t\to\pm\infty}{\text{s-lim}}\, e^{itT_2}e^{-itT_1}P_{1,ac}$$

are called the *wave operators* associated with the pair T_1, T_2. When $W^{(+)}(T_2,T_1)$ exists, it is a partial isometry with initial set $\mathcal{H}_{1,ac}$ and final set contained in $\mathcal{H}_{2,ac}$; moreover, it satisfies the so-called *intertwining relation*

$$T_2W^{(+)}(T_2,T_1)\supset W^{(+)}(T_2,T_1)T_1\ .$$

A similar result holds for $W^{(-)}(T_2,T_1)$ whenever it exists.

$W^{(+)}(T_2,T_1)$ or $W^{(-)}(T_2,T_1)$ is said to be *complete* if its final set is all of $\mathcal{H}_{2,ac}$. If either wave operator exists and is complete, then the part of T_1 in $\mathcal{H}_{1,ac}$ is unitarily equivalent to the part of T_2 in $\mathcal{H}_{2,ac}$.

Reconsidering the operators H_j and $-R_j(a)$, $j=1,\ 2$, we can now establish the main result of this section.

Theorem 1: *The wave operators* $W^{(\pm)}(-R_2(a),\ -R_1(a))$ *exist and are complete*; *furthermore*,

$$W^{(\pm)}(-R_2(a),\ -R_1(a))=W^{(\pm)}(H_2,H_1). \quad (1.4)$$

Proof: Lemma 1.2 yields the existence and completeness of $W^{(\pm)}(-R_2(a),-R_1(a))$ as consequences of a theorem by Kato on perturbations of the trace class; as for the identities (1.4), they follow from Kato's *invariance principle* for wave operators. See Kato's book,[4] Chap. X, Sec. 4.

2. EIGENFUNCTION EXPANSIONS IN THE PERIODIC CASE

From now on we shall assume that the real-valued function $q_1(x)\in C^0(R^3)\cap L^\infty(R^3)$ satisfies the identity

$$q_1(x+z)=q_1(x),\quad x\in R^3,$$

for all triplets z of integer numbers. The self-adjoint operator $H_1=H_0+Q_1$ defined from $-\Delta+q_1(x)$ in $L^2(R^3)$ is then the Hamiltonian describing the motion of an electron in an infinite periodic lattice.

For the spectral theory of H_1, we follow Odeh and Keller.[9] Thus, let Ω be the unit cube $\{x=(x_1,x_2,x_3)|\ 0\leq x_j\leq 1,\ j=1,2,3\}$ in R^3, and let Ω^* be the unit cube $\{k=(k_1,k_2,k_3)|\ 0\leq k_j\leq 1,\ j=1,2,3\}$ in the dual space of R^3. Given $k\in\Omega^*$, let $H_{1,k}$ be the self-adjoint realization of the operator

$$-\Delta-4\pi i\sum_{j=1}^{3}k_j\frac{\partial}{\partial x_j}+(4\pi^2|k|^2+q_1(x))$$

in $L^2(\Omega)$ with periodic boundary conditions.

Denoting by $\{\lambda_m(k)|\ m=1,2,\cdots\}$ the set of all the eigenvalues of $H_{1,k}$ repeated according to their multiplicities, we recall that the spectrum of the operator H_1 coincides with the set

$$\bigcup_{m=1}^{\infty}\{\lambda_m(k)\,|\,k\in\Omega^*\}.$$

For $k\in\Omega^*$ and $m=1,2,\cdots$, let us denote by $u_m(x,k)$, $x\in\Omega$, the orthonormalized eigenfunction of $H_{1,k}$ corresponding to the eigenvalue $\lambda_m(k)$.

By periodicity, $u_m(x,k)$ can be extended to the whole space. The bounded function

$$\psi_m(x,k)=e^{2\pi i k\cdot x}u_m(x,k), \quad x\in R^3,$$

is then a *generalized eigenfunction*[10] of H_1, which corresponds to $\lambda_m(k)$ and is called a *Bloch wave*.

The limit in the L^2 mean,

$$\tilde{f}_m(k)=\operatorname*{l.i.m.}_{C\to+\infty}\int_{|x|\le C}\psi_m(x,k)f(x)dx, \quad \text{a.e. } k\in\Omega^*,$$

exists for every $f\in L^2(R^3)$ and defines a bounded linear operator

$$\Psi_m: f(x)\to\tilde{f}_m(k)$$

from $L^2(R^3)$ to $L^2(\Omega^*)$. The adjoint Ψ_m^* of Ψ_m is given by

$$(\Psi_m^* g)(x)=\int_{\Omega^*}\psi_m^*(x,k)g(k)dk, \quad x\in R^3,$$

for $g\in L^2(\Omega^*)$. We have the expansion formulas

$$f(x)=\operatorname*{l.i.m.}_{\mu\to+\infty}\sum_{m=1}^{\mu}(\Psi_m^*\Psi_m f)(x),$$

$$=\operatorname*{l.i.m.}_{\mu\to+\infty}\sum_{m=1}^{\mu}\int_{\Omega^*}\psi_m^*(x,k)\tilde{f}_m(k)dk, \quad \text{a.e. } x\in R^3,$$

(*Bloch representation*) and

$$\|f\|^2_{L^2(R^3)}=\sum_{m=1}^{\infty}\|\tilde{f}_m\|^2_{L^2(\Omega^*)} \tag{2.1}$$

(*Parseval's equality*) for functions in $L^2(R^3)$.

It can be easily shown that, given any Borel subset Γ of R^1, (2.1) implies

$$\|E_1(\Gamma)f\|^2_{L^2(R^3)}=\sum_{m=1}^{\infty}\int_{\lambda_m(k)\in\Gamma}|\tilde{f}_m(k)|^2dk.$$

Throughout the rest of this paper we shall assume that our operator H_1 is spectrally absolutely continuous, which is equivalent to saying that for each m the set $\{k\in\Omega^*:\lambda_m(k)\in\Gamma\}$ has measure zero whenever $\Gamma\subset R^1$ has measure zero.

Let us set

$$\mathcal{H}^{(m)}=(\Psi_m^*\Psi_m)(\mathcal{H}),$$

with

$$\mathcal{H}=L^2(R^3).$$

The operator Ψ_m has the following property:

(Ψ_m.1) *Ψ_m is a partial isometry from $\mathcal{H}$ onto $L^2(\Omega^*)$, with initial set $\mathcal{H}^{(m)}$. Moreover, given any Borel subset Γ of R^1, the formula*

$$(\Psi_m E_1(\Gamma)f)(k)=\chi_\Gamma(\lambda_m(k))(\Psi_m f)(k), \quad \text{a.e. } k\in\Omega^*, \tag{2.2}$$

is satisfied for every $f\in\mathcal{H}$.

In (2.2), χ_Γ denotes the characteristic function of Γ.

Passing to the perturbed operator $H_2=H_1+Q_2$, we assume that the perturbation is given by a real-valued function $q_2(x)$ such that

$$(1+|x|)^\alpha q_2(x)\in L^2(R^3)$$

for some $\alpha>3/2$. We set

$$q_{21}(x)=(1+|x|)^{-\alpha}$$

and

$$q_{22}(x)=q_{21}(x)^{-1}q_2(x),$$

denoting by Q_{2j} the maximal multiplication operator associated with $q_{2j}(x)$, so that $Q_2=Q_{21}Q_{22}$.

Let us introduce the weighted L^2 space

$$\mathcal{X}=\{f(x)\,|\,(1+|x|)^\alpha f(x)\in\mathcal{H}\}$$

with norm

$$\|f\|_{\mathcal{X}}=\|Q_{21}^{-1}f\|_{\mathcal{H}}$$

and its dual

$$\mathcal{X}^*=\{g(x)\,|\,(1+|x|)^{-\alpha}g(x)\in\mathcal{H}\}.$$

Obviously, each generalized eigenfunction $\psi_m(x,k)$ is in $\mathcal{X}^*$ and the operator Ψ_m has the following property:

(Ψ_m.2) *The formula*

$$(\Psi_m f)(k)=\int_{R^3}\psi_m(x,k)f(x)dx, \quad k\in\Omega^*,$$

is satisfied for every $f\in\mathcal{X}$.

Denoting by $W^{(\pm)}$ the operators $W^{(\pm)}(H_2,H_1)=W^{(\pm)}(-R_2(a),-R_1(a))$[see (1.4)], let us define $\Phi_m^{(\pm)}=\Psi_m W^{(\pm)*}$.

The next theorem yields a representation of $\Phi_m^{(\pm)}$ in terms of generalized eigenfunctions of H_2 lying in $\mathcal{X}$.

Theorem 2: For each m the operators $\Phi_m^{(\pm)}$ have the following properties:

($\Phi_m^{(\pm)}$.1) *$\Phi_m^{(\pm)}$ are partial isometries from $\mathcal{H}$ onto $L^2(\Omega^*)$ with initial sets $W^{(\pm)}(\mathcal{H}^{(m)})$ contained in the subspace $\mathcal{H}_{2,ac}$ of absolute continuity with respect to H_2. Furthermore, given any Borel subset Γ of R^1, the identities*

$$(\Phi_m^{(\pm)}E_2(\Gamma)f)(k)=\chi_\Gamma(\lambda_m(k))(\Phi_m^{(\pm)}f)(k), \quad \text{a.e. } k\in\Omega^*,$$

are satisfied for every $f\in\mathcal{H}$.

($\Phi_m^{(\pm)}$.2) *For almost every $k\in\Omega^*$ there exist two bounded linear operators $G_m^{(\pm)}(k)^*:\mathcal{X}^*\to\mathcal{X}^*$ such that the formulas*

$$(\Phi_m^{(\pm)}f)(k)=\int_{R^3}\phi_m^{(\pm)}(x,k)f(x)dx$$

are satisfied for every $f\in\mathcal{X}$, with

$$\phi_m^{(\pm)}(.,k)=G_m^{(\pm)}(k)^*(\psi_m(.,k)).$$

The functions $\phi_m^{(\pm)}(x,k)$, $x\in R^3$, are generalized eigenfunctions of H_2 pertaining to the eigenvalue $\lambda_m(k)$.

Proof: The proof of ($\Phi_m^{(\pm)}$.1) is immediate, once having taken into account (Ψ_m.1) and the intertwining property of $W^{(\pm)}$.

As for ($\Phi_m^{(\pm)}$.2), we pass again from $H_j=\int_{-\infty}^{+\infty}\lambda dE_j(\lambda)$ to

$$-R_j(a)=-\int_{-\infty}^{+\infty}\frac{1}{\lambda-a}dE_j(\lambda)=\int_{-\infty}^{+\infty}\lambda dF_j(\lambda), \; j=1,2.$$

By (1.1), (2.2) is equivalent to

$$(\Psi_m F_1(\Gamma)f)(k)=\chi_\Gamma(-[\lambda_m(k)-a]^{-1})(\Psi_m f)(k), \quad \text{a.e.}\, k\in\Omega^*, \tag{2.3}$$

for every $f\in\mathcal{H}$, Γ being any Borel subset of R^1.

For $\lambda\in R^1$, $\epsilon>0$, let us consider the operators

$V(-R(a)-\lambda\mp i\epsilon)^{-1}$, with $V=R_1(a)-R_2(a)$.

Lemma 2.1: *For* $j=1,\ 2$ *the following statements hold true*: $(\mathrm{I})_j$ *For every* $\lambda\in R^1$ *and* $\epsilon>0$, *the operators* $V(-R_j(a)-\lambda\mp i\epsilon)^{-1}$ *belong to the space* $\mathcal{B}(\mathcal{X})$ *of all bounded linear operators in* $\mathcal{X}$. $(\mathrm{II})_j$ *For almost every* $\lambda\in R^1$, *the limits*

$$Q_j^{(\pm)}(\lambda)=\operatorname*{s-lim}_{\epsilon\downarrow 0} V(-R_j(a)-\lambda\mp i\epsilon)^{-1},\quad j=1,\ 2,$$

exist in $\mathcal{B}(\mathcal{X})$.

For the moment let us accept this lemma as proved. We then apply a theorem by Kato and Kuroda[1] with respect to the self-adjoint operators $-R_1(a)$ and $-R_2(a)$ and to the partial isometries

$$\Phi_m^{(\pm)}=\Psi_m W^{(\pm)}(-R_2(a),\ -R_1(a))^*,$$

considering (2.3) instead of (2.2) for Ψ_m.

In this way we can set

$$G_m^{(\pm)}(k)=I+Q_1^{(\pm)}(-[\lambda_m(k)-a]^{-1}):\ \mathcal{X}\to\mathcal{X},$$

where k ranges all over Ω^* except for a set of measure zero. We thus obtain the operators $G_m^{(\pm)}(k)^*:\ \mathcal{X}^*\to\mathcal{X}^*$ which satisfy all the required properties.

Finally, a direct argument based on the part of the theorem already proved ensures that the functions $\phi_m^{(\pm)}(x,k)$ are solutions of

$$-\Delta_x\phi_m^{(\pm)}(x,k)+(q_1(x)+q_2(x))\phi_m^{(\pm)}(x,k)=\lambda_m(k)\phi_m^{(\pm)}(x,k)$$

in the sense of distributions.[2]

Proof of Lemma 2.1: Let us start by proving that both operators $Q_{21}^{-1}R_1(a)Q_2$ and $Q_{21}^{-1}R_2(a)Q_2$ belong to the Hilbert–Schmidt class $\mathcal{B}_2(\mathcal{H})$. Since the first equation in (1.3) yields

$$Q_{21}^{-1}R_2(a)Q_2=Q_{21}^{-1}R_1(a)Q_2-Q_{21}^{-1}R_1(a)Q_2R_2(a)Q_2,$$

it suffices to show that $Q_{21}^{-1}R_1(a)Q_2$ is a Hilbert–Schmidt operator. Now, it follows from (1.2) that the kernel $l(x,y)=(1+|x|)^\alpha g_1(x,y;a)q_2(y)$ of $Q_{21}^{-1}R_1(a)\ Q_2$ satisfies the inequality

$$|l(x,y)|\leq(1+|x|)^\alpha g_0(x-y;a+K)(1+|y|)^{-\alpha}|q_{22}(y)|,\qquad \text{a.e. } (x,y)\in R^3\times R^3, \tag{2.4}$$

with

$$g_0(x-y;\ a+K)=\exp(-|a+K|^{1/2}|x-y|)/4|x-y|,\qquad q_{22}\in\mathcal{H}.$$

According to a well-known inequality by Peetre, there exists a positive constant C such that

$$(1+|x|)^\alpha(1+|y|)^{-\alpha}\leq C(1+|x-y|)^\alpha,\quad (x,y)\in R^3\times R^3. \tag{2.5}$$

Upon inserting (2.5) into (2.4), it is evident that $l(x,y)\in L^2(R^3\times R^3)$, and therefore $Q_{21}^{-1}R_1(a)Q_2\in\mathcal{B}_2(\mathcal{H})$.

On the other hand, the second equation in (1.3) yields $R_2(a)Q_{21}=R_1(a)Q_{21}-R_2(a)Q_2R_1(a)Q_{21}$.

Since $R_2(a)Q_2$ is bounded and $R_1(a)Q_{21}$ is a Hilbert–Schmidt operator (see the proof of Lemma 1.2), it follows that $R_2(a)Q_{21}$ is a Hilbert–Schmidt operator as well.

Let us now remark that

$$\begin{aligned}Q_{21}^{-1}V(-R_1(a)-\lambda\mp i\epsilon)^{-1}Q_{21}&=Q_{21}^{-1}R_2(a)Q_2R_1(a)(-R_1(a)-\lambda\mp i\epsilon)^{-1}Q_{21}\\&=Q_{21}^{-1}R_2(a)Q_2(-R_1(a)-\lambda\mp i\epsilon)^{-1}R_1(a)Q_{21},\end{aligned}$$

because $R_1(a)$ commutes with $(-R_1(a)-\lambda\mp i\epsilon)^{-1}$. Thus, given any function f in $\mathcal{X}$, we have

$$\begin{aligned}\|V(-R_1(a)-\lambda\mp i\epsilon)^{-1}f\|_{\mathcal{X}}&=\|Q_{21}^{-1}V(-R_1(a)-\lambda\mp i\epsilon)^{-1}Q_{21}Q_{21}^{-1}f\|_{\mathcal{H}}\\&\leq\|Q_{21}^{-1}R_2(a)Q_2(-R_1(a)-\lambda\mp i\epsilon)^{-1}R_1(a)Q_{21}\|_{\mathcal{B}_2(\mathcal{H})}\|f\|_{\mathcal{X}},\end{aligned} \tag{2.6}$$

for every $\lambda\in R^1$, $\epsilon>0$, which proves $(\mathrm{I})_1$.

As for $(\mathrm{II})_1$, let us set

$A_1=Q_{21}^{-1}R_2(a)Q_2,\quad B_1=R_1(a)Q_{21},\quad \alpha_1(\mu)=A_1F_1(\mu)B_1.$
Given any decomposition of R^1 into intervals $\Gamma_k=[\lambda_k,\lambda_{k+1}[$, , $k=0,1,\ldots,n-1$, with $-\infty=\lambda_0<\lambda_1<\ldots<\lambda_{n-1}<\lambda_n=+\infty$, the $\mathcal{B}_2(\mathcal{H})$-valued function $\alpha_1(\mu)$ satisfies

$$\begin{aligned}\sum_{k=0}^{n-1}\|\alpha_1(\Gamma_k)\|_{\mathcal{B}_2(\mathcal{H})}&\leq\sum_{k=0}^{n-1}\|A_1F_1(\Gamma_k)\|_{\mathcal{B}_2(\mathcal{H})}\|F_1(\Gamma_k)B_1\|_{\mathcal{B}_2(\mathcal{H})}\\&\leq\left(\sum_{k=0}^{n-1}\|A_1F_1(\Gamma_k)\|^2_{\mathcal{B}_2(\mathcal{H})}\right)^{1/2}\left(\sum_{k=0}^{n-1}\|F_1(\Gamma_k)B_1\|^2_{\mathcal{B}_2(\mathcal{H})}\right)^{1/2}\\&\leq\|A_1\|_{\mathcal{B}_2(\mathcal{H})}\|B_1\|_{\mathcal{B}_2(\mathcal{H})}.\end{aligned}$$

Hence it follows from a theorem by Asano[11] that for a.e. $\lambda\in R^1$ the operators

$$A_1(-R_1(a)-\lambda\mp i\epsilon)^{-1}B_1=\int_{-\infty}^{+\infty}\frac{d\alpha_1(\mu)}{\mu-\lambda\mp i\epsilon}$$

converge in the norm of $\mathcal{B}_2(\mathcal{H})$ as $\epsilon\downarrow 0$. We then obtain $(\mathrm{II})_1$ by taking (2.6) into account.

The statements $(\mathrm{I})_2$ and $(\mathrm{II})_2$ can be demonstrated similarly. In fact, since $R_2(a)$ commutes with $(-R_2(a)-\lambda\mp i\epsilon)^{-1}$, we have

$$\begin{aligned}Q_{21}^{-1}V(-R_2(a)-\lambda\mp i\epsilon)^{-1}Q_{21}\\=Q_{21}^{-1}R_1(a)Q_2(-R_2(a)-\lambda\mp i\epsilon)^{-1}R_2(a)Q_{21},\end{aligned}$$

and consequently, given any $f\in\mathcal{X}$, we obtain

$$\begin{aligned}\|V(-R_2(a)-\lambda\mp i\epsilon)^{-1}f\|_{\mathcal{X}}\\\leq\|Q_{21}^{-1}R_1(a)Q_2(-R_2(a)-\lambda\mp i\epsilon)^{-1}R_2(a)Q_{21}\|_{\mathcal{B}_2(\mathcal{H})}\|f\|_{\mathcal{X}}\end{aligned}$$

which yields $(\mathrm{I})_2$. Setting

$$A_2=Q_{21}^{-1}R_1(a)Q_2,\quad B_2=R_2(a)Q_{21},\quad \alpha_2(\mu)=A_2F_2(\mu)B_2,$$

we can again apply Asano's theorem, this time with respect to the operators

$$A_2(-R_2(a)-\lambda\mp i\epsilon)^{-1}B_2=\int_{-\infty}^{+\infty}\frac{d\alpha_2(\mu)}{\mu-\lambda\mp i\epsilon},$$

thus proving $(\mathrm{II})_2$.

Remark: Let f be any element of $\mathcal{H}_{2,ac}$, and let $g^{(\pm)} = W^{(\pm)*}f$. From the identity

$$\|g^{(\pm)}\|^2_{\mathcal{H}} = \sum_{m=1}^{\infty} \|\Psi_m g^{(\pm)}\|_{L^2(\Omega^*)}$$

it follows that

$$\|f\|^2_{\mathcal{H}_{2,ac}} = \sum_{m=1}^{\infty} \|\Psi_m W^{(\pm)*} f\|^2_{L^2(\Omega^*)} = \sum_{m=1}^{\infty} \|\Phi_m^{(\pm)} f\|^2_{L^2(\Omega^*)}$$

This can be expressed by saying that both sets $\{\phi_m^{(+)}(\circ, k)\}$ and $\{\phi_m^{(-)}(\circ, k)\}$ of *distorted Bloch waves* are "complete" in $\mathcal{H}_{2,ac}$.

ACKNOWLEDGMENTS

I wish to thank Professor P.D. Lax, who suggested this problem to me, Professor S. Agmon, Professor T. Kato, Professor S.T. Kuroda, and Professor G. Talenti, to whom I submitted the various drafts of the paper, and Professor G. Gallavotti, who gave me ideas as to how to proceed.

[1]T. Kato and S.T. Kuroda, "Theory of simple scattering and eigenfunction expansions," in *Functional Analysis and Related Fields*, edited by F.E. Browder (Springer, Berlin, 1970), pp. 99-131.
[2]T. Kato and S.T. Kuroda, Rocky Mountain Math. 1, 127 (1971).
[3]S.T. Kuroda, "A stationary method of scattering and some applications," in *Proceedings of the International Conference on Functional Analysis and Related Topics, Tokyo, April* 1969 (University of Tokyo Press, Tokyo, 1970).
[4]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[5]N. Dunford and J.T. Schwartz, *Linear Operators* (Interscience, New York, 1963), Part II, Thm. XII. 2.9.
[6]J. Ginibre, J. Math. Phys. 6, 238 (1967).
[7]E. Nelson, J. Math. Phys. 5, 332 (1964).
[8]A. Erdélyi (Bateman manuscript project), *Tables of Integral Transforms* (McGraw-Hill, New York, 1953), Vol. I, Sec. 4.5.(28).
[9]F. Odeh and J.B. Keller, J. Math. Phys. 5, 1499 (1964).
[10]I.M. Gelfand and G.E. Shilov, *Generalized Functions* (Academic, New York, 1967), Vol. 3, Chap. IV.
[11]K. Asano, Proc. Jap. Acad. 43, 572 (1967).

Relativistic quantum mechanics and local gauge symmetry

P. Roman and J. P. Leveille

Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 30 May 1974)

The requirement that (either Abelian or non-Abelian) local symmetry transformations be globally and unitarily implementable kinematical symmetries of relativistic systems implies the emergence of a dynamical group which has been suggested in earlier studies. The group leads to a 4-velocity operator and to the Newton–Wigner position operator. Demanding gauge invariance of localization determines a unique interaction structure. Superselection rules for the gauge charges arise.

I. INTRODUCTION

The now generally accepted success of unified theories of weak, electromagnetic, and strong interactions, based on local gauge groups combined with spontaneous symmetry breaking, [1] makes it plausible that local gauge symmetries play a fundamental role in nature. However, no connection between the internal dynamical invariance (associated with the gauge group) and the conventional relativistic space—time symmetry (Poincaré group) seems to be present.

On the other hand, in a recent study[2] we showed that, in nonrelativistic quantum theory, the requirement that a local phase transformation be an automorphism of Hilbert space, leads in a very natural manner to the Galilean structure of nonrelativistic quantum dynamics. In other words, here the postulate of local gauge symmetry is not only intimately related to the space—time group of kinematical and inertial transformations (as was already pointed out in various specific contexts by Jauch, [3] Piron, [4] and Lévy-Leblond[5]) but in fact it essentially determines the complete dynamics. The so-determined dynamical group contains, besides the Euclidean kinematical transformations, the time displacements and the Galilean boosts. The generators of the latter play the role of the position operators. We also demonstrated that by additionally demanding gauge invariance of localization, a unique form of interaction, viz. the usual minimal interaction, emerges.

The purpose of the present work is to show that entirely analogous considerations in the relativistic case lead also to the emergence of a dynamical group from the postulate that local gauge symmetry be a unitarily implementable kinematical symmetry (in the sense of Jauch[3]). This relativistic dynamical group contains, besides the kinematical Poincaré transformations, a covariant development transformation subgroup with respect to what may be called historical time, [6] and an Abelian subgroup which may be called the group of relativistic Galilean boosts. As a matter of fact, the relativistic dynamical group so obtained coincides with the group $\tilde{\mathcal{G}}_5$ which, from completely different considerations and by way of a somewhat naive analogy, one of us suggested several years ago. [7] Even though $\tilde{\mathcal{G}}_5$ has been subsequently studied in considerable detail, [8–11] its physical content remained rather obscure. The present paper to a large extent remedies these shortcomings of interpretation. We also study here the consequences of a non-Abelian local gauge symmetry. Finally, we consider interacting systems subject to gauge arguments.

In much of this paper we follow closely the pattern used in our corresponding study[2] of the nonrelativistic case. Even though the present work is self-contained, the reader is urged to first study Ref. 2, mainly because now, for the sake of brevity, we do not repeat in detail the motivations for the various steps in the argument, provided they are analogous to the nonrelativistic case.

II. THE KINEMATICAL GROUP

We adopt the usual geometry of special relativity:

Assumption 1: The space of events is the homogeneous and isotropic Minkowski space $E_{3,1}$.

This immediately implies the existence of the Poincaré group $\mathcal{P}\equiv SO(3,1)\otimes T_4^P$ with[12,13]

$$[P_\mu, P_\nu]=0, \quad [J_{\mu\nu}, P_\sigma]=i(g_{\nu\sigma}P_\mu - g_{\sigma\mu}P_\nu),$$
$$[J_{\mu\nu}, J_{\rho\sigma}]=i(g_{\nu\rho}J_{\mu\sigma}-g_{\mu\rho}J_{\nu\sigma}-g_{\mu\sigma}J_{\rho\nu}+g_{\nu\sigma}J_{\rho\mu}). \tag{2.1}$$

This algebra can be realized on the Hilbert space of square-integrable wavefunctions $\psi(x)$ by setting

$$P_\mu \sim i\partial_\mu, \quad J_{\mu\nu}\sim i(x_\mu\partial_\nu - x_\nu\partial_\mu). \tag{2.2}$$

Recall here that the inner product in the Hilbert space is defined by

$$(\varphi, \psi)=i\int_{x^0}\varphi^*(x)\overleftrightarrow{\partial_0}\psi(x)d^3\mathbf{x}. \tag{2.3}$$

Next, following the familiar arguments[14,15] we stipulate that the phase of a wavefunction is a matter of convention, even when comparing phases at different world points. In view of Wigner's theorem, [16] this demand that *a local phase transformation be an automorphism of Hilbert space*, can be formalized by

Assumption 2: To every transformation

$$\psi(x)\to\exp[i\omega(x)]\,\psi(x) \tag{2.4}$$

with a differentiable $\omega(x)$, there corresponds in Hilbert space a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\,\psi)(x)=\exp[i\omega(x)]\,\psi(x). \tag{2.5}$$

Using (2.2) we find that

$$(\mathcal{U}P_\mu\mathcal{U}^{-1}\psi)(x)=\exp(i\omega)\,[i\partial_\mu\exp(-i\omega)\psi(x)]$$
$$=(i\partial_\mu+\partial_\mu\omega)\,\psi(x),$$

i.e., under a local phase transformation (2.4)

$$P_\mu\to P_\mu+\partial_\mu\omega. \tag{2.6}$$

Similarly, we find that

$$J_{\mu\nu}\to J_{\mu\nu}+(x_\mu\partial_\nu\omega - x_\nu\partial_\mu\omega). \tag{2.7}$$

It follows that, unless we enlarge the algebra of observables, (2.4) cannot be considered to be a unitarily implementable permutation of observables, i.e., local phase transformations could not be kinematical transformations in Jauch's sense.[3] Indeed, if $U=\exp(iF)$ (where, at this stage, F can be a function of only P_μ and $J_{\mu\nu}$), Eq. (2.6) would imply

$$P_\mu + \partial_\mu\omega = U P_\mu U^{-1} = P_\mu + i[F, P_\mu] + \cdots \tag{2.8}$$

which cannot be satisfied (unless $\omega = \text{const}$), since $\partial_\mu\omega$ is a c-number multiple of the identity. We therefore postulate

Assumption 3: The algebra of observables is large enough to guarantee that arbitrary local phase transformations with a differentiable $\omega(x)$ are kinematical transformations.

To satisfy this postulate, *it suffices to adjoin to the set* $\{P_\mu, J_{\mu\nu}\}$ *the identity operator I and four additional commuting operators* Q_μ which, in fact, generate the *linear* local phase transformations corresponding to $\omega(x) = c_\mu x^\mu$, c_μ constant. Indeed, if we take $\omega(x) = c_\mu x^\mu$ and write[17] $F = -lc^\mu Q_\mu$, then Eq. (2.8) is satisfied provided

$$[P_\mu, Q_\nu] = -il^{-1} g_{\mu\nu}. \tag{2.9}$$

Since the c_μ are linearly independent, we also have

$$[Q_\mu, Q_\nu] = 0. \tag{2.10}$$

These equations tell us that we can realize Q_μ by

$$Q_\mu \sim -l^{-1} x_\mu. \tag{2.11}$$

Then, with $U = \exp(-ilc^\sigma Q_\sigma)$ we have

$$U J_{\mu\nu} U^{-1} = J_{\mu\nu} - ilc^\sigma [Q_\sigma, J_{\mu\nu}] + \cdots,$$

so that comparing with (2.7), using (2.11) and noting that now $\partial_\rho\omega = c_\rho$, we see that consistency requires

$$[J_{\mu\nu}, Q_\sigma] = i(g_{\nu\sigma} Q_\mu - g_{\sigma\mu} Q_\nu). \tag{2.12}$$

If now $\omega(x) = \sum_{n=0}^\infty c(a_\mu^{(n)} x^\mu)^n$ is an arbitrary differentiable function, then the effect of the corresponding unitary transformation U [as defined by (2.5)] on the operator algebra is characterized by (2.6), (2.7), and

$$Q_\mu \to Q_\mu. \tag{2.13}$$

Since $\partial_\mu\omega$ and $x_\mu \partial_\nu \omega$ in (2.6) and (2.7) are power series in the x^μ and since the realization (2.11) holds, the rhs of (2.6), (2.7), (2.13) represent merely a permutation of the operator algebra, so that we indeed have a kinematical symmetry transformation. This concludes the proof that the algebra of observables generated by the set $\{P_\mu,\ J_{\mu\nu},\ Q_\mu,\ I\}$ is large enough to assure that all local phase transformations be kinematical transformations.

It should be noted that the "relativistic Heisenberg commutation relations" (2.9), the commutativity relations (2.10), and Eq. (2.12) (which says that Q_σ is a vector operator under the Lorentz group) arose as consistency requirements.

If we consider the special case of linear local phase transformations, $\omega(x) = c_\mu x^\mu$, then we have

$$\begin{aligned} &Q_\mu \to Q_\mu,\\ &P_\mu \to P_\mu + c_\mu,\\ &J_{\mu\nu} \to J_{\mu\nu} - l(c_\nu Q_\mu - c_\mu Q_\nu) \end{aligned} \tag{2.14}$$

describing the effect of this special transformation on the algebra of observables. From the foregoing we see that the Q_μ are analogous to the well-known boosts of Galilean systems. We therefore call the Q_μ "relativistic Galilean boosts," (RG boosts[18] for short); and then (2.14) tells us that the RG *boost transformations arise as particular local phase transformations.*

The algebra of observables is fully characterized by the Lie relations (2.1), (2.9), (2.10), (2.12). This algebra has been previously introduced, via a somewhat ad hoc argument, by Johnson[19] and also by Broyles[20] and it forms a subalgebra of a group first considered by Castell[21] and later, independently, by one of us in Ref. 7. The structure of the corresponding simply connected group is

$$K = SL(2,C)^J \otimes [T_4{}^P \otimes (T_4{}^Q \times T_1{}^{l^{-1}})]. \tag{2.15}$$

Here $SL(2,C)$ arises as the universal covering group of $SO(3,1)$. Consequently, the wavefunctions will be vector valued representations and should be properly labeled as

$$\psi(x) \equiv \psi^{k,c}_{s,s_3}(x), \tag{2.16}$$

where k, c characterize the irreducible unitary representations of $SL(2,C)$ and the state labels s, s_3 are associated with the Casimir operators of the reduction chain $SL(2,C) \supset SU(2) \supset SO(2)$. Correspondingly, the realization (2.2) of $J_{\mu\nu}$ must be changed to

$$J_{\mu\nu} \sim i(x_\mu \partial_\nu - x_\nu \partial_\mu) + \Sigma_{\mu\nu} \tag{2.17}$$

where $\Sigma_{\mu\nu}$ is an $SL(2,C)$ matrix. We define $SL(2,C)$ spin $T_{\mu\nu}$ as the difference between the total and "orbital" $SL(2,C)$ angular momentum 6-vectors, i.e., set

$$T_{\mu\nu} \equiv J_{\mu\nu} - l(Q_\mu P_\nu - Q_\nu P_\mu) = \Sigma_{\mu\nu}. \tag{2.18}$$

It is easily seen that the Casimir invariants of K are

$$C_1 = l^{-1} I, \tag{2.19a}$$

$$C_2 = \tfrac{1}{2} T_{\mu\nu} T^{\mu\nu}, \tag{2.19b}$$

$$C_3 = \tfrac{1}{4} \epsilon_{\mu\nu\rho\sigma} T^{\mu\nu} T^{\rho\sigma}. \tag{2.19c}$$

Here C_1 arose from linear phase transformations and indicates a superselection rule (about which we shall comment later on). Because of (2.18), the spectra of C_2 and C_3 are

$$C_2 = (k^2 + c^2 - 1), \tag{2.20a}$$

$$C_3 = 2ikc, \tag{2.20b}$$

where[21]

$$k = 0, 1/2, 1, \cdots$$

and

$$c = ia \quad \text{with } -\infty < a < +\infty. \tag{2.20c}$$

The state labels s, s_3 [being the eigenvalues of $\mathbf{T}^2$ where $\mathbf{T} \equiv (T_{23},\ T_{31},\ T_{12})$ and of $T_3 \equiv T_{12}$, respectively] have the spectra

$$s = k,\ k+1,\ k+2,\ \cdots, \tag{2.21a}$$

$$s_3 = -s,\ -s+1,\ \ldots,\ s-1,\ s. \tag{2.21b}$$

Thus, the irreducible unitary representations of K are characterized by specifying a scale l^{-1} of phase and a spin-tower (k, c); and the additional state labels determine spin and spin component. Since these are kinematical labels of a certain set of states, we call K the *kinematical group*.

III. THE DYNAMICAL GROUP

To introduce dynamics, we define a *development transformation* of an isolated system as a kinematical symmetry characterized by

$$P_\mu \to P_\mu, \quad J_{\mu\nu} \to J_{\mu\nu}, \quad Q_\mu \to f(Q_\nu, P_\nu, J_{\rho\sigma}).$$

This rather obvious definition is motivated by the requirement that the intrinsic development must be compatible with the geometry of events, i.e., that its generator be invariant under $\mathcal{P}$. It is also natural to desire that development transformations form an Abelian group (cf. Ref. 2). A simplicity requirement leads to the more specific

Assumption 4: Development transformations form a one-parameter Lie group T_1^S.

Then any development transformation σ will be represented by a unitary operator $\mathcal{U}_\sigma = \exp(i\sigma S)$. Concerning the generator S we make the rather weak

Assumption 5: S is contained in the algebra of observables generated by P_μ, Q_ν, $J_{\rho\sigma}$.

Combined with the Poincaré invariance requirement inherent in the definition of a development transformation, this tells us that the most general form of S is[23]

$$S = S(P^2,\ T_{\alpha\mu} T^{\beta\mu} P^\alpha P_\beta,\ I). \tag{3.1}$$

Now we observe that the relation "$A \sim B$ iff $B = \mathcal{U}_\sigma A \mathcal{U}_\sigma^{-1}$ for some σ" is an equivalence relation on the algebra of observables. It is therefore natural to define a *dynamical group* G by

Assumption 6: The kinematical group K is isomorphic to the quotient group modulo T_1^S of some group G, i.e., $K \approx G/T_1^S$.

This implies that the generators of K and S together must form a closed Lie algebra. Consequently, the rhs of (3.1) becomes unique[24] and we have

$$S = -(l/2)P^2 + (l/2)\mathcal{D}_1. \tag{3.2}$$

Here $\mathcal{D}_1$ is an arbitrary constant and the scale factor $-l/2$ has been chosen for convenience and to conform with the notation of Refs. 7–9.

From (3.2) and (2.9) we can find the Lie brackets of S. Together with the previously established brackets, we have a closed Lie algebra as follows:

$$[P_\mu, P_\nu] = 0, \quad [Q_\mu, Q_\nu] = 0, \tag{3.3a}$$

$$[P_\mu, Q_\nu] = -il^{-1} g_{\mu\nu}, \tag{3.3b}$$

$$[J_{\mu\nu}, J_{\rho\sigma}] = i(g_{\nu\rho} J_{\mu\sigma} - g_{\mu\rho} J_{\nu\sigma} - g_{\mu\sigma} J_{\rho\nu} + g_{\nu\sigma} J_{\rho\mu}), \tag{3.3c}$$

$$[J_{\mu\nu}, P_\sigma] = i(g_{\nu\sigma} P_\mu - g_{\sigma\mu} P_\nu), \tag{3.3d}$$

$$[J_{\mu\nu}, Q_\sigma] = i(g_{\nu\sigma} Q_\mu - g_{\sigma\mu} Q_\nu), \tag{3.3e}$$

$$[S, P_\mu] = 0, \quad [S, J_{\mu\nu}] = 0, \tag{3.3f}$$

$$[S, Q_\mu] = iP_\mu. \tag{3.3g}$$

The corresponding 16-parameter simply connected dynamical group has the structure

$$G = T_1^S \otimes K = T_1^S \otimes \{SL(2,C)^J \otimes [T_4^P \otimes (T_4^Q \times T_1^{l^{-1}})]\} \tag{3.4}$$

This is precisely the "relativistic quantum mechanical group" $\tilde{\mathcal{G}}_5$ first introduced by Castell[21] and, independently by one of us[25] in Ref. 7.

As has been already shown in Refs. 7 and 8, the Casimir invariants of $\tilde{\mathcal{G}}_5$ are

$$\mathcal{D}_0 = l^{-1} I, \tag{3.5a}$$

$$\mathcal{D}_1 = P^2 + 2l^{-1} S, \tag{3.5b}$$

$$\mathcal{D}_2 = \tfrac{1}{2} T_{\mu\nu} T^{\mu\nu}, \tag{3.5c}$$

$$\mathcal{D}_3 = \tfrac{1}{4} \epsilon_{\mu\nu\rho\sigma} T^{\mu\nu} T^{\rho\sigma}. \tag{3.5d}$$

Here the $T_{\mu\nu}$ are defined by (2.18), and (3.5c, d) are the Casimir invariants of the corresponding *internal* $SL(2,C)$ algebra. The irreducible unitary representations of $\tilde{\mathcal{G}}_5$ can be denoted by the symbol $(l \mid \mathcal{D}_1,\ k,\ c)$. They have been explicitly constructed and studied in detail in Ref. 9.

For the purpose of the following discussion it will be useful to write

$$\tilde{\mathcal{G}}_5 \equiv G = \mathcal{G}_5 \times T_1^{l^{-1}}, \tag{3.6}$$

so that $\tilde{\mathcal{G}}_5$ appears as the scalar central extension of an (abstract) 15-parameter "geometrical" group $\mathcal{G}_5$. In $\mathcal{G}_5$ we consider the $SL(2,C)^J$ part replaced by $SO(3,1)^J$. Denoting the parameters of T_1^S, T_4^P, T_4^Q, $SO(3,1)^J$ by σ, a, b, Λ, respectively, exponentiation of the Lie algebra leads to the composition law

$$(\sigma, a, b, \Lambda)(\bar{\sigma}, \bar{a}, \bar{b}, \bar{\Lambda}) = (\sigma + \bar{\sigma}, a + \Lambda\bar{a} + \bar{\sigma} b, b + \Lambda\bar{b}, \Lambda\bar{\Lambda}). \tag{3.7}$$

If we want to represent this abstract group on some homogeneous space, the simplest choice is to take the left coset space $\mathcal{G}_5/SO(3,1)^J \otimes T_4^Q$, whose elements [i.e., the cosets of $SO(3,1)^J \otimes T_4^Q$] can be characterized by the pair $(\bar{\sigma}, \bar{a})$. Then (3.7) gives the left action of $\mathcal{G}_5$ on the coset space as

$$(\sigma, a) \to (\bar{\sigma} + \sigma, \Lambda\bar{a} + a + \bar{\sigma} b). \tag{3.8}$$

Employing the mapping $(\bar{\sigma}, \bar{a}) \to (u, x)$, our homogeneous space may be identified with a five-dimensional space $E_{3,1}(x) \times E_1(u)$ and (3.8) gives

$$u \to u + \sigma,$$

$$x_\mu \to \Lambda^\nu{}_\mu x_\nu + a_\mu + u b_\mu. \tag{3.9}$$

This transformation group of endomorphisms of $E_{3,1}(x) \times E_1(u)$ represents the *active viewpoint* of $\mathcal{G}_5$. In Ref. 7 we actually *defined* $\mathcal{G}_5$ in this way. However, we were not able to give a completely satisfactory interpretation[26] for the fifth variable u. But now, we can interpret u in a purely group theoretical manner. The one-dimensional space $E_1(u)$ is introduced, not at the start of kinematical considerations, but rather it emerges as a *convenience* permitting a simple active characterization of the abstract dynamical group. As we already showed in another context,[27] one can use a different homogeneous space, for example $\mathcal{G}_5/SO(3,1)^J$ and then one is led to a representation of $\mathcal{G}_5$ on the tangent space $E_{3,1}(x) \times E_{3,1}(\xi)$, where no explicit concept corresponding to u arises.

Once, however, the choice has been made to use the homogeneous space as specified above, we are led, in a natural manner, to a *sequence* of incoherent Hilbert spaces and to the description of intrinsic dynamical development in terms of a progression by the parameter u. We define, for each u, a Hilbert space $\mathcal{H}_u$ of square-integrable functions[28] by setting

$$\psi(x;u) = \exp(-iuS)\psi(x). \qquad (3.10)$$

The total Hilbert space $\mathcal{H}$ is then a suitable direct integral of the "slices" $\mathcal{H}_u$. In view of (3.10), it is clear that u can be interpreted as *historical time* which does not pertain to a particle, nor to an event, but it rather parametrizes *a sequence of descriptions*, or equivalently, a sequence of sets of measurements on events. This is precisely the concept of historical time as discussed by Horwitz and Piron.[6] Historical time u is the exact analog of "universal time t" in nonrelativistic physics.[29] We wish to emphasize once again that it was not necessary, in our present framework based on local gauge symmetry, to introduce historical time from the outset, but it rather emerged naturally in a group theoretic analysis, similarly as nonrelativistic universal time was "deduced" in Ref. 2. We also note that now S plays the role of a dynamical development operator, analogous to H in Galilean physics. S generates the displacement in historical time for relativistic descriptions, while H generates the displacement in Newtonian universal time for nonrelativistic systems.

So far the observables P_μ, Q_μ, $J_{\mu\nu}$, S were realized on $\mathcal{H}_{u=0}$. We now ask for their realization by differential operators on all of $\mathcal{H}$. A glance at the commutation relations (3.3a)–(3.3g) tells us that we can set

$$\begin{aligned} &P_\nu \sim i\partial_\nu, \\ &Q_\nu \sim -l^{-1}x_\nu + iu\partial_\nu, \\ &J_{\mu\nu} \sim i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}, \\ &S \sim i\partial_u. \end{aligned} \qquad (3.11)$$

In particular, S assumes a double role: on each slice $\mathcal{H}_u$ it has the realization

$$S \sim \frac{l}{2}\Box + \frac{l}{2}\mathcal{D}_1,$$

[cf. (3.2)], whereas on $\mathcal{H}$ it is given by $i\partial_u$. This is emphasized if one applies the Casimir invariant $\mathcal{D}_1$ onto the function space $\psi(x;u)$. One obtains[30]

$$(\Box - 2l^{-1}i\partial_u)\psi(x;u) = 0. \qquad (3.12)$$

This is the analog of the nonrelativistic Schrödinger equation. If one introduces the Fourier transform

$$\varphi(r,p) \equiv \int \exp[i(ru + px)]\,\psi(x;u)\,du\,d^4x, \qquad (3.13)$$

then (3.12) becomes

$$(p^2 + 2l^{-1}r)\varphi(r,p) = 0. \qquad (3.14)$$

We note that $p^2 + 2l^{-1}r = 0$ defines the orbits of the representation (for $\mathcal{D}_1 = 0$).

Let us summarize. Postulating that local phase transformations be a kinematical symmetry for relativistic systems, we were led to the existence of the *event-position operators* Q_μ and to the *kinematical group* K generated by P_μ, Q_μ, $J_{\mu\nu}$, I. The latter contains Heisenberg-type commutation relations. Defining dynamical development as a set of transformations that leave P_μ, $J_{\mu\nu}$ unchanged but alter Q_μ, we arrived, on the basis of very simple and natural additional assumptions, at the *quantum mechanical relativistic dynamical group* $\tilde{\mathcal{G}}_5$ for isolated systems. This group can be considered as a group extension (by a non-Abelian group) of the Lorentz group, cf. Ref. 10. The Poincaré invariant development operator S can be interpreted as the generator of displacements in historical time. Its effect on the event-position operators Q_μ is characterized by Eq. (3.3g). In the next section we show how the Q_μ operators can be related to the Newton–Wigner position operators of a particle with fixed mass, how a 4-velocity emerges, and how, more generally, the $\tilde{\mathcal{G}}_5$ group can be interpreted in conventional terms.

IV. PHYSICAL INTERPRETATION

In Galilean physics, elements of the event space E_3 have a simple and direct interpretation: The event $\mathbf{x}$ means that (at some time t) there is a particle at $\mathbf{x}$. In the Einsteinian relativistic theory, however, the element x_μ of the event space $E_{3,1}$ represents a "world event" (at some historical time u) in the accustomed sense, which can be best interpreted as the intersection of two world lines associated with two *different* particles. Consequently, the operators Q_μ are not associated with individual particles, they only localize events. Corresponding to this circumstance we also find that the unitary irreducible representations of $\tilde{\mathcal{G}}_5$ do not characterize "elementary particles" in Wigner's sense.[31] Indeed, fixing the value of the first Casimir invariant to be, say,[30] $\mathcal{D}_1 = 0$, Eq. (3.5b) gives $P^2 = -2l^{-1}S$, and since $S \sim i\partial_u$ has a continuous spectrum, states with any value for $M^2 \equiv p^2$ will occur in the representation space, $-\infty < M^2 < +\infty$. In other words, M^2 is not a representation label but only a state label.[32] Accordingly, particles will be associated with certain subspaces with fixed M^2. In order to study particle properties, we must study the action of observables on these subspaces.

As was discussed in Ref. 9, a unitary irreducible representation of $\tilde{\mathcal{G}}_5$ is spanned by the basis states[33] $|r,p\rangle$ which obey the relations

$$P_\mu|r,p\rangle = p_\mu|r,p\rangle, \quad S|r,p\rangle = r|r,p\rangle.$$

The normalization is $\langle r',p'|r,p\rangle = \delta(r'-r)\delta(p'-p)$, where the $\tilde{\mathcal{G}}_5$-invariant inner product is defined by Eq. (3.7) of Ref. 9. If we fix the orbit by taking[30] $\mathcal{D}_1 = 0$, then $p^2 + 2l^{-1}r = 0$, so that for the "physical states" that obey the wave equation, the label r becomes redundant.[34] Thus, we introduce the states $|p\rangle$, obeying

$$(P_\mu P^\mu + 2l^{-1}S)|p\rangle = 0 \qquad (4.1)$$

and normalize them by[35]

$$\langle p|p'\rangle = \delta(p-p'). \qquad (4.2)$$

The behavior of these states under the action of the unitary operators corresponding to the various subgroups of $\tilde{\mathcal{G}}_5$ is easily found[36] from Eqs. (3.10) and (3.5a) of Ref. 9:

$$J_{\mu\nu}: \quad \mathcal{U}(\Lambda)|p\rangle = |\Lambda^{-1}p\rangle, \qquad (4.3a)$$

$$P_\mu: \quad \mathcal{U}(a)|p\rangle = \exp(ipa)|p\rangle, \qquad (4.3b)$$

$$Q_\mu : \quad \mathcal{U}(b)|p\rangle = |p - l^{-1}b\rangle, \tag{4.3c}$$

$$S \quad : \quad \mathcal{U}(\sigma)|p\rangle = \exp[-i(l/2)p^2\sigma]|p\rangle. \tag{4.3d}$$

In particular, from (4.3c) we get

$$\frac{d}{db^\mu}\,\mathcal{U}(b)|p\rangle\Big|_{b=0} \equiv iQ_\mu|p\rangle = \frac{d}{db^\mu}\,|p - l^{-1}b\rangle\Big|_{b=0},$$

from which follows that

$$Q_\mu|p\rangle = il^{-1}\,\frac{\partial}{\partial p^\mu}\,|p\rangle. \tag{4.4a}$$

In a similar manner, (4.3b) gives

$$P_\mu|p\rangle = p_\mu|p\rangle. \tag{4.4b}$$

We now introduce Dirac kets and wavefunctions by setting

$$|\psi\rangle = \int d^4p\,\varphi(p)|p\rangle. \tag{4.5}$$

Clearly, from (4.2),

$$\langle\psi|\psi\rangle = \|\varphi\|^2 = \int |\varphi(p)|^2\,d^4p. \tag{4.6}$$

Thus, in the momentum space the wave functions corresponding to the physical states are the square-integrable functions (with respect to Lebesgue measure in R^4).

Consider now "mass-shell states"

$$|\mathbf{p}, M\rangle \equiv |\mathbf{p}, \quad p_0 = \sqrt{\mathbf{p}^2 + M^2}\rangle, \tag{4.7}$$

which obey $P^2|\mathbf{p},M\rangle = M^2|\mathbf{p},M\rangle$ (where we take $M^2 > 0$). To make sure that the corresponding wavefunctions are square integrable (i.e., that these states really belong to our Hilbert space) and in fact to find the normalization of these states, we must proceed carefully.[37] Consider the set[38]

$$E = \{p\,|\,p = \Lambda\hat{p}, \quad \Lambda \in SL(2,C), \quad \hat{p} = (M,0,0,0)\}. \tag{4.8}$$

Let $\varphi(p)$ be a square-integrable wavefunction and define

$$\varphi_M(p) = (\chi_E\varphi)(p), \tag{4.9}$$

where χ_E is the characteristic function of E. That is,

$$\varphi_M(p) = \chi_E(p)\varphi(p) = \begin{cases}\varphi(p) & \text{if } p \in E,\\ 0 & \text{if } p \notin E.\end{cases}$$

Since

$$\|\varphi_M\|^2 = \int \chi_E(p)|\varphi(p)|^2\,d^4p \le \int |\varphi(p)|^2\,d^4p < \infty,$$

the φ_M are admissible wavefunctions. Actually, since the effect of χ_E is simply to put a mass-shell condition on $\varphi(p)$, we can calculate[39]

$$\|\varphi_M\|^2 = \int d^4p\,\chi_E|\varphi|^2 = \int d^4p\,\delta(p^2/M^2 - 1)\theta(p_0)|\varphi(p)|^2$$

$$= M^2 \int \frac{d^3\mathbf{p}}{2p^0}\,|\varphi(p_0,\mathbf{p})|^2. \tag{4.10a}$$

Here, and in the sequel, it is understood that

$$p_0 \equiv \sqrt{\mathbf{p}^2 + M^2}. \tag{4.10b}$$

We introduce the on-mass-shell Dirac kets

$$|\psi_M\rangle = \int d^4p\,\varphi_M(p)|p\rangle, \tag{4.11a}$$

which can be also written as

$$|\psi_M\rangle = M^2 \int \frac{d^3\mathbf{p}}{2p_0}\,\varphi(p_0,\mathbf{p})|\mathbf{p},M\rangle. \tag{4.11b}$$

We then have

$$\langle\psi_M|\psi_M\rangle = \|\varphi_M\|^2 = M^4 \int \frac{d^3\mathbf{p}'}{2p_0'}\,\varphi^*(p')\langle\mathbf{p}',M|\mathbf{p},M\rangle\,\varphi(p)\,\frac{d^3\mathbf{p}}{2p_0}.$$

Comparing this with (4.10a), we see that

$$\langle\mathbf{p},M|\mathbf{p}',M\rangle = M^{-2}\,2p_0\,\delta(\mathbf{p}-\mathbf{p}'). \tag{4.12}$$

Thus, the mass-shell states $|\mathbf{p},M\rangle$ have Poincaré invariant normalization.[40] Obviously they do not span the $\tilde{\mathcal{G}}_5$ representation space, but they span a representation space of the Poincaré group, and as we saw, they are bona fide $\tilde{\mathcal{G}}_5$-states. In view of this, Eqs. (4.4a, b) *hold true* for the $|\mathbf{p},M\rangle$ states. In summary we can say that "particles" correspond to the subspaces of the $\tilde{\mathcal{G}}_5$ Hilbert space which are spanned by the subset $|\mathbf{p},M\rangle$ (fixed M) of the set $|p\rangle$ of physical $\tilde{\mathcal{G}}_5$ states.

We can now study physical questions. The most general particle states (with mass M) are given by (4.11b), with (4.10b) understood. They obey

$$P^2|\psi_M\rangle = M^2|\psi_M\rangle \tag{4.13}$$

and they norm is given [cf. (4.12)] by

$$\langle\psi_M|\psi_M\rangle = M^2 \int \frac{d^3\mathbf{p}}{2p_0}\,|\varphi(p_0,\mathbf{p})|^2. \tag{4.14}$$

Consider the operator ${}^MZ_\mu \equiv M^{-1}Q_\mu$. Because of (3.3g),

$${}^M\dot{Z}_\mu \equiv -i[S, {}^MZ_\mu] = M^{-1}P_\mu. \tag{4.15}$$

Therefore, using (4.13),

$${}^M\dot{Z}_\mu\,{}^M\dot{Z}^\mu|\psi_M\rangle = |\psi_M\rangle. \tag{4.16}$$

Thus, on the particle subspace which consists of M-mass shell states, *${}^M\dot{Z}_\mu$ acts as the 4-velocity operator.*

Next, we wish to determine *localized particle states*. Since they must be well-defined particle states with sharp mass, we define

$$|\psi_M;\mathbf{x}\rangle \equiv \mathcal{U}(\mathbf{x})|\psi_M\rangle. \tag{4.17}$$

Since $\mathcal{U}(\mathbf{x}) = \exp(ix^kP_k) = \exp(-i\mathbf{x}\mathbf{P})$, from (4.11b) we get

$$|\psi_M;\mathbf{x}\rangle = \int \frac{d^3\mathbf{p}}{2p_0}\,\exp(-i\mathbf{x}\mathbf{p})\,\varphi(p_0,\mathbf{p})|\mathbf{p},M\rangle, \tag{4.18}$$

with (4.10b) understood. Adopting the Newton–Wigner[41] localization requirement, we demand that

$$\langle\psi_M;\mathbf{x}|\mathcal{U}(-\mathbf{a})|\psi_M;\mathbf{x}\rangle = 0 \quad \text{if } \mathbf{a} \ne 0.$$

Using (4.18) and (4.12), this gives easily

$$\int \frac{d^3\mathbf{p}}{2p_0}\,\exp(i\mathbf{a}\mathbf{p})\,|\varphi(p)|^2 = 0 \quad \text{if } \mathbf{a} \ne 0,$$

so that, assuming the usual regularity condition,[41] we have

$$\varphi(p) = \sqrt{2p_0} \tag{4.19}$$

(apart from an unessential constant of dimension[42] $[\text{length}]^{5/2}$) for the wavefunction of the localized state in (4.18). Therefore, from (4.18) we obtain, when using (4.4a) and doing a partial integration, the following:

$$Q_k|\psi_M;\mathbf{x}\rangle = -il^{-1}\int d^3\mathbf{p}\,\left(\frac{\partial}{\partial p^k}\,(\exp(-i\mathbf{x}\mathbf{p})/\sqrt{2p_0})\right)\,|\mathbf{p},M\rangle$$

$$= -il^{-1}\int \frac{d^3\mathbf{p}}{\sqrt{2p_0}}\quad (ix_k + \tfrac{1}{2}p_k/p_0^2)\exp(-i\mathbf{xp})\,|\mathbf{p},M\rangle$$

$$=\left(x_k - i\frac{P_k}{2P_0^2}\right) l^{-1}\int\frac{d^3\mathbf{p}}{\sqrt{2p_0}}\exp(-i\mathbf{xp})\,|\,\mathbf{p},M\rangle,$$

or, in view of (4.18) and (4.19),

$$\left(lQ_k + i\frac{P_k}{2P_0^2}\right)\ |\psi_M;\mathbf{x}\rangle = x_k\,|\psi_M;\mathbf{x}\rangle. \tag{4.20}$$

This tells us that

$${}^MW_k \equiv lQ_k + i\frac{P_k}{2P_0^2} \tag{4.21}$$

is precisely the Newton—Wigner position operator in configuration space and the states $|\psi_M;\mathbf{x}\rangle$, given by (4.18) with (4.19) and (4.10b) understood, *are the localized* N—W *one-particle states with mass M.* Note that these localized states are simultaneous eigenstates of $P^2 \equiv M^2$ and of ${}^MW_k$, as it should be. Using (4.4a, b), the N—W operator (4.21) can be realized in momentum space by

$${}^MW_k \sim i\frac{\partial}{\partial p^k} + i\frac{p_k}{2p_0^2}, \tag{4.22}$$

which is its familiar form.

The presence of l in (4.21) also sheds light on the meaning of this constant. It clearly determines the scale of length.[43] Turning to conventional c.g.s. units, one should identify $-l^{-1}$ with Planck's constant $\hbar$. This is also born out by Eqs. (4.4a, b). In conventional units, when P_μ is represented by p_μ, Q_μ ought to be represented by $-i\hbar\partial/\partial p^\mu$. Comparison with (4.4a) gives then indeed $l^{-1} = -\hbar$. The superselection rule connected with l merely selects the world with a particular "quantal scale." It is interesting to note that this is not the case for Galilean (nonrelativistic) physics. There the corresponding superselection rule selects a particular Galilean mass M.

Let us summarize. From the event-position operator Q_μ of $\tilde{\mathcal{G}}_5$ we can construct, for each given particle subspace, a 4-velocity operator ${}^M\dot{Z}_\mu$ and a Newton—Wigner position operator ${}^MW_k$. The former is in the Lie algebra of $\tilde{\mathcal{G}}_5$, the latter is a function in the enveloping algebra. In fact, the configuration space operator (4.21) emerges naturally and its nonlocality is explicit. The physical mass shell particle states and even the corresponding localized states are bona fide states in the Hilbert space of $\tilde{\mathcal{G}}_5$.

V. NON-ABELIAN GAUGE SYMMETRY

Since the dynamics of elementary particles appears to be governed by some *non-Abelian* local gauge group, one might ask: What happens if we replace our fundamental Assumption 2 concerning local phase symmetry by the more reasonable requirement of local non-Abelian gauge symmetry? In this section we show that even in this more general case the $\tilde{\mathcal{G}}_5$ structure again emerges, and the dynamical group is simply the direct product of $\tilde{\mathcal{G}}_5$ with the non-Abelian symmetry group.

Let A be a compact N-parameter (simple) Lie group with Hermitian generators I^a ($a=1,2,\ldots,N$) and with the Lie algebra[44]

$$[I^a, I^b] = id_{abc}I^c. \tag{5.1}$$

The underlying space of events will be thought of as $E_{3,1}\times A$ and the corresponding Lie algebra of generators consists of the Poincaré relations (2.1) together with (5.1) and

$$[I^a, P_\mu]=0,\quad [I^a, J_{\mu\nu}]=0. \tag{5.2}$$

This algebra can be realized on a Hilbert space of A-vector valued functions

$$\Phi = \psi(x)\,\chi^{B_1,\ldots,B_s}_{r_1,\ldots,r_\tau}, \tag{5.3}$$

where the B_i ($i=1,\ldots,s$) denote the Casimir invariants of A and the r_α ($\alpha=1,\ldots,\tau$) are state labels for a fixed representation of A. One may think of $\chi^{B_1,\ldots,B_s}$ as a column which, for a p-dimensional representation, has p rows. The realization of $A\times\mathcal{P}$ is given by (2.2) and

$$I^a \sim t^a\quad (a=1,\ldots,N) \tag{5.4}$$

where the t^a are $p\times p$ matrices.

We now replace our Assumption 2 by the new locality requirement that the local version of the symmetry group A be a kinematical symmetry, i.e., we demand:

Assumption 2′: To every transformation[45]

$$\Phi \to \exp[i\omega^a(x)t^a]\,\Phi \tag{5.5}$$

with a differentiable set of functions $\omega^a(x)$ ($a=1,\ldots,N$) there corresponds in Hilbert space a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\,\Phi)(x) = \exp[i\omega^a(x)t^a]\,\Phi. \tag{5.6}$$

From the realizations (2.2) and (5.4) it now follows that under a local transformation (5.5)

$$P_\mu \to P_\mu + t^a\partial_\mu\omega^a, \tag{5.7a}$$

$$J_{\mu\nu}\to J_{\mu\nu} + t^a(x_\mu\partial_\nu\omega^a - x_\nu\partial_\mu\omega^a), \tag{5.7b}$$

$$I^a \to [\exp(i\omega^b k^b)]_{ac}I^c. \tag{5.7c}$$

In (5.7c) the k^b ($b=1,\ldots,N$) are the $N\times N$ matrices of the adjoint representation of A, i.e.,

$$(k^b)_{ef} = id_{bef}. \tag{5.8}$$

As was the case for local phase symmetry, we must enlarge the algebra of observables to ensure that (5.7a, b, c) represent a permutation of the observables. Somewhat surprisingly, all we have to do is to include in the algebra the already familiar Q_μ operators which, are now taken to satisfy the Lie relations

$$[P_\mu, Q_\nu] = -il^{-1}g_{\mu\nu},\quad [Q_\mu, Q_\nu]=0, \tag{5.9a}$$

$$[I^a, Q_\mu]=0. \tag{5.9b}$$

To see this, take $\omega^a(x) = c^a_\mu x^\mu$ (with constant c^a_μ) and set

$$\mathcal{U} = \exp[-ilc^a_\mu F^\mu_a], \tag{5.10}$$

where the F^μ_a ($\mu=0,1,2,3;\ a=1,\ldots,N$) are dimensionless. Then, from (5.7a)

$$P_\mu + t^a c^a_\mu \equiv \mathcal{U}P_\mu\mathcal{U}^{-1} = P_\mu - ilc^a_\nu[F^\nu_a, P_\mu] + \cdots. \tag{5.11}$$

If we put

$$F^\nu_a = I^aQ^\nu, \tag{5.12}$$

with Q^ν obeying (5.9a), then (5.11) is satisfied [if (5.2) and (5.4) is taken into account]. From the algebra of P_μ, Q_μ, I^a it now follows that, as before, Q_μ can be realized as

$$Q_\mu \sim -l^{-1}x_\mu, \tag{5.13}$$

Then, from (5.7b), with $\omega^a = c^a_\mu x^\mu$ we get

$$J_{\mu\nu} + t^a(x_\mu c^a_\nu - x_\nu c^a_\mu) \equiv \mathcal{U} J_{\mu\nu}\mathcal{U}^{-1} \tag{5.14}$$
$$= J_{\mu\nu} - ilc^a_\rho [F^\rho_a, J_{\mu\nu}] + \cdots = J_{\mu\nu} - ilc^a_\rho I^a [Q^\rho, J_{\mu\nu}] + \cdots,$$

so that we must have

$$[J_{\mu\nu}, Q_\sigma] = i(g_{\nu\sigma}Q_\mu - g_{\sigma\mu}Q_\nu), \tag{5.15}$$

as in the Abelian case. Finally, with (5.10), (5.12), (5.1), and (5.8) we find

$$\mathcal{U} I^a \mathcal{U}^{-1} = I^a - ilc^b_\rho [F^\rho_b, I^a] + \cdots$$
$$= I^a - ilc^b_\rho Q^\rho [I^b, I^a] + \cdots$$
$$= I^a + ic^b_\rho x^\rho (k^b)_{ac} I^c + \cdots = [\exp(ic^b_\rho x^\rho k^b)]_{ac} I^c,$$

so that, for the special choice of ω^b, Eq. (5.7c) is also satisfied. In conclusion, we note that, in addition to Eqs. (5.7a, b, c) we also have the transformation law

$$Q_\mu \to Q_\mu \tag{5.16}$$

for the behavior of Q_μ under local gauge transformations.

Since, for a general local A-transformation we can set

$$\omega^a(x) = \sum_{n=0}^{\infty} c[{}^{(n)}M^a_\mu x^\mu]^n \quad (a = 1, \ldots, N),$$

it now follows (as in the Abelian case) that arbitrary local transformations are kinematical symmetry transformations of the operator algebra generated by $\{P_\mu, J_{\mu\nu}, I^a, Q_\mu, I\}$. This concludes the proof of our assertion that it is sufficient to enlarge the original set by the Q_μ. Note that Q_μ is invariant under A [cf. (5.9b)] and otherwise it obeys the usual Lie relations.

It may be interesting to point out that the factorization (5.12) of F^ν_a is forced on us, because otherwise we would not have a closed Lie algebra. Indeed, from (5.14) we have

$$[F^\rho_a, J_{\mu\nu}] = -il^{-1}(g^\rho_\mu x_\nu - g^\rho_\nu x_\mu) I^a,$$

and even if we now admitted the "additional" operators Q_μ realized by $-l^{-1}x_\mu$, the rhs of this equation would belong to the *enveloping* algebra. The only solution of the problem[46] is the factorization (5.12).

On the other hand, since the first Casimir invariant of the global group A can be written as

$$B_1 = I^a I^a,$$

Eq. (5.12) may be solved to give

$$Q_\mu = B_1^{-1} I^a F^a_\mu. \tag{5.17}$$

One may consider this as a definition of Q_μ in terms of I^a and of the generators F^a_μ of linear local symmetry transformations. From this viewpoint, the RG boosts are generated, in the case of a non-Abelian symmetry, by *combining a global A-transformation with a specific (linear) local gauge transformation.*

Because of (5.9b), the kinematical group has the structure

$$K = A \times \{SL(2,C)^J \otimes [T_4^P \otimes (T_4^Q \times T_1^{I^{-1}})]\}. \tag{5.18}$$

The Casimir invariants are those given by Eqs. (2.19a, b, c) and the $B_1, \ldots, B_s$.

Since intrinsic development must be compatible not only with Poincaré transformations but also with the global A-transformations, development transformations are now characterized by

$$P_\mu \to P_\mu, \quad J_{\mu\nu} \to J_{\mu\nu}, \quad I^a \to I^a, \quad Q_\mu \to f(Q_\nu, P_\nu, J_{\rho\sigma}, I^a).$$

Using the same arguments as in Sec. III, the most general form of the generator is again given[47] by Eq. (3.1). From Assumptions 5 and 6 we then once again obtain the form of S as specified by Eq. (3.2). The entire Lie algebra is now specified by Eqs. (3.3a)–(3.3g) plus (5.1), (5.2), (5.9b) as well as the obvious relations $[S, I^a] = 0$. In other words, the structure of the dynamical group is

$$G = A \times \tilde{\mathcal{G}}_5 = A \times (\mathcal{G}_5 \otimes T_1^{I^{-1}}). \tag{5.19}$$

The Casimir invariants are given by (3.5a)–(3.5d) to which we have to add the B_i $(i = 1, \ldots, s)$.

Choosing the homogeneous space $A \times \mathcal{G}_5/SO(3,1)^J \otimes T_4{}^Q$ with elements $(\overline{\alpha}, \overline{\sigma}, \overline{a})$ (where $\overline{\alpha}$ stands for the parameters of A) and making the identification $(\overline{\alpha}, \overline{\sigma}, \overline{a}) \to (g, u, x)$ (where $g \in A$), we obtain the active representation of G as a group of endomorphisms of $A(g) \times E_{3,1}(x) \times E_1(u)$ given by the transformations

$$g \to \exp(i\alpha^b k^b) g,$$
$$u \to u + \sigma, \tag{5.20}$$
$$x_\mu \to \Lambda^\nu{}_\mu x_\nu + a_\mu + u b_\mu.$$

On the total Hilbert space $\mathcal{H}$ we again have the realizations (3.11) by differential operators, amended by $I^a \sim t^a$.

VI. INTERACTING SYSTEMS

In this section we shall follow closely the arguments presented in Sec. IV of Ref. 2 for obtaining a unique form of interactions. We concentrate on the non-Abelian symmetry.

From the realizations of the generators on $\mathcal{H}$ we obtain the effect of a local gauge transformation

$$\Phi(x;u) \to \exp[i\omega^a(x)t^a]\,\Phi(x;u) \tag{6.1}$$

as being given by

$$P_\mu \to P_\mu + t^a \partial_\mu \omega^a, \tag{6.2a}$$
$$Q_\mu \to Q_\mu + u t^a \partial_\mu \omega^a, \tag{6.2b}$$
$$J_{\mu\nu} \to J_{\mu\nu} + t^a(x_\mu \partial_\nu \omega^a - x_\nu \partial_\mu \omega^a), \tag{6.2c}$$
$$S \to S, \tag{6.2d}$$
$$I^a \to [\exp(i\omega^b k^b)]_{ac} I^c. \tag{6.2e}$$

Equation (6.2b) tells us that the event-space position operators Q_μ, and hence the associated particle position operators ${}^M W_k$ of Sec. IV, are not invariant under local gauge transformations. Since there is no reason why localization should depend on an arbitrary gauge, we stipulate

Assumption 7: Local gauge transformations commute with the Q_μ.

To satisfy the requirement that $Q_\mu \to Q_\mu$ under an arbitrary local transformation with a differentiable $\omega^a(x)$, we must obviously modify the realization (3.11) of the Q_μ. We set

$$Q_\mu \sim -l^{-1}x_\mu + iu\partial_\mu - ut^a W^a_\mu(x), \tag{6.3}$$

where the W^a_μ represent some vector fields. This modification is suggested by the requirements that (a) when the "interaction" characterized by W^a_μ is switched off, we recover (3.11), (b) when considering Q_μ on the slice $u=0$, we recover (2.11), (c) the modified Q_μ is still a scalar relative to the global group A. We now calculate, for an arbitrary local gauge transformation,

$$\begin{aligned}(\mathcal{U}Q_\mu\mathcal{U}^{-1}\Phi)(x;u) &= \exp(it^b\omega^b)[(-l^{-1}x_\mu + iu\partial_\mu) \\ &\quad \times \exp(-it^b\omega^b)\Phi(x;u)] \\ - u(\mathcal{U}t^a W^a_\mu \mathcal{U}^{-1}\Phi)(x;u) &= (-l^{-1}x_\mu + iu\partial_\mu + ut^b\partial_\mu\omega^b)\Phi(x;u) \\ &\quad - u(\mathcal{U}t^a W^a_\mu\mathcal{U}^{-1}\Phi)(x;u).\end{aligned}$$

Therefore, $Q_\mu \to Q_\mu$ provided

$$t^a W^a_\mu \to \exp(-it^b\omega^b)\, t^a W^a_\mu \exp(it^b\omega^b) + t^a\partial_\mu\omega^a. \tag{6.4}$$

In summary: The gauge-independence of localization can be achieved if an interaction with a set of vector fields W^a_μ $(a=1,\ldots,N)$ is introduced, where these gauge fields transform under a local gauge transformation in the familiar manner[48] as given by (6.4).

In passing we note that the modification of Q_μ does not affect the gauge behavior of $P_\mu, J_{\mu\nu}, S, I^a$, so that the Eqs. (6.2) still hold except (6.2b) which is replaced by $Q_\mu \to Q_\mu$.

In order to find the explicit form of the interaction, we use (6.3) and $S \sim i\partial_u$, $P_\mu \sim i\partial_\mu$ to compute that

$$[S, Q_\mu] = i(P_\mu - t^a W^a_\mu). \tag{6.5}$$

By transforming with $\exp(iuS)$, this tells us that on the slice $u=0$, $[\bar S, \bar Q_\mu] = i(\bar P_\mu - t^a \bar W^a_\mu)$. Since $\bar W^a_\mu$ is a power series in $\bar Q_\mu$, we easily find[49] that

$$\bar S = -\frac{l}{2}\bar P^2 + \frac{l}{2}\bar P^\nu t^a \bar W^a_\nu + \frac{l}{2} t^a \bar W^a_\nu \bar P^\nu + \bar N,$$

where N is an arbitrary scalar function of $\bar Q_\mu$. This expression can be trivially completed to the "square" of $\bar P_\mu - t^a\bar W^a_\mu$ (modulo an additive scalar), and finally transforming with $\exp(-iuS)$ we obtain, on the slice $\mathcal{H}_u$,

$$S = -\frac{l}{2}(P_\mu - t^a W^a_\mu)(P^\mu - t^a W^\mu_a) + V, \tag{6.6}$$

where V is an arbitrary function of Q_μ.

Since in the $\tilde{\mathcal{G}}_5$ theory S plays the role of a covariant relativistic Hamiltonian (relative to development in historical time), Eq. (6.6) describes a unique gauge invariant structure for interacting systems.[50] From (6.5) we see that, in the presence of interactions, the momentum is no longer P_μ but rather

$$\Pi_\mu \equiv P_\mu - t^a W^a_\mu. \tag{6.7}$$

This is invariant under local gauge transformations, and so is

$$S = -\frac{l}{2}\Pi_\mu \Pi^\mu + V. \tag{6.8}$$

The realization of (6.6) on the function space replaces the wave equation (3.12) by one containing the familiar "covariant derivatives,"

$$[-(i\partial_\mu - t^a W^a_\mu)(i\partial^\mu - t^a W^\mu_a) - 2l^{-1}V - 2l^{-1}i\partial_u]\Phi(x;u) = 0. \tag{6.9}$$

We can go one step further. Similarly, as we did in ref. 2 for the nonrelativistic system, we may also now decide to make the superselection rule connected with the incoherence of the slices $\mathcal{H}_u$ an explicit stipulation. In other words, we can generalize Assumption 2′ to become

Assumption 8: To every transformation

$$\Phi(x;u) \to \exp[it^a\omega^a(x,u)]\Phi(x;u) \tag{6.10}$$

with a differentiable $\omega^a(x,u)$ there corresponds in the Hilbert space $\mathcal{H}$ a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\Phi)(x;u) = \exp[it^a\omega^a(x,u)]\Phi(x;u). \tag{6.11}$$

Under these generalized, historical time dependent gauge transformations the formal behavior of $P_\mu, Q_\mu, J_{\mu\nu}, I^a$ is the same as before. However, S is no longer invariant: from $S \sim i\partial_u$ it follows that

$$S \to S + t^a\partial_u\omega^a. \tag{6.12}$$

Therefore, under u-dependent local gauge transformations Eq. (6.6) or (6.8) is inconsistent: The lhs transforms according to (6.12) but the rhs is unchanged. This can be remedied if we restrict the so far arbitrary V field to transform under a general local gauge transformation according to the law

$$V \to V + t^a\partial_u\omega^a. \tag{6.13}$$

We remark that the wave equation (6.9) can now be considered as arising from the free wave equation (3.12) by replacing all derivatives with covariant derivatives:

$$\begin{aligned}\partial_\mu &\Rightarrow \partial_\mu + it^a W^a_\mu, \\ \partial_u &\Rightarrow \partial_u - iV.\end{aligned} \tag{6.14}$$

The wave equation (6.9) is invariant under the simultaneous transformations

$$\begin{aligned}&\Phi \to \exp[-it^a\omega^a(x,u)]\Phi, \\ &t^a W^a_\mu \to \exp(-it^b\omega^b)\, t^a W^a_\mu \exp(it^b\omega^b) + t^a\partial_\mu\omega^a, \\ &V \to V + t^a\partial_u\omega^a.\end{aligned} \tag{6.15}$$

Whereas vector gauge fields are familiar and in fact most desirable, it may be difficult to ascertain, at the present time, the meaning and significance of the scalar gauge field V. However, it is easy to get rid of V altogether, viz. by performing a gauge transformation with

$$t^a\omega^a(x,u) = -\int_0^u V\, du. \tag{6.16}$$

Then

$$S \to S - V \equiv \tilde S, \qquad \Pi_\mu \to \Pi_\mu \equiv \tilde \Pi_\mu \tag{6.17}$$

and (6.8) becomes

$$\tilde S = -\frac{l}{2}\tilde\Pi_\mu \tilde\Pi^\mu. \tag{6.18}$$

Once we chose this particular gauge, we are no longer permitted to perform gauge transformations with a u-dependent ω^a. But in this special gauge, the remaining gauge transformations are an invariance property of the dynamics, in the sense that $\tilde{S}\to\tilde{S}$.

When the special gauge (6.16) is taken, then, in view of (6.15), the wave equation (6.9) assumes the form

$$[-(i\partial_\mu - t^aW^a_\mu)(i\partial^\mu - t^aW^\mu_a) - 2l^{-1}i\partial_u]\Phi(x;u)=0, \quad (6.19)$$

where of course W^a_μ means the vector fields in the special gauge. Separation of variables is obtained by setting

$$\Phi(x;u)=\Psi(x)\,\Theta(u) \quad (6.20)$$

and we have

$$\Theta(u)=\exp(i\tfrac{1}{2}lm^2u) \quad (6.21)$$

and

$$[-(i\partial_\mu - t^aW^a_\mu)(i\partial^\mu - t^aW^a_\mu)+m^2]\Psi(x)=0. \quad (6.22)$$

Here the *separation constant* m^2 appeared as the eigenvalue[51] of $-2l^{-1}i\partial_u=-2l^{-1}S=\Pi_\mu\Pi^\mu$, and it therefore represents the squared mass in the presence of the interaction. Equation (6.22) must be looked upon as an *eigenvalue equation* for m^2 with the boundary value condition that $\Psi(x)$ be Poincaré-normalizable. Thus, (6.22) will give rise to a *mass spectrum*. To actually perform the calculation, one ought to know W^a_μ. This can be done if one sets up an invariant Lagrangian formulation including the gauge fields and thus obtains field equations for the latter (coupled to Ψ). This problem [and the inclusion of $SL(2,C)$ spin] will be considered at a later time.

VII. CHARGE SUPERSELECTION RULES

As was discussed in Sec. V of Ref. 2, the gauge behavior of the vector field gives rise to a charge superselection rule for Galilean systems. A similar situation arises for $\tilde{\mathcal{G}}_5$.

Take a local A-transformation with the special phase $\omega^a(x)=c^a_\mu x^\mu$ and for ease of calculation, assume that the constants c^a_μ are infinitesimal. Then, because of (6.4),

$$W^a_\mu \to W^a_\mu + d_{abe}\,c^b_\nu x^\nu W^e_\mu + c^a_\mu.$$

Denote that part[52] of the corresponding unitary operator which acts on functionals $\mathcal{R}$ of W^a_μ, by

$$\mathcal{U}=\exp[-ilc^\mu_e\mathcal{K}^e_\mu]. \quad (7.1)$$

(Here $\mathcal{K}^e_\mu$ is dimensionless). Then one must have

$$d_{abe}\,c^b_\nu x^\nu W^e_\mu + c^a_\mu = -ilc^\nu_e[\mathcal{K}^e_\nu, W^a_\mu]. \quad (7.2)$$

Set

$$\mathcal{K}^e_\nu=\mathcal{L}^e_\nu+I^eQ_\nu, \quad (7.3)$$

and stipulate that

$$[\mathcal{L}^e_\nu, W^a_\mu]=il^{-1}\delta_{ea}g_{\nu\mu}. \quad (7.4)$$

Since Q_ν and W^a_μ commute and since W^a_μ transforms under the adjoint representation for the global group, we have

$$[I^eQ_\nu, W^a_\mu]=id_{eaf}W^f_\mu Q_\nu. \quad (7.5)$$

Substituting (7.3) into (7.2) and using (7.4), (7.5) we see that (7.2) is satisfied on the slice $u=0$, because there, on account of (6.3), $Q^\nu\sim -l^{-1}x^\nu$. Therefore, to accomodate the gauge transformation of W^a_μ at $u=0$, we need only to adjoin the new observables $\mathcal{L}^e_\nu$, whose action on the functionals $\mathcal{R}$ of W^a_μ is defined by (7.4). We can realize $\mathcal{L}^e_\nu$ by setting

$$\mathcal{L}^e_\nu\sim il^{-1}\frac{\delta}{\delta W^\nu_e}.$$

For arbitrary slices u, $\mathcal{L}^e_\nu$ is obtained by transforming this with $\exp(-iuS)$.

Since the rhs of (7.4) is in the center of the algebra, we have a new superselection rule for the complete system generated by $P_\mu, Q_\mu, J_{\mu\nu}, S, W^a_\mu$. The superselection rule corresponds to that of the "charge" associated with the gauge coupling. Indeed, using an explicit coupling constant γ, we must replace W^a_μ by $\gamma\hat{W}^a_\mu$. Then (7.4) gives

$$[\mathcal{L}^e_\nu, \hat{W}^a_\mu]=il^{-1}\gamma^{-1}\delta_{ea}g_{\nu\mu}, \quad (7.6)$$

so that[53] the "supersymmetry" observable is γ^{-1}.

It is significant that, contrary to the Poincaré framework, the $\tilde{\mathcal{G}}_5$ theory automatically explains "charge superselection rules."

[1]For a review see, for example, E.S. Abers and B.W. Lee, Physics Reports 9, 1 (1973) and S. Weinberg, Rev. Mod. Phys. 46, 255 (1974).
[2]P. Roman and J.P. Leveille, J. Math. Phys. 15, 1760 (1974).
[3]J.M. Jauch, Helv. Phys. Acta 37, 284 (1964).
[4]C. Piron, Found. Phys. 2, 287 (1972).
[5]J.-M. Lévy-Leblond, Commun. Math. Phys. 4, 157 (1967) and 6, 286 (1967); Ann. Phys. (N.Y.) 57, 481 (1970).
[6]Concerning a precise formulation of this concept, see L.P. Horwitz and C. Piron, Helv. Phys. Acta 46, 361 (1973).
[7]J.J. Aghassi, P. Roman and R.M. Santilli, Phys. Rev. D 1, 2753 (1970).
[8]J.J. Aghassi, P. Roman, and R.M. Santilli, J. Math. Phys. 11, 2297 (1970).
[9]J.J. Aghassi, P. Roman, and R.M. Santilli, Nuovo Cimento 5 A, 551 (1971).
[10]R.M. Santilli, Particles and Nuclei 1, 81 (1970).
[11]P.L. Huddleston, M. Lorente, and P. Roman, Found. Phys. (to be published).
[12]In this paper $\otimes$ and $\times$ will indicate semidirect and direct products, respectively.
[13]The signature is (+;−,−,−).
[14]C.N. Yang and R. Mills, Phys. Rev. 96, 191 (1954).
[15]R. Utiyama, Phys. Rev. 101, 1597 (1956).
[16]E.P. Wigner, *Gruppentheorie und ihre Anwendung* (Vieweg, Braunschweig, 1931); English translation by J.J. Griffin (Academic, New York, 1959).
[17]The arbitrary constant l is inserted for dimensional reasons. Since the c_μ have inverse length dimension and F is dimensionless, and since we use units $\hbar=c=1$, the constant l has dimension of length.
[18]In Ref. 8 and 9 we used the term "zest".
[19]J.E. Johnson, Phys. Rev. 181, 1755 (1969); D 3, 1735 (1971).
[20]A.A. Broyles, Phys. Rev. D 1, 979 (1970).
[21]L. Castell, Nuovo Cimento A 49, 285 (1967).
[22]We consider only unitary representations and select, for definiteness, the principal series.
[23]$T_{\mu\nu}P^\mu P^\nu$ and $\epsilon_{\mu\nu\rho\sigma}T^{\mu\nu}P^\rho P^\sigma$ vanish by symmetry. $T_{\mu\nu}T^{\mu\nu}$ and $\epsilon_{\mu\nu\rho\sigma}T^{\mu\nu}T^{\rho\sigma}$ are, in any irreducible representation, multiples of the identity, so they can be disregarded. It is also interesting to observe that $-\frac{1}{2}T_{\mu\nu}T^{\mu\nu}P_\sigma P^\sigma + T_{\alpha\mu}T^{\beta\mu}P^\alpha P_\beta \equiv W^2$ is the

squared Pauli—Lubanski vector associated with $SL(2,C)$ spin; so we can say that S depends on P^2, W^2, I.

[24] S cannot depend on $T_{\alpha\mu}T^{\beta\nu}P^{\alpha}P_{\beta}$ because, for example, $[Q_{\nu}, T_{\alpha\mu}T^{\beta\mu}P^{\alpha}P_{\beta}] = il^{-1}(T_{\nu\mu}T^{\beta\mu} + T^{\beta\mu}T_{\nu\mu})P_{\beta}$ is not in the Lie algebra. Higher powers of P^2 are ruled out similarly.

[25] In Ref. 7 a slightly different notation is used and there is a minor error in displaying the structure.

[26] In Ref. 8 we claimed that u is essentially the proper time. However, closer inspection of the calculation shows that this is erroneous. In fact, u is the proper time in $E_{3,2}$ which is related to $\mathcal{G}_5$ by the circumstance that the latter arises from $SO(3,2)\otimes T_5$ by means of a contraction, cf. Refs. 8 and **11**.

[27] P. Roman, J.J. Aghassi, and P.L. Huddleston, J. Math. Phys. **13**, 1852 (1972). Omit in Eqs. (2.1) the last two. For the corresponding result in the nonrelativistic case, cf. Eqs. (1.3a)—(1.3d) in that paper and for a deeper analysis, Appendix A of Ref. 2.

[28] The inner product for each $\mathcal{H}_u$ is of course given by $i\int\varphi^{\dagger}(x;u)\overleftrightarrow{\partial_0}\varphi(x;u)d^3\mathbf{x} \equiv i\int\varphi^*(x)\overleftrightarrow{\partial_0}\varphi(x)d^3\mathbf{x}$.

[29] A concept closely related to historical time was previously suggested by R.P. Feynman, Phys. Rev. **76**, 749 (1949).

[30] Since representations with different $\mathcal{D}_1$ are ray-equivalent (cf. Ref. 9), one might take $\mathcal{D}_1 = 0$ without restricting generality.

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

[32] This is analogous to the Galilean case where the energy $E(0 \le E < \infty)$ is not a representation label but only a state label.

[33] Since the meaning of the k, c labels as specifying spin-towers has been well clarified in Ref. 9, we now supress these labels.

[34] This is analogous to Galilean physics where for states $|E,\mathbf{p}\rangle$ obeying the wave equation, the label E is redundant.

[35] This norm is $\mathcal{G}_5$ – invariant, which can be seen from the fact that $\int dr d^4p\,\delta(p^2 + 2l^{-1}r)\delta(p-p') = l/2$ where the coefficient of $\delta(p-p')$ is the invariant measure for $\mathcal{G}_5$.

[36] Note that $|p\rangle \sim [|r,p\rangle|_{r=-p^2/2}l^{-1}$.

[37] The naive approach, i.e., to use (4.2) directly, fails, because it would give the undefined expression

$$\langle\mathbf{p},M|\mathbf{p}',M\rangle = \delta(\mathbf{p}-\mathbf{p}')\delta(\sqrt{\mathbf{p}^2+M^2} - \sqrt{\mathbf{p}'^2+M^2}) = \delta(\mathbf{p}-\mathbf{p}')\delta(0).$$

[38] In (4.8), $M > 0$ and $SL(2,C)$ means the orthochronous connected component.

[39] The unusual form of δ and the ensuing factor M^2 is connected with the fact that $\varphi(p)$ has dimension [length]^2 and that $\|\varphi\|^2$ should be dimensionless.

[40] The unusual factor M^{-2} arises because of dimensional reasons (cf. Ref. 39). Observe that the dimension of $|\mathbf{p},M\rangle$ is [length]^2, exactly as is the case for all $\mathcal{G}_5$ states $|p\rangle$.

[41] T.D. Newton and E.P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).

[42] Recall that $\varphi(p)$ must have dimension [length]^2.

[43] A similar attitude concerning the meaning of the constant on the rhs of "relativistic Heisenberg commutation relations" is implicit in the work of L. Castell, Nuovo Cimento A **46**, 1 (1966). See also Ref. 21.

[44] Upper or lower index-positions for the indices a,b, etc. referring to the group A are chosen in the sequel simply by printing convenience and have no significance.

[45] Summation over a $(a = 1, \ldots, N)$ is understood.

[46] There is no problem with other commutators: the relations. $[P_{\mu}, F_{\nu}^{a}] = -il^{-1}I^{a}g_{\rho}$ and $[F_{\mu}^{b}, I^{a}] = id_{bac}F_{\mu}^{c}$ would be acceptable.

[47] Note that in any irreducible representation the B_i are multiples of the identity.

[48] As is well known, (6.4) implies that under the global group A the fields W_u^a transform under the adjoint representation.

[49] Use $[\bar{P}_u, \bar{Q}_{\nu}] = il^{-1}g_{u\nu}$.

[50] We confined ourselves to representation with $k=0$, $c=0$. When spin-towers are considered, demanding gauge invariance of $T_{\mu\nu}$ leads to a modification thereof, and correspondingly, additional interaction terms depending on $SL(2,C)$ spin arise.

[51] The eigenfunction is Θ. The last equality follows from (6.8) since in the gauge used, $V = 0$.

[52] The part which acts on functions of P_{μ}, Q_{μ}, $J_{\mu\nu}$, S is of course given by (5.10).

[53] The factor l^{-1} has no significance because, as discussed at the end of Sec. IV, it is essentially just $-\hbar$.

Asymptotic solutions of second-order linear equations with three transition points

Ali H. Nayfeh

Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
(Received 19 November 1973; final revised manuscript received 22 February 1974)

A uniformly valid asymptotic expansion is obtained for the regular solution of a class of second-order linear differential equations with three transition points—a turning point and two regular singular points. The solution is found by matching three different solutions obtained using the Langer Transformation. The matching yields the eigenvalues and the eigenfunctions.

I. INTRODUCTION

We seek asymptotic solutions for large λ to the differential equation

$$\frac{d}{dx}\left((x+a)(b-x)\frac{dy}{dx}\right)+\left(\frac{p(x)}{(x+a)(b-x)}+\lambda x^n g(x)(x+a)^{m+1}(b-x)^{k+1}\right)y=0 \quad (1)$$

that are regular on the interval $[-a, b]$, where a and b are positive numbers, $p(x)$ and $g(x)$ are regular functions and $g(x)>0$ on $[-a, b]$, and n, m, and k are integers such that $n \geq 0$, $m \geq -2$, and $k \geq -2$. For $n \neq 0$, Eq. (1) has three transition points—$x=0$ is a turning point and $x=-a$ and $x=b$ are regular singular points.

The special case $p(x)=2(1-x^2)$, $g(x)=1$, $a=b=1$, $n=-m=-k=1$ describes stationary waves of small amplitude on the surface of a liquid sphere of unit radius whose center of mass is undergoing a constant acceleration.[1] Harper, Chang, and Grube[2] obtained a second-order asymptotic solution to this special case by using the method of matched asymptotic expansions (e.g., Chap. 4 of Ref. 3). The same technique was used by Jeffreys[4] to treat a problem with two simple turning points. The solution was represented by five asymptotic expansions valid on the intervals $[-1, -1+\delta_1]$, $[-1+\delta_1, -\delta_2]$, $[-\delta_2, \delta_3]$, $[\delta_3, 1-\delta_4]$, and $[1-\delta_4, 1]$, where the δ_i are small positive numbers. The five expansions were then matched to determine the eigenvalues and the eigenfunctions.

Nayfeh (Sec. 7.3.3 of Ref. 3) used a combination of the Langer transformation (e.g., Sec. 7.3.2 of Ref. 3) and the method of matched asymptotic expansions to obtain a uniformly valid asymptotic solution to a problem with two simple turning points. Rather than use the procedure of Ref. 2 and represent the solution by five asymptotic expansions, Nayfeh[3] represented the solution by only two expansions. Nayfeh[5] used the method of multiple scales (e.g., Sec. 6.4.4 of Ref. 3) to analyze the case of two simple turning points, while Nayfeh[6] used a combination of the Langer transformation and the method of matched asymptotic expansions to analyze a problem with two transition points—a turning point and a regular singular point of any order.

Problems with multiple transition points were also treated by using the Olver transformation (e.g., Sec. 7.3.2 of Ref. 3). The solution is represented by a single uniformly valid expansion by relating it to the solution of an equation which approximates the original equation. Using this approach, Olver,[7] Moriguchi,[8] and Pike[9] treated problems with two turning points. Problems with several turning points were treated by Evgrafov and Fedoryuk,[10] Hsieh and Sibuya,[11] Sibuya,[12] and Lynn and Keller[13] among others.

In this paper, we determine an asymptotic solution to Eq. (1) by using a combination of the Langer transformation and the method of matched asymptotic expansions. We prefer to use this technique rather than the Olver transformation because there exists no solution yet to the related equation. Thus, we represent the solution of the general problem by three expansions valid on the intervals $[-a, -\delta_2]$, $[-a+\delta_1, b-\delta_4]$, and $[\delta_3, b]$. Then, we match these expansions to determine the eigenvalues.

Before carrying out the expansions, it is more convenient to remove the first derivative in Eq. (1) by introducing the transformation

$$y(x)=u(x)(x+a)^{-1/2}(b-x)^{-1/2}. \quad (2)$$

The result is

$$\frac{d^2u}{dx^2}+\left(\frac{f(x)}{(x+a)^2(b-x)^2}+\lambda q(x)\right)u=0, \quad (3)$$

where

$$f(x)=p(x)+\tfrac{1}{4}(a+b)^2,\qquad q(x)=x^n g(x)(x+a)^m(b-x)^k. \quad (4)$$

II. AN EXPANSION VALID ON $[-a, -\delta_2]$

To determine an expansion valid near the singular transition point $x=-a$, we note that, as $x\to -a$, Eq. (3) tends to

$$\frac{d^2u}{dx^2}+\left((-1)^n\lambda q_1(x+a)^m+\frac{r_1}{(x+a)^2}\right)u=0, \quad (5)$$

where

$$q_1=a^n g(-a)(b+a)^k \quad\text{and}\quad r_1=f(-a)(b+a)^{-2}. \quad (6)$$

Hence, an asymptotic solution valid near $x=-a$ can be obtained by relating this solution to the solution of the "related" equation

$$\frac{d^2v}{dz^2}+\left((-1)^n\lambda z^m+\frac{r_1}{z^2}\right)v=0. \quad (7)$$

To relate the solutions of Eq. (3) to the solutions of Eq. (7), we introduce the transformation (e.g., Sec. 7.3.9 of Ref. 3)

$$z=\phi(x),\qquad v=u(x)[\phi'(x)]^{1/2}, \quad (8)$$
$$\beta^{-1}\phi^\beta=\int_{-a}^{x}[q(-\xi)]^{1/2}\,d\xi=G_1(x)$$

in Eq. (3) and obtain

$$\frac{d^2v}{dz^2}+\left((-1)^n\lambda z^m+\frac{r_1}{z^2}\right)v=F_1(x)v, \tag{9}$$

where primes denote differentiation with respect to x and

$$F_1=\frac{r_1}{\phi^2}-\frac{1}{\phi'^2}\left(\frac{f(x)}{(x+a)^2(b-x)^2}+\frac{3\phi''^2}{4\phi'^2}-\frac{\phi'''}{2\phi'}\right), \tag{10a}$$

$$\beta=(m+2)/2.$$

As $x\to -a$,

$$\beta^{-1}\phi^\beta\to\beta^{-1}q_1^{1/2}(x+a)^\beta \tag{10b}$$

so that

$$\phi\to(q_1)^{1/2\beta}(x+a), \quad \phi'\to(q_1)^{1/2\beta},$$

and

$$F_1=O[(x+a)^{-1}].$$

Hence, a first-approximation to Eq. (9) is given by Eq. (7), whose general solution is

$$v=z^{1/2}[\tilde{c}_1J_\nu(\lambda^{1/2}\beta^{-1}z^\beta)+\tilde{c}_2J_{-\nu}(\lambda^{1/2}\beta^{-1}z^\beta)] \quad \text{for even } n \tag{11}$$

and

$$v=z^{1/2}[\tilde{c}_1I_\nu(\lambda^{1/2}\beta^{-1}z^\beta)+\tilde{c}_2I_{-\nu}(\lambda^{1/2}\beta^{-1}z^\beta)] \quad \text{for odd } n \tag{12}$$

where $\tilde{c}_1$ and $\tilde{c}_2$ are arbitrary constants and

$$\nu=(1-4r_1)^{1/2}/(2+m). \tag{13}$$

In what follows, we restrict our analysis to the case $r_1\leq\frac{1}{4}$ so that ν is real.

In order that y be regular at $x=-a$, Eqs. (2), (8), (11), and (12) show that $\tilde{c}_2=0$. Hence,

$$u_1(x)=c_1[G_1(x)]^{1/2}[q(x)]^{-1/4}J_\nu[\lambda^{1/2}G_1(x)][1+o(1)] \quad \text{as } \lambda\to\infty \tag{14}$$

for even n and

$$u_1(x)=c_1[G_1(x)]^{1/2}[q(-x)]^{-1/4}I_\nu[\lambda^{1/2}G_1(x)][1+o(1)] \quad \text{as } \lambda\to\infty \tag{15}$$

for odd n, where c_1 and c_2 are arbitrary constants. These expansions, although valid at $x=-a$, they break down as $x\to 0$ if $n\neq 0$. Thus, they are valid only on the interval $[-a,-\delta_2]$. An expansion valid near $x=0$ is obtained in the next section.

III. AN EXPANSION VALID ON $[-a+\delta_1, b-\delta_4]$

As $x\to 0$, Eq. (3) tends to

$$\frac{d^2u}{dx^2}+\lambda q_0x^nu=0, \tag{16}$$

where $q_0=a^mb^kg(a)$. Therefore, an asymptotic solution to Eq. (3) valid near $x=0$ can be obtained by relating it to the solutions of

$$\frac{d^2v}{dz^2}+\lambda z^nv=0. \tag{17}$$

To do this, we introduce the transformation

$$z=\phi(x), \quad v=u(x)[\phi'(x)]^{1/2},$$
$$\frac{2}{n+2}\phi^{(n+2)/2}=\int_0^x[q(\xi)]^{1/2}d\xi=G_0(x) \tag{18}$$

in Eq. (3) and obtain

$$\frac{d^2v}{dz^2}+\lambda z^nv=F_0v, \tag{19}$$

where

$$F_0=-\frac{1}{\phi'^2}\left(\frac{f(x)}{(x+a)^2(b-x)^2}+\frac{3\phi''^2}{4\phi'^2}-\frac{\phi'''}{2\phi'}\right). \tag{20}$$

As $x\to 0$, $\phi=O(x)$, $\phi'=O(1)$, and $F_0=O(1)$. Since λ is large, v is given approximately by Eq. (17) whose general solution is

$$v=z^{1/2}\left[\tilde{c}_3J_\mu\left(\frac{2\lambda^{1/2}}{n+2}z^{(n+2)/2}\right)+\tilde{c}_4J_{-\mu}\left(\frac{2\lambda^{1/2}}{n+2}z^{(n+2)/2}\right)\right] \tag{21a}$$

for all z if n is even and for $z\geq 0$ if n is odd, and

$$v=\xi^{1/2}\left[\tilde{c}_4I_{-\mu}\left(\frac{2\lambda^{1/2}}{n+2}\xi^{(n+2)/2}\right)-\tilde{c}_3I_\mu\left(\frac{2\lambda^{1/2}}{n+2}\xi^{(n+2)/2}\right)\right] \tag{21b}$$

for $z=-\xi<0$ if n is odd. Here, $\tilde{c}_3$ and $\tilde{c}_4$ are arbitrary constants and

$$\mu=(n+2)^{-1}. \tag{21c}$$

Note that the solution (21b) is an analytic continuation of the solution (21a). To see this, we express J_μ and $J_{-\mu}$ in terms of their power series expansions, use Eq. (21c), let $\tau=\lambda^{1/2}/(n+2)$, and rewrite Eq. (21a) as

$$v=\tilde{c}_3\tau^\mu z\sum_{m=0}^\infty\frac{(-1)^m\tau^{2m}z^{m(n+2)}}{m!\,\Gamma(m+\mu+1)}+\tilde{c}_4\tau^{-\mu}\sum_{m=0}^\infty\frac{(-1)^m\tau^{2m}z^{m(n+2)}}{m!\,\Gamma(m-\mu+1)}, \tag{22a}$$

which is an entire function of z, and hence it is defined over the whole complex z plane. Thus, we let $z=-\xi$ in Eq. (22a), use the fact that n is an odd integer, and obtain

$$v=-\tilde{c}_3\tau^\mu\xi\sum_{m=0}^\infty\frac{\tau^{2m}\xi^{m(n+2)}}{m!\,\Gamma(m+\mu+1)}+\tilde{c}_4\tau^{-\mu}\sum_{m=0}^\infty\frac{\tau^{2m}\xi^{m(n+2)}}{m!\,\Gamma(m-\mu+1)}, \tag{22b}$$

which is simply the power series representation of Eq. (21b). Therefore,

$$u_0=[G_0(x)]^{1/2}[q(x)]^{-1/4}\{c_3J_\mu[\lambda^{1/2}G_0(x)] \quad c_4J_{-\mu}[\lambda^{1/2}G_0(x)]\}[1+o(1)] \quad \text{as } \lambda\to\infty. \tag{23a}$$

for all x if n is even and for $x>0$ if n is odd, and

$$u_0=[G_0(-x)]^{1/2}[q(-x)]^{-1/4}\{c_3I_{-\mu}[\lambda^{1/2}G_0(-x)]-c_4I_\mu[\lambda^{1/2}G_0(-x)]\}[1+o(1)] \text{ as } \lambda\to\infty \tag{23b}$$

for $x<0$ if n is odd. Although this expansion is valid at $x=0$, it breaks down as $x\to -a$ or b. Thus, it is valid only on the interval $[-a+\delta_1, b-\delta_4]$. An expansion valid near $x=b$ is obtained in the next section.

IV. AN EXPANSION VALID ON $[\delta_3, b]$

As $x \to b$, Eq. (3) tends to

$$\frac{d^2u}{dx^2} + \left(\lambda q_2 (b-x)^k + \frac{r_2}{(b-x)^2}\right)u = 0, \tag{24}$$

where

$$q_2 = b^n g(b)(a+b)^m \text{ and } r_2 = f(b)(a+b)^{-2}. \tag{25}$$

Hence, an asymptotic solution for Eq. (3) valid near $x=b$ can be obtained by relating it to the solutions of

$$\frac{d^2v}{dz^2} + \left(\lambda z^k + \frac{r_2}{z^2}\right)v = 0. \tag{26}$$

This is accomplished by using the transformation

$$z = \phi(x), \quad v = u(x)[\phi'(x)]^{1/2}, \qquad \beta^{-1}\phi^\beta = \int_x^b [q(\xi)]^{1/2} d\xi = G_2(x), \tag{27}$$

where $\beta = (2+k)/2$.

Introducing the transformation (27) into Eq. (3), we obtain

$$\frac{d^2v}{dz^2} + \left(\lambda z^k + \frac{r_2}{z^2}\right)v = F_2 v, \tag{28}$$

where

$$F_2 = \frac{r_2}{\phi^2} - \frac{1}{\phi'^2}\left(\frac{f(x)}{(x+a)^2(b-x)^2} + \frac{3\phi''^2}{4\phi'^2} - \frac{\phi'''}{2\phi'}\right). \tag{29}$$

As $x \to b$, $\phi = O[(b-x)]$, $\phi' = O(1)$, and $F_2 = O[(b-x)^{-1}]$. Hence, a first approximation to v is given by Eq. (26) whose general solution is

$$v = z^{1/2}[\tilde{c}_5 J_\gamma(\lambda^{1/2}\beta^{-1}z^\beta) + \tilde{c}_6 J_{-\gamma}(\lambda^{1/2}\beta^{-1}z^\beta)], \tag{30}$$

where $\tilde{c}_5$ and $\tilde{c}_6$ are arbitrary constants and

$$\gamma = (1-4r_2)^{1/2}/(2+k). \tag{31}$$

In what follows, we restrict our analysis to the case $r_2 \leq \frac{1}{4}$ so that γ is real.

In order that y be regular at $x=b$, Eqs. (2), (3), and (30) show that $\tilde{c}_6 = 0$. Hence,

$$u_2(x) = c_5[G_2(x)]^{1/2}[q(x)]^{-1/4} J_\gamma[\lambda^{1/2}G_2(x)][1+o(1)] \quad \text{as } \lambda \to \infty \tag{32}$$

Although this expansion is valid at $x=b$, it breaks down as $x \to 0$. In order to obtain a uniformly valid expansion on the interval $[-a, b]$, we match the three expansions obtained in this and the preceding two sections.

V. MATCHING WHEN n IS ODD

Since $u_1(x)$ and $u_0(x)$ are valid over on the interval $-a < -a+\delta_1 \leq x \leq -\delta_2 < 0$, they have a large overlapping region which allows their matching. To match these expansions, we fix x in this overlap interval and expand both u_0 as given by Eq. (23b) and u_1 for large λ. The result is

$$u_1 = c_1(2\pi)^{-1/2}[\lambda q(-x)]^{-1/4}\exp[\lambda^{1/2}G_1(x)][1+o(1)], \tag{33}$$

$$u_0 = (2\pi)^{-1/2}[\lambda q(-x)]^{-1/4}\{(c_4-c_3)\exp[\lambda^{1/2}G_0(x)] + [c_4\exp(i\pi\mu) - c_3\exp(-i\pi\mu)] \times \exp(-\tfrac{1}{2}i\pi)\exp[-\lambda^{1/2}\tilde{G}_0(x)]\}\,[1+o(1)], \tag{34}$$

where

$$\tilde{G}_0(x) = \int_x^0 [q(-\xi)]^{1/2} d\xi = G_0(-x). \tag{35}$$

The expansions (33) and (34) are two WKBJ approximations of $u(x)$ on the same interval $-a+\delta_1 \leq x \leq -\delta_2$. Hence, they must be identical. This is so if, and only if,

$$c_4 = c_3, \qquad 2c_3 \sin\mu\pi = c_1 \exp\left(\lambda^{1/2}\int_{-a}^0 [q(-x)]^{1/2}dx\right). \tag{36}$$

To match u_0 and u_2, we note that they overlap on the interval $[\delta_3, b-\delta_4]$. Thus, we fix x in this overlap interval and expand both u_0 as given by Eq. (23a) and u_2 for large λ. The result is

$$u_0 = 2c_3(2/\pi)^{1/2}[\lambda q(x)]^{-1/4}\cos(\tfrac{1}{2}\pi\mu)\cos[\lambda^{1/2}G_0(x) - \tfrac{1}{4}\pi] \times [1+o(1)], \tag{37}$$

$$u_2 = c_5(2/\pi)^{1/2}[\lambda q(x)]^{-1/4}\cos[\lambda^{1/2}G_2(x) - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\gamma] \times [1+o(1)]. \tag{38}$$

Since Eqs. (37) and (38) represent u over the same interval $[\delta_3, b-\delta_4]$, they must be identical.

If we let

$$\Delta = \lambda^{1/2}(G_0+G_2) - \tfrac{1}{2}(1+\gamma)\pi = \int_0^b [q(x)]^{1/2}dx - \tfrac{1}{2}(1+\gamma)\pi, \tag{39}$$

then

$$\lambda^{1/2}G_0 - \tfrac{1}{4}\pi = \Delta - (\lambda^{1/2}G_2 - \tfrac{1}{4}\pi - \tfrac{1}{2}\gamma\pi) = \Delta - \alpha. \tag{40}$$

Equating Eqs. (37) and (38), using Eqs. (39) and (40), and equating the coefficients of $\sin\alpha$ and $\cos\alpha$ to zero, we obtain

$$\sin\Delta = 0 \text{ or } \Delta = j\pi, \quad j = 1, 2, 3, \cdots, \tag{41}$$

$$c_5 = 2c_3\cos\tfrac{1}{2}\pi\mu\,\cos j\pi. \tag{42}$$

Combining Eqs. (39) and (41), we find that the eigenvalues are

$$\lambda = \pi^2\left((j + \tfrac{1}{2} + \tfrac{1}{2}\gamma)\Big/\int_0^b [q(x)]^{1/2}dx\right)^2. \tag{43}$$

For the special case, $a=b=1$, $p(x) = 2(1-x^2)$, $g(x)=1$, $n=-m=-k=1$, Eq. (43) reduces to the first-order solution of Ref. 2. We emphasize again that we represented the solution by only three expansions which were matched, whereas the solution was represented by five expansions in Ref. 2.

VI. MATCHING WHEN n IS EVEN

To match u_1 as given by Eq. (14) with u_0, we fix x in the interval $[-a+\delta_1, -\delta_2]$, expand u_1 and u_0 for large λ, equate the results, and obtain

$$c_1\cos(\lambda^{1/2}G_1 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\nu) = c_3\cos(\lambda^{1/2}G_0 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\mu) + c_4\cos(\lambda^{1/2}G_0 - \tfrac{1}{4}\pi + \tfrac{1}{2}\pi\mu) \tag{44}$$

To match u_0 and u_2, we fix x in the interval $[\delta_3, b-\delta_4]$, expand u_0 and u_2 for large λ, equate the results, and obtain

$$c_3\cos(\lambda^{1/2}G_0 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\mu) + c_4\cos(\lambda^{1/2}G_0 - \tfrac{1}{4}\pi + \tfrac{1}{2}\pi\mu) = c_5\cos(\lambda^{1/2}G_2 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\gamma). \tag{45}$$

If we let

$$\Delta_1 = \lambda^{1/2}(G_1 - G_0) + \tfrac{1}{2}\pi(\mu - \nu), \tag{46}$$

then

$$\lambda^{1/2}G_0 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\mu = \alpha - \Delta_1, \tag{47}$$

$$\lambda^{1/2}G_0 - \tfrac{1}{4}\pi + \tfrac{1}{2}\pi\mu = \alpha - (\Delta_1 - \pi\mu), \tag{48}$$

where

$$\alpha = \lambda^{1/2}G_1 - \tfrac{1}{4}\pi - \tfrac{1}{2}\pi\nu. \tag{49}$$

Substituting Eqs. (46)—(49) into Eq. (44) and equating the coefficients of $\cos\alpha$ and $\sin\alpha$ on both sides, we obtain

$$c_3\cos\Delta_1 + c_4\cos(\Delta_1 - \pi\mu) = c_1, \tag{50}$$

$$c_3\sin\Delta_1 + c_4\sin(\Delta_1 - \pi\mu) = 0. \tag{51}$$

Similarly, we obtain from Eq. (45) the following relationships:

$$c_3\cos\Delta_2 + c_4\cos(\Delta_2 + \mu\pi) = c_5, \tag{52}$$

$$c_3\sin\Delta_2 + c_4\sin(\Delta_2 + \mu\pi) = 0, \tag{53}$$

where

$$\Delta_2 = \lambda^{1/2}(G_0 + G_2) - \tfrac{1}{2}\pi - \tfrac{1}{2}\pi(\mu + \gamma). \tag{54}$$

In order that Eqs. (51) and (53) have a nontrivial solution,

$$\sin\Delta_1\sin(\Delta_2 + \mu\pi) - \sin\Delta_2\sin(\Delta_1 - \mu\pi) = 0,$$

which gives

$$\sin(\Delta_1 + \Delta_2) = 0.$$

Hence,

$$\Delta_1 + \Delta_2 = j\pi, \quad j = 1, 2, 3, \cdots. \tag{55}$$

Substituting for Δ_1 and Δ_2 from Eqs. (46) and (54) into Eq. (55), then substituting for G_1 and G_2 from Eqs. (8) and (27) into the resulting expression, and solving for λ, we obtain

$$\lambda = \pi^2\left([j + \tfrac{1}{2}(1 + \gamma + \nu)]\Big/\int_{-a}^{b}[q(x)]^{1/2}dx\right)^2. \tag{56}$$

Once λ is known, we can solve Eq. (53) to determine c_4 as a function of c_3, and then Eqs. (50) and (52) to determine c_1 and c_5 in terms of c_3.

VII. SUMMARY

A general procedure is presented for the determination of approximate solutions of linear differential equations with multiple transition points. The procedure is a combination of the Langer transformation and the method of matched asymptotic expansions. It is applied to a class of second-order differential equations with three transition points—a turning point of any order and two regular singular points.

The solution is represented by three different regular asymptotic expansions. Each expansion is valid on an open interval containing one of the transition points but excluding the other two. These expansions were then matched to relate the arbitrary constants and determine the eigenvalues. Adding these expansions and subtracting their common parts, one can determine a so-called composite expansion, which is a single uniformly valid expansion.

[1]E. Y. Harper, G. W. Grube, and I-Dee Chang, J. Fluid Mech. **52**, 565 (1972).
[2]E. Y. Harper, I-Dee Chang, and G. W. Grube, J. Math. Phys. **12**, 1955 (1971).
[3]A. H. Nayfeh, *Perturbation Methods* (Wiley-Interscience, New York, 1973).
[4]H. Jeffreys, *Asymptotic Approximations* (Oxford U. P., Oxford, 1962).
[5]A. H. Nayfeh, J. Math. Phys. **6**, 1946 (1965).
[6]A. H. Nayfeh, J. Math. and Phys. **46**, 349 (1967).
[7]F. W. J. Olver, J. Res. Natl. Bur. St. B **63**, 131 (1967).
[8]H. Moriguchi, J. Phys. Soc. Japan **14**, 1771 (1959).
[9]E. R. Pike, Quart. J. Mech. Appl. Math. **17**, 369 (1964).
[10]M. A. Evgrafov and M. B. Fedoryuk, Usp. Mat. Nauk **21**, 3 (1966).
[11]P. F. Hsieh and Y. Sibuya, J. Math. Anal. Appl. **16**, 84 (1966).
[12]Y. Sibuya, Acta. Math. **119**, 235 (1967).
[13]R. Y. S. Lynn and J. B. Keller, Comm. Pure Appl. Math. **23**, 379 (1970).

A geometric interpretation of classical relativistic electrodynamics

Richard Petti

Lamp Manufacturing Research Laboratory, Lighting Research and Technical Services Operation, General Electric Co., Nela Park, Cleveland, Ohio 44112

(Received 15 March 1974; revised manuscript received 16 May 1974)

A solution is offered for this problem: Describe the observables of classical electrodynamics with connections on fiber bundles without using nonobservable entities, either in computations or in conceptual development. The solution employs a connection on the affine frame bundle of space–time. Comparisons are made with other geometric interpretations of electrodynamics.

INTRODUCTION

Many of the descriptions of classical electrodynamics by connections on fiber bundles are closely related and can be derived one from the other. The Kaluza–Klein theory is representative of one such group of theories,[1,2] and was the first historically in this group. The purpose of this paper is to present a geometrical description of classical electrodynamics which employs only (classical) observables, both computationally and conceptually. The space–time metric and electromagnetic field define a connection on the affine frame bundle of space–time without recourse to a gauge-dependent fiber metric, which is employed by the Kaluza–Klein-type theories given in the references. If desired, variational functionals can be expressed as tensors contracted with respect to a (different) gauge-dependent fiber metric which is defined in Sec. II.B.2.

Section I discusses mathematical foundations; Sec. II gives the physical interpretation and compares this theory with several other geometrical theories of electrodynamics, including the semiclassical Yang–Mills–Utiyama[3,4,5] description.

I. MATHEMATICAL FOUNDATIONS

A. Fiber bundles

Let M be a T_2 C^∞ n-manifold. Then we can construct: $L(M)=$ linear frame bundle [principal bundle with group $Gl(n;R)=$ homogeneous linear group]; $T(M)=$ linear tangent bundle; $A(M)=$ affine frame bundle [principal bundle with group $A(n;R)=$ inhomogeneous linear group]; $AT(M)=$ total affine tangent bundle [an $(n+1)$-vector bundle with principal bundle $A(M)$]. There are natural inclusions $L(M)\subset A(M)$, $T(M)\subset AT(M)$, and $Gl(n;R)\subset A(n;R)$ [induced by $L(M)\subset A(M)$]. The orbits in R^{n+1} induced by the action of $A(n;R)$ are disjoint hypersurfaces which fill R^{n+1}. Parameterizing the family of orbits in a way that respects the addition in R^{n+1}, we can construct associated fiber bundles $A_\sigma T(M)$, $\sigma\in R$, with principal bundle $A(M)$ such that:

$$AT(M)=\bigcup_{\sigma\in R} A_\sigma T(M),$$
$$A_\sigma T(M)+A_\tau T(M)=A_{\sigma+\tau}T(M).$$

Each $A_\sigma T(M)$ is a fiber bundle whose fiber is a flat affine manifold of dimension n; since they all have the same principal bundle, all $A_\sigma T(M)$ are bundle-isomorphic. The inclusion $L(M)\subset A(M)$ defines a vector bundle structure on $A_\sigma T(M)$; this structure will not be parallel when the connections are introduced, but it does give an explicit vector-bundle isomorphism $A_\sigma T(M)\approx A_\tau T(M)$.

B. Connections

Let $\{\tilde\omega\}=$ set of all C^∞ connection 1-forms on $A(M)$, $\{\omega\}=$ set of all C^∞ connection 1-forms on $L(M)$, $\{K\}$ $=$ set of all C^∞ $(1,1)$ linear tensor fields on M. There is a natural one-to-one correspondence[6]:

$$\{\tilde\omega\}\approx\{\omega\}\times\{K\}.$$

The curvature form $\tilde\Omega$ of a connection form $\tilde\omega$ defines an affine curvature tensor $\tilde R^i{}_{\hat j k l}$ on M:

$$\tilde R^i{}_{\hat j kl}=\left[\begin{array}{c|c} R^i{}_{jkl} & K^i{}_{l;k}-K^i{}_{k;l} \\ \hline 0\cdots 0 & 0\end{array}\right],$$

where $R^i{}_{jkl}=$ curvature tensor on M of ω, and where ";" denotes covariant differentiation (with respect to ω, since K is a linear tensor). All indices run from 0 to n, but for unhatted indices (which indicate linear tensors), the last index value indicates a zero coefficient.

Given $\tilde\omega$, we can define affine geodesics.

Notation: for $u\in T_x(M)$, $\overset{\sigma}{\hat u}\in A_\sigma T_x(M)$ denotes the corresponding element under the $Gl(n;R)$-bundle isomorphism $T(M)=A_0T(M)\approx A_\sigma T(M)$, $\sigma\in R$. In coordinates, $\overset{\sigma}{\hat u}$ $=(u^0,\ldots,u^{n-1},\sigma)$.

Definition: A σ-geodesic, $\sigma\in R$, is a C^∞ curve in M such that

$$\tilde\nabla_u\overset{\sigma}{\hat u}=0,$$

where

$u=$ tangent vector to the curve.

$\tilde\nabla=$ covariant differentiation with respect to $\tilde\omega$.

In local coordinates, this is:

$$u^j u^i{}_{,j}+u^j u^k\Gamma^i_{jk}+\sigma u^j K^i{}_j=0,$$

where Γ^i_{jk} is a Christoffel symbol for the connection form ω. For $\sigma=0$, these are the geodesics of the linear connection ω. For σ, $\tau\in R$, σ, $\tau\neq 0$, σ-geodesics differ from τ-geodesics only by a nonzero scaling factor of the path parameter. (If we own a metric and restrict u by $u^2=\pm1$, then σ-geodesics are in general distinct from τ-geodesocs, $\sigma\neq\tau$.)

C. Metric generalized affine connections

Given a manifold M with metric tensor g, a metric generalized affine connection on M is one such that:

(i) The corresponding linear connection, ω, is the Levi–Civita connection of g.

(ii) The corresponding (1, 1) tensor field K is subject to: $u\cdot(\tilde{\nabla}_u u)=0$, where $u=$ tangent vector field along any σ-geodesic, $\sigma\in R$.

Condition (i) says that parallel translation of linear vectors and tensors preserves inner products. Condition (ii) is equivalent to requiring the tangent vector field along a σ-geodesic to be of constant length, which is equivalent to requiring K to be antisymmetric as a (0, 2) tensor field.

Remarks: (1) A linear metric tensor field g and *a* linear antisymmetric (0, 2) tensor field K uniquely determine a metric generalized affine connection in the sense defined here.

(2) The metric g is a nondegenerate bilinear form on the tangent spaces over M, but it is degenerate as a bilinear form on the total affine tangent fibers over M.

D. The Jacobi equation for σ-geodesics of a metric connection

Let $f:[0,1]\to M$ be a σ-geodesic, $\sigma\in R$;

$u=$ tangent vector field along f;

$X=$ a variational vector field of σ-geodesics along f (X is a linear vector field);

$\dot{X}=\tilde{\nabla}_u X$.

Then X satisfies

$$-(\tilde{\nabla}_u^2 X)^i=R^i{}_{jkl}u^jX^ku^l+\sigma K^i{}_{l;k}X^ku^l+\sigma K^i{}_l\dot{X}^l$$

Proof: Write out the equation in local coordinates, and use the σ-geodesic equation to derive the result for a geodesic variation which gives rise to X. If $K_{[il;k]}=0$ (antisymmetric part $=0$), then the Jacobi equation can be written in the form

$$-(\tilde{\nabla}_u^2 X)^i=\tilde{R}_{kj}{}^i{}_l X^k\overset{\sigma}{\hat{u}}{}^j u^l+\sigma K^i{}_l\dot{X}^l.$$

From the Jacobi equation, we can obtain a geometric interpretation of the affine Ricci curvature as the average second-order convergence of σ-geodesics around f.

II. DESCRIPTION AND INTERPRETATION OF ELECTRODYNAMICS IN GENERAL RELATIVITY IN TERMS OF GENERALIZED AFFINE CONNECTIONS

Convention: $g_{ij}\sim(+,-,-,-)$, $M=T_2\ C^\infty$ 4-manifold.

A. The motion of charged matter

1. An isolated charged particle

For a space–time M with metric g_{ij} and electromagnetic field K_{ij}, construct the metric generalized affine connection determined by g and K. Consider an isolated charged particle in the space–time, where

$m=$ particle mass,

$e=$ particle charge,

$u=$ unit 4-velocity (timelike: $u^2=+1$).

Define the "affine momentum vector" of the particle by:

$$\hat{v}=(\overset{e}{\widehat{mu}}).$$

In local coordinates on the affine tangent bundle,

$$\hat{v}=(mu^0,\ mu^1,\ mu^2,\ mu^3,\ e).$$

The equation of particle motion is

$$\tilde{\nabla}_u\hat{v}=0,$$

that is, $u^j(mu^i)_{,j}+u^jmu^k\Gamma^i_{jk}+eK^i{}_ju^j=0$, that is, the particle trajectory is a σ-geodesic, where $\sigma=e/m$.

2. A perfect charged fluid

Construct the metric generalized affine connection determined by g and K as above. Consider a perfect charged fluid described by:

$N=$ particle density function,

$p=$ hydrostatic pressure,

$m=$ particle mass,

$e=$ particle charge,

$u=$ mean 4-velocity, $u^2=+1$.

Define the affine energy–momentum tensor to be

$$P^{\hat{i}j}=N\hat{v}^iu^j+p(u^iu^j-g^{ij}),$$

where $\hat{v}=(\overset{e}{\widehat{mu}})$ as before. Note that $g^{4j}=0$, $j=0,\ 1,\ 2,\ 3,\ 4$. The flux density of $\hat{i}$-momentum through the hypersurface $\{x^j=\text{const}\}$ is given by $P^{\hat{i}j}$. This has the standard meaning for $\hat{i}$, $j=0,\ 1,\ 2,\ 3$. For $\hat{i}=4$, we have $P^{4j}=$ charge flux density through the hypersurface $x^j=$const. For $j=4$, $P^{\hat{i}4}$ is identically zero.

The fluid dynamic equations are:

(i) particle conservation: $(Nu^j)_{;j}=0$;

(ii) affine momentum conservation: $P^{\hat{i}j}{}_{;j}=0$.

Applying (i) to (ii), and writing (ii) in local coordinates, we have

$$Nu^j(mu^{\hat{i}})_{,j}+Nmu^ju^k\Gamma^{\hat{i}}_{jk}+NeK^{\hat{i}}{}_ju^j+(p(u^iu^j-g^{ij}))_{;j}=0,\quad \hat{i}=0,\ 1,\ 2,\ 3,$$

$$(Neu^j)_{;j}=0,\quad \hat{i}=4.$$

3. Tidal forces

Perform an Einstein elevator experiment with six identical particles of charge-to-mass ratio $\sigma=e/m$. If the particles are placed symmetrically along three orthogonal spacelike directions, then their average second-order convergence is

$$-(\tilde{\nabla}_u^2X_{(i)})^i=\tilde{R}_{jl}\overset{\sigma}{\hat{u}}{}^ju^l=R_{jl}u^ju^l-\sigma J_lu^l,$$

where $J_l\equiv K_{lm}{}^{;m}$, and where $u\equiv$ center-of-mass 4-velocity.

4. Generalizations

We can generalize concerning when vectors and tensors are linear and when they are affine. Space–time itself is described by linear vectors and tensors, while charged matter is described by affine tensors, and electromagnetic fields relate the two.

Examples:

(a) Charged particle trajectory:

$$\tilde{\nabla}_u\hat{v}=0.$$

Here $\hat{v}$ describes the particle and u describes the space–time path.

(b) $P^{\hat{i}j}$ is the flux density of $\hat{i}$-momentum through the x^j-surface element.

(c) Particle conservation $(Nu^j)_{;j}=0$ involves the flux of a scalar quantity through a space–time boundary.

(d) Affine Ricci curvature describes average gravitational tidal forces with its purely linear part, and electromagnetic tidal forces with its partly-affine components.

B. The field equations

1. The standard field equations

The field equations are those of standard general relativity, when all tensors are interpreted as linear tensors. The equations are:

(i) $-G^{\hat{i}j}+2P^{\hat{i}j}=0$,

(ii) $dK=0$ (i.e., $K_{[ij;m]}=0$),

where $P^{\hat{i}j}$ is the affine energy–momentum tensor for matter, and does not include electromagnetic stress energy. That is,

$$P^{\hat{i}j}=\left[\begin{array}{c|c} P^{ij} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} \\ \hline J^j & 0\end{array}\right],$$

where P^{ij} is the linear momentum tensor for matter, and J^j is the electric current. The remaining term is

$$G^{\hat{i}j}=\left[\begin{array}{c|c} \begin{matrix} R^{ij}-\frac{1}{2}g^{ij}R \\ \\ -2(K^i{}_aK^{aj}+\frac{1}{4}g^{ij}K_{ab}K^{ab})\end{matrix} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} \\ \hline 2K^{jm}{}_{;m} & 0\end{array}\right].$$

When $dK=0$, then $G^{\hat{i}j}{}_{;j}\equiv 0$ identically:

$$G^{\hat{i}j}{}_{;j}=(R^{ij}-\tfrac{1}{2}g^{ij}R)_{;j}-2K^i{}_aK^{ab}{}_{;b}+2K^i{}_jK^{jm}{}_{;m}\equiv 0,$$

$$\hat{i}=0,\ 1,\ 2,\ 3,$$

$$G^{\hat{i}j}{}_{;j}=2K^{jm}{}_{;m;j}\equiv 0, \quad \hat{i}=4.$$

This is a generalization of the contracted second Bianchi identity for metric linear connections.

For metric generalized affine connections, the equation

$$R_{[i|j|kl]}=0 \quad \text{(antisymmetrize } i,\ k,\ l)$$

is equivalent to

$$R_{[i|j|kl]}=0 \quad \text{and} \quad K_{[kl;i]}=0.$$

Although $dK=0$ is not an identity of the theory and must be postulated as a field equation, it is similar in form to an extension of the first Bianchi identity for metric linear connections. Magnetic charge cannot be ruled out in principle (as in the rest of classical physics), but there is no room for it to enter with the geometric significance of electric charge into the affine momentum of charged matter or into the Ricci curvature. This point is discussed further in the comparisons with other theories.

This concludes the theory except for a separate construction of variational functionals using a gauge-dependent fiber metric that respects the affine structure, but ignores the linear structure of the σ-affine tangent fibers.

2. The field equations from a variational principle

The metric g_{ij} on the linear tangent spaces of M can be extended to the affine tangent spaces by ignoring the affinity of the vectors:

$$g_{i4}=g_{4i}=0 \quad (i=0,\ldots,4).$$

By moving the zero vector of g to $-\sigma\overset{1}{A}$, for some linear vector field A on M, a bilinear form $\tilde{g}$ can be defined on the total affine tangent spaces of M:

$$\tilde{g}(\overset{\sigma}{\hat{v}},\overset{\tau}{\hat{w}})=g(v-\sigma A,w-\tau A)=(v^i-\sigma A^i)g_{ij}(w^j-\tau A^j),$$

that is, $\tilde{g}$ is a bilinear form on each σ-affine tangent fiber whose zero has been moved to $\sigma\overset{1}{A}$. The gauge-invariant metric structure of the σ-affine tangent fibers is that of a flat four-dimensional Lorentz manifold without a linear structure (unless $\sigma=0$).

The condition $\tilde{\nabla}\tilde{g}=0$ does not generally admit single-valued solutions for A:

$$A_{i,j}-A_m\Gamma^m_{ji}=K_{ji}.$$

However, $\tilde{\nabla}g_{\hat{i}[j;k]}=0$ is for A equivalent to

$$dA=\tfrac{1}{2}(A_{j,i}-A_{i,j})=K_{ij}$$

which has solutions (for $dK=0$) and is physically useful. This condition will be imposed on $\tilde{g}$, but the absence of a meaningful four-dimensional interpretation of the condition leaves the relation between A and K in an unsatisfactory state. By considering quantum mechanical complex phase, the Yang–Mills–Utiyama approach gives a satisfactory interpretation of $dA=K$ as the definition of K as a curvature.

$\tilde{g}$ is a degenerate form on the total affine fibers; in order to get familiar results for linear vectors ($\sigma=0$), define $\tilde{g}^{\hat{i}\hat{j}}$ by: $\tilde{g}^{i\hat{4}}=-A^i$ $(i=0,\ldots 3)$, $\tilde{g}^{\hat{4}\hat{4}}=A\cdot A$.

The basic variational functionals of electrodynamics can be expressed using $\tilde{g}$:

(a) Sourceless field equations, $\tilde{R}\equiv\tilde{g}^{\hat{i}j}\tilde{R}_{\hat{i}j}=R+A\cdot J$:

$$\delta\int\tilde{R}\sqrt{|g|}\,d^4x=0.$$

(b) Action of a charged particle, $\hat{w}\equiv(\overset{\sigma}{mu})$:

$$\tilde{g}(\hat{w},u)=(mu^i-eA^i)g_{ij}u^j.$$

(c) Field equations for a collisionless charged fluid come from variation of: $\tilde{R}+2N\tilde{g}(\hat{w},u)$.

C. Comparisons with other theories

1. The Kaluza–Klein theory interprets the observables of classical electrodynamics with a connection on the tangent bundle of a 5-manifold with a cylindrical gauge-dependent metric. The Maxwell equation $dK=0$ results from the cylindrical restriction on the 5-geometry, which is itself of dubious significance. The other Maxwell equation and the Lorentz force equation arise

from a variational principle, and the Lorentz force can also be derived from parallel translation of the 5-momentun of a charged particle. Gauge transformations arise from 5-dimensional coordinate transformations.

The Einstein—Mayer variant of the Kaluza—Klein theory employs a 5-vector bundle over a 4-manifold with a gauge-dependent fiber metric. The equation $dK=0$ must be separately postulated as a field equation; the resulting theory is basically the same as the Kaluza—Klein theory.

2. The interpretation of electrodynamics presented in this paper is a solution to this problem: "Describe the observables of classical electrodynamics with connections on fiber bundles without using nonobservable entities, either in computations or in conceptual development." The content of this theory can be boiled down to: *The parallel-translation law for classical physics preserves the flat four-dimensional Lorentzian structure of the (σ-affine) tangent fibers, but not generally their linear structure.* The results are similar to those of the Kaluza—Klein theory: The Lorentz force follows from parallel translation of the affine momentum of charged matter; Ricci curvature describes average tidal forces on charged matter. $dK=0$ must be postulated as a separate field equation; however, electric charge-mass ratio has been interpreted as the relative strength of curvature rotations and affine curvature translations of affine momentum, and there is no room for an analogous geometrical role to be played by magnetic charge.

A gauge-dependent metric can be introduced to express variational functions; however, the relation $dA=K$ between $\tilde{g}$ and K is not well-justified conceptually. The theory can stand without $\tilde{g}$ if it is satisfactory to write down ad hoc variational functionals.

3. The Yang—Mills—Utiyama approach solves the following problem: "Describe semiclassical electrodynamics with connections on fiber bundles using quantum mechanical complex waves and 4-vector potential." Introducing a connection on a $U(1)$-bundle, $dA=K$ defines K as the curvature, and $dK=0$ is the Bianchi identity. The other Maxwell equation and the Lorentz force follow from a variational principle. Changing the coordinate splitting of the $U(1)$-bundle gives rise to gauge changes. The methods of this theory are applicable to a class of gauge-invariant field theories with internal symmetry algebras. In particular, the description of compound states by tensor products yields a plausible explanation of charge quantization.

ACKNOWLEDGMENTS

I would like to acknowledge the useful comments I received from Professor R.K. Sachs and Professor A. Trautman.

[1]The Kaluza-Klein, Einstein-Mayer, and projective theories of relativity are very similar. P.G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, New York, 1942).
[2]M.-A. Tonnelat, *Les Theories Unitaire de l'Electromagnetisme et de la Gravitation* (Gauthier-Villars, Paris, 1965).
[3]C.N. Yang and R.L. Mills, Phys. Rev. **96**, 191 (1954).
[4]R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[5]A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.

Summation of regularized perturbative expansions for singular interactions

D. Bessis, L. Epele*, and M. Villani†

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP n°2 91190 Gif-sur-Yvette, France
(Received 13 May 1974)

In this paper we give a first application of a general method whose mathematical aspects will be fully developed in a forthcoming article. We are concerned with strongly singular perturbative series. Here we shall restrict ourselves to the most general two-body repulsive singular potential for which a regularization exists. Various extensions of this case are discussed in the conclusion. We show that, knowing only a finite number of regularized Born terms, it is possible to construct an upper bound to the exact phase shifts and that this upper bound is the best possible for the given regularization. The method uses the construction of the $[N/N]$ Padé approximation indifferently on the regularized partial waves of the K or T matrix and exploits the fact that the approximate corresponding phase shifts have an absolute minimum as a function of the regularization parameter (cutoff). Three numerical examples are provided which show, even for very large phase shifts, an excellent convergence.

INTRODUCTION

One is often faced with a perturbative expansion in some parameter λ, whose terms are given by diverging quantities, for instance diverging multiple integrals. It is therefore natural to regularize those quantities by introducing some cutoff parameter ϵ, and therefore consider the formal regularized perturbative expansion

$$T_\epsilon(\lambda)=\lambda T_\epsilon^{(1)}+\lambda^2 T_\epsilon^{(2)}+\cdots+\lambda^n T_\epsilon^{(n)}+\cdots \tag{1}$$

with the condition

$$\epsilon\to+0, \quad T_\epsilon^{(n)}\to\infty. \tag{2}$$

The basic problem is how to obtain meaningful approximations to the true value when one has at disposal only the knowledge of a *finite* number of regularized perturbative terms.

In principle, one has to sum the complete series for fixed ϵ, and then send the cutoff to its limit zero. But practically this procedure is, in realistic cases, impossible to do. We propose a different approach, which is based on the following idea: We replace the regularized perturbative expansion stopped at order $2N$ by its $[N/N]$ diagonal Padé approximation (PA),[1-4] and look for extremal values of the PA in the cutoff parameter ϵ. We expect that these values will tend, when $N\to\infty$, to the exact solution. The mathematical proof of such property will be given only for a specific case, namely, for the perturbative expansion of the mean value of the resolvent of an unbounded self-adjoint Hilbert space linear operator (Ref. 5). Such philosophy and the proof of its usefulness has been already given for a special class of regularized series: the so-called Stieltjes series with infinite moments (Ref 6) with a physical application to the scattering length of singular potentials (Ref. 7).

In Ref. 5, we generalize and extend the method to a more general class of series with infinite moments. In this work, we give an idea of this generalization restricting ourselves to a more specific physical problem: the regularized Born series produced by a singular two-body positive potential, at an arbitrary positive energy and angular momentum. The physical interest of this problem is two fold:

(1) On one hand, its similarities with nonrenormalizable field theory perturbative expansions can suggest methods to treat the trouble of this kind of series. In this respect, we observe that the PA method being entirely algebraic, can always be applied, at least formally, to any regularized series.

(2) It provides a very efficient and systematic way of computing the phase shifts of potentials as singular as one likes. Although our proof in the scattering region applies only to positive potentials, numerical experiences reported in Sec. 5 show that the convergence is also very rapidly achieved for changing sign singular potentials.

In this paper, we show that the approximated regularized phase shifts, computed from the $[N/N]$ PA on the T or K matrix, at a given energy, angular momentum, and coupling constant, enjoy an extremal property in the cutoff: They have an absolute minimum in the cutoff which, when $N\to\infty$, tends towards the true physical phase shift. Furthermore, the sequence of these minimums provides a set of monotonic decreasing upper bounds to the physical phase shift.

In our demonstration, we have not explicitly made use of the variational Lippman–Schwinger principle for the phase shifts; however, it would be possible to deduce all our results from this principle, taking into account that the $[N/N]$ PA on the K or T matrix is the solution of this principle, when one chooses the perturbative ansatz (the so-called Cini–Fubini ansatz), to use this principle (Ref. 8).

In Sec. 1, the reader will find the precise class of singular potential for which the complete proof is given.

In Sec. 2, we give the notations and equations for the T and K matrices for completeness.

In Sec. 3, we show a remarkable property of monotony of the approximate phase shifts of the PA on the K or T matrices.

In Sec. 4, we demonstrate the extremal properties in the cutoff parameter of the approximate phase shifts built on the PA on the K or T matrices and the convergence of these extremal values towards the exact result.

In Sec. 5, we discuss various numerical examples and show the extreme rapidity of convergence even for very large phase shifts (typically 300°).

In the conclusion, we come back on the general aspects of the method and discuss various generalizations.

1. THE CLASS OF POTENTIALS

Although our method applies to very general potentials, we shall restrict ourselves, in this paper to a class for which complete proofs can be derived without much effort.

In the scattering region we shall consider spherically symmetric potentials V singular in the origin restricted in the following ways (for the definition of regular and singular potentials, see Ref. 9):

(1) $V>0;$ (1.1)

(2) such that a unitary S matrix exists for a given angular momentum and a given energy;

(3) such that a regularized version $V^{(\epsilon)}$ exists with the properties:

(a) $V^{(\epsilon)}$ is a regular potential

(b) when $\epsilon_1>\epsilon_2$, $0\leqslant V^{(\epsilon_1)}\leqslant V^{(\epsilon_2)}$, (1.2)

(c) when $V^{(\epsilon)}\to V$, $\epsilon\to 0$ in the weak sense, (1.3)

then $\delta_l^{(\epsilon)}(E)\to\delta_l(E)$, $\epsilon\to 0$, (1.4)

where $\delta_l^{(\epsilon)}(E)$ and $\delta_l(E)$ are the scattering phase shifts relative to $V^{(\epsilon)}$ and V, at energy E and angular momentum l.

Such potentials which fulfill all these conditions have been extensively analyzed in the literature (see Ref. 9).

Among the possible regularizations fulfilling conditions (3b) we can use the so-called "θ regularization"

$$V^{(\epsilon)}(r)=\theta(r-\epsilon)V(r). \tag{1.5}$$

2. THE K AND T MATRICES AND THEIR BORN SERIES

We consider a two-body Hamiltonian

$$H=H_0+\lambda V\quad(\lambda>0). \tag{2.1}$$

By going into partial waves, it is convenient to introduce the scalar product

$$\langle\varphi_1|\varphi_2\rangle=\int_0^\infty\varphi_1^*(r)\varphi_2(r)\,dr. \tag{2.2}$$

The T matrix is given by

$$\exp[i\delta_l(k,\lambda)]\sin\delta_l(k,\lambda)=T_l(k,\lambda)=-k\langle\varphi_l(k)|\lambda V|\psi_l^+(k,\lambda)\rangle, \tag{2.3}$$

with $k=\sqrt{E}$, where E is the energy and $|\varphi_l(k)\rangle$ is the spherical free wave of momentum k,

$$\langle r|\varphi_l(k)\rangle=rj_l(kr), \tag{2.4}$$

and $|\psi_l^+(k,\lambda)\rangle$ is the outgoing spherical wave, solution of

$$|\psi_l^+(k,\lambda)\rangle=|\varphi_l(k)\rangle+\lambda G_l^+(k)V|\psi_l^+(k,\lambda)\rangle \tag{2.5}$$

and

$$\langle r|G_l^+(k)|r'\rangle=-krr'j_l(kr_<)h_l^+(kr_>) \tag{2.6}$$

$$=\langle r|G_l^P(k)|r'\rangle-i\,krr'j_l(kr)j_l(kr') \tag{2.7}$$

with

$$\langle r|G_l^P(k)|r'\rangle=-krr'j_l(kr_<)n_l(kr_>). \tag{2.8}$$

Therefore, we can write

$$G_l^+(k)=G_l^P(k)-ik|\varphi_l(k)\rangle\langle\varphi_l(k)|. \tag{2.9}$$

From which we deduce by a standard argument the relation between the T and K matrices:

$$K_l(k,\lambda)=T_l(k,\lambda)/[1+iT_l(k,\lambda)], \tag{2.10}$$

where $K_l(k,\lambda)$ is defined by

$$K_l(k,\lambda)=-k\langle\varphi_l(k)|\lambda V|\psi_l^P(k,\lambda)\rangle=tg\delta_l(k,\lambda) \tag{2.11}$$

and $|\psi_l^P(k,\lambda)\rangle$ is solution of the equation

$$|\psi_l^P(k,\lambda)\rangle=|\varphi_l(k)\rangle+\lambda G_l^P(k)V|\psi_l^P(k,\lambda)\rangle. \tag{2.12}$$

All these equations are valid as well as for a regular potential as for a singular one fulfilling our previous statement.

In the case in which $V>0$, we can rewrite equation (2.11) and (2.12) in the following way:

$$K_l(k,\lambda)=-k\lambda\langle\varphi_l(k)|\sqrt{V}\sqrt{V}|\psi_l^P(k,\lambda)\rangle \tag{2.13}$$

and

$$\sqrt{V}|\psi_l^P(k,\lambda)\rangle=\sqrt{V}|\varphi_l(k)\rangle+\lambda\sqrt{V}G_l^P(k)\sqrt{V}\sqrt{V}|\psi_l^P(k,\lambda)\rangle \tag{2.14}$$

and deduce

$$K_l(k,\lambda)=-k\lambda\langle\varphi_l(k)|\sqrt{V}[1-\lambda\sqrt{V}G_l^P(k)\sqrt{V}]^{-1}\sqrt{V}|\varphi_l(k)\rangle. \tag{2.15}$$

Therefore, $K_l(k,\lambda)$ appears as the mean value in the state $\sqrt{V}|\varphi_l(k)\rangle$ of the resolvent of a symmetric operator $\sqrt{V}G_l^P(k)\sqrt{V}$. When V is a regular potential, $K_l(k,\lambda)$ is a meromorphic function of λ with poles on the real axis and negative residues (extended Stieltjes function); furthermore, the function has a finite radius of convergence around the origin. On the contrary, when V is singular, we expect to have a cut running from $-\infty$ to zero in the λ plane, supplemented by an infinite number of poles for λ positive [corresponding to a phase shift passing through $(2p+1)\pi/2$]. The Born series does not exist anymore due to the cut which extends up to the origin. By regularizing the singular potential V, we obtain a family of potentials $V^{(\epsilon)}$ for which we can define a family of Born series, for the K or T matrix (in the following we shall drop the dependence on the angular momentum):

$$K^{(\epsilon)}(k,\lambda)=\lambda K_B^{(\epsilon)}(k)+\lambda^2K_{B2}^{(\epsilon)}(k)+\cdots, \tag{2.16}$$

$$T^{(\epsilon)}(k,\lambda)=\lambda T_B^{(\epsilon)}(k)+\lambda^2T_{B2}^{(\epsilon)}(k)+\cdots. \tag{2.17}$$

For fixed ϵ, we can construct the $[N/N]$ Padé approximation to these Born series. We shall define an approximate phase shift $\delta_\epsilon^N(k,\lambda)$ by

$$\delta_\epsilon^N(k,\lambda)=\arctan[N/N]_{K_\epsilon}(k,\lambda), \tag{2.18}$$

where the arctan is followed by continuity from zero for $\lambda=0$.

If instead we start from the $[N/N]$ PA on the T matrix, it is easily shown that the previous phase shift will fulfill the expected relation

$$\exp[i\delta_\epsilon^N(k,\lambda)\sin\delta_\epsilon^N(k,\lambda)=[N/N]_{T_\epsilon}(k,\lambda). \tag{2.19}$$

This results from the fact that T and K are connected by the homographical transformation (2.10), for which Padé approximations are covariant (Ref. 10).

Therefore, for the calculation of $\delta_\epsilon^N(k,\lambda)$ it is a matter of indifference whether K or T matrix Born expansions are used. For reasons of convenience we shall, now on, use only the K matrix Padé approximants.

We remind the reader that for every fixed $\epsilon>0$, the phase shift $\delta_\epsilon^N(k,\lambda)$ as $N\to\infty$, tends to $\delta_\epsilon(k,\lambda)$ which is the true phase shift for the regularized regular potential $\lambda V^{(\epsilon)}$(Refs. 11, 12).

3. A PROPERTY OF THE [N/N] PADÉ APPROXIMANT ON THE K MATRIX FOR A POSITIVE POTENTIAL

We shall first recall the following theorem (see Refs. 12, 13):

Theorem: Let H be a symmetric operator and $|\varphi\rangle$ a vector with finite norm which is in the field of any power of H. We set

$$R_\varphi(z)=\langle\varphi|\frac{1}{1-zH}|\varphi\rangle, \tag{3.1}$$

$$\mu_k=\langle\varphi|H^k|\varphi\rangle<\infty. \tag{3.2}$$

The Padé approximation $[N-1/N]_{R_\varphi}(z)$ constructed out of the first $2N$ moments μ_k $(k=0,1,2,\ldots,2N-1)$ is equal to the resolvent of the finite rank symmetric operator P_NHP_N

$$[N-1/N]_{R_\varphi}(z)=\langle\varphi|\frac{1}{1-zP_NHP_N}|\varphi\rangle \tag{3.3}$$

where P_N is the projector on the N-dimensional space $\mathcal{E}^{(N)}$ spanned by the vectors $\{|\varphi\rangle;H|\varphi\rangle;H^2|\varphi\rangle;\ldots;H^{N-1}|\varphi\rangle\}$ supposed to be linearly independent. In the special case where this set of vectors is linearly dependent $[N-1/N]_{R_\varphi}(z)$ is equal to $R_\varphi(z)$ itself. We can now state the following:

Theorem: The potential (nonlocal) producing the $[N/N]$ Padé approximant to the K_l matrix of the regular potential $\bar V$ is, for $\bar V>0$,

$$\bar V_N=\bar V^{1/2}P_N\bar V^{1/2}, \tag{3.4}$$

where P_N is the projector onto the space $\mathcal{E}^{(N)}$ spanned by the vectors

$$\{(\bar V^{1/2}G_l^P(k)\bar V^{1/2})^p\bar V^{1/2}|\varphi_l(k)\rangle\},\quad p=0,1,\ldots,N-1. \tag{3.5}$$

Proof: The exact K matrix for the potential $\bar V$ is, recalling formula (2.15),

$$K_l(\lambda)=-k\lambda\langle\varphi_l(k)|\bar V^{1/2}[1-\lambda\bar V^{1/2}G_l^P(k)\bar V^{1/2}]^{-1}\bar V^{1/2}|\varphi_l(k)\rangle; \tag{3.6}$$

using the previous theorem, we have

$$\begin{aligned}K_l^N(\lambda)&\equiv[N/N]_{K(\lambda)}(\lambda)\\&=-k\lambda\langle\varphi_l(k)|\bar V^{1/2}\\&\quad\times[1-\lambda P_N\bar V^{1/2}G_l^P(k)\bar V^{1/2}P_N]^{-1}\bar V^{1/2}|\varphi_l(k)\rangle;\end{aligned} \tag{3.7}$$

setting

$$\begin{aligned}P_N\bar V^{1/2}&=A,\\\bar V^{1/2}P_N&=A^+,\\\bar V^{1/2}P_N\bar V^{1/2}&=A^+A=\bar V_N>0;\end{aligned} \tag{3.8}$$

and taking into account that $P_N\bar V^{1/2}|\varphi_l(k)\rangle=\bar V^{1/2}|\varphi_l(k)\rangle$, we can rewrite $K_l^N(\lambda)$ on the form

$$K_l^N(\lambda)=-k\lambda\langle\varphi_l(k)|A^+[1-\lambda AG_l^P(k)A^+]^{-1}A|\varphi_l(k)\rangle. \tag{3.9}$$

Now we have the identity

$$A^+[1-\lambda AG_l^P(k)A^+]^{-1}A=A^+A[1-\lambda G_l^P(k)A^+A]^{-1} \tag{3.10}$$

and, therefore,

$$K^N(\lambda)=-k\langle\varphi_l(k)|\lambda\bar V_N[1-\lambda G_l^P(k)\bar V_N]^{-1}|\varphi_l(k)\rangle \tag{3.11}$$

which proves the theorem.

This gives a compact and simple form of a previous statement which can be found in Ref. 11.

It is clear that the sequence of potentials $V_\epsilon^N=\sqrt{V^{(\epsilon)}}P_N\times\sqrt{V^{(\epsilon)}}$ is a monotonous sequence of positive operators, because the space $\mathcal{E}^{(N+1)}$ on which P_{N+1} projects, always contains the space $\mathcal{E}^{(N)}$ on which P_N projects. Therefore, we can write

$$0\leq\lambda V_\epsilon^1\leq\lambda V_\epsilon^2\leq\cdots\leq\lambda V_\epsilon^N<\lambda V_\epsilon^{N+1}<\cdots\leq\lambda V^{(\epsilon)}. \tag{3.12}$$

Now by the extension to nonlocal potentials of a well-known theorem (see Ref. 14), we deduce, that the phase shifts $\delta_\epsilon^N(k,\lambda)$ for fixed ϵ, k, and λ form a monotonous decreasing sequence, converging to $\delta_\epsilon(k,\lambda)$

$$\delta_\epsilon^1(k,\lambda)\geq\delta_\epsilon^2(k,\lambda)\geq\cdots\geq\delta_\epsilon(k,\lambda)\geq\delta_\epsilon^{N+1}(k,\lambda)\geq\cdots\geq\delta_\epsilon(k,\lambda). \tag{3.13}$$

4. CONSTRUCTION OF THE BEST CONVERGING UPPER BOUND FOR THE PHASE SHIFT OF A SINGULAR POTENTIAL STARTING FROM THE KNOWLEDGE OF A FINITE NUMBER OF REGULARIZED BORN TERMS

The exact phase shift is given by the limit

$$\delta(k,\lambda)=\lim_{\epsilon\to0}\lim_{N\to+\infty}\delta_\epsilon^N(k,\lambda). \tag{4.1}$$

It is not possible to invert those two limits, even using Padé approximants, because one sees easily that

$$\lim_{\epsilon\to0}\delta_\epsilon^N(k,\lambda)=0. \tag{4.2}$$

What we shall show is that it is possible to replace this double limit by a simple one, by connecting ϵ and N in such a way that for

$$\begin{aligned}&N\to\infty,\\&\epsilon(N)\to0.\end{aligned} \tag{4.3}$$

The most remarkable property of such connection is that it is unique, nonambiguous, and produces at a given N the best approximation, in contradistinction with other proposals (Ref. 9).

We shall work, from now on, at a given energy, given angular momentum, and given coupling constant, and study the property of the phase shifts as a function of the cutoff ϵ.

The exact phase shift δ_ϵ is, as a function of ϵ, a monotonous increasing function of ϵ. This results from the fact that, with our choice of regularization, we have

$$\begin{aligned}&0<\epsilon_2<\epsilon_1,\\&0<V^{(\epsilon_1)}<V^{(\epsilon_2)}<V,\end{aligned} \tag{4.4}$$

and, therefore,

$$\delta(k,\lambda)\leq\delta_{\epsilon_2}(k,\lambda)\leq\delta_{\epsilon_1}(k,\lambda)\leq0. \tag{4.5}$$

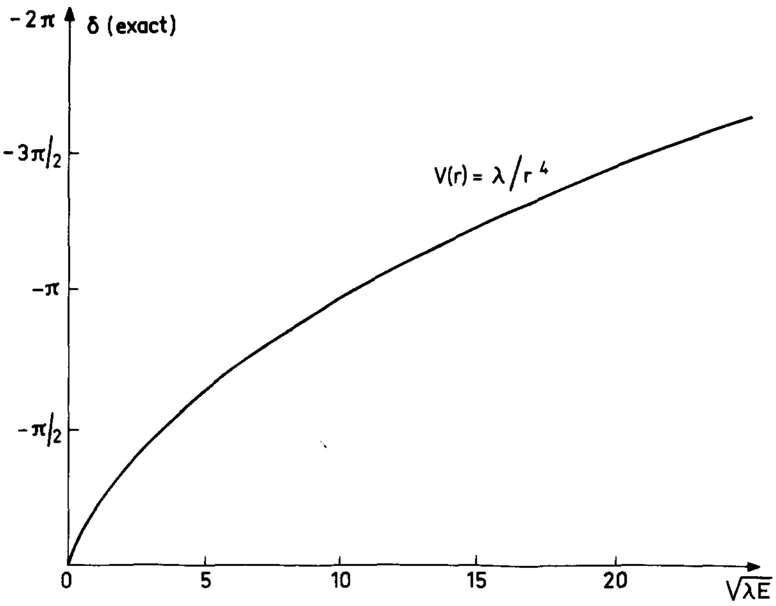

FIG. 1. For the potential $V_1(r)=\lambda/r^4$. We have plotted the exact s-wave phase shift as a function of $\sqrt{\lambda E}$.

Since the phase shift $\delta_\epsilon^N(k,\lambda)$ is bounded from below by $\delta_\epsilon(k,\lambda)$ [see (3.13)] as a function of ϵ and from above by zero, it has therefore a least lower bound that we shall call $\bar{\delta}^N(k,\lambda)$. The value of ϵ for which this least lower bound is obtained is called ϵ_N (should the least lower bound be obtained for more than one value of ϵ, we shall take ϵ_N as the smallest). This procedure defines a unique nonambiguous, and precise set of values $\{\epsilon_N, \bar{\delta}^N(k,\lambda)\}$. We shall show that the sequence $\bar{\delta}^N(k,\lambda)$ is monotonous decreasing in N:

$$\delta^1(k,\lambda)\geq\delta^2(k,\lambda)\geq\cdots\geq\bar{\delta}^N(k,\lambda)\geq\bar{\delta}^{N+1}(k,\lambda)\geq\cdots\geq\delta(k,\lambda). \tag{4.6}$$

This results from the fact that whatever ϵ,

$$\delta_\epsilon^{N+1}(k,\lambda)\leq\delta_\epsilon^N(k,\lambda)\ \text{(as proved in Sec. 3)}, \tag{4.7}$$

while the last inequality results from the fact that $\delta_\epsilon(k,\lambda)$ reaches its least lower bound at $\epsilon=0$.

We shall prove now that the sequence $\{\epsilon_N,\bar{\delta}^N(k,\lambda)\}$ which provides already an upper bound for the exact phase shift, tends, when $N\to\infty$, to the point $\epsilon=0$ and $\delta(k,\lambda)$. The sequence $\bar{\delta}^N(k,\lambda)$ being monotonous decreasing and bounded from below necessarily has a limit for $N\to+\infty$:

$$\bar{\delta}(k,\lambda)\geq\delta(k,\lambda). \tag{4.8}$$

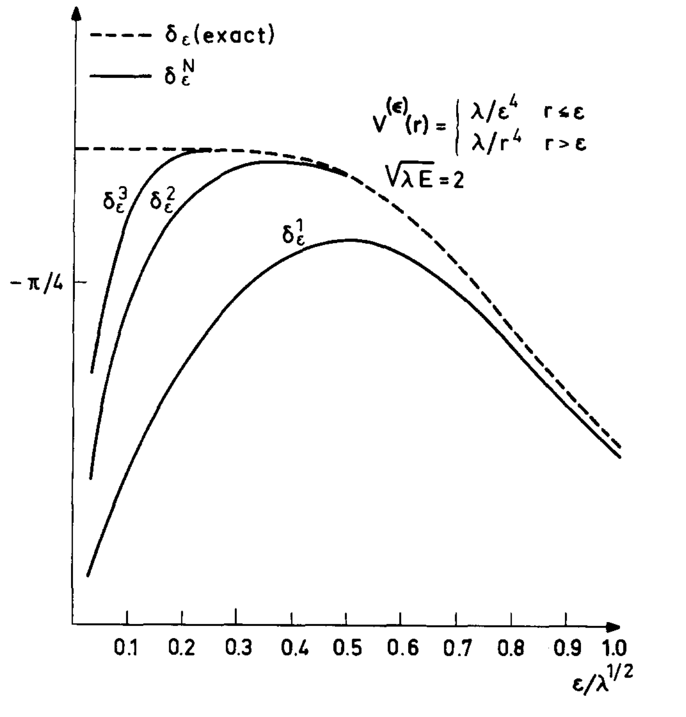

FIG. 2. For the potential $V_1(r)=\lambda/r^4$. We have plotted the approximated regularized s-wave phase shift δ_ϵ^N as a function of the cutoff $\epsilon\lambda^{-1/2}$, for the value $\sqrt{\lambda E}=2$.

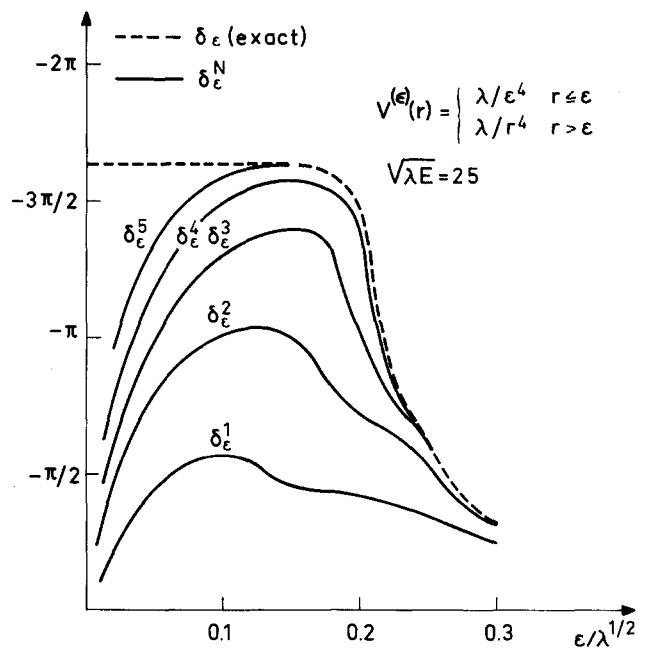

FIG. 3. The same as Fig. 2, but for the value $\sqrt{\lambda E}=25$.

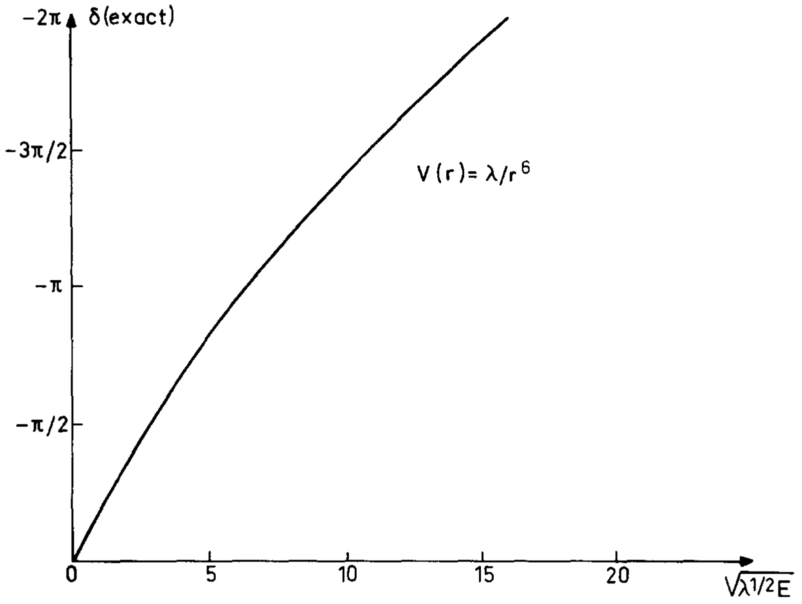

FIG. 4. The same as Fig. 1, but for the potential $V_2(r)=\lambda/r^6$.

We shall show now that $\bar{\delta}(k,\lambda)=\delta(k,\lambda)$. Let us suppose that this is false and prove that it is impossible. In fact, if $\bar{\delta}(k,\lambda)$ is strictly larger than $\delta(k,\lambda)$, then there exists an $\epsilon>0$ for which

$$\lim_{N\to\infty}\delta_\epsilon^N(k,\lambda)>\bar{\delta}(k,\lambda)>\delta_\epsilon(k,\lambda) \tag{4.9}$$

and therefore, for this positive ϵ, the PA do not converge to the correct value $\delta_\epsilon(k,\lambda)$, which is wrong. It is now clear that the sequence ϵ_N goes to zero when $N\to\infty$ due to the monotonous increasing nature of $\delta_\epsilon(k,\lambda)$ in ϵ. In general, there is no reason for the sequence ϵ_N to be monotonous, in contradistinction to the sequence $\bar{\delta}^N(k,\lambda)$ in which we are interested. In fact, numerical experience shows that while $\bar{\delta}^N(k,\lambda)$ tends very rapidly (geometrically) to its limit, the sequence of ϵ_N is slowly converging, but this fact is irrelevant and from practical point of view it turns out to be very favorable as explained in the following paragraph.

Furthermore, $\bar{\delta}^N(k,\lambda)$ provides inside the family of positive potential having the same first $2N$ regularized

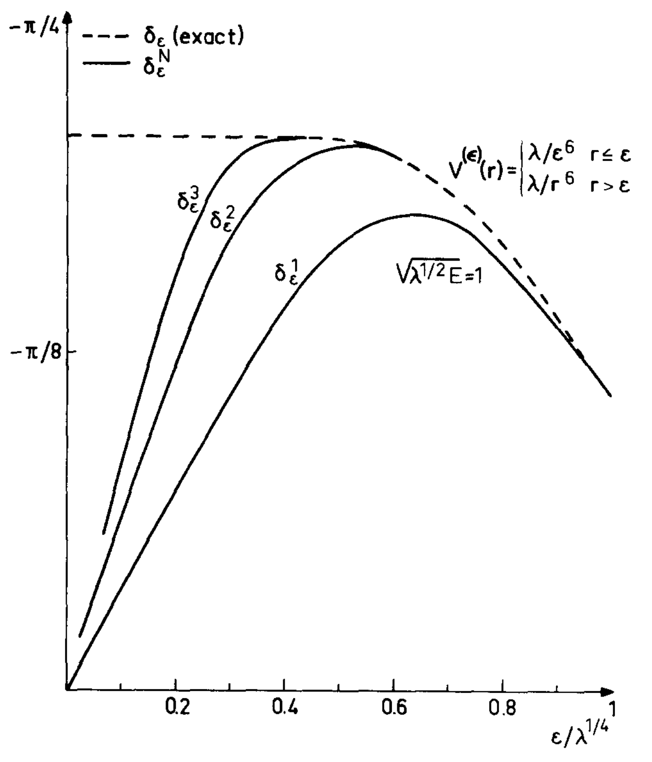

FIG. 5. The same as Fig. 2, but for the potential $V_2(r)=\lambda/r^6$ and $\sqrt{\lambda^{1/2}E}=1$.

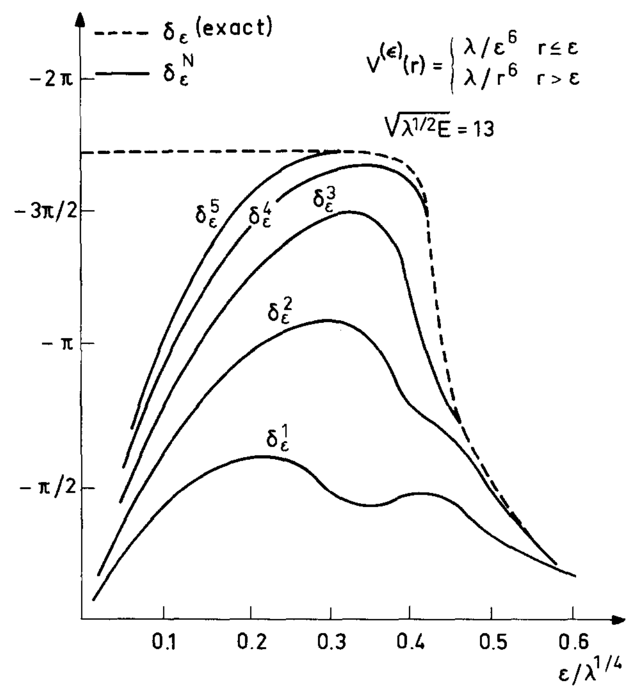

FIG. 6. The same as Fig. 2, but for $\sqrt{\lambda^{1/2}E}=13$.

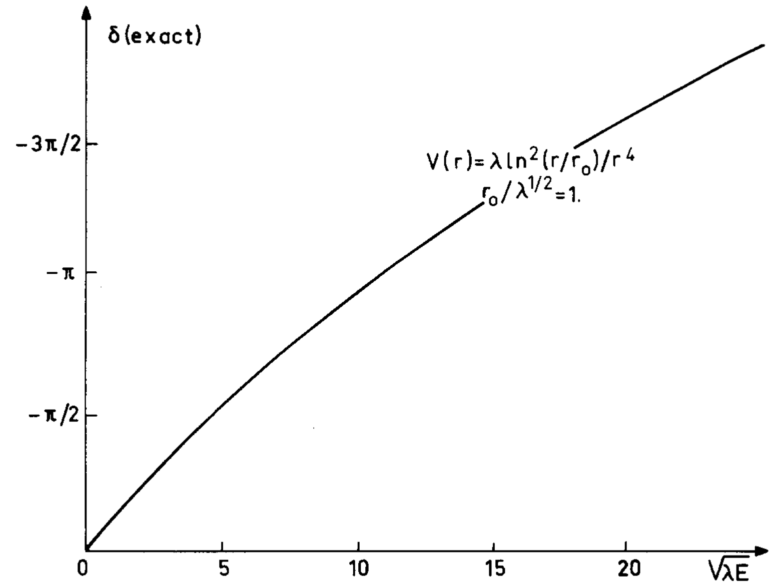

FIG. 7. The same as Fig. 1, but for the potential $V_3(r)=(\lambda/r^4)\log^2(r/r_0)$.

Born terms the *best* upper bound for the phase shift, because it is produced by the potential

$$V^N_{\epsilon_N}=\sqrt{V^{(\epsilon_N)}}\,P_N\,\sqrt{V^{(\epsilon_N)}}. \qquad (4.10)$$

5. EXAMPLES

We have considered three types of potentials fulfilling our statements (Ref. 9):

$$V_1(r)=\lambda/r^4, \qquad (5.1)$$

$$V_2(r)=\lambda/r^6, \qquad (5.2)$$

$$V_3(r)=\lambda\frac{\log^2 r/r_0}{r^4}. \qquad (5.3)$$

Furthermore, we have, also, looked into a changing sign potential

$$V_4(r)=\lambda\frac{\log r/r_0}{r^4} \qquad (5.4)$$

and found surprisingly that our method applies as well in this case. This allows one to think that our method extends much beyond singular potentials which do not change sign.

We report here, first, the results for the potential

$$V_1(r)=\lambda/r^4. \qquad (5.5)$$

Besides the θ-regularization we have also introduced the c-regularization (c as constant) defined by

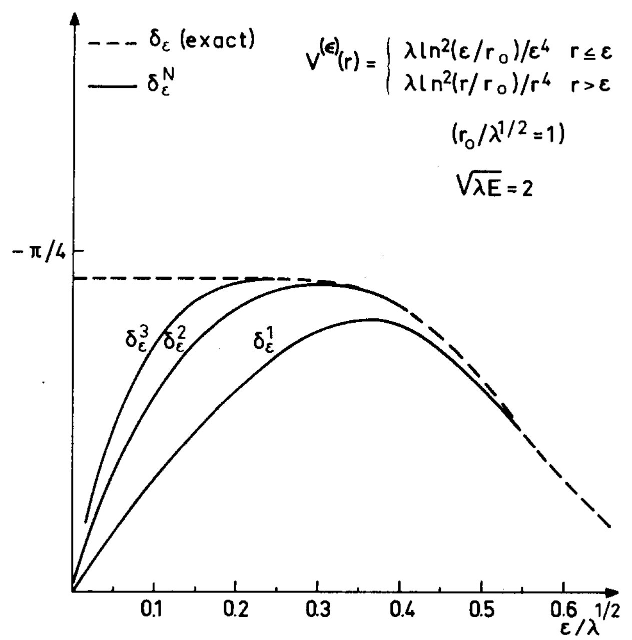

FIG. 8. The same as Fig. 2, but for the potential $V_3(r)$ and $\sqrt{\lambda E}=2$.

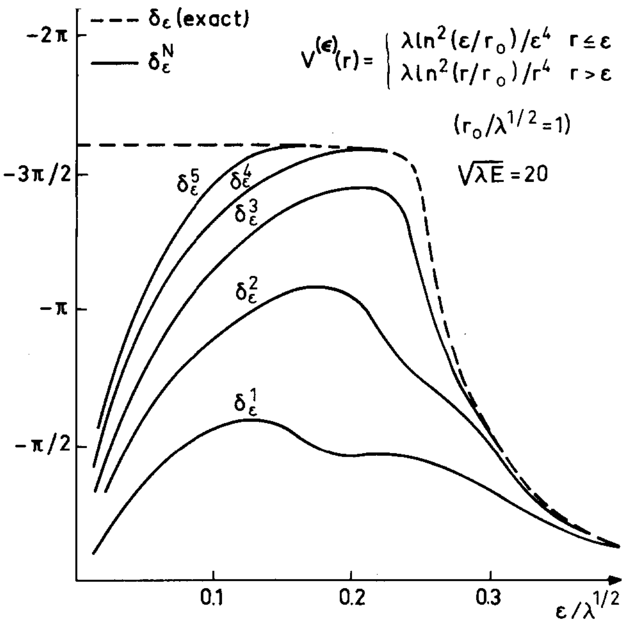

FIG. 9. The same as Fig. 2, but for the potential $V_3(r)$ and $\sqrt{\lambda E}=20$.

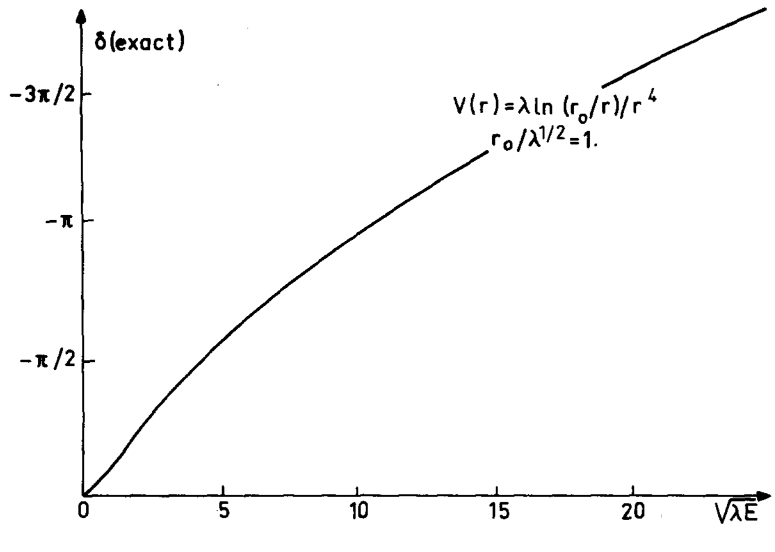

FIG. 10. The same as Fig. 1, but for the changing sign potential $V_4(r) = (\lambda/r^4)\log(r/r_0)$.

$$V^{(\epsilon)} = \begin{cases} V(\epsilon) & \text{for } r \leq \epsilon \\ V(r) & \text{for } r \geq \epsilon. \end{cases} \tag{5.6}$$

These regularizations fulfill all the previous necessary restrictions of Sec. 1(3a–c): Equation (4.1) is valid due to the general arguments of Cornille (Ref. 15).

We have done the calculations with both regularizations and found no difference in the results. We report here the results only for the c-regularization. We have considered only the s-wave case.

Due to dimensional considerations the phase shift for the regularized potential,

$$V_1^{(\epsilon)}(r) = \begin{matrix} \lambda/\epsilon^4 & \text{for } r \leq \epsilon \\ \lambda/r^4 & \text{for } \epsilon \geq r \end{matrix} \tag{5.7}$$

depends only in the s wave on the reduced variables:

$$\delta(\epsilon, E, \lambda) = \delta(\epsilon\lambda^{-1/2}; \sqrt{\lambda E}). \tag{5.8}$$

The Padé approximants on the T matrix have been computed by expanding in powers of λ. Then all manipulations have been done at fixed λ and E.

In Fig. 1 we have drawn the exact s-wave phase shift for the potential λ/r^4 as a function of λ. In Fig. 2, we see that for $\sqrt{\lambda E} = 2$ the PA [1/1] involving only the first and second regularized Born terms gives

$$\bar{\delta}^1 = -(0.282)\pi \tag{5.9}$$

to be compared with the exact value

$$\delta = -(0.347)\pi. \tag{5.10}$$

The second approximation and third approximation gives

$$\bar{\delta}^2 = -(0.338)\pi, \quad \bar{\delta}^3 = -(0.346)\pi. \tag{5.11}$$

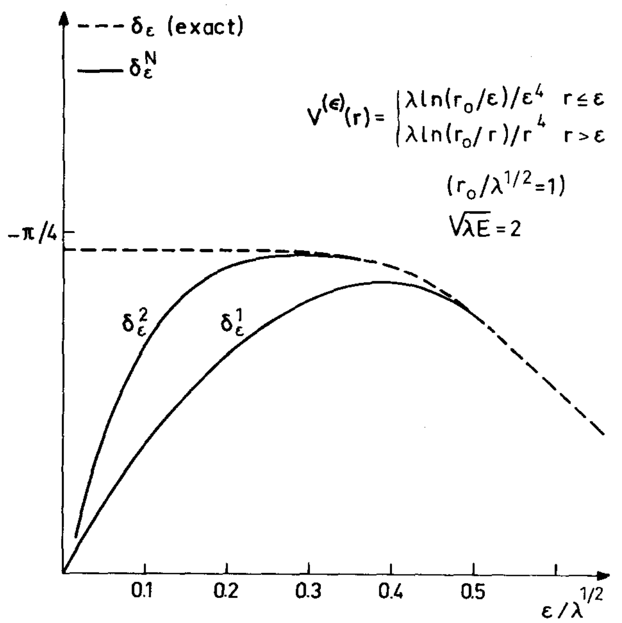

FIG. 11. The same as Fig. 2, but for the potential $V_4(r)$ and $\sqrt{\lambda E} = 2$.

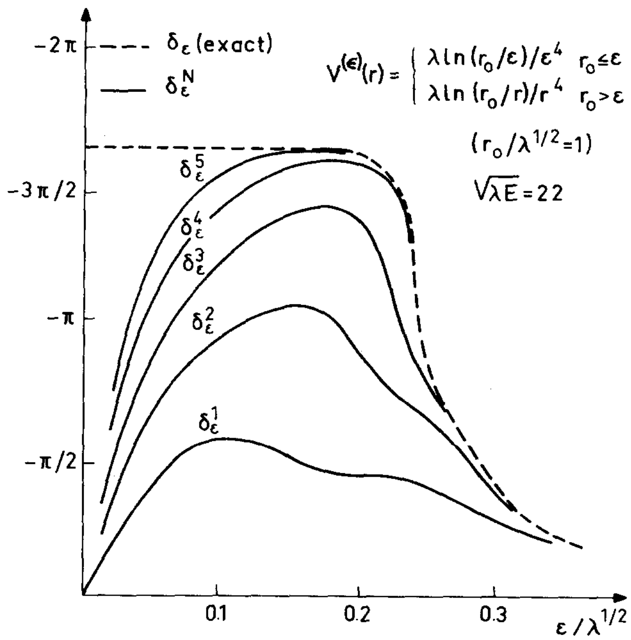

FIG. 12. The same as Fig. 2, but for the potential $V_4(r)$ and $\sqrt{\lambda E} = 22$.

In Fig. 3, we report the same results, but for a high value of the coupling (or the energy), $\sqrt{\lambda E}=25$:

$$\bar{\delta}^1=-(0.572)\pi,\quad \bar{\delta}^2=-(1.04)\pi,\quad \bar{\delta}^3=-(1.40)\pi,$$
$$\bar{\delta}^4=-(1.585)\pi,\quad \bar{\delta}^5=-(1.631)\pi,\quad \delta=-(1.640)\pi. \tag{5.12}$$

We see that the convergence is slightly slower, but the calculation of $\bar{\delta}^5$ still takes an unsignificant time on the computer. We must acknowledge that the calculation of $\bar{\delta}^5$ corresponds to analytically continuing by the PA the regularized Taylor series in the coupling constant to 4.68 radius of convergence !

The numerical calculation has been made in the following way: For the exact phase shift corresponding to $\delta_\epsilon(k,\lambda)$ ($\epsilon>0$) we have used the Volterra equation fulfilled by the Jost solution. To get the exact phase shift $\delta(k,\lambda)$ we have numerically extrapolated $\delta_\epsilon(k,\lambda)$ for $\epsilon\to+0$. This can be done without difficulty due to the general theorem on the convergence in ϵ and to the extremely strong stability of the function $\delta_\epsilon(k,\lambda)$ for ϵ small.

For the Padé approximants, the approximated phase shift $\delta_\epsilon^N(k,\lambda)$ has been obtained from the regularized Born series on the K matrix. This last Born series being deduced from the Taylor series of the Jost function.

On Fig. 3, one notices a very interesting fact: The maximum of the $\delta_\epsilon^N(k,\lambda)$ in ϵ, are, when N increases, still obtained for rather large value of ϵ. As a consequence the calculation is very reliable and do not present any difficulty for the precision required (3×10^{-3}). This would not have been the case, if the points of maximum would have gone to zero too rapidly, due to the difficulty of computing integrals nearly singular.

In Figs. 4, 5, and 6, we report results very analogous to previous ones, but for the potential λ/r^6.

In Figs. 7, 8, and 9 we give the results for the potential $(\lambda/r^4)\log^2 r/r_0$. This potential is interesting because for it, the peratisation method fails completely.

In Figs. 10, 11, and 12 we show the results for a changing sign potential $-(\lambda/r^4)\log r/r_0$. We see that nothing is changed in the behaviour of the curves.

CONCLUSION

In this paper we have given a rigorous proof in a particular physical context of a very general procedure, which was explained in the Introduction.

One can remark, looking to formula (2.15), that the values for which the phase shift goes through an odd multiple of $\pi/2$ correspond to the eigenvalues in λ of the operator $\sqrt{V}\,G_l^P(k)\,\sqrt{V}$. Therefore, we also compute for these values of the phase shift approximated discrete eigenvalues of an operator having an unbounded continuous spectrum (corresponding to negative values of the coupling constant for which one has physically a collapse). By making use of this remark and also of a theorem of Ref. 5, it is not very difficult to obtain a rigorour proof for a similar method applied to the calculation of the bound states of an arbitrary singular potential repulsive in the origin but not of definite sign. This will be the object of a forthcoming paper.

The generalization of these procedures to the N-body problem, can be done for certain types of interactions. The interested reader will find in Ref. 5 the method and the rigorous tools.

ACKNOWLEDGMENTS

Two of us (L.E. and M.Y.) would like to thank Professor C. De Dominicis for the kind hospitality and the facilities he provided us in the Saclay Theoretical Division. The authors wish to thank also H. Cornille for helpful discussions.

* Fellow of Consejo Nacional de Investigaciones Cientificas y Technicas, under CTE contract with the CEA. On leave of absence from the University of La Plata, La Plata, Argentina.
† On leave of absence from the University of Bari, Bari, Italy.

[1] G.A. Baker Jr. and J.L. Gammel, editors, *The Padé Approximant in Theoretical Physics* (Academic, New York, 1970).
[2] P.R. Graves-Morris, editor, *Padé Approximants*, University of Kent Summer School, July 1972, The Institute of Physics, London and Bristol.
[3] P.R. Graves-Morris, editor, *Padé Approximants and Their Applications*, Proceedings of the Conference held at the University of Kent, 17—21 July (1972) (Academic, New York, 1973).
[4] J.L. Basdevant, D. Bessis, J. Zinn-Justin, Nuovo Cimento 60A, 185 (1969).
[5] D. Bessis and M. Villani, "Perturbative-variational approximations to the spectral properties of semi-bounded Hilbert space operators, based on the moment problem with finite or diverging moments. Application to quantum mechanical systems," Saclay preprint DPh-T/74/27.
[6] G. Fogli, M.F. Pellicoro, and M. Villani, Nuovo Cimento 6A, 79 (1971).
[7] G. Garibotti, M.F. Pellicoro, and M. Villani, Nuovo Cimento 66, 749 (1970).
[8] D. Bessis, "Topics in the theory of Padé approximants" in Ref. 2.
[9] W.M. Frank, D.J. Land, and R.M. Spector, Rev. Mod. Phys. 43, 36 (1971).
[10] See Appendix of Ref. 4.
[11] J. Nuttall, Phys. Rev. 157, 1312 (1967).
[12] G.A. Baker Jr., "Convergence of Padé approximants using the solution of linear functional equations," BNL 18382 AMD 665 (November 1973).
[13] D. Masson, "Hilbert space and the Padé approximants," of Ref. 1.
[14] R. Sugar and R. Blankenbecler, Phys. Rev. 136, B 472 (1964).
[15] H. Cornille, Nuovo Cimento 43, 786 (1966).

Phase properties of some photon states with nonzero energy density

J. P. Provost, F. Rocca, and G. Vallee

Physique Theorique, Universite de Nice,[*†] *06034 Nice Cedex, France*

M. Sirugue

Physique Théorique, C.N.R.S. Marseille, France

(Received 22 April 1974)

We describe some photon states with nonzero energy density in the whole space; these states are obtained by taking a finite number of photons within a finite box and letting the volume and the number of photons go to infinity according to usual procedure in statistical mechanics. In such a limit we describe an observable phase operator; we investigate its properties both in the case of free field and in the case of coupling with prescribed classical sources. Finally we give a quantum description of uniform static field.

INTRODUCTION

In a previous paper,[1] we studied the problem of the existence of an observable phase operator for a system of massive bosons. Let us briefly recall the classical origin of the problem: let p and q be the classical canonical variables of a system with one degree of freedom; equivalently, one can consider the following canonical variables:

$$n=\tfrac{1}{2}(p^2+q^2), \tag{1.1}$$

$$\varphi=\arctan(p/q). \tag{1.2}$$

This transformation leads to trivial Hamilton's equations for the harmonic oscillator.

Quantization of the problem has been studied in a lot of papers (see Refs. 2,3,4 and the references of these two last papers). If we assume there exists two self-adjoint operators N and Φ such that

$$N\Phi-\Phi N\subseteq i\mathbb{1}, \tag{1.3}$$

then we have the following uncertainty relation:

$$\Delta\Phi\cdot\Delta N\geq\tfrac{1}{2}. \tag{1.4}$$

It has been recognized that (1.4) leads to contradictions from a physical point of view; in any case, the existence of a self-adjoint operator Φ satisfying (1.3) is incompatible with a lower bounded N.

In the previous papers on this question the only systems which have been considered were one-dimensional systems (except in Refs. 1 and 5). In Ref. 1 we dealt with systems with more than one degree of freedom and we showed that the existence of an observable phase operator is incompatible with an observable number operator. This operator has been used to build a simple mophase operator for systems with a finite number of degrees of freedom. Moreover, in the thermodynamical limit and for condensed systems of massive bosons we were able to exhibit explicitly an observable phase operator. This operator has been used to build a simple model of the Josephson's effect where the notion of phase plays an important role (see Refs. 5–7).

However, the first use of a phase operator was done by Dirac[8] for radiation field. This field of application is certainly much more natural than the one we considered in Ref. 1. Indeed, in classical theory one has a notion of phase for light beams. This notion is essential to define the coherence properties of light, and, for example, to describe interference phenomena.

On the other hand, one has to realize that for a photon field the notion of total number of particles has no physical meaning. Indeed, and this is a well-known argument, whatever be the detector it has a finite resolution and cannot detect the soft photons; hence there is a complete uncertainty on the number of photons. Consequently, for a field of photons in the whole space we are precisely in the same situation as for a massive Bose gas with finite density as far as the number of particles is concerned. It is tempting to look at the problem of the radiation field using the methods of Ref. 1 to define an observable phase operator for photons.

We describe in Sec. 2 an algebra for infinitely extended photon field which takes into account the Poincaré invariance, the gauge invariances of both kinds, and we give within this formalism the essential usual definitions of Fock representation.

In Sec. 3 we start from a system of N photons localized in a box of finite volume V. In a box, the allowed energy levels are discrete and it is possible to separate the first excited level from the zero energy level; so the notion of total number of excitations makes sense and we can work within the Fock representation.

We make the standard limiting procedure $N\to\infty$, $V\to\infty$, N/V being constant and define in this way a state of the photon field with finite energy density which is no longer a Fock state. This state is precisely a plane wave state. Within the associated representation there exists an observable phase operator, which is explicitly given. Notice the similarity of this situation to the condensation of a massive Bose gas below the critical temperature in the zero momentum mode.

In Sec. 4, we repeat that procedure in the case of a system of photons coupled with an external classical source. The limiting state has a phase which is not equally distributed but which tends to be concentrated around the phase of the radiated classical field.

Finally, in Sec. 5 we make the limit $k\to 0$ for the state we obtained in Sec. 3. In this way we describe an uniform static field.

2. THE ALGEBRA OF THE PHOTON FIELD

Classically a free photon field is described by a vector potential $x \to A_\mu(x)$, $\mu = 0,1,2,3$, $x = (x_0, \mathbf{x})$, satisfying the wave equation

$$\Box A_\mu(x) = 0. \tag{2.1}$$

In order to construct in quantum theory an algebra whose states will be the states of the photon field we shall make use of the explicitly covariant formalism which is in Ref. 9; we shall recall the essential definitions.

The Minkowski space M is the real space $\mathbb{R}^4$ endowed with the metric

$$x \cdot y = -x_0 y_0 + \mathbf{x} \cdot \mathbf{y} = x^\mu y_\mu; \tag{2.2}$$

its complexification M' is the complex space $\mathbb{C}^4$ of vectors $x = x_1 + i\, x_2 (x_1, x_2 \in M)$ with the scalar product

$$x \cdot y = x_1 \cdot y_1 + x_2 \cdot y_2 + i(x_1 \cdot y_2 - x_2 \cdot y_1). \tag{2.3}$$

Let us consider a function f from M to M' which satisfies the wave equation

$$\Box f(x) = 0; \tag{2.4}$$

it can be written

$$f(x) = (2\pi)^{-3/2} \int \exp(ik \cdot x) \tilde{f}(\mathbf{k})\, d\Omega(k) \tag{2.5}$$

where $d\Omega(k) = d^3k/2k_0$ is the invariant measure on the light cone ($k_0 = |\mathbf{k}|$). The space L is the linear space of the positive energy solutions of (2.4) which satisfy:

(i) $\sum_{\mu=0}^{3} \int |\tilde{f}_\mu(\mathbf{k})|^2 d\Omega(k) < \infty$,

(ii) $\partial^\mu f_\mu(x) = 0 \Leftrightarrow k \cdot \tilde{f}(\mathbf{k}) = 0$ (Lorentz condition).

On this space it is possible to define a positive semidefinite scalar product

$$(f,g) = i \int f^{\mu *}(x)\, \overleftrightarrow{\partial}_0 g_\mu(x)\, d^3x \tag{2.6}$$

$$= \int \tilde{f}^{\mu *}(\mathbf{k})\, \tilde{g}_\mu(k)\, d\Omega(k).$$

The one photon Hilbert space is the quotient $\bar{L} = L/L_0$ where L_0 is the subspace of isotropic vectors in L with respect to (2.6). The Fourier transform of functions in L_0 are of the form:

$$\tilde{f}_0(\mathbf{k}) = \lambda(k) k, \tag{2.7}$$

where $\lambda(k)$ is an arbitrary function.

For the sake of convenience we shall deal with L and we shall verify at each stage that our results are compatible with gauge invariance (see below).

As usually for bosons, we define the antisymmetric real bilinear form σ on L:

$$\sigma(f,g) = (2i)^{-1}[(f,g) - (g,f)]. \tag{2.8}$$

This symplectic form σ is degenerate on L_0.

The C^*-algebra $\overline{\Delta(L,\sigma)}$ is constructed according to standard procedure[10,11]: $\Delta(L,\sigma)$ is the $*$-algebra generated by the elements δ_f, $f \in L$, which satisfy the Weyl relation

$$\delta_f \delta_g = \exp[-i\sigma(f,g)]\delta_{f+g}. \tag{2.9}$$

The δ_f are unitary:

$$(\delta_f)^* = \delta_{-f}. \tag{2.10}$$

On this algebra $\Delta(L,\sigma)$ there exists a C^*-algebra norm for which any state of $\Delta(L,\sigma)$ [i.e., any linear positive normalized functional on $\Delta(L,\sigma)$] is continuous hence extends to a state of the closure $\overline{\Delta(L,\sigma)}$.

A state ω of $\overline{\Delta(L,\sigma)}$ is a Weyl state if the function $\lambda \to \omega(\delta_{f+\lambda g})$ is continuous at $\lambda = 0$. If π_ω is the representation of $\overline{\Delta(L,\sigma)}$ on $\mathcal{H}_\omega$ with cyclic vector Ω_ω deduced from the Weyl state ω by the usual Gelfand–Naimark–Segal (GNS) construction, the Stone theorem leads to

$$\pi_\omega(\delta_f) = \exp[iA_\omega(f)]. \tag{2.11}$$

$A_\omega(f)$ is the field operator smeared out with the function $f \in L$.

Creation and annihilation operators are defined by

$$A_\omega^\pm(f) = \tfrac{1}{2}[A_\omega(f) \mp iA_\omega(if)] \tag{2.12}$$

and satisfy the usual commutation relations; in particular,

$$[A_\omega^-(f), A_\omega^+(g)]_- = (f,g). \tag{2.13}$$

Due to the choice of L instead of $\bar{L}$, $\overline{\Delta(L,\sigma)}$ has a center which is precisely $\overline{\Delta(L_0)}$. The gauge invariance of the theory imposes to consider only states such that

$$\omega(\delta_{f+f_0}) = \omega(\delta_f) \quad \forall\ f_0 \in L_0. \tag{2.14}$$

This condition is equivalent to

$$\omega(\delta_{f_0}) = 1 \quad \forall\ f_0 \in L_0 \tag{2.15}$$

or $A_\omega(f_0) = 0$.

On $\overline{\Delta(L,\sigma)}$ is defined the compact group of $*$-automorphisms corresponding to gauge transformations of the first kind:

$$\alpha_\theta \delta_f = \delta_{e^{i\theta} f}. \tag{2.16}$$

If π is a representation of $\overline{\Delta(L,\sigma)}$ such that there exists a weakly continuous group of unitaries U_θ which implement α_θ, i.e.,

$$\pi(\alpha_\theta \delta_f) = U_\theta \pi(\delta_f) U_\theta^{-1}, \tag{2.17}$$

then the infinitesimal generator N of U_θ is a particle number for the representation π.[12]

We can also define the representation of the Poincaré group into the $*$-automorphisms of $\overline{\Delta(L,\sigma)}$ by

$$\alpha_d \delta_f = \delta_{f_d} \quad (f_d(x) = f(x-d),\ d \in M), \tag{2.18}$$

$$\alpha_\Lambda \delta_f = \delta_{f_\Lambda} \quad (f_\Lambda(x) = \Lambda f(\Lambda^{-1}x),\ \Lambda \in \mathcal{L}'). \tag{2.19}$$

Let us recall now some properties of the well-known Fock representation: It is the representation associated by the GNS construction with the state ω_F of $\overline{\Delta(L,\sigma)}$ defined by

$$\omega_F(\delta_f) = \exp[-(f,f)/2]. \tag{2.20}$$

Let $(\pi_F, \mathcal{H}_F, \Omega_F)$ be the GNS triplet associated with ω_F. ω_F is a Weyl state, so $\pi_F(\delta_f) = \exp[iA_F(f)]$. The operators $A_F(f) = A_F^-(f) + A_F^+(f)$ can be deduced from the field operator $A_F(x)$

$$A_{F_\mu}(x) = A_{F_\mu}^+(x) + A_{F_\mu}^-(x) \tag{2.21}$$

$$= (2\pi)^{-3/2} \int d\Omega(k)[a_\mu(\mathbf{k}) \exp(ik \cdot x) + a_\mu^+(\mathbf{k}) \exp(-ik \cdot x)]$$

by

$$A_F^-(f)=(f,A_F^-) \text{ and } A_F^+(f)=(A_F^-,f). \quad (2.22)$$

In (2.21), the operators $a_\mu(\mathbf{k})$ and $a_\mu^*(\mathbf{k})$ satisfy the canonical commutation relations

$$[a_\mu(\mathbf{k}), a_\nu^*(\mathbf{k}')]_- = g_{\mu\nu} 2k_0 \delta^3(\mathbf{k}-\mathbf{k}') \quad (k_0=|\mathbf{k}|) \quad (2.23)$$

and we have

$$A_F^-(f)|\Omega_F\rangle = 0 \cdot \quad \forall\, f\in L \quad (2.24)$$

which completes the connection with the usual formalism. Moreover, $\omega_F(\delta_{f_0})=1$, $\forall\, f_0\in L_0$; so, we have gauge invariance.

As ω_F is invariant with respect to the group of *-automorphisms (2.16), π_F is a representation with a particle number N_F. The spectrum of N_F is the set $\mathbb{Z}^+$ of positive integers.

We already mentioned in the introduction that the existence of such an observable particle number is not physically reasonable except within a box.

On the other hand, as shown in Ref. 1, there is no Hermitian phase operator Φ_F verifying the commutation relation $[N_F, \Phi_F]_- = i$. Indeed, the existence of an observable phase operator Φ in a representation π with respect to a particle number N implies that the spectrum of N be the whole set $\mathbb{Z}$ of integers.[1] In this case the representation is not quasi-equivalent to the Fock one.[12]

In the Fock representation the coherent states are well-known:

$$\omega_g(\delta_f)=\exp[-\tfrac{1}{2}(f,f)-2i\sigma(g,f)], \quad (2.25)$$

$$g\in L,$$

where the corresponding cyclic vector is

$$|\Omega_g\rangle = \exp[-iA_F(g)]|\Omega_F\rangle. \quad (2.26)$$

More generally, we shall consider in the following states defined by

$$\omega_G(\delta_f)=\exp[-\tfrac{1}{2}(f,f)+iG(f)] \quad (2.27)$$

where G is a real linear form on $\bar{L}$; then we shall call them coherent states even if G is not continuous with respect to the norm in $\bar{L}$. In this case ω_G is not quasi-equivalent to Fock state.

3. QUANTUM CONSTRUCTION OF THE PLANE WAVE STATE

We are interested in describing a plane wave extended in the whole space $\mathbb{R}^3$ as it is commonly defined in classical theory. We look for the corresponding state of $\overline{\Delta(L,\sigma)}$ in the quantum description. This state cannot be a Fock state since a plane wave has an infinite energy and the number of photons must be strictly infinite. Nevertheless, it carries a finite density of energy and a finite density of photons. This will allow us to induce its form by a method adapted from the classical calculation of Araki and Woods for the massive Bose gas.[13]

Let us consider N photons all in the same one-particle state: their wavefunction is the function $f_{V,\epsilon,\mathbf{k}}$ of L which is given at $t=0$ by

$$f_{V,\epsilon,\mathbf{k}}(\mathbf{x},0)=\frac{\chi(V)}{\sqrt{V}}\frac{\exp(i\mathbf{k}\cdot\mathbf{x})}{\sqrt{2k_0}}\epsilon \quad (k_0=|\mathbf{k}|), \quad (3.1)$$

$$\epsilon\in M',$$

where $\chi(V)$ is the conveniently regularized characteristic function of the volume $V\subset\mathbb{R}^3$.

$f_{V,\epsilon,\mathbf{k}}$ describes a photon of polarization ϵ, localized within the volume V at the time $t=0$. In the limit $V\to\infty$, this function will describe a photon of definite momentum $\mathbf{k}$. The natural representation for such an N-photon system is the Fock representation, and the corresponding vector in $\mathcal{H}_F$ is

$$|\Phi_{N,V,\epsilon,\mathbf{k}}\rangle = \frac{1}{\sqrt{N!}}A_F^+(f_{V,\epsilon,\mathbf{k}})^N|\Omega_F\rangle. \quad (3.2)$$

Denote by ω_V the state of $\overline{\Delta(L,\sigma)}$ associated with the vector (3.2) through the relation

$$\omega_V(\delta_f)=\langle\Phi_{N,V,\epsilon,\mathbf{k}}|\exp[iA_F(f)]|\Phi_{N,V,\epsilon,\mathbf{k}}\rangle. \quad (3.3)$$

A calculation similar to the one done in Ref. 13 gives

$$\omega_V(\delta_f)=\exp[-(f,f)/2]\,L_N(|(f_{V,\epsilon,\mathbf{k}},f)|^2) \quad (3.4)$$

where L_N is the Nth Laguerre polynomial.

In order to get the plane wave situation, we take the thermodynamical limit $N\to\infty$, $V\to\infty$, keeping constant the energy density $w=Nk_0/V$. In this limit the scalar product $(f_{V,\epsilon,\mathbf{k}},f)$ behaves like

$$\frac{(2\pi)^{3/2}}{\sqrt{2k_0}}\frac{1}{\sqrt{V}}\epsilon\cdot\tilde{f}(\mathbf{k}).$$

So we obtain ($\tilde{f}_\epsilon=\epsilon\cdot\tilde{f}$)

$$\text{th-lim } \omega_V(\delta_f)=\exp[-(f,f)/2]\lim_{N\to\infty}L_N\left((2\pi)^3\frac{w}{2Nk_0^2}|\tilde{f}_\epsilon(\mathbf{k})|^2\right)$$
$$=\exp[-(f,f)/2]\,J_0(\lambda|\tilde{f}_\epsilon(\mathbf{k})|) \text{ with } \lambda=(2\pi)^{3/2}(\sqrt{2w}/k_0). \quad (3.5)$$

Formula (3.5) defines a Weyl state $\omega_{k,\epsilon}$ of $\overline{\Delta(L,\sigma)}$.

Rewrite (3.5) as

$$\omega_{k,\epsilon}(\delta_f)=\frac{1}{2\pi}\int_0^{2\pi}d\theta\exp\{-\tfrac{1}{2}(f,f)+i\lambda[\mathrm{Re}\tilde{f}_\epsilon(\mathbf{k})\cos\theta$$
$$+\mathrm{Im}\tilde{f}_\epsilon(\mathbf{k})\sin\theta]\}. \quad (3.6)$$

$\omega_{k,\epsilon}$ is a convex combination of Weyl states; precisely it is a convex combination of coherent states which are not quasiequivalent to Fock state.

It is obvious to verify that the state $\omega_{k,\epsilon}$ does not depend on the time origin where the initial condition (3.1) was given. Furthermore, $\omega_{k,\epsilon}$ is translation invariant and satisfies the covariance property

$$\omega_{k,\epsilon}\circ\alpha_\Lambda=\omega_{\Lambda^{-1}k,\Lambda^{-1}\epsilon}, \quad \Lambda\in\mathcal{L}^{\uparrow}. \quad (3.7)$$

We have gauge invariance $\omega_{k,\epsilon}(\delta_{f_0})=1$, $f_0\in L_0$, since $(f_0,f_0)=0$ and $\epsilon\cdot\tilde{f}_0(\mathbf{k})=0$ from (2.7) and Lorentz condition.

Finally, $\omega_{k,\epsilon}$ is invariant with respect to the gauge transformation of the first kind (2.16); hence the corresponding cyclic representation $\pi_{k,\epsilon}$ has a particle number N. From (3.6), $\pi_{k,\epsilon}$ is not quasiequivalent to the Fock representation; so the spectrum of N is the whole set $\mathbb{Z}$ of integers.[12]

Explicit construction of $\pi_{k,\epsilon}$ can be made following Ref. 13. Let $\mathcal{H}$ and Ω be the space representation and the cyclic vector. Then

$$\mathcal{H}=\mathcal{H}_F\otimes\mathcal{M} \tag{3.8}$$

where $\mathcal{M}$ is the Hilbert space of square-integrable functions on the unit circle with respect to the Haar measure $d\theta/2\pi$,

$$|\Omega\rangle=|\Omega_F\rangle\otimes|\chi^\circ\rangle \tag{3.9}$$

where χ° is the constant function equal to one on the unit circle,

$$\pi_{k,\epsilon}(\delta_f)=\pi_F(\delta_f)\otimes\exp\{i\lambda[\mathrm{Re}\tilde{f}_\epsilon(\mathrm{k})C+\mathrm{Im}\tilde{f}_\epsilon(\mathrm{k})S]\} \tag{3.10}$$

with the following definitions of C and S:

$$\left.\begin{aligned}(C\chi)(\theta)&=\cos\theta\,\chi(\theta)\quad&(3.11)\\(S\chi)(\theta)&=\sin\theta\,\chi(\theta)\quad&(3.12)\end{aligned}\right\}\chi\in\mathcal{M}.$$

Creation and annihilation operators are obtained from (2.11), (2.12), and (3.10):

$$A^+(f)=A_F^+(f)\otimes\mathbb{1}+\mathbb{1}\otimes\tfrac{1}{2}\lambda\tilde{f}_\epsilon(\mathrm{k})(C-iS), \tag{3.13}$$

$$A^-(f)=A_F^-(f)\otimes\mathbb{1}+\mathbb{1}\otimes\tfrac{1}{2}\lambda\tilde{f}_\epsilon^*(\mathrm{k})(C+iS). \tag{3.14}$$

Using these formulas and (2.22), one can deduce the expression of the field operator components in a normalized basis (e_μ), $\mu=0,1,2,3$ of M':

$$A_\mu^\pm(x)=A_{F,\mu}^\pm(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{w/2}\,\frac{\exp(\mp ikx)}{k_0}(\epsilon\cdot e_\mu)(C\mp iS). \tag{3.15}$$

One verifies that $A_\mu(x)$ satisfies the free field equation. The particle number is

$$N=N_F\otimes\mathbb{1}+\mathbb{1}\otimes\left(i\frac{d}{d\theta}\right). \tag{3.16}$$

According to Ref. 1, there exists an observable phase operator Φ defined by

$$e^{i\Phi}=\mathbb{1}\otimes(C+iS) \tag{3.17}$$

with

$$[N,\Phi]=i. \tag{3.18}$$

Two remarks on the operator $e^{i\Phi}$:

(i) Its mean value in the cyclic vector Ω is zero. We shall say in this case that the phase is equally distributed.

(ii) Its eigenvectors are improper elements of $\mathcal{H}$: their $\mathcal{M}$ components are Dirac measures on the unit circle.

Using (3.9) and (3.15), one can calculate the different correlation functions:

$$\langle\Omega|\Omega\rangle=1, \tag{3.19}$$

$$\langle\Omega|A_\mu^+(x)|\Omega\rangle=\langle\Omega|A_\mu^-(x)|\Omega\rangle=0, \tag{3.20}$$

$$\begin{aligned}\langle\Omega|\prod_{i=1}^n A_{\mu_i}^+(x_i)\prod_{j=1}^m A_{\mu_j}^-(x_j)|\Omega\rangle&=0\quad\text{if } n\neq m\\&=\prod_{i=1}^n\mathcal{A}_{\mu_i}^*(x_i)\prod_{j=1}^n\mathcal{A}_{\mu_j}(x_j)\quad\text{if } n=m,\end{aligned} \tag{3.21}$$

where

$$\mathcal{A}_\mu(x)=\sqrt{w/2}\,\frac{\exp(ik\cdot x)}{k_0}(\epsilon\cdot e_\mu)\quad(k_0=|\mathrm{k}|). \tag{3.22}$$

$\mathcal{A}(x)$ is precisely the 4 potential corresponding to a plane electromagnetic wave of momentum k and polarization ϵ.

Formula (3.21) shows that the state $\omega_{k,\epsilon}$ ensures a complete factorization of even correlation functions. This result has to be connected with the conditions of complete coherence for the electromagnetic field as given and discussed in Ref 14.

The problem of finding the states of the electromagnetic field which satisfy (3.21) has been already solved in two different ways. Firstly, Glauber showed in Refs. 14 and 15 that, in the Fock representation for finite systems, the solutions are the coherent states up to an integration over their argument, which corresponds to the decomposition (3.6). Secondly, if one works within a classical theory where $A_\mu(x)$ is a random function and the mean value stands for an average with respect to some probability density, one gets classical plane waves (see for example Ref. 16). At this point, it is tempting to make the connection between these two approaches by identifying the amplitude and the phase of the plane wave, respectively, with the modulus and the argument of the coherent state. However, this leads to troubles since the relationship between the phase and the number of particles is not as expected (see, for example, Ref. 3).

We have here escaped to these difficulties by going to a representation of CCR which is not quasiequivalent to the Fock one. Within our formalism it is possible to exhibit a state which ensures the full coherence and such that the modulus and the phase of the field are both observables.

4. PLANE WAVE LIMIT IN PRESENCE OF A CLASSICAL CURRENT

We want now to accomplish the same program as in Sec. 3 but in the case where the electromagnetic field is coupled to a prescribed c-number current distribution.

This study is the natural extension of the previous one in the sense that it will allow to reach more concrete physical situations with specific phase properties.

We shall work in the radiation gauge and in the Heisenberg picture. The field equation is then

$$\Box\mathbf{A}(\mathbf{r},t)=\mathbf{J}(\mathbf{r},t), \tag{4.1}$$

where $\mathbf{J}(\mathbf{r},t)$ is the transverse component of the current. In the Fock representation we write the Fourier decomposition of the field operator:

$$\begin{aligned}\mathbf{A}_F(\mathbf{r},t)=(2\pi)^{-3/2}\int d\Omega(k)[&\mathbf{a}(\mathbf{k},t)\exp(i\mathbf{k}\cdot\mathbf{r})\\&+\mathbf{a}^+(\mathbf{k},t)\exp(-i\mathbf{k}\mathbf{r})]\end{aligned} \tag{4.2}$$

with the commutation relations (2.23) at equal time for the operators $a_i(\mathbf{k},t)$ and $a_j^+(\mathbf{k},t)$.

Similarly, we shall use

$$\mathbf{J}(\mathbf{r},t)=(2\pi)^{-3/2}\int d\Omega(k)\mathbf{J}(\mathbf{k},t)\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{4.3}$$

We choose the solution of (4.1) which reduces to a free field at time $t=-\infty$. We get (see, for example, Ref. 17)

$$\mathbf{a}(\mathbf{k},t)=\mathbf{a}(\mathbf{k})\exp(-ik_0t)+\mathbf{g}(\mathbf{k},t)\quad(k_0=|\mathbf{k}|) \tag{4.4}$$

with

$$\mathbf{g}(\mathbf{k},t)=\frac{i}{2k}\int_{-\infty}^t dt'\exp[-ik_0(t-t')]\mathbf{J}(\mathbf{k},t')\quad(k_0=|\mathbf{k}|). \tag{4.5}$$

The operators $A_F^{\pm}(f)$ are defined as before by formula (2.22), where f belongs to L, but now they explicitly depend on time, since $\mathbf{A}_F(\mathbf{r},t)$ no longer verifies the free field equation (2.1). We shall write them $A_{F,t}^{\pm}(f)$; from (2.22) and (4.4), we have

$$A_{F,t}^{-}(f)=A_{F,(-\infty)}^{-}(f)+(f,g),$$
$$A_{F,t}^{+}(f)=A_{F,(-\infty)}^{+}(f)+(g,f), \qquad (4.6)$$

where (f,g) is defined by (2.6) but only the spatial components of f and g are concerned due to our choice of gauge, and it explicitly depends on time.

Note from (4.4) that the Fock algebras of the field operators are the same in the both free and interacting cases. This justifies that we continue to work in the interacting case with the photon algebra $\overline{\Delta(L,\sigma)}$.

Let us now approach the plane wave limit as we did in Sec. 3. Namely, we consider the Fock vector state at time t_0:

$$|\Phi_{N,V,\epsilon,\mathbf{k}}^{t_0}\rangle=A_{F,t_0}^{+}(f_{V,\epsilon,\mathbf{k}})^N|\Omega_F\rangle \qquad (4.7)$$

and we calculate the thermodynamical limit of

$$\omega_V^{t_0}(\delta_f)=\frac{\langle\Phi_{N,V,\epsilon,\mathbf{k}}^{t_0}|\exp[iA_{F,t}(f)]|\Phi_{N,V,\epsilon,\mathbf{k}}^{t_0}\rangle}{\langle\Phi_{N,V,\epsilon,\mathbf{k}}^{t_0}|\Phi_{N,V,\epsilon,\mathbf{k}}^{t_0}\rangle} \qquad (4.8)$$

keeping constant the energy density $w=Nk_0/V$.

Obviously, the state defined by (4.7) is no longer an N photon state. It appears, however, as a natural generalization in the interacting case of the previous one (3.2): Starting from the same initial condition at time $t=-\infty$, the same construction is performed at time t_0.

In the present situation $|\Omega_F\rangle$ is no longer the vacuum for the operators $A_{F,t}^{-}(f)$, but a coherent vector state for these ones.

Indeed, from (4.6)

$$A_{F,t}^{-}(f)|\Omega_F\rangle=(f,g)|\Omega_F\rangle. \qquad (4.9)$$

As in Sec. 3, through a calculation similar to the one given in Ref. 13, we obtain $(g_\epsilon(\mathbf{k},t)=\epsilon\cdot g(\mathbf{k},t))$

$$\text{th-lim}\,\omega_V^{t_0}(\delta_f)=\exp[-(f,f)/2]\exp[2i\mathrm{Re}(f,g)]\times\frac{J_0(\lambda[(\tilde{f}_\epsilon(\mathbf{k})-ig_\epsilon(\mathbf{k},t_0))(\tilde{f}_\epsilon^*(\mathbf{k})-ig_\epsilon^*(\mathbf{k},t_0))]^{1/2})}{J_0(i\lambda|g_\epsilon(\mathbf{k},t_0)|)} \qquad (4.10)$$

which reduces to (3.5) when $\mathbf{J}=0$.

As before, (4.10) defines a Weyl state $\omega_{\mathbf{k},\epsilon}^{t_0}$ of $\overline{\Delta(L,\sigma)}$ and we have gauge invariance since

$$\omega_{\mathbf{k},\epsilon}^{t_0}(\delta_{f_0})=1 \quad \forall f_0\in L_0.$$

We can rewrite (4.10):

$$\omega_{\mathbf{k},\epsilon}^{t_0}(\delta_f)=\exp[-(f,f)/2]\exp[2i\mathrm{Re}(f,g)]\int_0^{2\pi}d\mu(\theta)\times\exp\{i\lambda[\mathrm{Re}\tilde{f}_\epsilon(\mathbf{k})\cos\theta+\mathrm{Im}\tilde{f}_\epsilon(\mathbf{k})\sin\theta]\} \qquad (4.11)$$

with

$$d\mu(\theta)=\frac{d\theta}{2\pi}\,\frac{\exp\{\lambda[\mathrm{Re}g_\epsilon(\mathbf{k},t_0)\cos\theta+\mathrm{Im}g_\epsilon(\mathbf{k},t_0)\sin\theta]\}}{J_0(i\lambda|g_\epsilon(\mathbf{k},t_0)|)}. \qquad (4.12)$$

The corresponding space representation as previously in the free case is $\mathcal{H}=\mathcal{H}_F\otimes\mathcal{M}$. On the contrary, the cyclic vector[18] is notably changed:

$$|\Omega^{t_0}\rangle=|\Omega_F\rangle\otimes|\chi_{\mathbf{k},\epsilon}^{t_0}\rangle \qquad (4.13)$$

with

$$\chi_{\mathbf{k},\epsilon}^{t_0}(\theta)=\frac{\exp[\frac{1}{2}\lambda|g_\epsilon(\mathbf{k},t_0)|\cos(\theta-\varphi)]}{[J_0(i\lambda|g_\epsilon(\mathbf{k},t_0)|)]^{1/2}} \qquad (4.14)$$

$$(\varphi=\mathrm{Arg}g_\epsilon(\mathbf{k},t_0)).$$

Creation and annihilation operators in this representation space are

$$A_t^{+}(f)=A_{F,t}^{+}(f)\otimes\mathbb{1}+\mathbb{1}\otimes\frac{\lambda}{2}\tilde{f}_\epsilon(\mathbf{k})(C-iS), \qquad (4.15)$$

$$A_t^{-}(f)=A_{F,t}^{-}(f)\otimes\mathbb{1}+\mathbb{1}\otimes\frac{\lambda}{2}\tilde{f}_\epsilon^*(\mathbf{k})(C+iS), \qquad (4.16)$$

and the field operator is

$$A_i^{\pm}(\mathbf{r},t)=A_{F,i}^{\pm}(\mathbf{r},t)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{w/2}\times\frac{\exp[\mp i(\mathbf{k}\cdot\mathbf{r}-k_0t)]}{k_0}(\epsilon\cdot e_i)(C\mp iS). \qquad (4.17)$$

This expression (4.17) is very similar to formula (3.15). However, one must keep in mind that $\mathbf{A}_F(\mathbf{r},t)$ is no longer a free field, but verifies the coupled equation (4.1). This ensures that $\mathbf{A}(\mathbf{r},t)$ given by (4.17) also verifies (4.1).

The particle number and its associated phase operator are still defined through (3.16) and (3.17). The major difference between the free and the interacting case concerns the cyclic vector, specially on the question of the phase properties.

More precisely, let us calculate the mean value

$$\langle\Omega^{t_0}|e^{i\Phi}|\Omega^{t_0}\rangle=\langle\chi_{\mathbf{k},\epsilon}^{t_0}|(C+iS)|\chi_{\mathbf{k},\epsilon}^{t_0}\rangle=\exp(i\varphi)\frac{I_1(\lambda|g_\epsilon(\mathbf{k},t_0)|)}{I_0(\lambda|g_\epsilon(\mathbf{k},t_0)|)} \qquad (4.18)$$

where $I_n(x)$ are the real Bessel functions of complex argument which are defined by

$$I_n(x)=(-i)^nJ_n(ix), \quad x\in\mathbb{R},$$

and whose asymptotic behaviour is

$$I_n(x)\sim(2\pi x)^{-1/2}e^x \quad \forall n, \text{ when } x\to+\infty. \qquad (4.19)$$

The relation (4.18) shows that the phase is no longer equally distributed as it was in the free case. This is due to the fact that the function $\chi_{\mathbf{k},\epsilon}^{t_0}(\theta)$ has a maximum for $\theta=\varphi$. The larger $\lambda|g_\epsilon(\mathbf{k},t_0)|$ the more important is the peak. When $\lambda|g_\epsilon(\mathbf{k},t_0)|$ tends to infinity, the ratio I_1/I_0 tends to one, according to (4.19); so

$$\langle\Omega^{t_0}|e^{i\Phi}|\Omega^{t_0}\rangle\to\exp(i\varphi) \text{ when } \lambda|g_\epsilon(\mathbf{k},t_0)|\to\infty. \qquad (4.20)$$

Since $\exp(i\Phi)$ is unitary, this implies that the corresponding dispersion tends to zero.

We consider now the correlation functions of the state $\omega_{\mathbf{k},\epsilon}^{t_0}$. As the phase in this interacting case is not equally distributed, the mean value of the field operator doesn't vanish:

$$\langle\Omega^{t_0}|A_i^{-}(\mathbf{r},t)|\Omega^{t_0}\rangle=g_i(\mathbf{r},t)+\sqrt{w/2}\,\frac{\exp[i(\mathbf{k}\cdot\mathbf{r}-k_0t)]}{k_0}(\epsilon\cdot e_i)\times\exp(i\varphi)\frac{I_1(\lambda|g_\epsilon(\mathbf{k},t_0)|)}{I_0(\lambda|g_\epsilon(\mathbf{k},t_0)|)}. \qquad (4.21)$$

The ratio $I_1(\lambda|g_\epsilon(\mathbf{k},t_0)|)/I_0(\lambda|g_\epsilon(\mathbf{k},t_0)|)$ is zero for $\lambda|g_\epsilon|=0$, hence we recover the free field situation for $\mathbf{J}=0$. This ratio approaches one in the limit $\lambda|g_\epsilon|\to\infty$. In this case, (4.21) appears like the superposition of the classical field $\mathbf{g}(\mathbf{r},t)$ radiated by the current distribution with a plane wave of momentum $\mathbf{k}$, energy density w and whose phase φ is completely determined by the time parameter t_0 (more precisely, φ is the phase of the classical radiation field at time t_0).

The n-point correlation functions no longer factorize except in the limit $\lambda|g_\epsilon|\to\infty$. Moreover, the odd functions are different of zero. Namely,

$$\langle\Omega^{t_0}|\prod_{l=1}^{p}A^+_{i_l}(x_l)\prod_{m=1}^{q}A^-_{i_m}(x_m)|\Omega^{t_0}\rangle \xrightarrow[\lambda|g_\epsilon|\to\infty]{} \prod_{l=1}^{p}\mathcal{G}^*_{i_l}(x_l)\prod_{m=1}^{q}\mathcal{G}_{i_m}(x_m) \tag{4.22}$$

with

$$\mathcal{G}_i(x)=g_i(x)+\sqrt{w/2}\,\frac{\exp(ik\cdot x)}{k_0}(\epsilon\cdot e_i)\exp(i\varphi). \tag{4.23}$$

The state we have constructed exhibits nontrivial phase properties. It was already clear in Sec. 3 that one can build a lot of such states with phase properties, more precisely states where the phase is not equally distributed. The interesting feature of the last procedure is that this phase distribution comes from nontrivial physical situations. Indeed, one can, for example, use the previous analysis to devise a simple phenomenological model of laser above threshold in the spirit of Ref. 19, where the external current is the electronic current in the laser cavity. One can also use the previous results to build a simple quantum model of interferences far from sources. We shall come back to these applications in a forthcoming paper.

5. CONDENSATION AT $k=0$: THE STATIC FIELD

In Sec. 3, the momentum $\mathbf{k}$ was strictly different of zero. Indeed, starting with N photons of momentum $\mathbf{k}=0$ in a box, we could not recover a finite density of energy in the thermodynamical limit. However, a uniform static field is actually a physical situation where there exists a finite density of energy and whose quantum description needs an infinite number of photons of momentum $\mathbf{k}$ strictly equal to zero. So, it seems that such a situation can be only obtained in considering the limit $\mathbf{k}=0$ after having performed the thermodynamical limit on a finite free system with photons of momentum $\mathbf{k}\neq 0$.

We shall accomplish this procedure on the electromagnetic tensor $F_{\mu\nu}$ of the representation in $\mathcal{H}$, which is explicitly built from the field operators (3.15):

$$\begin{aligned}F_{\mu\nu}(x)&=\partial_\mu A_\nu(x)-\partial_\nu A_\mu(x)\\&=F_{F,\mu\nu}(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{2w}\,(\sin(k\cdot x)C+\cos(k\cdot x)S)\\&\quad\times((\epsilon\cdot e_\mu)\alpha_\nu-(\epsilon\cdot e_\nu)\alpha_\mu)\end{aligned} \tag{5.1}$$

where

$\alpha_\nu=k_\nu/k_0,\quad \nu=0,1,2,3.$

Then, we take the limit $\mathbf{k}=0$ in (5.1), keeping constant the four quantities α_ν.

It becomes

$$F^{(0)}_{\mu\nu}(x)=F_{F,\mu\nu}(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{2w}\,((\epsilon\cdot e_\mu)\alpha_\nu-(\epsilon\cdot e_\nu)\alpha_\mu)S. \tag{5.2}$$

We derive now the electric and magnetic field operators

$$\begin{aligned}E_i^{(0)}(x)&=F^{(0)}_{0i}(x)\\&=E_{F,i}(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{2w}((\epsilon\cdot e_i)\alpha_0-(\epsilon\cdot e_0)\alpha_i)S,\end{aligned} \tag{5.3}$$

$$\begin{aligned}B_i^{(0)}(x)&=\tfrac12\epsilon_{ijk}F^{(0)}_{kj}(x)=B_{F,i}(x)\otimes\mathbb{1}+\mathbb{1}\\&\quad\otimes\epsilon_{ijk}\sqrt{2w}\,((\epsilon\cdot e_k)\alpha_j-(\epsilon\cdot e_j)\alpha_k)S.\end{aligned} \tag{5.4}$$

In the radiation gauge, the physical meaning of formulas (5.3) and (5.4) is more apparent. Indeed, we can rewrite in this gauge:

$$\mathbf{E}(x)=\mathbf{E}_F(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{2w}\boldsymbol{\epsilon}S, \tag{5.5}$$

$$\mathbf{B}(x)=\mathbf{B}_F(x)\otimes\mathbb{1}+\mathbb{1}\otimes\sqrt{2w}(\boldsymbol{\epsilon}\wedge\boldsymbol{\alpha})S. \tag{5.6}$$

The $\mathcal{M}$ components of the operators $\mathbf{E}(x)$ and $\mathbf{B}(x)$ keep the features of a plane wave which propagates in the $\boldsymbol{\alpha}$ direction with the polarization $\boldsymbol{\epsilon}$.

The averages in $|\Omega\rangle$ of the density energy operator and of the density momentum operator are easily calculated and we obtain the expected values

$$\langle\Omega|\mathbf{E}^+(x)\cdot\mathbf{E}^-(x)+\mathbf{B}^+(x)\mathbf{B}^-(x)|\Omega\rangle=w, \tag{5.7}$$

$$\langle\Omega|\mathbf{E}^+(x)\wedge\mathbf{B}^-(x)-\mathbf{B}^+(x)\wedge\mathbf{E}^-(x)|\Omega\rangle=w\boldsymbol{\alpha}. \tag{5.8}$$

Further, the mean values $\langle\Omega|\mathbf{E}(x)|\Omega\rangle$ and $\langle\Omega|\mathbf{B}(x)|\Omega\rangle$ are clearly equal to zero.

These last results show that the representation in $\mathcal{H}=\mathcal{H}_F\otimes\mathcal{M}$ with cyclic vector $|\Omega\rangle$ and field operator deduced from (5.2) is the good one to describe a uniform static electromagnetic field.

*Equipe de Recherche Associée au C.N.R.S.
†Postal Address: Physique Théorique, I.M.S.P. Parc Valrose, 06034 Nice Cedex, France.

[1] F. Rocca and M. Sirugue, Commun. Math. Phys. **34**, 111 (1973).
[2] M. Sirugue and A. Verbeure, "On the quantum analogue of the classical canonical transformation," Preprint Universiteit Leuven (1972).
[3] P. Carruthers and M.M. Nieto, Rev. Mod. Phys. **40**, 411 (1968).
[4] H.C. Volkin, J. Math. Phys. **14**, 1965 (1973).
[5] M. Viano and F. Rocca, Nuovo Cimento **20B**, 133 (1974).
[6] R.P. Feynman *Lectures on Physics* (Addison Wesley, Reading, Mass., 1965), Vol. III.
[7] M.M. Nieto, Phys. Rev. **167**, 416 (1968).
[8] P.A.M. Dirac, Proc. Roy. Soc. (London) **A114**, 243 (1927).
[9] D. Kastler, *Introduction à l'electrodynamique quantique* (Dunod, Paris, 1961).
[10] J. Manuceau, Ann. Inst. Henri Poincaré **VIII** (2), 139 (1968).
[11] J. Manuceau, M. Sirugue, D. Testard, and A. Verbeure, Commun. Math. Phys. **32**, 231 (1973).

[12]J.M. Chaiken, Commun. Math. Phys. **8**, 164 (1968).
[13]H. Araki and E.J. Woods, J. Math Phys. **4**, 637 (1963).
[14]R.J. Glauber, Phys. Rev. **130**, 2529 (1963).
[15]R.J. Glauber, Phys. Rev. **131**, 2766 (1963).
[16]C.L. Mehta, E. Wolf, and A.P. Balachandran, J. Math. Phys. **7**, 133 (1966).
[17]N. Kroll, in *Optics and Quantum Electronics* (Les Houches, 1964), edited by C. De Witt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965).
[18]The cyclicity of $|\Omega^{t_0}\rangle$ follows, as in the free case, from the fact that $\mathbb{1}\otimes(C+iS)$ belongs to the algebra (see Ref. 13) and that $|\Omega_F\rangle$ and $|\chi^{t_0}\rangle$ are cyclic in respectively $\mathcal{H}_F$ and $\mathcal{H}$. That property comes for $|\chi^{t_0}\rangle$ from the decomposition:

$$\chi^0=\sum_n a_n(C+iS)^n\chi^{t_0}$$

with

$$a_n=i^n\exp(-in\varphi)J_n(i\tfrac{1}{2}\lambda|g_\epsilon|)[J_0(i\lambda|g_\epsilon|)]^{1/2}.$$

[19]R.J. Glauber, in *Optics and Quantum Electronics* (Les Houches, 1964), edited by C. De Witt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965).

The semiclassical fermion μ-space density in three dimensions

N. L. Balazs and G. G. Zipfel Jr.*

Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11790
(Received 25 June 1974)

An approximation is constructed to the phase-space, or Wigner, distribution function for a three-dimensional, dense Fermi gas in a spherically symmetric potential well. Near the surface separating the classically forbidden region from the classically allowed region, quantum oscillations occur. The oscillations are expressed in terms of a universal function.

I. INTRODUCTION

In paper I[1] we investigated the Wigner transform $f(p, q)$ of the density matrix at absolute zero for a collection of N independent fermions in an external, one-dimensional potential field. The following results were established. Consider the function f displayed in the (two-dimensional) p-q space which can be thought of as the semiclassical μ-space; then

(1) The classical limit of f gives the well-known Fermi–Dirac step-function distribution, $f=1/2\pi\hbar$, $H<E_F$; $f=0$, $H>E_F$, where E_F is the Fermi energy and H the classical one-particle Hamiltonian.

(2) In the semiclassical approximation there are undulations superimposed on the step function. The extrema of these undulations lie on constant energy curves which are separated from each other by a characteristic energy $\epsilon(E)$; the last and largest maximum occurs below the Fermi energy, at $E_F - 2.33\,\epsilon(E_F)$. At the Fermi energy, f is 1/3 of its average interior value, and it drops exponentially to zero for energies larger than E_F. The characteristic energy $\epsilon(E)$ can be expressed in two approximately equivalent ways. If the Hamiltonian exhibits two turning points, we obtain $\epsilon(E)=\hbar\omega(E)(J(E)/\hbar)^{1/2}$, where $J(E)$ is the classical action and $\omega(E)$ the classical frequency associated with the energy E. If the system has one turning point, we obtain $\hbar^{2/3}/(mV_E')^{1/3}$, where V_E' is the derivative of the potential evaluated at the classical turning point associated with the energy E. The functional forms are given as follows. If there is only one turning point [I, Eq. (8)],

$$f(p, q)=\frac{1}{2\pi\hbar}\int_{(H-E_F)/\epsilon}^{\infty} dt\,\mathrm{Ai}(t) \tag{1}$$

where $\mathrm{Ai}(t)$ is the Airy function of argument t, and

$$\epsilon=(\hbar^{2/3}/2m^{1/3})V'^{2/3}.$$

If there are two turning points [I. Eq. (21)],

$$f(p, q)=\frac{(-1)^N}{2\pi\hbar}\int_{4J(H)/2\pi\hbar}^{\infty} dt\,\exp(-t/2)L_N^{(1)}(t), \tag{2}$$

$J(H)=$ classical action with E replaced by $H(p, q)$.

The functional forms immediately show that at a given point p, q, f exhibits a branch point as a function of $\hbar$, for $\hbar=0$. This enables one to interpret the undulations around the sharp edge of the classical distribution as a diffraction pattern in the μ-space.

In the present note we shall generalize these results for three-dimensional spherically symmetrical potentials, and show that all the previous results carry over, both qualitatively and quantitatively. The simplicity of our results is due to the fact that the Wigner transform of the angular part of the density matrix associated with the total angular momentum $\sim\hbar(l+1/2)$ reduces to a high degree of approximation to the delta function $\delta(\hbar(l+1/2)-p_\perp q)$ (see Eq. 7a), where $p_\perp$ is the magnitude of the momentum perpendicular to the coordinate vector $\mathbf{q}$ which has the magnitude of q.

II. THE SETTING OF THE PROBLEM

Let $\langle\mathbf{x}|\rho|\mathbf{x}'\rangle$ be the singlet density matrix in the coordinate representation of a collection of N independent fermions immersed in the spherically symmetrical potential V, the system being in its ground state at absolute zero. The Wigner function associated with the singlet density matrix is given as

$$f(\mathbf{p},\mathbf{q})=\frac{1}{(2\pi)^3}\int d^3z\langle\mathbf{q}+\mathbf{z}/2|\rho|\mathbf{q}-\mathbf{z}/2\rangle\exp(-i\mathbf{pz}), \tag{3}$$

where $\mathbf{q}$ and $\mathbf{p}$ are vectors; $\mathbf{q}$ and $\mathbf{z}$ are defined through the transformation

$$(\mathbf{x}+\mathbf{x}')/2=\mathbf{q},$$

$$\mathbf{x}-\mathbf{x}'=\mathbf{z}.$$

The units are so chosen that $m=1$, $\hbar=1$.

In the present case, ρ is defined as

$$\langle\mathbf{x}|\rho|\mathbf{x}'\rangle=\sum_{n,l,m}\frac{\chi_{nl}(r)}{r}\frac{\chi_{nl}(r')}{r'}Y_{lm}(\theta,\phi)Y_{lm}(\theta',\phi') \tag{4}$$

where the summation extends over all quantum numbers nlm for which $E_{nl}<E_F$; $\mathbf{x}$, $\mathbf{x}'$ have the coordinates (r,θ,ϕ), (r',θ',ϕ') in spherical polars. The wavefunction associated with the energy E_{nl} is given by $(\chi_{nl}(r)/r)Y_{lm}(\theta,\phi)$; n is the *radial* quantum number (usually denoted by n_r); l, m are the azimuthal and magnetic quantum numbers; $Y_{lm}(\theta,\phi)$ are normalized spherical harmonics; χ_{nl} is normalized as $\int^\infty\chi_{nl}^2\,dr=1$.

The summation over m can immediately be performed using the addition formula

$$P_m(\cos\theta)P_n(\cos\theta')+2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}$$

$$\times P_n^m(\cos\theta)p_n^m(\cos\theta')\cos m(\phi-\phi')=P_n(\cos\Theta)$$

where Θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{x}'$, with angular coordinates θ, ϕ and θ', ϕ'.

This way we find

$$\langle\mathbf{x}|\rho|\mathbf{x}'\rangle=\sum_n{}'\sum_l{}'\frac{(l+1/2)}{2\pi}\frac{\chi_{nl}(r)}{r}\frac{\chi_{nl}(r')}{r'}P_l(\cos\Theta). \tag{5}$$

To induce Wigner's transformation on (5) we must express r, r', and Θ as functions of the vectors $\mathbf{q}$ and $\mathbf{z}$. In this we are greatly facilitated by observing that in the lowest approximation only small values of z ($<q$) will be of importance, if we neglect zero angular momentum

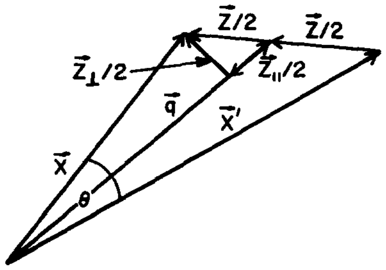

FIG. 1. Transformation from $\mathbf{x},\mathbf{x}'$ to $\mathbf{q},\mathbf{z}$.

states. If we restore h in the exponential, we see that only z values of order h/p will contribute significantly, while we expect from the uncertainty relations that $q > h/p \sim z$. The omission of the zero angular momentum states will introduce an error for small values of q where these states contribute exclusively to the density. For large values of q their contribution is small compared to the nonzero angular momentum states. In what follows we can consider the $l=0$ states omitted from the sum; as we shall see later it does not matter, since for $p \neq 0$ the $l=0$ terms will not contribute anyway.

For small values of $\mathbf{z}$, r, r', and Θ are simple functions of $\mathbf{q}$ and $\mathbf{z}$. Introduce $z_\parallel$ and $z_\perp$, the magnitudes of the vector-projections of $\mathbf{z}$ along and perpendicular to $\mathbf{q}$. Then for small values of $\mathbf{z}$, $r \approx q + z_\parallel/2$, $r' \approx q - z_\parallel/2$, $\Theta \approx z_\perp/q$, as is immediately evident from Fig. 1. Introduce similarly $p_\parallel$ and $p_\perp$ as the magnitudes of the vector-projections of $\mathbf{p}$ along and perpendicular to $\mathbf{q}$, and ϕ, the angle between the perpendicular vector-projections of $\mathbf{z}$ and $\mathbf{p}$ (Fig. 2). This enables us to write $\mathbf{p}\cdot\mathbf{z}$ as $p_\parallel z_\parallel + p_\perp z_\perp \cos\phi$. Finally, we observe that for small Θ and $l \neq 0$, $P_l(\cos\Theta) \approx J_0[(l+1/2)\Theta]$.

From (5) and (4) we get

$$\begin{aligned} f(p,q) &= \frac{1}{(2\pi)^3} \sum_n{}' \sum_l{}' \int_{-\infty}^{\infty} dz_\parallel \int_0^{\infty} dz_\perp z_\perp \int_0^{2\pi} d\phi \frac{(l+1/2)}{2\pi} \\ &\quad \times J_0[(l+1/2)z_\perp/q] \exp(-i(p_\parallel z_\parallel + p_\perp z_\perp \cos\phi)) \\ &\quad \times \frac{\chi_{nl}(q+z_\parallel/2)\chi_{nl}(q-z_\parallel/2)}{q^2} = \frac{1}{(2\pi)^3}\sum_n{}'\sum_l{}' \\ &\quad \times \left(\int_{-\infty}^{\infty} dz_\parallel\, e^{-ip_\parallel z_\parallel} \chi_{nl}(q+z_\parallel/2)\chi_{nl}(q-z_\parallel/2)\right) \\ &\quad \times \frac{(l+1/2)}{q^2} \int_0^{\infty} dz_\perp z_\perp J_0[(l+1/2)z_\perp/q] \\ &\quad \times \int_0^{2\pi} \frac{d\phi}{2\pi} \exp(-ip_\perp z_\perp \cos\phi). \end{aligned} \tag{6}$$

Here two further approximations have been made. In the denominator we approximated rr' by q^2, and replaced the q dependent limits of the z integrations by $\pm\infty$. In principle the z integrations must take into account that r is defined for positive values only. However, to the extent that our expansion in the smallness of z makes sense the results must be insensitive to the finiteness of these limits, and can be replaced by $\pm\infty$.

The ϕ integration can be performed using the formula $\int_0^{2\pi} \exp(-it\cos\phi)\, d\phi/2\pi = J_0(t)$, where $J_0(t)$ is the Bessel function of order zero. The $z_\perp$ integration can now be performed using the formula

$$\int_0^\infty dt\, t J_0(\lambda t) J_0(\lambda' t) = \delta(\lambda - \lambda')/\sqrt{\lambda\lambda'}$$
$$(\text{with } \lambda = (l+1/2)/q,\ \ \lambda' = p),$$

which expresses the completeness of the Bessel function J_0 in the Fourier–Bessel series. This way we obtain our basic expression

$$f(\mathbf{p},\mathbf{q}) = \frac{1}{(2\pi)^3} \sum_{nl}{}' F_{nl}(q,p_\parallel)\delta(l+1/2-p_\perp q) \tag{7a}$$

with

$$F_{nl}(q,p_\parallel) = \int_{-\infty}^{\infty} dz_\parallel \exp(-ip_\parallel z_\parallel)\chi_{nl}(q+z_\parallel/2)\chi_{nl}(q-z_\parallel/2). \tag{7b}$$

(The other factors which appear in the sum cancel if we replace them with those values which are allowed by the delta function.)

In the subsequent sections we shall evaluate (6) using various approximations in (7b).

III. THE CLASSICAL APPROXIMATION

Approximate $\chi_n(q)$ by normalized WKB wavefunctions

$$\chi_{n,l}(q) = \left(\frac{2}{\pi}\frac{\partial E_{nl}}{\partial n}\right)^{1/2} \left(\frac{1}{k_\parallel(q)}\right)^{1/2} \sin(S_\parallel(q) + \pi/4),$$

with

$$S_\parallel(q) = \int_q^{q_1} k_\parallel(\eta)\, d\eta, \quad k_\parallel(\eta)^2 = 2\left(E_{nl} - V(q) - \frac{(l+1/2)^2}{2q^2}\right). \tag{8}$$

(Thus q_1 is the outer turning point, while E_{nl} is specified by the Bohr–Sommerfield quantization rules.)

Then,

$$\chi_{nl}(q+z_\parallel/2)\chi_{nl}(q-z_\parallel/2) = \frac{1}{\pi}\frac{\partial E_{nl}}{\partial n}\frac{1}{k_\parallel(q)}\cos(k_\parallel q). \tag{9}$$

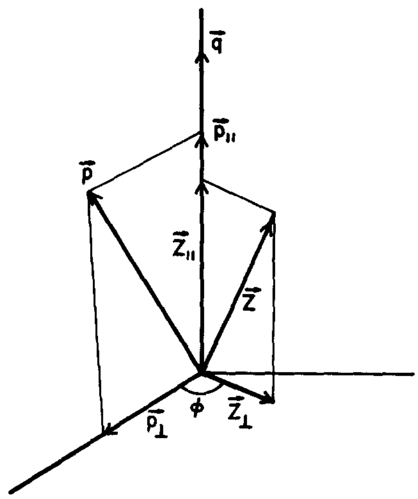

FIG. 2. Coordinate system used for the approximate evaluation of the Wigner function.

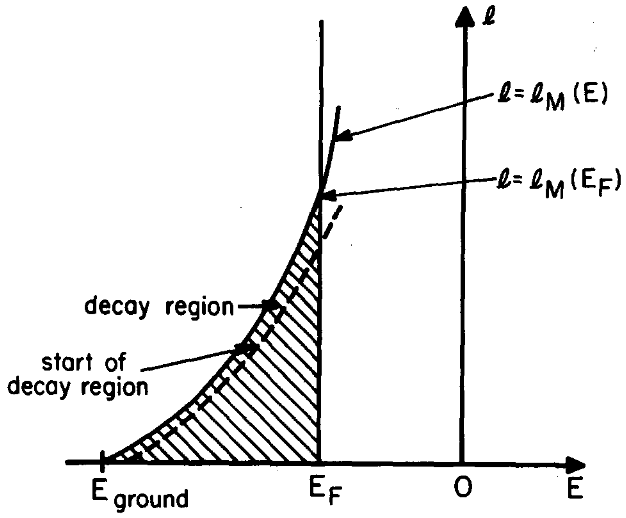

FIG. 3. The cross-hatched region gives the region of integration. The upper boundary is the boundary between the classically permitted and excluded regions.

Inserting this in (7b), linearizing $S_{\parallel}(q)$, and neglecting fast oscillating terms in q, we obtain

$$F_{nl}(q,p)=\frac{\partial E_{nl}}{\partial n}\delta[\tfrac{1}{2}p_{\parallel}^2-\tfrac{1}{2}k_{\parallel}^2(q)] \tag{10}$$

giving (7a) as

$$f_{\mathrm{WKB}}(p,q)=\sum_{nl}\frac{1}{(2\pi)^3}\delta(\tfrac{1}{2}p_{\parallel}^2-\tfrac{1}{2}k_{\parallel}^2)\,\delta(l+\tfrac{1}{2}-p_{\perp}q). \tag{11}$$

Replace the sums with integrals and define the limits of the integral so that $E_n<E_F$. Since χ_{nl} is zero in the classically excluded region, q should be restricted to lie in the classically permitted region. The first restriction is accomplished by inserting the step function $\theta(E_F-E)$, while the second restriction is secured by inserting the step function $\theta[E-(l+1/2)/2q^2-V(q)]$, which makes $k_{\parallel}$ positive or zero; $\theta(t)=1,\ t>0;\ \theta(t)=0,\ t<0$. The l integration can be immediately performed, resulting in the substitution of $p_{\perp}q$ for $(l+1/2)$ in the integrand. We obtain this way

$$\begin{aligned}f_{\mathrm{WKB}}(p,q)&=\frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}dE\,\delta[\tfrac{1}{2}p_{\parallel}^2-(E-\tfrac{1}{2}p_{\perp}^2-V)]\theta(E_F-E)\\&=\frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}dE\,\delta[H(p,q)-E]\theta(E_F-E)\\&=\frac{1}{(2\pi)^3}\theta[E_F-H(p,q)],\end{aligned} \tag{12}$$

which is the usual Thomas-Fermi result, $H(p,q)$ being the classical Hamiltonian.

IV. AIRY APPROXIMATION

The WKB approximation used in (8) breaks down at the classical turning points. However, in the vicinity of a turning point we may use, after Langer and others, an Airy type approximation for the wave function; this gives the WKB approximation far from the turning point and goes smoothly through the turning point itself. This approximation is given by

$$\chi_{nl}(q)=\left(\frac{\partial E_{nl}}{\partial n}\right)^{1/2}\left(\frac{2}{k_{\parallel}}\right)^{1/2}(3S_{\parallel}/2)^{1/6}\,\mathrm{Ai}[(-3S_{\parallel}/2)^{2/3}] \tag{13}$$

normalized in such a manner as to give the WKB function in the interior [I. Eq. (3)].

The so resulting F_{nl} has been evaluated in I

$$F_{nl}=\frac{\partial E_{nl}}{\partial n}\sigma_{\parallel}\mathrm{Ai}\left[\sigma_{\parallel}\left(\tfrac{1}{2}p_{\parallel}^2+\tfrac{1}{2}\frac{(l+1/2)^2}{q^2}+V(q)-E\right)\right],$$
$$\sigma_{\parallel}=2^{5/3}(3S_{\parallel}/2)^{2/3}/k_{\parallel}^2. \tag{14}$$

[Compare with integrand in I. Eq. (7), substituting there for V, $V+(l+1/2)^2/2q^2$, and multiplying the result with $\partial E_{nl}/\partial n$ to change the normalization.]

If we substitute (13) into (7a) and change the summation into an integration, we find that

$$f(p,q)=\frac{1}{(2\pi)^3}\int dl\int dE\sigma_{\parallel}\mathrm{Ai}\left[\sigma_{\parallel}\left(\tfrac{1}{2}p_{\parallel}^2+\tfrac{1}{2}\frac{(l+1/2)^2}{q^2}+V-E\right)\right]$$
$$\times\delta(l+1/2-p_{\perp}q). \tag{15}$$

The integration is extended over the cross-hatched region in the E, l plane as given in Fig. 3. The boundary curve $l=l_n(E)$ gives the maximum angular momentum possible for a given E. These values can be easily found from a potential energy diagram (Fig. 4) in which we plot $V+(l+1/2)^2/2q^2$ against q for several values of l. Given an E line, that l will be the $l_M(E)$ associated with this E whose potential curve just touches the E line.

We proceed now to analyze the integrand and show (a) the slowly varying scale factor can be approximated by a constant and (b) the integrand is exponentially small around the $l=l_M(E)$ curve which enables us to extend the limits of the integration. Consider q, p, and l fixed, and investigate the variation of the integrand with E. The function $\mathrm{Ai}(t)$ is oscillatory for $t<0$ decaying exponentially for $t>0$; it exhibits a maximum near $t\approx-1$. As discussed in I (p. 142) $\sigma_{\parallel}$ is a slowly varying function

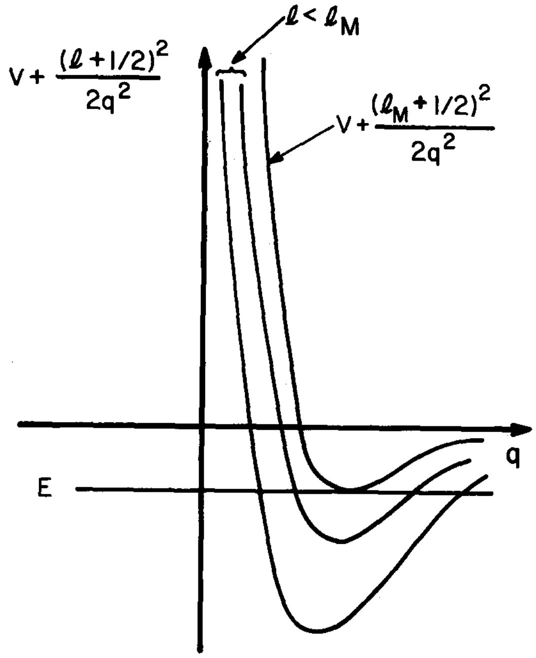

FIG. 4. Potential energy diagram defining l_M.

of E, hence we can replace it by its value at the maximum of Ai, denoted by $\sigma_{\parallel(1)}$. If we put the argument of the Ai function equal to zero, we find the equation of that curve which separates in the E, l plane the decay region from the rest. Considering $p_\parallel$ and q fixed, this gives an $l(E)$ curve which lies below the $l_M(E)$ curve since l_M gives the *maximum* value of l permitted for a given E. This curve is given by the broken line in Fig. 3. Similarly, the Airy function will be exponentially damped if the energy is below the ground state energy. Consequently, we can extend the domain of integration over the quadrant $E<E_F$, $l>0$ with a small error. Interchange the order of integration and introduce in place of E the new integration variable

$$t=\sigma_{\parallel(1)}\left(\tfrac{1}{2}p_\parallel^2+\frac{(l+1/2)^2}{2q^2}+V(q)-E\right).$$

We immediately obtain

$$f(p,q)=\frac{1}{(2\pi)^3}\int_{(H-E_F)\sigma_{\parallel(1)}}^{\infty}dt\,\mathrm{Ai}(t), \tag{16}$$

with $H(p,q)=\frac{1}{2}(p_\parallel^2+p_\perp^2)+V=\frac{1}{2}p^2+V$.

This is of the same form as the one obtained in I for the one-dimensional problem, replacing the one-dimensional weight $(2\pi)^{-1}$ with the three-dimensional one $(2\pi)^{-3}$. Thus the detailed analysis of this expression provided in I (and summarized in the Introduction) serves for the present case as well.

V. OSCILLATOR MAPPING

In (7b) the wavefunctions $\chi_{n,l}$ have two turning points if $l\neq 0$, and consequently their approximate form can be found by mapping the radial part of the Schrödinger equation onto the wave equation associated with the harmonic oscillator potential, i.e., the parabolic cylinder equation, (I, Sec. V).

The following differences arise if we compare the one-dimensional case discussed in I with the present problem. (a) In the one-dimensional case, the mapping function $z_n(x)$ (mapping the coordinate x of the actual equation onto the coordinate of the parabolic cylinder equation z) contained only the quantum number n, labeling the energy associated with the wavefunction to be mapped. Here the mapping function $\zeta_{nl}(r)$ depends both on the radial and azimuthal quantum numbers. (b) In the one-dimensional case, the Wigner transform introduced the variable q; here the q appearing in the argument of F_{nl} is the magnitude of the vector $\mathbf{q}$. Once these alterations are noted, the results of I immediately carry over and we find

$$F_{nl}(q,p_\parallel)=\sum_{nl}{}'' \omega_{nl}(\zeta'_{nl})^{-2}\frac{1}{4\pi^3}(-1)^n\exp\left[-2\left(\frac{2(H_l-E_n)}{(\zeta'_{nl})^2}\right)+n+1/2\right]\;L_n\left[4\left(\frac{2(H_l-E_n)}{(\zeta'_{nl})^2}+n+1/2\right)\right]$$

where $\omega_{nl}=\partial E_{nl}/\partial n$ and $H_l=\frac{1}{2}p_\parallel^2+[(l+1/2)^2/2q^2]+V(q)$.

The summations can be approximately evaluated using the following observations to simplify the limits. The function $\exp(-t/2)L_n(t)$ is oscillatory for small t and decays exponentially for large t. Between these two regions is the turning point region where the function can be approximated by an Airy function. Hence in this region the considerations of the previous section on l_M apply and we can extend the limits of the l sum from 1 to infinity, while the n sum can be formally converted into an integration over E from $-\infty$ to E_F. The l sum then simply substitutes $p_\perp q$ for $l+1/2$. This converts H into the Hamiltonian $\frac{1}{2}p^2+V(q)$. The n sum can now be approximately evaluated as in I and we finally obtain

$$f(p,q)=(-1)^N\frac{1}{(2\pi\hbar)^3}\int_{4J(H)/2\pi\hbar}^{\infty}dt\exp(-t/2)L_N^{(1)}(t), \tag{17}$$

where $J(H)$ is the classical action evaluated for that energy E which is the value of the Hamiltonian for the p, q point in question. [see Eq. (21) of I].

VI. CONCLUDING REMARKS

Two semiclassical approximations to the Wigner function in six-dimensional (p, q) space have been developed for the problem of fermions in a spherically symmetric potential well. Equations (16) and (17) are the final results of these approximations. Similarly to the one-dimensional case,[1] the Wigner function oscillates in the vicinity of the boundary separating the classically permitted and excluded regions of phase space. The extrema of the oscillations lie on constant energy surfaces. In the classically excluded region the Wigner function falls off exponentially. When the classical limit is taken the expressions (16) and (17) approach the Thomas-Fermi distribution.

In a recent article Lieb and Simon[2] have broken down the spatial density distribution into five regions. It is interesting to speculate that the transition region described by them can be described by our results. The behavior of the spatial density formed by integrating $f(\mathbf{p},\mathbf{q})$ over $\mathbf{p}$ can be quite complex.

ACKNOWLEDGMENT

We wish to acknowledge the National Science Foundation for its support of this work.

*Present address: Bell Telephone Laboratories, Whippany, N.J. 07981.

[1]N. L. Balazs and G.G. Zipfel, Jr., Ann. Phys. 77, 139 (1973).
[2]E.H. Lieb and B. Simon, Phys. Rev. Lett. 31, 681 (1973).

Baker–Campbell–Hausdorff formulas

R. Gilmore

Physics Department, University of South Florida, Tampa, Florida 33620
(Received 19 July 1974)

Baker–Campbell–Hausdorff formulas can be constructed simply by matrix multiplication. Examples are given.

I. INTRODUCTION

Baker[1]—Campbell[2]—Hausdorff[3] (BCH) formulas[4–8] provide a powerful and elegant tool for solving many problems[9–16] of physical interest. However, the complexity[17–19] of the usual expansions for $Z(X, Y)$, where $e^X e^Y = e^{Z(X,Y)}$, has prevented even more widespread application of these formulas. The operator $Z(X, Y)$ can be computed simply and in closed form by matrix multiplication whenever X and Y are operators in a finite-dimensional Lie algebra. The procedure appears to be not widely recognized. This is very surprising, since both Baker[20] and Hausdorff[21] pointed this out explicitly in their original works. In the following sections we give several examples of the matrix construction of BCH formulas.

II. EXAMPLES

A. Example 1

Within the context of the Foldy[22] model of a superfluid[23] system, the ground state wavefunction is a direct product of single mode states $\Pi|\psi_k\rangle$, where

$$|\psi_k\rangle = \exp[-\tfrac{1}{2}\theta(k)(b_k^\dagger b_{-k}^\dagger - b_k b_{-k})]\ |{}^j_0\rangle_k$$

and the operators are as defined by Solomon.[24] The bilinear products of boson operators $b_k^\dagger b_{-k}^\dagger$, $b_k b_{-k}$, $\frac{1}{2}(b_k^\dagger b_k + b_{-k}^\dagger b_{-k} + 1)$ obey $su(1,1)$ commutation relations,[9,24] and have a faithful 2×2 matrix representation: $b_k^\dagger b_{-k}^\dagger \to M_{12}$, $b_k b_{-k} \to -M_{21}$, $\frac{1}{2}(b_k^\dagger b_k + b_{-k}^\dagger b_{-k} + 1) \to \frac{1}{2}(M_{11} - M_{22})$, where M_{ij} is an $n\times n$ matrix with $+1$ at the intersection of row i and column j and 0 elsewhere, and with $n \geqslant \max(i, j)$. By simple matrix mutiplication it is possible to verify[25] the 2×2 matrix equation

$$\begin{aligned}\exp[-\tfrac{1}{2}\theta(M_{12}+M_{21})] &= \exp(-\tanh\tfrac{1}{2}\theta\, M_{12})\exp\{(-2\ln\cosh\tfrac{1}{2}\theta)\tfrac{1}{2}(M_{11}-M_{22})\}\\ &\quad\times\exp[\tanh\tfrac{1}{2}\theta\ (-M_{21})].\end{aligned}$$

This BCH formula is valid for the bilinear boson products (since it is valid for their faithful matrix representatives), and can be used to construct[26,27] the single mode ground states, which are coherent states[13,28]:

$$\begin{aligned}|{}^{j}_{\theta(k)}\rangle &= \exp[-\tanh\tfrac{1}{2}\theta(k)\, b_k^\dagger b_{-k}^\dagger]\ |{}^j_0\rangle\\ &\quad\times\exp\{[-2\ln\cosh\tfrac{1}{2}\theta(k)](-j)\}\\ &= \{\cosh\tfrac{1}{2}\theta(k)\}^{2j}\sum_{n=0}^{\infty}\left(\frac{n!\,\Gamma(n-2j)}{\Gamma(0-2j)}\right)^{1/2}\frac{[-\tanh\frac{1}{2}\theta(k)]^n}{n!}\ |{}^j_n\rangle,\end{aligned} \tag{1}$$

where $-2j = |\Delta_k| + 1 = 1, 2, \cdots$ describes an irreducible representation[24] of $SU(1,1)$ and the nonnegative integer $|\Delta_k| = |n_k - n_{-k}|$ is the difference between the number of bosons in modes k and $-k$.

B. Example 2

The number, creation, annihilation, and identity operators[29] $n = a^\dagger a$, $a^\dagger$, a, and $I = [a, a^\dagger]$ have a faithful non-Hermitian 3×3 matrix representation[30]: $\Gamma(n) = M_{22}$, $\Gamma(a^\dagger) = M_{23}$, $\Gamma(a) = M_{12}$, $\Gamma(I) = M_{13}$. The product $\exp\{Na^\dagger a\}\exp\{Ra^\dagger + La\}$ can be computed in this representation and is found to be

$$\mathcal{M} = \exp\{NM_{22}\}\exp\{RM_{23} + LM_{12}\} = \begin{bmatrix}1 & L & \frac{1}{2}LR\\ 0 & e^N & e^N R\\ 0 & 0 & 1\end{bmatrix}. \tag{2}$$

A similarity transformation can be performed on $\mathcal{M}$ which brings it to the form

$$S\mathcal{M}S^{-1} = \exp(NM_{22} + DM_{13}), \tag{3}$$

where $S = \exp(rM_{23} + lM_{12})$, $r = R(1 - e^{-N})^{-1}$, $l = L(1 - e^{+N})^{-1}$, and $D = -\frac{1}{2}RL\coth\frac{1}{2}N$.

Since this calculation is valid in a faithful 3×3 nonunitary group representation, it is valid in all representations. In particular, it is valid in the infinite-dimensional unitary representation carried by the Hilbert space spanned by the harmonic oscillator eigenstates $|n\rangle$, where $a^\dagger a|n\rangle = n|n\rangle$. In this representation,

$$\begin{aligned}\mathrm{Tr}[\exp(Na^\dagger a)\exp(Ra^\dagger + La)] &= \mathrm{Tr}[\exp(Na^\dagger a - \tfrac{1}{2}RL\coth\tfrac{1}{2}NI)]\\ &= [\exp(-\tfrac{1}{2}RL\coth(\tfrac{1}{2}N)](1 - e^N)^{-1}, \quad N < 0.\end{aligned} \tag{4}$$

For the thermodynamic average $\langle\exp(i\Delta k\cdot x)\rangle_{\mathrm{Th}}$ with respect to the Hamiltonian $H = \hbar\omega a^\dagger a$, we set $N = -\beta\hbar\omega$, $i\Delta kx = i\Delta k\ (\hbar/2m\omega)^{1/2}(a^\dagger + a)$ and we recover the Debye[31]—Waller[32] factor $\exp[-\frac{1}{2}(\Delta k)^2\langle(\Delta x)^2\rangle]$, where $\langle(\Delta x)^2\rangle = (\hbar/2m\omega)\coth\frac{1}{2}\beta\hbar\omega$ is obtained by considering $\langle\exp(i\Delta k\cdot x)\rangle$ as a thermodynamic generating function. The notation is standard.[11,12]

C. Example 3

With the notation as in Example 2,

$$\exp[\Gamma(Na^\dagger a + Ra^\dagger + La + DI)] = \begin{pmatrix}1 & L(e^N-1)/N & D + LR(e^N - 1 - N)/N^2\\ 0 & e^N & R(e^N-1)/N\\ 0 & 0 & 1\end{pmatrix}. \tag{5}$$

The product of two such group elements[30] is a group element and may be expressed in the form (5). In fact

$$\begin{aligned}&\exp[\Gamma(Na^\dagger a + Ra^\dagger + La + DI)]\\ &\quad\times\exp[\Gamma(\eta a^\dagger a + \rho a^\dagger + \lambda a + \delta I)]\\ &\qquad= \exp[\Gamma(N'a^\dagger a + R'a^\dagger + L'a + D'I)].\end{aligned} \tag{6}$$

The parameters N', R', L', D' can be computed simply by matrix multiplication. All the usual exponential operator products for the harmonic oscillator can be derived in this way.

The trace of the group element on the right-hand side of (6) can be computed in the infinite-dimensional unitary representation following the procedure described in Example 2:

$$\mathrm{Tr}[\exp(Na^\dagger a + Ra^\dagger + La + DI) \quad \exp(\eta a^\dagger a + \rho a^\dagger + \lambda a + \delta I)]$$
$$= e^{D''}(1 - e^{N+\eta})^{-1}, \quad \mathrm{Re}(N+\eta) < 0,$$

$$D'' = D + \delta - \left(\frac{LR}{N} + \frac{\lambda\rho}{\eta}\right) + \left(\frac{L}{N} - \frac{\lambda}{\eta}\right)\left(\frac{R}{N} - \frac{\rho}{\eta}\right)\frac{(e^N-1)(e^\eta-1)}{(e^N e^\eta - 1)}. \tag{7}$$

This generating function possesses all the expected symmetries and contains all the familiar generating functions as special cases or limits.

III. COMMENTS

1. The matrix construction of BCH formulas discussed by Baker[20] and Hausdorff[21]

$$e^X e^Y = e^Z \Rightarrow e^{\mathrm{Ad}X} e^{\mathrm{Ad}Y} = e^{\mathrm{Ad}Z} \tag{8}$$

is valid in the regular or adjoint[9] representation. It is generally possible to find smaller $n\times n$ faithful matrix representations $\Gamma^{(n)}$ of a Lie algebra ($n <$ dim algebra) in which the computation is easier. This procedure was used in all three examples above. Under certain conditions[33] this smaller representation may fail to define $Z(X, Y)$ at all. Under these conditions the adjoint representation must be used. When the operator $Z(X, Y)$ constructed in the adjoint representation is not defined uniquely by the (algebraic) matrix equations, continuity (topological) arguments can be used to define $Z(X, Y)$ uniquely.

2. Mutually commuting generators of nonsemisimple groups cannot generally be simultaneously diagonalized in a finite dimensional representation [e.g., $\Gamma(a^\dagger a) = M_{22}$, $\Gamma(I) = M_{13}$]. However, in a unitary representation acting on a Hilbert space representing physically realizable states, mutually commuting generators will always be simultaneously diagonalizable.

3. If g is a Lie algebra of $n\times n$ matrices with $R, S, T, U, \cdots \in g$ and if a is the $n\times 1$ column vector col(a_1, $a_2 \cdots, a_n$) and $a^\dagger$ is its Hermitian adjoint, where the operators a_i, $a_j^\dagger$ obey $[a_i, a_j^\dagger] = \delta_{ij}$, $[a_i, a_j] = 0 = [a_j^\dagger, a_i^\dagger]$, then it is an easy matter to show[34] that $[a^\dagger Ra, a^\dagger Sa] = a^\dagger[R, S]a$. The Lie algebra of bilinear operator products of the form $a^\dagger ga$ is isomorphic with the matrix algebra g itself. Thus, if the matrix equation $\exp(R)\exp(S) = \exp(U)$ is valid, so also is the operator equation $\exp(a^\dagger Ra) \times \exp(a^\dagger Sa) = \exp(a^\dagger Ua)$. This observation has been used to construct the angular momentum algebra $su(2)$ and its properties,[35] the representations of $SU(n)$,[36] coherent states for multilevel atomic systems,[14] and ground states for superfluid[27] and superconducting[37] systems.

4. If H (Hamiltonian) is an element in a finite-dimensional Lie algebra g spanned by generators X_i, $i = 1, 2, \cdots, n$, then $\mathrm{Tr}\, e^{-\beta H}\,\Pi \exp(\alpha^i X_i)$ is a thermodynamic generating function that can be used to compute all moments of the operators X_i, and expectation values of operator products in arbitrary order. The trace may be computed in a simple faithful finite-dimensional representation $\Gamma^f(\exp g)$; it is an analytic continuation of a character[38] function $\chi(f, \beta H, \alpha)$. In the Hilbert space on which H acts through $\Gamma^\lambda(H)$, $\mathrm{Tr}\, e^{-\beta H}\,\Pi\exp(\alpha^i X_i)$ is the analytic continuation of the character function $\chi(\lambda, \beta H, \alpha)$. When g is semisimple, $\chi(f, \beta H, \alpha)$ uniquely determines $\chi(\lambda, \beta H, \alpha)$; when g is compact, all characters are known.[39] For example, when $g = su(2)$, the first and second moments of $\ln\chi(j, \beta H, \alpha\cdot J)$ give the parallel[40] and transverse Brillouin functions when $[\beta H, \alpha\cdot J] = 0$ and $(\beta H, \alpha\cdot J) = 0$, respectively.[41]

IV. CONCLUSION

We have repeated an observation made over sixty years ago by Baker and Hausdorff: That is, that a BCH formula for $Z(X, Y)$ can be constructed simply by matrix multiplication in the adjoint representation when X and Y belong to a finite-dimensional Lie algebra. This construction has been illustrated by application to several examples of physical interest. Some implications and caveats were discussed.

[1] H. F. Baker, Proc. London Math. Soc (2) 3, 24 (1905).
[2] J. E. Campbell, Proc. London Math. Soc. (1) 28, 381 (1897), 29, 14 (1897).
[3] F. Hausdorff, Leipz. Ber. 58, 19 (1906).
[4] A linearized version of this formula was first studied by F. Schur, Math. Ann. 35, 161 (1890); Leipz. Ber. 42, 1 (1890). This formula was also studied by Lie, Engel, and Poincaré (Refs. 5—7). An accessible derivation is given by Helgason.[8]
[5] S. Lie, Leipz. Ber. 42, 485 (1890); *Transformations gruppen, Vol. III* (Leipzig, 1893; reprinted: Chelsea, New York, 1970), Sec. 144, pp. 788—800.
[6] F. Engel, Leipz. Ber. 43, 308 (1891); 44, 52 (1892).
[7] H. Poincaré, C.R. Acad. Sci. (Paris) 128, 1065 (1899); Cambridge Phil. Trans. 18, 220 (1899); Rend. Circ. Mat. Palermo 15, (1901); collected in *Oeuvres de Henri Poincaré* (Gauthier-Villars, Paris, 1934), Vol. III, pp. 169—321.
[8] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), cf. pp. 36—40.
[9] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[10] R. J. Glauber, Phys. Rev. 130, 2529 (1963); 131, 2766 (1963).
[11] G. H. Weiss and A. A. Maradudin, J. Math. Phys. 3, 771 (1962).
[12] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York, 1963).
[13] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A 6, 2211 (1972).
[14] F. T. Arecchi, R. Gilmore, and D. M. Kim, Lett. Nuovo Cimento 6, 219 (1973).
[15] W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill, New York, 1964).
[16] R. Engleman and P. Levi, J. Math. Phys. 4, 105 (1963).
[17] Ref. 1, p. 42; Ref. 2, pp. 15 *et seq.*; Ref. 3, Eq. 26, p. 30.
[18] E. B. Dynkin, Mat. Sb. 25, 155 (1949) (in Russian).
[19] N. Jacobson, *Lie Algebras* (Wiley, New York, 1962), pp. 170—74, Eqs. 39, 40.
[20] Ref. 1, pp. 25, 45—46.
[21] Ref. 3, p. 36.
[22] W. H. Bassichis and L. L. Foldy, Phys. Rev. 133, A935 (1964).
[23] N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947).
[24] A. I. Solomon, J. Math. Phys. 12, 390 (1971).
[25] For similar calculations, cf. Ref. 8, p. 316, Ref. 9, pp. 149—58, 461—62.
[26] For similar calculations, cf. Ref. 9, pp. 168—75, Ref. 13, Sec. 3 and Appendix A.

[27]R. Gilmore, Quart. J. Florida Acad. Sci. 36, 2 (1973).
[28]R. Gilmore, Rev. Méx. Fis. 23, 143 (1974).
[29]A. Messiah, *Quantum Mechanics, Vol. II* (North-Holland, Amsterdam, 1962).
[30]These are the infinitesimal generators for the linear Lie group described in Willard Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972), p. 205, Problem 5.3.
[31]P. Debye, Ann. Physik 43, 49 (1914); E. Schrödinger, Physik. Z. 15, 79, 497 (1914).
[32]I. Waller, Z. Physik 51, 213 (1928).
[33]These involve certain operator products in noncompact groups. Cartan's counterexample provides an illustration of such a failure. Cf. Ref. 9, pp. 118–19, exercise 12, also p. 389.
[34]Ref. 9, p. 505, Exercise 21.
[35]J. Schwinger, "On Angular Momentum," in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965).
[36]G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).
[37]R. Gilmore, "Coherent States in Modern Physics," in *Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf (Plenum, New York, 1973).
[38]M. Hamermesh, *Group Theory and Its Applications to Physical Problem* (Addison-Wesley, Reading, Mass., 1962).
[39]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, N. J., 1931).
[40]L. Brillouin, J. de Phys. 8, 74 (1927).
[41]The commutator and the Cartan–Killing metric are defined in Ref. 9, pp. 94 and 248, respectively.

On the existence of weakly retarded and advanced Green's functions*

Anton Z. Capri

Theoretical Physics Institute, Physics Department, University of Alberta, Edmonton, Alberta, Canada
(Received 14 June 1974)

By considering a model field equation that contains the acausal propagation features of a spin 3/2 field we show that depending on the "external field" one can either have weakly retarded fundamental solutions or not.

I. INTRODUCTION

It is by now clear from the work of Velo and Zwanziger[1] and Shamaly and Capri[2] that for a field of spin 3/2, acausal propagation must result if this field is coupled in a gauge-invariant manner to an externally applied electromagnetic field F. Thus, the characteristic determinant for a Rarita–Schwinger spin 3/2 field minimally coupled to an electromagnetic field contains a factor

$$n^2 + g^2(F^4{}_n)^2 = n_0^2(1 - g^2\mathbf{B}^2) - \mathbf{n}^2 - g^2[(\mathbf{n}\times\mathbf{E})^2 - (\mathbf{n}\cdot\mathbf{B})^2]. \tag{1}$$

Clearly in any region of space–time where $\mathbf{B}\neq 0$ acausal propagation occurs and an equation with such characteristics cannot have strictly retarded or advanced fundamental solutions. The term strictly retarded (advanced) is used to mean that the support of the fundamental solution is contained in the forward (backward) light cone.

It was this lack of strictly retarded (advanced) fundamental solutions that motivated Wightman[3] to introduce the concept of weakly retarded (advanced) fundamental solutions. He furthermore showed that if such weakly retarded (advanced) fundamental solutions exist then it is possible, in spite of the nonlocal nature of the field, to formulate a satisfactory scattering theory for the coupled field equation. It thus is of interest to study the existence of weakly retarded (advanced) fundamental solutions for equations with this acausality property. To initiate this study we have constructed a model equation which has these properties and is nevertheless exactly soluble.

II. THE MODEL

A field equation of the form

$$f(t,\mathbf{x})\frac{\partial^2\psi}{\partial t^2} + \frac{1}{2}f_t(t,\mathbf{x})\frac{\partial\psi}{\partial t} - (\nabla^2 - m^2)\psi = 0 \tag{2}$$

has the acausality properties mentioned previously. Here $f(t,\mathbf{x})$ may be thought of as $(1 - g^2\mathbf{B}^2)$. The term $f_t(t,\mathbf{x}) = (\partial/\partial t)f(t,\mathbf{x})$ is added to make the equation soluble. We furthermore assume that $f(t,\mathbf{x})$ is C^∞ and that $f(t,\mathbf{x}) \to 1$ as $t\to\pm\infty$ or $|\mathbf{x}|\to\infty$. This is analogous to assuming that $\mathbf{B}$ vanishes for large $|\mathbf{x}|$ and $|t|$. There are two separate cases to consider depending on whether $f(t,\mathbf{x}) > 0$ or else if there is a region for which $f(t,\mathbf{x}) < 0$. We consider these cases separately.

Case 1: $f(t,\mathbf{x}) > 0$

In this case Eq. (2) is strictly hyperbolic and we can further subdivide the problem according to whether $f(t,\mathbf{x}) \geq 1$ or not. The reason for this is that the normals to the characteristic surfaces n_μ are given by

$$f(t,\mathbf{x})\, n_0^2 - \mathbf{n}^2 = 0. \tag{3}$$

So for $f(t,\mathbf{x}) \geq 1$ the normals are spacelike or lightlike and the characteristic surfaces lie inside the light cones. Thus the elementary solutions in this case are strictly retarded or advanced. For $0 < f(t,x) \leq 1$ in some region we find that weakly retarded (advanced) fundamental solutions exist, but they are not strictly retarded (advanced).

Case 2: $f(t,\mathbf{x}) < 0$, $a(\mathbf{x}) < t < b(\mathbf{x})$

In this case the equation ceases to be hyperbolic in the region between the hyperplanes $t = a(\mathbf{x})$ and $t = b(\mathbf{x})$ and not even weakly retarded (advanced) fundamental solutions exist.

III. PROOF OF EXISTENCE OF WEAKLY RETARDED FUNDAMENTAL SOLUTIONS

Since the case $f(t,\mathbf{x}) \geq 1$ falls in the class of well-established results we assume specifically that $0 < f(t,\mathbf{x}) \leq 1$ such that $\lim_{t\to\infty} f(t,\mathbf{x}) = 1$. Then as stated before, the characteristic surfaces lie outside the light cones and are asymptotically parallel to the light cones.

We begin by recalling the definition of weakly retarded.[3] Since the argument for the weakly advanced fundamental solutions is identical to that for the weakly retarded, we restrict ourselves henceforth to the retarded case.

Definition: Let $g_1(x), g_2(x) \in \mathcal{S}^{(4)}$ the space of C^∞ functions of four variables of rapid decrease, then a fundamental solution $G_R(x;y)$ is weakly retarded if for every positive integer n and every vector l not in the future light cone

$$\left|\iint g_1(x)\, G_R(x+\tau l; y)\, g_2(y)\, d^4x\, d^4y\right| \leq \frac{C(g_1, g_2, n, l)}{1+\tau^n} \tag{4}$$

for some constant $C(g_1, g_2, n, l)$.

This is an asymptotic form of the notion of retarded. Clearly, G_R is weakly retarded if any vector not in the future light cone when extended to infinity eventually lies outside the support of G_R. We shall now show that this is indeed the case if $f(t,\mathbf{x}) > 0$ and $\lim_{t\to\infty} f(t,\mathbf{x}) = 1$.

To find the "retarded" elementary solution of

$$f(t,\mathbf{x})\frac{\partial^2 G}{\partial t^2} + \frac{1}{2}f_t(t,\mathbf{x})\frac{\partial G}{\partial t} - (\nabla^2 - m^2)G = \delta(t-t')\,\delta(\mathbf{x}-\mathbf{x}') \tag{5}$$

we begin by defining

$$\tau=\tau(t,\mathbf{x})\equiv\int_0^t \frac{ds}{[f(s,\mathbf{x})]^{1/2}}. \tag{6}$$

τ is then a monotone increasing function of t and therefore, since $f(t,\mathbf{x})$ is C^∞, has a unique C^∞ inverse h. Thus

$$t=h(\tau,\mathbf{x}). \tag{7}$$

Differentiating with respect to t yields

$$1=h_\tau\tau_t=h_\tau[f(t,\mathbf{x})]^{-1/2}$$

so that

$$\partial h/\partial\tau=h_\tau=[f(t,\mathbf{x})]^{1/2}. \tag{8}$$

Changing variables from t to τ Eq. (5) becomes

$$(\partial G/\partial\tau^2)-(\nabla^2-m^2)G=\delta[h(\tau,\mathbf{x})-h(\tau',\mathbf{x})]\,\delta(\mathbf{x}-\mathbf{x}') \tag{9}$$

or

$$(\partial^2G/\partial\tau^2)-(\nabla^2-m^2)G=[h_\tau(\tau,\mathbf{x})]^{-1}\,\delta(\tau-\tau')\,\delta(\mathbf{x}-\mathbf{x}').$$

The "retarded" solution is

$$G_R(\tau,\tau';\mathbf{x},\mathbf{x}')=[h_\tau(\tau,\mathbf{x})]^{-1}\,\Delta_R(\tau-\tau',\mathbf{x}-\mathbf{x}'),$$

where Δ_R is the usual Klein—Gordon retarded Green's function. Changing from τ,τ' back to t,t' we get

$$G_R(t,t';\mathbf{x},\mathbf{x}')=[f(t,\mathbf{x})]^{-1/2}\,\Delta_R\left(\int_{t'}^t \frac{ds}{[f(s,\mathbf{x})]^{1/2}},\ \mathbf{x}-\mathbf{x}'\right). \tag{10}$$

Choosing $t'=0$, $\mathbf{x}'=0$, the support of this G_R is the forward cone with vertex at the origin and defined by

$$\int_0^t \frac{ds}{[f(s,\mathbf{x})]^{1/2}}=|\mathbf{x}|. \tag{11}$$

As stated before, in order that G_R be weakly retarded it is sufficient that any point with $t=(1-\epsilon)a$, $|\mathbf{x}|=a$ $(0<\epsilon<1)$ should be outside the support of G_R for $a\to\infty$.

This is achieved if

$$\lim_{a\to\infty}\left([(1-\epsilon)\int_0^a \frac{ds}{[f(s,\mathbf{x})]^{1/2}}-a]\right)<0 \tag{12}$$

or equivalently if

$$\lim_{a\to\infty}\left([\frac{1}{a}\int_0^a \frac{ds}{[f(s,\mathbf{x})]^{1/2}}]\right)<\frac{1}{1-\epsilon}. \tag{13}$$

Using L'Hôpital's Rule this condition becomes

$$\lim_{a\to\infty}[f(a,\mathbf{x})]^{-1/2}<\frac{1}{1-\epsilon} \tag{14}$$

which is true since $\lim_{a\to\infty}f(a,\mathbf{x})=1$. Thus in this case we have a weakly retarded fundamental solution which, for $f(t,\mathbf{x})<1$, over any open domain, is clearly not a strictly retarded solution.

IV. LACK OF WEAKLY RETARDED SOLUTIONS

In this case we assume that $f(t,\mathbf{x})$ is negative over the hypervolume $a(\mathbf{x})<t<b(\mathbf{x})$ with *simple* zeroes on the hyperplanes $t=a(\mathbf{x})$ and $t=b(\mathbf{x})$. Thus

$$\begin{aligned} &f(t,\mathbf{x})>0 \quad \text{for} \quad t<a(\mathbf{x}) \quad \text{or} \quad t>b(\mathbf{x}),\\ &f(t,\mathbf{x})<0 \quad \text{for} \quad a(\mathbf{x})<t<b(\mathbf{x}),\\ &f(t,\mathbf{x})=0 \quad \text{for} \quad t=a(\mathbf{x}) \quad \text{or} \quad t=b(\mathbf{x}).\end{aligned} \tag{15}$$

In analogy to Eq. (6) we now define

$$\tau=\tau_-(t,\mathbf{x})=\alpha(\mathbf{x})+\int_{a(\mathbf{x})}^t \frac{ds}{[f(s,\mathbf{x})]^{1/2}},\ \text{for } t<a(\mathbf{x}), \tag{16}$$

$$\tau=\tau_0(t,\mathbf{x})=\alpha(\mathbf{x})+\int_{a(\mathbf{x})}^t \frac{ds}{[-f(s,\mathbf{x})]^{1/2}},\ \text{for } a(\mathbf{x})<t<b(\mathbf{x}), \tag{17}$$

$$\tau=\tau_+(t,\mathbf{x})=\beta(\mathbf{x})+\int_{b(\mathbf{x})}^t \frac{ds}{[f(s,\mathbf{x})]^{1/2}},\ \text{for } t>b(\mathbf{x}), \tag{18}$$

where

$$\beta(\mathbf{x})=\alpha(\mathbf{x})+\int_{a(\mathbf{x})}^{b(\mathbf{x})} \frac{ds}{[-f(s,\mathbf{x})]^{1/2}}. \tag{19}$$

And $\alpha(\mathbf{x})$ is determined by the value of τ at $t=0$. Thus if $a(\mathbf{x})<0\,b(\mathbf{x})$ we can choose

$$\alpha(\mathbf{x})=-\int_{a(\mathbf{x})}^0 \frac{ds}{[-f(s,\mathbf{x})]^{1/2}} \tag{20}$$

and then $\tau=0$ for $t=0$. Similarly if $0<a(\mathbf{x})$ we choose

$$\alpha(\mathbf{x})=-\int_{a(\mathbf{x})}^0 \frac{ds}{[f(s,\mathbf{x})]^{1/2}}$$

and for $0>b(\mathbf{x})$ we choose

$$\alpha(\mathbf{x})-\int_{a(\mathbf{x})}^{b(\mathbf{x})} \frac{ds}{[-f(s,\mathbf{x})]^{1/2}}-\int_{b(\mathbf{x})}^0 \frac{ds}{[(s,\mathbf{x})]^{1/2}}.$$

All this is really equivalent to choosing

$$\tau=\int_0^t \frac{ds}{[|f(s,\mathbf{x})|]^{1/2}}.$$

Thus the functions $\alpha(\mathbf{x})$, $\beta(\mathbf{x})$ are so chosen as to make $\tau(t,\mathbf{x})$ continuous in t with $\tau(0,\mathbf{x})=0$. Again $\tau_\pm$ and τ_0 are real, monotone C^∞ functions of t, in their respective domains of definition, and therefore have unique monotone C^∞ inverses. Thus

$$t=h^{(-)}(\tau,\mathbf{x}),\ \text{for } \tau<\alpha(\mathbf{x}), \tag{21}$$

$$t=h^{(0)}(\tau,\mathbf{x}),\ \text{for } \alpha(\mathbf{x})<\tau<\beta(\mathbf{x}), \tag{22}$$

$$t=h^{(+)}(\tau,\mathbf{x}),\ \text{for } \tau>\beta(\mathbf{x}). \tag{23}$$

Taking partial derivatives with respect to t yields

$$\frac{\partial}{\partial\tau}h^{(\pm)}(\tau,\mathbf{x})=h_\tau^{\pm}(\tau,\mathbf{x})=[f(t,\mathbf{x})]^{-1/2},\ \tau<\alpha(\mathbf{x}) \text{ or } \tau>\beta(\mathbf{x}), \tag{24}$$

$$\frac{\partial}{\partial\tau}h^{(0)}(\tau,\mathbf{x})=h_\tau(\tau,\mathbf{x})=[-f(t,\mathbf{x})]^{-1/2},\ \alpha(\mathbf{x})<\tau<\beta(\mathbf{x}). \tag{25}$$

The equation for the Green's function now becomes

$$\frac{\partial^2G_-}{\partial\tau^2}-(\nabla^2-m^2)G_-=(h_\tau^{(-)})^{-1}\,\delta(\tau-\tau')\,\delta(\mathbf{x}-\mathbf{x}'),\ \tau<\alpha(\mathbf{x}), \tag{26}$$

$$\frac{\partial^2G_0}{\partial\tau^2}+(\nabla^2+m^2)G_0=-(h_\tau^{(0)})^{-1}\,\delta(\tau-\tau')\,\delta(\mathbf{x}-\mathbf{x}'),\quad \alpha(\mathbf{x})<\tau<\beta(\mathbf{x}), \tag{27}$$

$$\frac{\partial^2G_+}{\partial\tau^2}-(\nabla^2-m^2)G_+=(h_\tau^{(+)})^{-1}\,\delta(\tau-\tau')\,\delta(\mathbf{x}-\mathbf{x}'),\ \tau>\beta(\mathbf{x}). \tag{28}$$

To obtain the matching conditions at $\tau=\alpha(\mathbf{x})$, $\tau=\beta(\mathbf{x})$ we integrate Eq. (5) from $t=a(\mathbf{x})-\epsilon$ to $t=a(\mathbf{x})+\epsilon$ [also

$b(\mathbf{x})-\epsilon$ to $b(\mathbf{x})+\epsilon]$ and take the limit $\epsilon\to 0$ for $t\neq t'$, $\mathbf{x}\neq\mathbf{x}'$. Recalling that the zeroes in $f(t,\mathbf{x})$ are simple we have that in the vicinity of $t=a(\mathbf{x})$ $[t=b(\mathbf{x})]$ that

$$f(t,\mathbf{x})\simeq[t-a(\mathbf{x})]f_t[a(\mathbf{x}),\mathbf{x}]$$

{respectively, $f(t,\mathbf{x})\simeq[t-b(\mathbf{x})]f_t[b(\mathbf{x}),\mathbf{x}]$}. This leads to the conditions

$$\lim_{t\to a(\mathbf{x})}[G_-(t,t';\mathbf{x},\mathbf{x}')-G_0(t,t';\mathbf{x},\mathbf{x}')]=0, \tag{29}$$

$$\lim_{t\to b(\mathbf{x})}[G_+(t,t';\mathbf{x},\mathbf{x}')-G_0(t,t';\mathbf{x},\mathbf{x}')]=0. \tag{30}$$

This in turn implies that

$$\lim_{\tau\to\alpha(\mathbf{x})}[G_-(\tau,\tau';\mathbf{x},\mathbf{x}')-G_0(\tau,\tau';\mathbf{x},\mathbf{x}')]=0, \tag{31}$$

$$\lim_{\tau\to\beta(\mathbf{x})}[G_+(\tau,\tau';\mathbf{x},\mathbf{x}')-G_0(\tau,\tau';\mathbf{x},\mathbf{x}')]=0. \tag{32}$$

It is implicitly understood here that t (respectively, τ) approach their limits from inside the domain of definition of the functions involved. Furthermore since the various G's diverge like reciprocal square roots at the zeroes of $f(t,\mathbf{x})$ [see the solutions Eqs. (34), (35), and (36)] the above relations do not imply the continuity of G.

We shall now show that the set of equations (26), (27), and (28) subject to the matching conditions (31) and (32) have no weakly retarded solution. The most general solutions of (26), (27), (28) are

$$G_-=cG_R+(1-c)G_A, \tag{33}$$

where

$$G_R(\tau,\tau';\mathbf{x},\mathbf{x}')=[f(t,\mathbf{x})]^{-1/2}\Delta_R(\tau-\tau',\mathbf{x}-\mathbf{x}'), \tag{34}$$

$$G_A(\tau,\tau';\mathbf{x},\mathbf{x}')=[f(t,\mathbf{x})]^{-1/2}\Delta_R(\tau-\tau';\mathbf{x}-\mathbf{x}'), \tag{35}$$

$$G_0(\tau,\tau';\mathbf{x},\mathbf{x}')=2\pi m[-f(t,\mathbf{x})]^{-1/2}\times\frac{K_1\{m[(\tau-\tau')^2+(\mathbf{x}-\mathbf{x}')^2]^{1/2}\}}{[(\tau-\tau')^2+(\mathbf{x}-\mathbf{x}')^2]^{1/2}}, \tag{36}$$

$$G_+=dG_R+(1-d)G_A, \tag{37}$$

where c, d are constants.

To obtain a weakly retarded solution requires that $c=d=1$. Otherwise we always have asymptotically a superposition of strictly advanced and retarded fundamental solutions. It is, however, impossible to achieve $c=1$ or $d=1$ since with say $c=1$ and the point $(\tau-\tau',\mathbf{x}-\mathbf{x}')$ in the past light cone G_- vanishes. On the other hand, $G_0\neq 0$ in any finite domain. This makes it impossible to achieve the conditions (29) or (31). Similarly with $d=1$ G_+ vanishes for any point $(\tau-\tau',\mathbf{x}-\mathbf{x}')$ in the past light cone and therefore for the same reason as before the solutions G_+ and G_0 cannot be matched according to (30) and (32). Thus not even weakly retarded fundamental solutions exist.

V. CONCLUSIONS

We have studied a differential operator of the form

$$f(t,\mathbf{x})\frac{\partial^2}{\partial t^2}+\frac{1}{2}f_t(t,\mathbf{x})\frac{\partial}{\partial t}-(\nabla^2-m^2)$$

where $f(t,\mathbf{x})$ is C^∞ and approaches 1 for large $|t|$ or $|\mathbf{x}|$. This operator contains the essential features of acausality expected in the wave operator for a spin 3/2 field coupled to an external electromagnetic field that vanishes for large $|t|$ or $|\mathbf{x}|$. The main propagation properties of this operator depend on $f(t,\mathbf{x})$ and fall into two categories expressed in terms of the fundamental solutions.

(1) If $f(t,\mathbf{x})>0$ and $f(t,\mathbf{x})\to 1$ for $t\to\pm\infty$ then weakly retarded and advanced fundamental solutions exist. These solutions are furthermore strictly retarded and advanced, respectively, if $f(t,\mathbf{x})\geq 1$.

(2) If for some open hypervolume $a(\mathbf{x})<t<b(\mathbf{x})$, $f(t,\mathbf{x})<0$ then not even weakly retarded or advanced fundamental solutions exist.

*Work supported in part by the National Research Council of Canada.

[1]G. Velo and D. Zwanziger, Phys. Rev. 186, 1337 (1969).
[2]A. Shamaly and A. Z. Capri, Ann. Phys. (N.Y.) 74, 503 (1972).
[3]A. S. Wightman, "Relativistic Wave Equations as Singular Hyperbolic Systems," Princeton University preprint.

The algebra and group deformations $I^m\,[SO(n)\otimes SO(m)] \Rightarrow SO(n,m)$, $I^m\,[U(n)\otimes U(m)] \Rightarrow U(n,m)$, and $I^m\,[Sp(n)\otimes Sp(m)] \Rightarrow Sp(n,m)$ for $1 \leqslant m \leqslant n$

Kurt Bernardo Wolf and Charles P. Boyer

Centro de Investigación en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México D. F., Mexico

(Received 14 September 1973; revised manuscript received 28 March 1974)

We discuss a class of deformations of the inhomogeneous classical algebras $i^m\ [k(n)\oplus k(m)]$ to $k(n,m)$ for $1\leqslant m\leqslant n$. This generalizes the previously known expansions $i\,k(n)\Rightarrow k(n,1)$. As the title indicates, this is done explicitly for the orthogonal, unitary, and symplectic cases. We construct the corresponding deformed groups $K(n,m)$ as multiplier representations on the space of functions over the rank m coset space $K(n-m)\backslash K(n)$. This method allows us to build a principal series of unitary representations of $K(n,m)$. The contractions of the deformed algebras and groups are considered.

I. INTRODUCTION

The concepts of expansion and deformation of a Lie algebra are being actively developed in the mathematical physics literature. The motivation in studying this subject is twofold: first, it is, loosely speaking, the inverse of contraction and second, it allows one to build dynamical algebras for systems whose symmetry algebras are realized as a set of operators acting on a definite homogeneous space.

The first deformations treated were $iso(n)\Rightarrow so(n,1)$,[1] $i[u(n)\oplus u(1)]\Rightarrow u(n,1)$,[2] and $i[sp(n)\oplus sp(1)]\Rightarrow sp(n,1)$,[3] and their noncompact versions,[4] which constitute a family shown by Gilmore[5] to have a rank 1 coset space in the Cartan decomposition.[6] These deformations were constructed by the use of an algorithm in which the noncompact generators are produced by commuting the Casimir operator of a classical algebra with an element of a normal Abelian algebra transforming as a vector—a rank 1 tensor. A second family was treated in Ref. 7, which uses the algorithm with an Abelian algebra transforming as a second-rank symmetric tensor under the classical algebra. This produced the deformations of representations of $i_2so\Rightarrow sl(n,\mathbb{R})$, $i_2u(n)\oplus u(1)\Rightarrow sl(n,\mathbb{C})\oplus u(1)$, and $i_2sp(n)\oplus sp(1)\Rightarrow sl(n,\mathbb{Q})\oplus sp(1)$, which can only be realized on a rank 1 homogeneous space. As an example, for the hydrogen atom system where the symmetry algebra is $so(4)$ and the homogeneous space is the 3-sphere projected out of momentum space, the first kind of deformation yields $so(4,1)$[8] as a dynamical algebra while the second kind yields $sl(4,\mathbb{R})$.[9]

In Sec. II we show that the first family of deformations can be generalized by considering an Abelian ideal transforming as a set of m orthogonal vectors and gives rise to the deformations[10] $i^m[so(n)\oplus so(m)]\Rightarrow so(n,m)$, $i^m[u(n)\oplus u(m)]\Rightarrow u(n,m)$, and $i^m[sp(n)\oplus sp(m)]\Rightarrow sp(n,m)$. We will see that the set of vectors x_μ^α forming the Abelian ideal is also isomorphic to the rank m homogeneous spaces, respectively, $SO(n-m)\backslash SO(n)$, $U(n-m)\backslash U(n)$, and $Sp(n-m)\backslash Sp(n)$ for $1\leqslant m\leqslant n$, and contains the first family mentioned for the special case $m=1$. To the best of the authors' knowledge, this is the first application of the above deformation algorithm to coset spaces of rank greater than one. A quite different expansion, however, has been given by Mukunda.[11]

Corresponding to deformations of Lie algebra representations there are deformations of representations of the Lie groups[12] which gives rise to multiplier representations of the type developed by Bargmann[13] and Gel'fand and collaborators,[14] and generalized by Mackey.[15] The deformations $ISO(n)\Rightarrow SO(n,1)$[16] and $IU(n)\otimes U(1)\Rightarrow U(n,1)$[17] realized as multiplier representations on the real and complex spheres have been used to find the unitary irreducible representation (UIR) matrix elements of the principal series of the $SO(n,1)$ and $U(n,1)$ groups and have also been applied to the supplementary series[18] of the former. This method has been further used to develop a complete solution to the "missing label" problem in the noncanonical chain reduction $SO(n,1)\supset SO(1,1)\otimes SO(n-1)$.[19] The corresponding group representations for the second family of deformations has also been developed.[7]

In Sec. III we carry the deformation over to the corresponding groups and study the "nonrigid" action of the deformed group on the rank m homogeneous spaces, and show that one can thus obtain a principal series of UIRs of the deformed group. In Sec. IV we touch upon the inverse problem of contraction.[20]

II. DEFORMATIONS OF THE CLASSICAL ALGEBRAS

A. General construction

Consider the classical Lie algebra $k(n)$, which can be $so(n)$, $u(n)$, or $sp(n)$, the metric-preserving algebra of the sphere $S_{n-1}^{\mathbb{F}}$ on a field $\mathbb{F}$, i.e., the real $\mathbb{R}$, complex $\mathbb{C}$ or quaternionic $\mathbb{Q}$ fields, respectively, given by $x_\mu x_\mu^*=1$, where the asterisk $*$ stands for the involutive automorphism of the field, identity for $\mathbb{R}$, complex conjugation for $\mathbb{C}$ and quaternionic conjugation for $\mathbb{Q}$. Summation over repeated indices is implied and all middle

Greek letters range from 1 to n. The dimension of $S_{n-1}^{\mathbb{F}}$ is $n \dim \mathbb{F} - 1$. Let the index ω range over the corresponding components of the field $\mathbb{F}$, i.e., $\omega \equiv 0$ for $\mathbb{R}$, $\omega = 0,1$ for $\mathbb{C}$, and $\omega = 0,1,2,3$ for $\mathbb{Q}$. The $n \dim \mathbb{F}$ quantities $x_\mu{}^\omega$ transform as the components of a vector under commutation with the elements $M_{\mu\nu}^\omega$ of $k(n)$ and can be adjoined to them to construct the inhomogeneous classical algebra $ik(n)$, semidirect sum of $k(n)$ with an $[n \dim \mathbb{F}]$-dimensional Abelian ideal. It has been known[1-4] that out of the $x_\mu{}^\omega$ and the second-order Casimir operator $\Psi_k = \frac{1}{2} M_{\mu\nu}^\omega M_{\mu\nu}^\omega$ of $k(n)$ one can build the $n \dim \mathbb{F}$ operators

$$M_{\mu,n+1}^\omega = \tfrac{1}{2}[\Psi_k, x_\mu^\omega] + \tau x_\mu^\omega, \quad \tau \in \mathbb{C}, \tag{2.1}$$

which are elements of the enveloping algebra of $ik(n)$. We can verify that together with the generators of $k(n)$, they close into a $k(n,1)$ algebra $[so(n,1)$, $u(n,1)$, and $sp(n,1)$, respectively, in the last two cases, though, one has to add[6] to the former set the commutator between two operators (2.1), producing $u(1)$ and $sp(1)$ subalgebras which commute with the original compact ones].

We now introduce the action of the algebra $k(n)$ on a set of mn-vectors $x_\mu{}^\alpha$ $(\alpha = 1, \ldots, m)$. Such vectors can be taken as orthonormal since $k(n)$ commutes with the vector space scalar product $x_\mu{}^\alpha x_\mu{}^{\beta *}$. We introduce the constraints

$$x_\mu{}^\alpha x_\mu{}^{\beta *} = \delta_{\alpha,\beta}. \tag{2.2}$$

Such a choice of vectors can be conveniently thought of as an $n \times m$ rectangular matrix which is a submatrix of the $n \times n$ matrix self-representation of the Lie group generated by $k(n)$. Equation (2.2) represents $\frac{1}{2}m(m-1) \times \dim \mathbb{F} + m$ restrictions since $x_\mu^\alpha x_\mu^{\beta *} = (x_\mu^\beta x_\mu^{\alpha *})^*$ and $x_\mu^\alpha x_\mu^{\alpha *}$ is real. The number of independent components of the matrix $\mathbf{x}$ is thus $m[(n - \frac{1}{2}\{m-1\}) \dim \mathbb{F} - 1]$. The $nm \dim \mathbb{F}$ quantities $x_\mu{}^{\alpha\omega}$, however, can form the generators of an Abelian algebra which, when added in semidirect sum to $k(n)$ produces what we shall call the $i^m k(n)$ algebra. Our algorithm now generalizes (2.1) in constructing the operators

$$M_{\mu,n+\alpha}^\omega = \tfrac{1}{2}[\Psi_k, x_\mu^{\alpha\omega}] + \tau x_\mu^{\alpha\omega}, \quad \tau \in \mathbb{C}. \tag{2.3}$$

Moreover, building the commutators $[M_{\mu,n+\alpha}^\omega, M_{\nu,n+\beta}^{\omega'}]$, we see that we still obtain some extra operators $M_{n+\alpha,n+\beta}^{\omega''}$ which close onto a $k(m)$ algebra commuting with the original $k(n)$, and all of these, together with (2.3) form a $k(n,m)$ algebra. The free parameter τ, it has to be noted, must be the same for all $M_{\mu,n+\alpha}^\omega$ (i.e., it cannot have indices μ, α, or ω), or the resulting operators will not close onto an algebra of finite dimension. We will now write the results for the classical groups considered, using for consistency the relations as presented in Ref. 7.

B. $i^m\,[so(n) \oplus so(m)] \Rightarrow so(n,m)$

The generators of $so(n)$ are $M_{\mu\nu}$ with the commutation relations[21]

$$[M_{\mu\nu}, M_{\rho\sigma}] = g_{\nu\rho} M_{\mu\sigma} - g_{\mu\rho} M_{\nu\sigma} - g_{\nu\sigma} M_{\mu\rho} + g_{\mu\sigma} M_{\nu\rho}, \tag{2.4}$$

with $g_{\mu\nu} = \delta_{\mu\nu}$ for $\mu, \nu = 1, \ldots, n$. The generators of the normal Abelian subalgebra are $x_\mu{}^\alpha$ $(\alpha = 1, \ldots, m)$ satisfying (2.2) and

$$[M_{\mu\nu}, x_\rho{}^\alpha] = \delta_{\nu\rho} x_\mu{}^\alpha - \delta_{\mu\rho} x_\nu{}^\alpha. \tag{2.5}$$

Our deformation algorithm (2.3) now takes the form

$$M_{\mu,n+\alpha} = \tfrac{1}{2}[\Psi_{so}, x_\mu^\alpha] + \tau x_\mu^\alpha = x_\nu^\alpha M_{\nu\mu} + (-\tfrac{1}{2}[n-1] + \tau) x_\mu^\alpha. \tag{2.6}$$

Moreover, the commutator of two of the generators (2.6) will bring in the generators

$$M_{n+\alpha,n+\beta} = x_\mu{}^\alpha x_\nu{}^\beta M_{\mu\nu} \tag{2.7}$$

which close onto an $so(m)$ algebra and commute with the $M_{\mu\nu}$'s verifying that (2.6), (2.7), and the $M_{\mu\nu}$'s satisfy (2.4) with $g_{n+\alpha,n+\beta} = -\delta_{\alpha\beta}$ $(\alpha, \beta = 1, \ldots, m)$. Furthermore, one can show that the $so(m)$ subalgebra of generators (2.7) acts on the *column* indices of x_μ^α as

$$[M_{n+\alpha,n+\beta}, x_\mu{}^\gamma] = -(\delta_{\beta\gamma} x_\mu{}^\alpha - \delta_{\alpha\gamma} x_\mu{}^\beta), \tag{2.8}$$

i.e., as a vector with respect to the *upper* index. Notice that (2.5) and (2.8), however, have opposite signs. This will be shown in the next section to correspond to group actions from left and right.

C. $i^m\,[u(n) \oplus u(m)] \Rightarrow u(n,m)$

It is convenient to deal with the "complex" form of the generators of $u(n)$ given by $C_{\mu\nu}$ with the commutation relations

$$[C_{\mu\nu}, C_{\rho\sigma}] = g_{\nu\rho} C_{\mu\sigma} - g_{\mu\sigma} C_{\rho\nu}. \tag{2.9}$$

The Abelian generators are $z_\mu{}^\alpha$ and $z_\mu{}^{\alpha *}$ satisfying (2.2) in its form $z_\mu{}^\alpha z_\mu{}^{\beta *} = \delta_{\alpha\beta}$ and

$$[C_{\mu\nu}, z_\rho{}^\alpha] = \delta_{\nu\rho} z_\mu{}^\alpha, \tag{2.10}$$

$$[C_{\mu\nu}, z_\rho^{\alpha *}] = -\delta_{\mu\rho} z_\nu^{\alpha *}, \tag{2.11}$$

forming the $i^m u(n)$ algebra. The Casimir operator $\Psi_u = -2 C_{\mu\nu} C_{\nu\mu}$ now leads us to write (2.3) in the form

$$C_{\mu,n+\alpha} = \tfrac{1}{4}[\Psi_u, z_\mu^\alpha] + \tau' z_\mu^\alpha = -z_\nu^\alpha C_{\mu\nu} + (-\tfrac{1}{2}n + \tau') z_\mu^\alpha, \tag{2.12a}$$

$$C_{n+\alpha,\mu} = -\tfrac{1}{4}[\Psi_u, z_\mu^{\alpha *}] + \tau'^* z_\mu^{\alpha *} = -z_\nu^{\alpha *} C_{\nu\mu} + (\tfrac{1}{2}n + \tau'^*) z_\mu^{\alpha *}, \tag{2.12b}$$

and out of the commutators of (2.12) we find

$$C_{n+\alpha,n+\beta} = z_\mu^{\alpha *} z_\nu^\beta C_{\mu\nu} - (\tau' + \tau'^*) \delta_{\alpha\beta} \tag{2.13}$$

which together with (2.12) and the $C_{\mu\nu}$'s, close onto $u(n,m)$ with the commutator (2.9). Again we see that the z_μ^γ and $z_\mu^{\gamma *}$ transform as vectors with respect to the upper index under (2.13), i.e.,

$$[C_{n+\alpha,n+\beta}, z_\mu^\gamma] = \delta_{\alpha\gamma} z_\mu^\beta, \tag{2.14a}$$

$$[C_{n+\alpha,n+\beta}, z_\mu^{\gamma *}] = -\delta_{\beta\gamma} z_\mu^{\alpha *}. \tag{2.14b}$$

Writing $z_\mu^\alpha = x_\mu^{\alpha 0} + i x_\mu^{\alpha 1}$ and $z_\mu^{\alpha *} = x_\mu^{\alpha 0} - i x_\mu^{\alpha 1}$, the "real" form of the $u(n,m)$ generators can be written as

$$M_{ab}^0 = C_{ab} - C_{ba}, \tag{2.15a}$$

$$M_{ab}^1 = -i(C_{ab} + C_{ba}), \tag{2.15b}$$

for $a, b = 1, \ldots, n+m$. We can see that the results for the $so(n,m)$ algebra become a special case of those of the $u(n,m)$ algebra when we consider the subset of M^0's. Indeed, in the form (2.15) the $u(n,m)$ subalgebra of $sp(n,m)$ will become apparent in the next subsection.

D. $i^m\,[sp(n)\oplus sp(m)] \Rightarrow sp(n,m)$

As the symplectic algebra is not as well known as the other two classical algebras, we refer the reader to Ref. 7 where the explicit form of the generators is given. There are two isomorphic sets which we denote by $\{M^0_{\mu\nu}, M^{i+}_{\mu\nu}\}$ and $\{M^0_{\mu\nu}, M^{i-}_{\mu\nu}\}$ $(i=1,2,3)$, which correspond to left and right action with respect to quaternion multiplication. The commutation relations of the $sp(n)$ generators are

$$[M^0_{\mu\nu}, M^\omega_{\rho\sigma}] = g_{\nu\rho}M^\omega_{\mu\sigma} - g_{\mu\rho}M^\omega_{\nu\sigma} + g_{\nu\sigma}M^\omega_{\rho\mu} - g_{\mu\sigma}M^\omega_{\rho\nu}, \tag{2.16a}$$

$$[M^i_{\mu\nu}, M^0_{\rho\sigma}] = g_{\nu\rho}M^i_{\mu\sigma} - g_{\mu\rho}M^i_{\nu\sigma} - g_{\nu\sigma}M^i_{\rho\mu} + g_{\mu\sigma}M^i_{\rho\nu}, \tag{2.16b}$$

$$[M^i_{\mu\nu}, M^i_{\rho\sigma}] = -g_{\nu\rho}M^0_{\mu\sigma} - g_{\mu\rho}M^0_{\nu\sigma} + g_{\nu\sigma}M^0_{\rho\mu} + g_{\mu\sigma}M^0_{\rho\nu} \quad \text{(no sum)}, \tag{2.16c}$$

$$[M^i_{\mu\nu}, M^j_{\rho\sigma}] = \epsilon_{ijk}(g_{\nu\rho}M^k_{\mu\sigma} + g_{\mu\rho}M^k_{\nu\sigma} + g_{\nu\sigma}M^k_{\rho\mu} + g_{\mu\sigma}M^k_{\rho\nu}). \tag{2.16d}$$

The normal Abelian subalgebra is generated by $x^{\alpha\omega}_\mu$ satisfying (2.2) which, componentwise, yields $x^{\alpha\omega}_\mu \lambda^{\beta\omega}_\mu = \delta_{\alpha\beta}$ and $-x^{\alpha 0}_\mu x^{\beta i}_\mu + x^{\alpha i}_\mu x^{\beta 0}_\mu - \epsilon_{ijk}x^{\alpha j}_\mu x^{\beta k}_\mu = 0$. The semidirect sum algebra $i^m sp(n)$ is given by (2.18) plus

$$[M^0_{\mu\nu}, x^{\alpha\omega}_\rho] = \delta_{\nu\rho}x^{\alpha\omega}_\mu - \delta_{\mu\rho}x^{\alpha\omega}_\nu, \tag{2.17a}$$

$$[M^{i\pm}_{\mu\nu}, x^{\alpha 0}_\rho] = \pm(\delta_{\nu\rho}x^{\alpha i}_\mu - \delta_{\mu\rho}x^{\alpha i}_\nu), \tag{2.17b}$$

$$[M^{i\pm}_{\mu\nu}, x^{\alpha i}_\rho] = \mp(\delta_{\nu\rho}x^{\alpha 0}_\mu + \delta_{\mu\rho}x^{\alpha 0}_\nu) \quad \text{(no sum)}, \tag{2.17c}$$

$$[M^{i\pm}_{\mu\nu}, x^{\alpha j}_\rho] = \epsilon_{ijk}(\delta_{\nu\rho}x^{\alpha k}_\mu + \delta_{\mu\rho}x^{\alpha k}_\nu). \tag{2.17d}$$

For this case, the algorithm (2.3) takes the form

$$\begin{aligned} M^{0\pm}_{\mu,n+\alpha} &= \tfrac{1}{2}[\Psi^\pm_{sp}, x^{\alpha 0}_\mu] + \tau x^{\alpha 0}_\mu \\ &= x^{\alpha 0}_\nu M^0_{\nu\mu} \pm x^{\alpha i}_\nu M^{i\pm}_{\nu\mu} + (-2n-1+\tau)x^{\alpha 0}_\mu, \end{aligned} \tag{2.18a}$$

$$\begin{aligned} M^{i\pm}_{\mu,n+\alpha} &= \tfrac{1}{2}[\Psi^\pm_{sp}, x^{\alpha i}_\mu] + \tau x^{\alpha i}_\mu \\ &= x^{\alpha i}_\nu M^0_{\nu\mu} \mp x^{\alpha 0}_\nu M^{i\pm}_{\nu\mu} + \epsilon_{ijk}x^{\alpha j}_\nu M^{k\pm}_{\nu\mu} \\ &\quad + (-2n-1+\tau)x^{\alpha i}_\mu, \end{aligned} \tag{2.18b}$$

and again, from their commutators we extract

$$\begin{aligned} M^0_{n+\alpha,n+\beta} &= x^{\alpha\omega}_\mu x^{\beta\omega}_\nu M^0_{\mu\nu} \\ &\quad + (\pm x^{\alpha 0}_\mu x^{\beta i}_\nu \mp x^{\alpha i}_\mu x^{\beta 0}_\nu - \epsilon_{ijk}x^{\alpha j}_\mu x^{\beta k}_\nu)M^{i\pm}_{\mu\nu}, \end{aligned} \tag{2.19a}$$

$$\begin{aligned} M^{i\pm}_{n+\alpha,n+\beta} &= (x^{\alpha 0}_\mu x^{\beta i}_\nu - x^{\alpha i}_\mu x^{\beta 0}_\nu \mp \epsilon_{ijk}x^{\alpha j}_\mu x^{\beta k}_\nu)M^0_{\mu\nu} \\ &\quad \pm (x^{\alpha 0}_\mu x^{\beta 0}_\nu - x^{\alpha j}_\mu x^{\beta j}_\nu)M^{i\pm}_{\mu\nu} \\ &\quad + (\pm[x^{\alpha k}_\mu x^{\beta i}_\nu + x^{\alpha i}_\mu x^{\beta k}_\nu] - \epsilon_{ijk}[x^{\alpha 0}_\mu x^{\beta j}_\nu + x^{\alpha j}_\mu x^{\beta 0}_\nu])M^{k\pm}_{\mu\nu}, \end{aligned} \tag{2.19b}$$

which after some calculation can be seen to close, together with (2.21) and the $M^\omega_{\mu\nu}$'s, onto $sp(n,m)$. Moreover, the $Sp(m)$ subalgebra generated by (2.19) transforms the $x^{\alpha\omega}_\mu$ as

$$[M^0_{n+\alpha,n+\beta}, x^{\gamma\omega}_\rho] = -(\delta_{\beta\gamma}x^{\alpha\omega}_\rho - \delta_{\alpha\gamma}x^{\beta\omega}_\rho), \tag{2.20a}$$

$$[M^{i\pm}_{n+\alpha,n+\beta}, x^{\gamma 0}_\rho] = \pm(\delta_{\beta\gamma}x^{\alpha i}_\rho - \delta_{\alpha\gamma}x^{\beta i}_\rho), \tag{2.20b}$$

$$[M^{i\pm}_{n+\alpha,n+\beta}, x^{\gamma i}_\rho] = \mp(\delta_{\beta\gamma}x^{\alpha 0}_\rho - \delta_{\alpha\gamma}x^{\beta 0}_\rho) \quad \text{(no sum)}, \tag{2.20c}$$

$$[M^{i\pm}_{n+\alpha,n+\beta}, x^{\gamma j}_\rho] = -\epsilon_{ijk}(\delta_{\beta\gamma}x^{\alpha k}_\rho - \delta_{\alpha\gamma}x^{\beta k}_\rho), \tag{2.20d}$$

i.e., as the $M^{\mp}$'s on the column indices, acting from the opposite side, tensor- and quaternionwise, on the rectangular matrix x^α_μ.

It should be noticed that throughout this section we have never used any explicit realization of the algebra generators as operators on a homogeneous space. The only restriction has been that in each case we have only one continuous parameter τ, i.e., we can deform only along *one* direction.

III. MULTIPLIER REPRESENTATIONS

In this section we shall discuss representations of the groups $K(n,m)$ whose infinitesimal generators correspond to the Lie algebra representations presented in the last section. Rather than integrate directly these representations, we construct multiplier representations of $K(n,m)$ over the compact homogeneous space $X \equiv \{x^\alpha_\mu \in \mathbb{F} : x^\alpha_\mu x^\beta_\mu = \delta_{\alpha\beta}\}$ by generalizing the projective transformations on spheres used previously[16,17] for $SO(n,1)$ and $U(n,1)$ to projective transformations on X. Since such transformations will map X into itself, we are assured of the boundedness of the representations. Then by an appropriate choice of multiplier functions we obtain unitary representations. We then find the infinitesimal generators by the usual one-parameter subgroup method and it is seen that these correspond precisely to the formal representations of the Lie algebras $k(n,m)$ obtained through the deformation procedure of the previous section.

A. The group action

Given a realization of a compact classical group $K(n)$ of general element g by an $n\times n$ matrix $\mathbf{g} = \|g_{\mu\nu}\|$, $g_{\mu\nu} \in \mathbb{F}$ the action of $K(n)$ on the space of infinitely differentiable functions over the homogeneous space X can be written as

$$F(x^\alpha_\mu) \xrightarrow{g} F(g^{-1}_{\mu\nu}x^\alpha_\nu). \tag{3.1a}$$

In the case when the field is the noncommutative quaternion field $\mathbb{Q}$, we have the possibility of a related though distinct action[7]

$$F(x^\alpha_\mu) \xrightarrow{g^{(*)}} F(x^\alpha_\nu g^{-1}_{\mu\nu}{}^*) \tag{3.1b}$$

which is still from the left tensorwise, but from the right quaternionwise. The action (3.1a) for $Sp(n)$ is generated by the set of operators $\{M^0_{\mu\nu}, M^{i+}_{\mu\nu}\}$, while (3.1b) is generated by $\{M^0_{\mu\nu}, M^{i-}_{\mu\nu}\}$. For the $SO(n)$ groups (3.1a) and (3.1b) are the same and for $U(n)$ we have the complex conjugate representation of the group, an involutive automorphism of the algebra given by $M^i_{\mu\nu} \leftrightarrow -M^i_{\mu\nu}$. Only for $Sp(n)$ is it necessary to explicitly point out the difference.

In our case there are m orthonormal n-vectors forming an $n\times m$ rectagular matrix $\mathbf{x}$ satisfying (2.2). Consider first the vector x^1_μ. Equation (2.2) says it has to lie on the unit sphere $S^{\mathbb{F}}_{n-1}$. Now, x^2_μ is orthogonal to it, and thus constrained to lie on an $S^{\mathbb{F}}_{n-2}$ sphere orthogonal to x^1_μ. We follow the process up to x^m_μ and thus find that the space X is isomorphic with the product of the m spheres $S^{\mathbb{F}}_{n-1}\otimes S^{\mathbb{F}}_{n-2}\otimes\cdots\otimes S^{\mathbb{F}}_{n-m}$. This is also isomorphic to the homogeneous space $K(n-m)\backslash K(n)$ since a point $(x_0)^\alpha_\mu = \delta_{\mu\alpha}$ has $K(n-m)$ as its stability subgroup. The measure $d\mu(\mathbf{x})$ on X is induced by the Haar measure of $K(n)$ and is thus invariant under the action (3.1), the metric-preserving group of the manifold of $K(n)$. The transformations (3.1) can thus be called *rigid*.

B. Nonrigid transformations

The boost elements of the group $K(n,m)$, whose generators will be shown to be the noncompact operators (2.3) [concretely (2.6), (2.12), and (2.18)], are seen to produce *nonrigid* transformations of the space X, i.e., $d\mu(\mathbf{x})$ is not invariant under the general $K(n,m)$ action and a multiplier function is needed to obtain unitary representations.[13,14] Furthermore, the multiplier is generated by the inhomogeneous part of the $M^{\omega}_{\mu,n+\alpha}$'s, i.e., additive terms in $x^{\alpha\omega}_{\mu}$ with no derivative operators. Hence we obtain unitary representations of $K(n,m)$ in the form

$$F(\mathbf{x}) \xrightarrow{G} T^{\sigma}(G)F(\mathbf{x}) = \mu_{\sigma}(\mathbf{x},\mathrm{G})F(\mathbf{x}'(\mathbf{x},\mathrm{G})), \tag{3.2a}$$

where $G \in K(n,m)$ given by its $(n+m)\times(n+m)$ matrix representation G and, as will be shown in Secs. III.E to III.G.

$$\sigma = -\tfrac{1}{2}(n+m)\dim\mathbb{F} + 1 + i\rho, \quad \rho \text{ real}, \tag{3.2b}$$

and the action of the group on X, $\mathbf{x}'(\mathbf{x},\mathrm{G})$, will be given explicitly in Sec. III.D. Furthermore, as will be shown in Sec. III.F, the multiplier function $\mu_{\sigma}(\mathbf{x},\mathrm{G})$ enters into the Jacobian of the nonrigid transformation as

$$J \equiv \frac{d\mu(\mathbf{x}')}{d\mu(\mathbf{x})} = |\mu_{\sigma}(\mathbf{x},\mathrm{G})|^2. \tag{3.3}$$

C. The other rigid transformations in $K(n,m)$

The compact generators $M^{\omega}_{n+\alpha,n+\beta}$ produced out of commuting the operators (2.3) which close onto the $k(m)$ subalgebra of $k(n,m)$ [concretely (2.7), (2.13), and (2.19)] will be seen to correspond to the infinitesimal generators of a compact $K(m)$ subgroup of $K(n,m)$. It was shown in Sec. II that these generators transform the upper index of the x^{α}_{μ} in the same way (with opposite sign) as the original $k(n)$ algebra. We let the group action of $K(m)$ on the rectangular matrix $\mathbf{x}$ be from the right, i.e.,

$$F(x^{\alpha}_{\mu}) \xrightarrow{h} F(x^{\beta}_{\mu}h_{\beta\alpha}), \tag{3.4a}$$

where $\mathrm{h} = \|h_{\alpha\beta}\|$ is the $m\times m$ matrix realization of $h \in K(m)$ and, in the case when the x's and h's belong to the noncommutative field $\mathbb{Q}$, we have corresponding to the action (3.1b)

$$F(x^{\alpha}_{\mu}) \xrightarrow{h(*)} F(h^{*}_{\beta\alpha}x^{\beta}_{\mu}), \tag{3.4b}$$

for which the measure on X is again invariant and the transformations (3.4), therefore, rigid. Here we shall work only with the action (3.1a) and correspondingly (3.4a). The actions (3.1b)–(3.4b) do not bring in any fundamentally new features.

The space $X \approx K(n-m)\backslash K(n)$ is also isomorphic to a homogeneous space of the deformed group $K(n,m)$ which is elucidated in the Iwasawa decomposition[15] $K(n,m) \approx [K(n)\otimes K(m)]AN$, where A is an Abelian subgroup formed by m commuting boosts and N is a nilpotent subgroup. Now the stability subgroup of the point $(x_0)^{\alpha}_{\mu} = \delta_{\mu\alpha}$ can be shown[12] to be $H = \hat{K}AN$, where

$$\hat{\mathrm{K}} = \begin{pmatrix} \mathrm{K}(m) & 0 & 0 \\ 0 & \mathrm{K}'(n-m) & 0 \\ 0 & 0 & \mathrm{K}(m) \end{pmatrix}.$$

We can thus write $X \approx K(n-m)\backslash K(n) \approx H\backslash K(n,m)$.

D. General transformations of the homogeneous space

We shall now give explicitly the boost action (3.2) on the $n\times m$ rectangular matrix space X, generalizing the projective transformations used in Refs. 16 and 17 to matrix form. Split the matrix realization of $G \in K(n,m)$ as

$$\mathrm{G} = \begin{pmatrix} \mathbf{g} & \mathbf{b} \\ \tilde{\mathbf{b}} & \mathbf{h} \end{pmatrix}, \tag{3.5}$$

where $\mathbf{g}$ and $\mathbf{h}$ are $n\times n$ and $m\times m$ submatrices which contain the $K(n)$ and $K(m)$ subgroups of $K(n,m)$ and $\mathbf{b}$ and $\tilde{\mathbf{b}}$ are rectangular $n\times m$ and $m\times n$ matrices. We propose the action of G on X to be given by

$$\mathbf{x} \xrightarrow{G} \mathbf{x}' = (\mathbf{g}^{(-1)}\mathbf{x} + \mathbf{b}^{(-1)})(\tilde{\mathbf{b}}^{(-1)}\mathbf{x} + \mathbf{h}^{(-1)})^{-1} \tag{3.6}$$

where $\mathbf{g}^{(-1)}, \ldots, \mathbf{h}^{(-1)}$ are the *submatrices* of G^{-1} in the decomposition (3.5). We emphasize that $\mathbf{g}^{(-1)}, \ldots, \mathbf{h}^{(-1)}$ are *not* the inverses of the submatrices $\mathbf{g}, \ldots, \mathbf{h}$.

The action (3.6) is seen to give the correct composition law $F(\mathbf{x}''(\mathbf{x}'(\mathbf{x},\mathrm{G}_1),\mathrm{G}_2)) = F(\mathbf{x}'(\mathbf{x},\mathrm{G}_1\mathrm{G}_2))$ and reduce to the action (3.1a) and (3.4a) when $G \in K(n)$ and $K(m)$, respectively. Moreover, it can be verified that (3.6) preserves the restrictions (2.2) and hence maps the space X onto itself. Now consider an infinitesimal transformation through

$$\mathrm{G} \approx \mathbf{1} + \epsilon\Gamma, \quad \epsilon \ll 1, \quad \text{where } \Gamma = \begin{pmatrix} \gamma & \beta \\ \tilde{\beta} & \eta \end{pmatrix}.$$

The action on X is then given by

$$\mathbf{x}'(\mathbf{x}, 1+\epsilon\Gamma) \approx \mathbf{x} - \epsilon(\gamma\mathbf{x} + \beta - \mathbf{x}\eta - \mathbf{x}\tilde{\beta}\mathbf{x}). \tag{3.7}$$

It can be finally verified, after some computation, that the infinitesimal generators M in $F(\mathbf{x}'(\mathbf{x}, 1+\epsilon\Gamma)) = (\mathbb{1} - \epsilon M)F(\mathbf{x})$ are exactly the homogeneous part of the generators of $K(n,m)$ found by the deformation process in Sec. II when we use the explicit forms[7]

$$M^{0}_{\mu\nu} = x^{\alpha\omega}_{\mu}\partial^{\alpha\omega}_{\nu} - x^{\alpha\omega}_{\nu}\partial^{\alpha\omega}_{\mu}, \tag{3.8a}$$

$$\begin{aligned} M^{i\pm}_{\mu\nu} &= \pm(x^{\alpha i}_{\mu}\partial^{\alpha 0}_{\nu} + x^{\alpha i}_{\nu}\partial^{\alpha 0}_{\mu} - x^{\alpha 0}_{\mu}\partial^{\alpha i}_{\nu} - x^{\alpha 0}_{\nu}\partial^{\alpha i}_{\mu}) \\ &\quad - \epsilon_{ijk}(x^{\alpha j}_{\mu}\partial^{\alpha k}_{\nu} + x^{\alpha j}_{\nu}\partial^{\alpha k}_{\mu}), \end{aligned} \tag{3.8b}$$

where $\partial^{\alpha\omega}_{\mu} \equiv \partial/\partial x^{\alpha\omega}_{\mu}$. We have the freedom to add to the generators (3.8) a spin part induced by the subgroup $K(n-m)$, the centralizer of the boosts. These would arise if we consider tensor-valued functions over the coset space X.

E. The multiplier function

The inhomogeneous part of the boost generators is obtained when we consider the full representation (3.2) with the multiplier function

$$\mu_{\sigma}(\mathbf{x},\mathrm{G}) = [\mathrm{DET}(\tilde{\mathbf{b}}^{(-1)}\mathbf{x} + \mathbf{h}^{(-1)})]^{\sigma/2}, \tag{3.9}$$

where the determinant symbol DET of an $m\times m$ matrix A of in general noncommuting quaternionic elements $A_{\alpha\beta}$ is to be taken as the ordinary determinant of a $2m\times 2m$ matrix constructed representing the quaternions involved as 2×2 submatrices $A^{0}_{\mu\nu}\sigma_{\sigma} - iA^{k}_{\mu\nu}\sigma_k$, where σ_k are the Pauli matrices and $\sigma_0 = 1$. This construction can also be used for the real and complex cases where we have $A^{\omega}_{\mu\nu}$ for $\omega \equiv 0$ and $\omega = 0,1$. In these two cases DET A

$=|\det \mathbf{A}|^2$. Notice that the trace of such a matrix is $\mathrm{TR}\ \mathbf{A}=2A^0_{\mu\mu}$.

In order to show that (3.9) is the correct multiplier, one can verify the corresponding composition law $\mu(\mathbf{x},\mathrm{G}_1)\mu(\mathbf{x}'(\mathbf{x},\mathrm{G}_1),\mathrm{G}_2)=\mu(\mathbf{x},\mathrm{G}_1\mathrm{G}_2)$ while $\mu(\mathbf{x},\mathrm{G}_0)=1$ when $\mathrm{G}_0\in K(n)\otimes K(m)$ and thus for the group identity. Notice also that due to (3.6), (3.9) cannot be zero. Again, the consideration of infinitesimal transformations yields for the multiplier function in (3.2)

$$\mu_\sigma(\mathbf{x},1+\epsilon\Gamma)\approx[\mathrm{DET}(1-\epsilon[\tilde{\beta}\mathbf{x}+\eta])]^{\sigma/2}$$
$$\approx 1-\tfrac{1}{2}\sigma\epsilon\mathrm{TR}(\tilde{\beta}\mathbf{x}+\eta)=1-\sigma\epsilon(\tilde{\beta}\mathbf{x})^0_{\alpha\alpha}. \quad (3.10)$$

Consideration of a particular infinitesimal boost given by one $\tilde{\beta}_{\alpha\mu}=1$ and all others zero, shows that the coefficient of ϵ is σx^α_μ, which is precisely the inhomogeneous part of all boost generators seen in the last section with

$$\sigma=-\tfrac{1}{2}(n+1)\dim\mathbb{F}+1+\tau. \quad (3.11)$$

F. The transformation Jacobian

Having found the multiplier function $\mu_\sigma(\mathbf{x},\mathrm{G})$ in (3.2), we shall show that the Jacobian function (3.3) is closely related to it. Instead of starting directly with the transformation (3.6), it will prove easiest to show that the infinitesimal Jacobian has a form related to (3.10).

First notice that not all $x^{\alpha\omega}_\mu$'s are independent, but obey the restrictions (2.2). We start, therefore, with $nm\dim\mathbb{F}$ *independent* quantities $y^{\alpha\omega}_\mu$ which are made to undergo the transformation (3.6) induced by the x's. We shall show that the transformation Jacobian $J'\equiv\partial(y')/\partial(y)$ is equal to (3.3). Indeed, parametrize $y^{\alpha\omega}_\mu$ through (i) the in general quaternionic quantities $r_{\alpha\beta}\equiv y^\alpha_\mu y^{\beta *}_\mu$ of which there are $\frac{1}{2}m(m-1)\dim\mathbb{F}+m$ independent components $r^\omega_{\alpha\beta}$. Since $r_{\alpha\beta}=r^*_{\beta\alpha}$ while $r_{\alpha\alpha}$ is real, (ii) the independent parameters in $x^{\alpha\omega}_\mu$ chosen so that they satisfy (2.2) and which can be written in terms of the quaternionic Euler angles.[22] Now, the Jacobian J' is independent of the $r_{\alpha\beta}$ since they are invariant under $K(n,m)$ transformations. Hence J' only depends on the $x^{\alpha\omega}_\mu$ and is thus the Jacobian J in (3.3).

The explicit calculation of the infinitesimal $J'=J$ proceeds rather easily: From (3.6) we find $\mathbf{y}'(\mathbf{y},1+\epsilon\Gamma)$ and of these we need only the diagonal elements $\partial y'^{\alpha\omega}_\mu/\partial y^{\alpha\omega}_\mu$ (no sum). The Jacobian then reduces to

$$J\approx 1-\epsilon[(n+m)\dim\mathbb{F}-2]\mathrm{TR}(\tilde{\beta}\mathbf{x})^0 \quad (3.12)$$

which is directly comparable with (3.10) and (3.11), and assures us the form (3.3).

G. Unitary representations of $K(n,m)$ on X

We can obtain unitary representations of the group $K(n,m)$ on the space of infinitely differentiable functions[23] over $X\approx K(n-m)\backslash K(n)$ completing then with respect to the norm induced by the inner product

$$(F_1,F_2)_X=\int_X d\mu(\mathbf{x})F_1(\mathbf{x})^*F_2(\mathbf{x}) \quad (3.13)$$

when we introduce the group action through the operators $T^\sigma(G)$ as in (3.2). The choice of a complete and orthonormal set of functions $\{\Phi_n(\mathbf{x})\}$ on X allows the construction of the representation matrix elements as

$$D^\sigma_{n'n}(G)=(\Phi_{n'},T^\sigma(G)\Phi_n)_X. \quad (3.14)$$

Due to the relation between the multiplier function and the transformation Jacobian, for

$$\tau=-\tfrac{1}{2}(m-1)\dim\mathbb{F}+i\rho,\quad \rho \text{ real}, \quad (3.15)$$

the representations are unitary, i.e., $(T^\sigma(G)F_1, T^\sigma(G)F_2)_X=(F_1,F_2)_X$. They correspond to a principal degenerate series of representations[22,24] of $K(n,m)$ characterized by the value of σ given by (3.2b). Had we used the freedom allowed by the addition of a "spin" part to the generators, our functions F would be tensor-valued and the inner product (3.13) would include an inner product in an additional finite-dimensional vector space. In this way we can describe less degenerate representations where the additional labels are induced by the subgroup $K(n-m)$. This in no way hinders our construction since $K(n-m)$ is the boosts' centralizer in $K(n)\otimes K(m)$.

IV. CONTRACTIONS

A. Of the algebra

The representations of the algebras $k(n,m)$ in Sec. II can be labelled by $k(n,m)_\tau$. By a contraction of these representations we mean to divide some of the generators by τ and let $|\tau|\to\infty$. It is then seen that we can effect essentially two kinds of contractions, one with respect to the $k(n)$ subalgebra considering $M^\alpha_{\mu\nu}$, $\tau^{-1}M^\omega_{\mu,n+\alpha}$, and $\tau^{-1}M^\omega_{n+\alpha,n+\beta}$ and letting $|\tau|\to\infty$, thereby contracting $k(n,m)$ back to $i^m[k(n)\oplus 0]$, where 0 denotes the identity representation of $k(m)$. Another contraction considers $M^\omega_{\mu\nu}$, $\tau^{-1}M^\omega_{\mu,n+\alpha}$, and $M^\omega_{n+\alpha,n+\beta}$ for $|\tau|\to\infty$ [i.e., with respect to the $k(n)\oplus k(m)$ subalgebra]. In this case we obtain an $i^m[k(n)\oplus k(m)]$ algebra where the boost generators x^α_μ transform as vectors with respect to the lower index under $k(n)$ and with respect to the upper index under $k(m)$.

B. Of the group

The contraction of the representations of $K(n,m)$ given by (3.2) proceeds through considering group transformations $G(\epsilon)$ approaching the identity for $\epsilon\to 0$ in the boost elements and letting the representation parameter $\rho\to\infty$ such that $\epsilon\rho=\xi$, a finite number. As regards the boost generators, the group action (3.6) colapses to the identity while the multiplier function in (3.2) becomes, using (3.10),

$$\lim_{\epsilon\to 0}\mu_\sigma(\mathbf{x},\mathrm{G}(\epsilon))\approx\lim_{\epsilon\to 0}[1-2\epsilon(\tilde{\beta}\mathbf{x})^0_{\alpha\alpha}]^{i\rho/2}=\exp[-i\xi(\tilde{\beta}\mathbf{x})^0_{\alpha\alpha}], \quad (4.1)$$

thus the boost action in the direction (μ,α) becomes multiplication by $\exp(-i\xi x^\alpha_\mu)$ and the group contracts to $I^m[K(n)\otimes K(m)]$.

[1] M.A. Melvin, Bull. Am. Phys. Soc. 7, 493 (1962); 8, 356 (1963); A. Sankaranarayanan, Nuovo Cimento 38, 1441 (1965); A. Sankaranarayanan and R.H. Good Jr., Phys. Rev. 140, B509 (1965); A. Bohm, Phys. Rev. B145, 1212 (1966).
[2] J. Rosen and P. Roman, J. Math. Phys. 7, 2072 (1966); J. Rosen, Nuovo Cimento 46B, 1 (1966); J. Rosen, J. Math. Phys. 9, 1305 (1968); A. Chakrabarti, J. Math. Phys. 9, 2087 (1968).
[3] J.G. Nagel and K.T. Shah, J. Math. Phys. 11, 1483 (1970); J.G. Nagel, J. Math. Phys. 11, 1779 (1970).

[4]J.G. Nagel, Ann. Inst. H. Poincaré 13, 1 (1970).
[5]R. Gilmore, J. Math. Phys. 13, 883 (1972).
[6]A remark is in order here concerning the difference between expansions and deformations. Definitions for both terms are given in Ref. 7. We prefer to speak of deformations instead of expansions since we wish to emphasize the close connection with contractions. Expansions on the other hand have in general case little to do with contractions. Thus $u(n,1)$ is a deformation of $i[u(n)\otimes u(1)]$ and an expansion—but not a deformation—of $iu(n)$, where similar remarks hold for the more general case, i.e., $k(n,m)$ is an expansion of $i^m[k(n)\otimes k(m)]$.
[7]C.P. Boyer and K.B. Wolf, J. Math. Phys. 14, 1853 (1973).
[8]M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 (1966).
[9]C.P. Boyer and K.B. Wolf, Lett. Nuovo Cimento 8, 458 (1973).
[10]The first of these deformations was suggested as a private communication to one of the authors (K.B.W.) by Prof. Yuval Ne'eman of Tel-Aviv University, Israel, and has remained unpublished but for a footnote (p. 1045) in K.B. Wolf, Suppl. Nuovo Cimento 5, 1041 (1967).
[11]N. Mukunda, J. Math. Phys. 10, 897 (1969).
[12]C.P. Boyer, Rev. Mex. Fisica 23, 99 (1974).
[13]V. Bargmann, Ann. Math. 48, 568 (1947).
[14]I.M. Gel'fand and M.A. Naĭmark, *Unitäre Darstellungen der Klassischen Gruppen* (Academie Verlag, Berlin, 1957); I.M. Gel'fand and M.I. Graev, Am. Math. Soc. Transl. (Ser. 2) 37, 351 (1964).
[15]G.W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1967).
[16]K.B. Wolf, J. Math. Phys. 12, 197 (1971); C.P. Boyer, J. Math. Phys. 12, 1599 (1971); C.P. Boyer and F. Ardalan, J. Math. Phys. 12, 2070 (1971).
[17]K.B. Wolf, J. Math. Phys. 13, 1634 (1972).
[18]C.P. Boyer, J. Math. Phys. 14, 609 (1973).
[19]C.P. Boyer and K.B. Wolf, J. Math Phys. 15, 560 (1974).
[20]E. Inönü and E.P. Wigner, Proc. Natl. Acad. Sci. 39, 510 (1953); E.J. Saletan, J. Math. Phys. 2, 1 (1961); M. Lévy-Nahas, J. Math. Phys. 8, 1211 (1967); M. Lévy-Nahas and R. Seneor, Commun. Math. Phys. 9, 242 (1968); R. Hermann, Commun. Math. Phys. 5, 131 (1967); 6, 157 (1967).
[21]M. Gourdin, *Unitary Symmetries* (North-Holland, Amsterdam, 1967).
[22]P. Pajas and R. Rączka, J. Math. Phys. 9, 1188 (1968); P. Pajas, J. Math. Phys. 10, 1777 (1969).
[23]R.L. Anderson and K.B. Wolf, J. Math. Phys. 11, 3176 (1970).
[24]For $SO(n,m)$ see J. Fischer, J. Niederle, and R. Rączka, J. Math. Phys. 7, 816 (1966); R. Rączka, N. Limić, and J. Niederle, J. Math. Phys. 7, 1861 (1966); N. Limić, J. Niederle, and R. Rączka, J. Math. Phys. 7, 2026 (1966); N. Limić, J. Niederle, and R. Rączka, J. Math. Phys. 8, 1079 (1967); and for $U(n,m)$ see J. Fischer and R. Rączka, Commun. Math. Phys. 3, 233 (1966); 4, 8 (1967); M.I. Graev, Am. Math. Soc. Transl. 66, 1 (1968); T. Hirai, J. Math. Soc. Japan 22, 134 (1970).

Canonical transforms. II. Complex radial transforms

Kurt Bernardo Wolf

C. I. M. A. S., Universidad Nacional Autónoma de México, México 20, D. F., Mexico
(Received 27 June 1974)

Continuing the line of development of Paper I [J. Math. Phys. **15**, 1295 (1974)], we enlarge the concept of canonical transformations in quantum mechanics in two directions: first, by allowing the definition of a canonical transformation to be made through the preservation of an so(2,1) algebra, rather than the usual Heisenberg algebra, and providing the bridge between the classical and quantum mechanical descriptions, and, second, through the complexification of the transformation group. In this paper we study specifically the transformations which can be interpreted as the radial part of n-dimensional complex linear transformations in Paper I. We show that we can build Hilbert spaces of analytic functions with a scalar product defined through integration over half the complex plane of a variable which has the meaning of a complex radius. A unitary mapping to the ordinary Hilbert space $\mathcal{L}'^2_{r^{n-1}}(0,\infty)$ is provided with a kernel involving a Bessel function. Special cases of this are shown to be the Barut–Girardello, one-dimensional Bargmann and Hankel transforms. The transform kernels provide a series of representations of a subsemigroup of $SL(2,\mathbb{C})$ and allow the construction of coherent states for the harmonic oscillator with an extra centrifugal force. We present a hyperdifferential operator realization of these transforms which yields new Baker–Campbell–Hausdorff and special function relations.

1. INTRODUCTION

In the article which started this series (Ref. 1, henceforth refered to as I), we described complex linear transformations between the quantum-mechanical operators of position $\hat{\mathbf{x}}$ and momentum $\hat{\mathbf{p}}$, and a new pair of quantities given by

$$\hat{\eta} = a\hat{\mathbf{x}} + b\hat{\mathbf{p}}, \qquad \hat{\boldsymbol{\xi}} = c\hat{\mathbf{x}} + d\hat{\mathbf{p}}, \tag{1.1a}$$

with the unimodularity condition

$$ad - bc = 1, \tag{1.1b}$$

which ensures that (1.1a) is a canonical transformation in the sense that

$$[\hat{\mathbf{x}}, \hat{\mathbf{p}}] = i\mathbb{1} \Leftrightarrow [\hat{\eta}, \hat{\boldsymbol{\xi}}] = i\mathbb{1}. \tag{1.1c}$$

The motivation for such a program was the observation that particular complex transformations (1.1) have been fruitful: Bargmann[2,3] considered (1.1a) with

$$a = 2^{-1/2} = d, \quad b = -i2^{-1/2} = c \tag{1.2}$$

and the ensuing formalism has been applied to the coherent-state description of quantum optics.[4] Equations (1.1) for a, b, c, d real have provided unitary representations[5,6] of $SL(2,\mathbb{R})$ and, when continued into some regions of the complex plane of the parameters, have been used to relate and evaluate matrix elements of n-body systems subject to Gaussian-potential interactions relevant for the nuclear cluster model.[7]

In I we showed that: (i) The three examples given above are particular cases of a *canonical transform* (1.1) for $a, b, c, d \in \mathbb{C}$, the complex field, between the Hilbert space $\mathcal{H} \equiv \mathcal{L}^2(\mathbb{R})$ of square-integrable functions over the real line $\mathbb{R}$ and spaces $\mathcal{F}_{(a,b,c,d)}$ isomorphic to the Bargmann space of entire analytic functions in $\mathbb{C}$ with the well-known scalar product and decrease conditions.[2] (ii) A unitary transformation between $\mathcal{H}$ and $\mathcal{F}$ could be implemented [for $\mathrm{Im}(a/b) \geq 0$ and b real when $a = 0$] which contained the Bargmann transform for (1.2) and the Moshinsky–Quesne transform[5] for $a, b, c, d \in \mathbb{R}$. (iii) The transform kernels provided a representation of a subsemigroup of $SL(2,\mathbb{C})$ for $a, b, c, d \in \mathbb{C}$ subject to certain conditions.[8] (iv) A realization of these transforms through hyperdifferential operators was given, defined at least on spaces of entire functions. The defining conditions for $\mathcal{F}_{(a,b,c,d)}$ were to find a scalar product where $\hat{\eta}$ and $\hat{\xi}$ had the hermiticity properties derived from (1.1a) and the self-adjointness of $\hat{x}$ and $\hat{p}$ and were represented in the Schrödinger realization η and $-i\,\partial/\partial\eta$ on functions of $\eta \in \mathbb{C}$. The results were seen as a step towards exploiting the fact that quantum mechanics, being a richer structure than classical mechanics, and making use of the complex field in an essential way, should be amenable to a wider class of canonical transformation—defined through (1.1c)—than have been generally considered,[9] introducing scalar products more general than the usual Dirac integral over $\mathbb{R}$.

Among the extensions foreseen in I were to consider *n-dimensional* transformations (1.1) where $\hat{\mathbf{x}} \equiv (\hat{x}_j)$, $j = 1, \ldots, n$ etc. were n-vectors, but a, b, c, d remained (complex) multiples of the unit matrix. Equation (1.1c) now takes the familiar form $[\hat{x}_j, \hat{p}_k] = i\delta_{jk}$, etc. The "angular" properties, as given by the angular momentum operators in any of the subspaces, remain invariant under (1.1a) since the unimodularity condition (1.1b) insures that

$$L_{jk} \equiv \hat{x}_j\hat{p}_k - \hat{x}_k\hat{p}_j = \hat{\eta}_j\hat{\xi}_k - \hat{\eta}_k\hat{\xi}_j. \tag{1.3}$$

The "radial" part of (1.1) is displayed through the three equations

$$\eta^2 = a^2\hat{\mathbf{x}}^2 + 2ab\hat{\mathbf{x}}\cdot\hat{\mathbf{p}} + b^2\hat{\mathbf{p}}^2 - inab, \tag{1.4a}$$

$$\hat{\eta}\cdot\hat{\boldsymbol{\xi}} = ac\hat{\mathbf{x}}^2 + (ad+bc)\hat{\mathbf{x}}\cdot\hat{\mathbf{p}} + bd\hat{\mathbf{p}}^2 - inbc, \tag{1.4b}$$

$$\boldsymbol{\xi}^2 = c^2\hat{\mathbf{x}}^2 + 2cd\hat{\mathbf{x}}\cdot\hat{\mathbf{p}} + d^2\hat{\mathbf{p}} - incd. \tag{1.4c}$$

Seen classically, the canonical transformation can be described setting $\mathbf{x}^2 = r^2$, $\mathbf{x}\cdot\mathbf{p} = rp_r$, where the Poisson bracket $\{r, p_r\} = 1$, so that r and p_r are canonically conjugate quantities and $\mathbf{p}^2 = p_r^2 + p_\theta^2/r^2$, p_θ being the (constant) angular momentum. Correspondingly $\eta^2 = \rho^2$, $\boldsymbol{\eta}\cdot\boldsymbol{\xi} = \rho p_\rho$, $\boldsymbol{\xi}^2 = p_\rho^2 + p_\theta^2/\rho$. Equations (1.4) then read

$$\rho = [a^2r^2 + 2abrp_r + b^2(p_r^2 + p_\theta^2/r^2)]^{1/2}, \tag{1.5a}$$

$$p_\rho = [acr^2 + (ad+bc)rp_r + bd(p_r^2 + p_\theta^2/r^2)]/\rho, \tag{1.5b}$$

and the transformation of the pairs $(r,p_r) \to (\rho,p_\rho)$ can be checked to be canonical (i.e., $\{r,p_r\}=1 \Leftrightarrow \{\rho,p_\rho\}=1$). As the variable r takes values in $\mathbb{R}^+$ (the half-axis $[0,\infty)$), $\rho^2 \in \mathbb{C}$, and ρ will take values on half this region, which we can choose as

$$\mathbb{C}^+ \equiv \{\rho \in \mathbb{C} \mid \arg(\rho) \in [-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi)\}.$$

The transformation (1.5) is not particularly simple-looking, yet its quantum mechanical version will be seen to be implementable. This suggests that the definition of a quantum mechanical canonical transformation be made not in terms of the conservation of the Heisenberg algebra[9] as in (1.1c), which loses its meaning since the "quantization" of (1.5) is not well defined. The alternative, as suggested in this paper, is its definition in terms of the conservation of a higher algebra, in this case $so(2,1)$, which can be built out of the basic classical quantities.

The Schrödinger representation[10] of the operators $\hat{\mathbf{x}}^2$, $\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}$, and $\hat{\mathbf{p}}^2$ is

$$\hat{\mathbf{x}}f(r) = r^2 f(r), \tag{1.6a}$$

$$\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}f(r) = -ir\frac{d}{dr}f(r), \tag{1.6b}$$

$$\hat{\mathbf{p}}^2 f(r) = -\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{d}{dr} + \frac{\lambda}{r^2}\right)f(r), \quad \lambda \in \mathbb{R}, \tag{1.6c}$$

on the (at least twice-differentiable) elements of the Hilbert space $\mathcal{H}_n^+ \equiv \mathcal{L}^2_{r^{n-1}}(\mathbb{R}^+)$ of functions $f(r)$ on the positive half-axis with the scalar product

$$(f,g)_n = \int_0^\infty r^{n-1}\,dr\,f(r)^* g(r), \tag{1.7}$$

(the star indicates complex conjugation). The operators (1.6) are Hermitian between these elements and their domain can be enlarged through the usual adjunction procedure to self-adjoint operators in $\mathcal{H}_n^+$. The constant λ in (1.6c) comes from the spectrum of $L^2 = \frac{1}{2}\sum L_{ij}L_{ij}$ when acting on the $so(n)$-irreducible components of the functions, and has the values

$$\lambda = -l(l+n-2), \quad l=0,1,2,\cdots. \tag{1.8}$$

The statements concerning the hermiticity of $\hat{\mathbf{p}}^2$ continue to be valid, however, for arbitrary $\lambda \in \mathbb{R}$.

It is the purpose of this article to describe a family of Hilbert spaces $\mathcal{J}^+_{nl\,(a,b,c,d)}$ (the indices a,b,c,d will be suppressed) for which a Schrödinger representation parallel to (1.6) can be implemented for the new variables in (1.4), namely

$$\hat{\eta}^2\bar{f}(\rho) = \rho^2\bar{f}(\rho), \tag{1.9a}$$

$$\hat{\eta}\cdot\hat{\boldsymbol{\zeta}}\bar{f}(\rho) = -i\rho\frac{d}{d\rho}\bar{f}(\rho), \tag{1.9b}$$

$$\hat{\boldsymbol{\zeta}}^2\bar{f}(\rho) = -\left(\frac{d^2}{d\rho^2} + \frac{n-1}{\rho}\frac{d}{d\rho} + \frac{\lambda}{\rho^2}\right)\bar{f}(\rho) \tag{1.9c}$$

on functions of the complex variable ρ restricted to the region $\mathbb{C}^+$ [Eq. (1.6)]. In order that the total derivative with respect to a complex variable be well defined, the functions $\bar{f}$ will be *analytic* functions of ρ and $\partial\bar{f}(\rho)/\partial\rho^* = 0$. The measure for the defining scalar product in $\mathcal{J}^+_{nl}$,

$$(\bar{f},\bar{g})_{nl} = \int_{\mathbb{C}^+} d\mu_{nl}(\rho)\bar{f}(\rho)^*\bar{g}(\rho) \tag{1.10a}$$

is of the form

$$d\mu_{nl}(\rho) = \nu_{nl}(\rho,\rho^*)\,d\,\mathrm{Re}\rho\,d\,\mathrm{Im}\rho, \tag{1.10b}$$

where the weight function $\nu_{nl}(\rho,\rho^*)$ will be found from the hermiticity properties of (1.9)—(1.1) and the hermiticity of $\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}$. This will be performed in Sec. 2 and the characteristics of the Hilbert space $\mathcal{J}^+_{nl}$ ascertained. In Sec. 3 we will find the unitary transformation between $\mathcal{H}_n^+$ and $\mathcal{J}^+_{nl}$ as given by

$$\bar{f}(\rho) = \int_0^\infty r^{n-1}\,dr\,A_{nl}(\rho,r)f(r), \tag{1.11a}$$

$$f(r) = \int_{\mathbb{C}^+} d\mu_{nl}(\rho)A_{nl}(\rho,r)^*\bar{f}(\rho), \tag{1.11b}$$

through the transform kernel $A_{nl}(\rho,r)$ function of n,l and a,b,c,d. This *complex radial* transform will relate to the complex linear transform of I as the Hankel transform relates to the n-dimensional Fourier transform and, as will be shown, contains the Barut–Girardello transform[11] for the value (1.2) of the parameters[12] and the radial transform of Moshinsky, Seligman, and Wolf in Ref. 13 for a,b,c,d real. In Sec. 4 it is shown that this last transform is indeed regained when a,b,c,d become real and that the scalar product (1.10) collapses to the line integral (1.7). The one-dimensional Bargmann space[2] is also regained when $n=1$ as the direct sum of $\mathcal{J}^+_{10}$ and $\mathcal{J}^+_{11}$. We consider the interest of the complex radial transform to go beyond that of the mere description of the radial part of a known transform: As we will be mapping the radial wavefunctions of potentials of the harmonic oscillator + centrifugal potential ($\sim 1/r^2$) kind on functions of the type ρ^{2N+l}, coherent states for these systems can be defined. This is shown in Sec. 5. In Sec. 6 we make the composition of transforms and shown that the transform kernels provide a representation of a subsemigroup of $SL(2,\mathbb{C})$ in (1.1). Some conclusions of the role of complex canonical transformations in quantum mechanics are presented in Sec. 7. In two appendices we give a hyperdifferential operator realization of the transform (1.8) obtaining a new representation of the associated Laguerre functions and its direct relation to the n-dimensional complex linear transform.

2. THE SPACE $\mathcal{J}^+_{nl}$

We will construct a space $\mathcal{J}^+_{nl\,(a,b,c,d)}$ of functions $\bar{f},\bar{g}$ over $\rho \in \mathbb{C}^+$ endowed with a scalar product of the type (1.10) such that the operators $\hat{\eta}^2$, $\hat{\eta}\cdot\boldsymbol{\zeta}$, and $\boldsymbol{\zeta}^2$ have the Hermitian conjugation property obtained from inverting (1.4),

$$\begin{aligned}(\hat{\mathbf{x}}^2\bar{f},\bar{g})_{nl} &= ([d^2\hat{\eta}^2 - 2b\hat{\eta}\cdot\hat{\boldsymbol{\zeta}} + b^2\hat{\boldsymbol{\zeta}}^2 + indb]\bar{f},\bar{g})_{nl} \\ &= (\bar{f},\hat{\mathbf{x}}^2\bar{g})_{nl} = (\bar{f},[d^2\hat{\eta}^2 - 2b\hat{\eta}\cdot\hat{\boldsymbol{\zeta}} + b^2\hat{\boldsymbol{\zeta}}^2 + indb]\bar{g})_{nl}\end{aligned} \tag{2.1}$$

and similar companion equations for $(\hat{\mathbf{x}}\cdot\hat{\mathbf{p}})^\dagger = \hat{\mathbf{p}}\cdot\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}^2$ in the Schrödinger representation (1.9). Equation (2.1) and its companions can be turned into differential equations on the weight function $\nu_{nl}(\rho,\rho^*)$ in (1.10) through integration by parts, using, for $\rho = |\rho|\exp(i\theta)$, $d\,\mathrm{Re}\rho\,d\,\mathrm{Im}\rho = |\rho|\,d|\rho|\,d\theta$, and $d/d\rho = \frac{1}{2}\exp(-i\theta)[\partial/\partial|\rho| + (i\rho)^{-1}\partial/\partial\theta]$ so that $\partial\rho^*/\partial\rho = 0$ and, for analytic functions $A(\rho)$, $B(\rho)$,

$$\int_0^\infty |\rho|\, d|\rho| \int_\alpha^\beta d\theta A(\rho)\ \frac{d}{d\rho} B(\rho)$$
$$= -\int_0^\infty |\rho|\, d|\rho| \int_\alpha^\beta d\theta \left(\frac{d}{d\rho} A(\rho)\right) B(\rho)$$
$$+ \tfrac{1}{2}|\rho| \int_\alpha^\beta d\theta \exp(-i\theta) AB \Big|_{|\rho|\to\infty}$$
$$- \tfrac{1}{2} i \exp(-i\theta) \int_0^\infty d|\rho| AB(|_{\theta=\beta} - |_{\theta=\alpha}). \qquad (2.2)$$

By assuming the boundary integral terms vanish (the restrictions from this condition will be made explicit below), Eq. (2.1) yields the differential equation

$$b^{*2}\left[-\frac{\partial^2}{\partial\rho^{*2}} + \left(2i\frac{d^*}{b^*}\rho^* + \frac{n-1}{\rho^*}\right)\frac{\partial}{\partial\rho^*} + \frac{d^{*2}}{b^{*2}}\rho^{*2} + i\frac{d^*}{b^*}(2-n) - \frac{\lambda+n-1}{\rho^*}\right]\nu_{nl}(\rho,\rho^*)$$
$$= b^2\left[-\frac{\partial^2}{\partial\rho^2} + \left(-2i\frac{d}{b}\rho + \frac{n-1}{\rho}\right)\frac{\partial}{\partial\rho} + \frac{d^2}{b^2}\rho^2 - i\frac{d}{b}(2-n) - \frac{\lambda+n-1}{\rho^2}\right]\nu_{nl}(\rho,\rho^*), \qquad (2.3)$$

and similar ones (i.e., replacing $b\to a$, $d\to c$, etc) for the companions, with vanishing conditions for the boundary terms of $\rho\nu \bar{f}^*\bar{g}$, $\nu\bar{f}^*(\partial_\rho\bar{g})$, $(\partial_\rho\nu)\bar{f}^*\bar{g}$, and $\rho^{-1}\nu\bar{f}^*\bar{g}$ and similar ones replacing ρ and ρ^*. Notice that whereas in I we had two simultaneous first-order differential equations, here we have three second-order ones. Based on I, however, we can make the ansatz that

$$\nu_{nl}(\rho,\rho^*) = \exp\left(\frac{u}{2v}\rho^2\right)\exp\left(\frac{u^*}{2v}\rho^{*2}\right)\mu_{nl}(\rho\rho^*), \qquad (2.4)$$

where, as in I, we define

$$u \equiv a^*d - b^*c, \qquad (2.5a)$$

$$v \equiv 2\,\mathrm{Im}(b^*a). \qquad (2.5b)$$

We obtain the result that the three equations (2.3) yield a single differential equation for μ_{nl} which shows that $\mu_{nl}(\rho\rho^*) = (\rho\rho^*)^{n/2}\beta_{n/2+l-1}(\rho\rho^*/v)$, where β is a solution of Bessel's modified equation: I or K functions. The boundary integral over the semicircle at infinity appearing in the integration by parts of (2.3) will vanish for functions of less or equal growth than $\exp(\frac{1}{2}\rho^2/v)$ if we choose the MacDonald (or modified Hankel) function K. We find, with a specific choice of normalization, justified in Sec. 4 that

$$\nu_{nl}(\rho,\rho^*) = (2/\pi v)\exp[(1/2v)(u\rho^2 + u^*\rho^{*2})] \times (\rho\rho^*)^{n/2} K_{n/2+l-1}(\rho\rho^*/v). \qquad (2.6)$$

If we let $u = \omega\exp(i\varphi)$ be the polar representation of u, the behavior of (2.6) at the interval end points is

$$\nu_{nl}(\rho,\rho^*) \underset{|\rho|\to\infty}{\approx} (\tfrac{1}{2}\pi v)^{1/2}|\rho|^{n-1} \times \exp[-(1/v)|\rho|^2(1-\omega\cos\{\varphi+2\theta\})] \qquad (2.7a)$$

and

$$\nu_{nl}(\rho,\rho^*) \underset{|\rho|\to 0}{\approx} 2(2v)^{n/2+l-1}\Gamma(\tfrac{1}{2}n+l-1)|\rho|^{2(1-l)}, \quad l > -\tfrac{1}{2}n+1, \qquad (2.7b)$$

$$\nu_{nl}(\rho,\rho^*) \underset{|\rho|\to 0}{\approx} -2(\pi v)^{-1}|\rho|^{2(1-l)}\ln(|\rho|^2/v), \quad l = -\tfrac{1}{2}n+1. \qquad (2.7c)$$

As λ in (1.8) is invariant under the replacement $l \leftrightarrow -l-n+2$, only $l \geq -\frac{1}{2}n+1$ need be considered. Correspondingly, we have the property $K_\mu(z) = K_{-\mu}(z)$. The remaining boundary integrals over the imaginary axis will be made to vanish and the finiteness of $(\bar{f},\bar{g})_{nl}$ itself determined by restricting the space of functions. Consider

$$\bar{\phi}'_m(\rho) = c_m \exp[-(u/2v)\rho^2]\rho^m \qquad (2.8a)$$

for $m\in\mathbb{R}$ and c_m a normalization constant. In performing the scalar product $(\bar{\phi}'_m, \bar{\phi}'_{m'})_{nl}$ we can separate the integration of $\rho\in\mathbb{C}^+$ into a radial and angular part, the latter being $\int d\theta\exp[i(m'-m)\theta]$ over $[-\frac{1}{2}\pi, \frac{1}{2}\pi)$. This is zero if $m-m'$ is an even nonzero integer, and π if $m=m'$. In the last case, the remaining integral can be evaluated[14] and $\bar{\phi}'_m(\rho)$ normalized through (2.8a) setting

$$c_m = \vartheta_m[\tfrac{1}{2}(2v)^{n/2+m}\Gamma(\tfrac{1}{2}(n+l+m))\Gamma(\tfrac{1}{2}(m-l+2))]^{-1/2}, \qquad (2.8b)$$

where ϑ_m is an arbitrary phase and the arguments of the function reflect the fact that the integration is valid and the result finite for $m > l-2$ and $m > -n-l$. The latter is a consequence of the former for $l \geq -\frac{1}{2}n+1$. In checking the vanishing conditions for the boundary terms mentioned below Eq. (2.3), we come to the conclusion that these hold if $m-m'$ is an even integer. If we now write $m = l+\alpha+2N$ with $N=0,1,2,\cdots$ and $\alpha\in(-2,0]$ we can see that asking $\bar{\phi}'_m(\rho)$ to be in the invariant common domain of the three operators (1.9) forces $\alpha=0$. Hence an orthonormal basis for the space $\mathcal{J}^+_{nl}$ object of our construction is, with a specific choice of phase,

$$\bar{\phi}_N(\rho) = (-1)^N[\tfrac{1}{2}(2v)^{n/2}N!\,\Gamma(N+\tfrac{1}{2}n+l)]^{-1/2} \times \exp[-(u/2v)\rho^2][(2v)^{-1/2}\rho]^{2N+l}, \quad N=0,1,2,\cdots. \qquad (2.9)$$

Now, the basis (2.9) is *complete* in the Hilbert space $\mathcal{J}^+_{nl}$ of functions $\bar{f}$ of the type $\bar{f}(\rho) = \exp[-(u/2v)\rho^2]\rho^l$ times an entire function in $\rho^2/2v$ of growth (1.1) [or of growth $(2, 1/2v)$ in ρ] completed with respect to the norm induced by (1.10) with the weight function (2.6). The proof is the standard one[15] which proves that convergence in the norm implies pointwise convergence for these functions. Indeed, for

$$\bar{f}(\rho) = \exp[-(u/2v)\rho^2]\rho^l\sum_{N=0}^\infty f_N\rho^{2N} = \sum_{N=0}^\infty \alpha_N f_N \bar{\phi}_N(\rho), \qquad (2.10a)$$

$$\alpha_N = (-1)^N(2v)^{N+(n/2+l)/2}[\tfrac{1}{2}N!\,\Gamma(N+\tfrac{1}{2}n+l)]^{1/2}, \qquad (2.10b)$$

we have

$$\|\bar{f}\|^2_{nl} \equiv (\bar{f},\bar{f})_{nl} = \sum_{N=0}^\infty \alpha_N^2|f_N|^2. \qquad (2.10c)$$

Using the Schwartz inequality, we obtain

$$|\bar{f}(\rho)|^2 = |\exp[-(u/2v)\rho^2]\rho^l|^2\left|\sum_{N=0}^\infty f_n\rho^{2N}\right|^2$$
$$\leq |\textstyle\sum f_N\alpha_N|^2|\exp[-(u/2v)\rho^2]\rho^l|^2 \times |\textstyle\sum \alpha_{N'}^{-1}\rho^{2N'}|^2$$
$$\leq \textstyle\sum|f_N|^2\alpha_N^2|\exp[-(u/2v)\rho^2]\rho^l|^2\sum\alpha_{N'}^{-2}|\rho|^{2N'}$$

$$= \|\bar{f}\|^2_{nl} |\exp[-(u/2v)\rho^2]|^2 |\rho|^{2-n} \times v^{-1} I_{n/2+l-1}(|\rho|^2/v) = \|\bar{f}\|^2_{nl} K_{nl}(\rho,\rho) \quad (2.11)$$

[where the function $K_{nl}(\rho,\rho')$ will be defined below], and hence any Cauchy sequence of functions coverging in the norm to a function in $\mathcal{J}^+_{nl}$ implies the uniform convergence of the functions themselves on any compact set in $\mathbb{C}^+$. The reproducing kernel in the integral (1.10)—(2.6) is thus

$$K_{nl}(\rho,\rho') \equiv \sum_{N=0}^{\infty} \bar{\phi}_N(\rho)\bar{\phi}_N(\rho')^* = v^{-1}(\rho\rho'^*)^{1-n/2} \exp[-(1/2v)(u\rho^2 + u^*\rho'^{*2})] \times I_{n/2+l-1}(\rho\rho'^*/v), \quad (2.12)$$

and appears in the last number of (2.11).

Before closing this section, we will find an algebra of raising, lowering, and weight operators for the basis functions (2.9). Easiest to build, the raising operator is

$$\mathcal{R}\,\bar{\phi}_N(\rho) \equiv [-(1/2v)\rho^2]\bar{\phi}_N(\rho) = [(N+1)(N+\tfrac{1}{2}n+l)]^{1/2}\bar{\phi}_{N+1}(\rho). \quad (2.13a)$$

Its Hermitian conjugate under the scalar product (1.10) is the lowering operator

$$\mathcal{L}\,\bar{\phi}_N(\rho) \equiv -\left[\tfrac{1}{2}v\frac{d^2}{d\rho^2} + \left(u\rho + \tfrac{1}{2}v\frac{n-1}{\rho}\right)\frac{d}{d\rho} + \left(\frac{u^2}{2v}\rho^2 + \tfrac{1}{2}nu + \tfrac{1}{2}v\frac{\lambda}{\rho^2}\right)\right]\bar{\phi}_N(\rho) = [N(N+\tfrac{1}{2}n+l-1)]^{1/2}\bar{\phi}_{N-1}(\rho). \quad (2.13b)$$

The weight operator

$$\mathcal{N}\bar{\phi}_N(\rho) \equiv \left(\rho\frac{d}{d\rho} + \frac{u}{v}\rho^2 + \tfrac{1}{2}n\right)\bar{\phi}_N(\rho) = (2N+\tfrac{1}{2}n+l)\bar{\phi}_N(\rho) \quad (2.13c)$$

completes the set of generators of an $so(2,1)$ algebra with commutation relations

$$[\mathcal{N},\mathcal{R}] = 2\mathcal{R}, \quad [\mathcal{N},\mathcal{L}] = -2\mathcal{L}, \quad [\mathcal{R},\mathcal{L}] = -\mathcal{N}. \quad (2.14)$$

3. THE TRANSFORM BETWEEN $\mathcal{H}^+_n$ AND $\mathcal{J}^+_{nl}$

The transform kernel $A_{nl}(\rho,r)$ in (1.11) can be calculated if we ask for the conditions (1,4), (1.6), and (1.9) to hold, namely, that if $\bar{f}(\rho)$ is the transform of $f(r)$, then $\rho^2\bar{f}(\rho)$ be the transform of

$$\left[a^2r^2 + 2iabr\partial_r - b^2\left(\partial_r^2 + \frac{n-1}{r}\partial_r + \frac{\lambda}{r^2}\right) + niab\right]f(r).$$

Similar conditions stem from $-i\rho\partial_\rho$ and $-\{\partial^2_\rho + [(n-1)/\rho]\partial_\rho + \lambda/\rho^2\}$. By partial integration in (1.11) these can be turned into three second-order differential equations for $A_{nl}(\rho,r)$. From I we make the ansatz that $A_{nl}(\rho,r)$ have the form

$$\exp[(i/2b)(ar^2 + d\rho^2)]B_{nl}(\rho r)$$

whereupon the three differential equations for $A_{nl}(\rho,r)$ yield a single one for $B_{nl}(\rho r)$ as $(\rho r)^{1-n/2}$ times a solution of Bessel's equation. If $f(r)$ belongs to the space $\mathcal{H}^+_n$ with scalar product (1.7), for (1.11a) to be integrable we must require $\mathrm{Im}(a/b) \geq 0$ (i.e., $v \geq 0$) for the exponent and the Bessel function as solution for $B_{nl}(\rho r)$. With a specific choice for phase and normalization to be justified below and in Sec. 4, we write

$$A_{nl}(\rho,r) = b^{-1}\vartheta_{n,l}\exp[(i/2b)(ar^2 + d\rho^2)](\rho r)^{1-n/2} \times J_{n/2+l-1}(\rho r/b) \quad (3.1a)$$

with

$$\vartheta_{n,l} = \exp[-i\tfrac{1}{2}\pi(\tfrac{1}{2}n+l)]. \quad (3.1b)$$

The calculation of the explicit form of the orthonormal basis transform to (2.9) can be simplified if we look for the eigenfunctions of the weight operator (2.13c) which through (1.4) becomes

$$\mathcal{N}\phi_N(r) = v^{-1}\left[|a|^2r^2 - \tfrac{1}{2}i\,\mathrm{Re}(ab^*)r\frac{d}{dr} - |b|^2\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{d}{dr} + \frac{\lambda}{r^2}\right) - \tfrac{1}{4}in\,\mathrm{Re}(ab^*)\right]\phi_N(r) = (2N+\tfrac{1}{2}n+l)\phi_N(r), \quad (3.2)$$

plus normalization under (1.7) and a phase to satisfy Eq. (3.4) below. The result is, if we denote the phase of b by $\exp[i\arg b]$ with $\arg b \in [-\pi,\pi)$,

$$\phi_N(r) = \vartheta_N\{2N!\,[\mathrm{Im}(a/b)]^{n/2+l}/\Gamma(N+\tfrac{1}{2}n+l)\}^{1/2} \times \exp[-\tfrac{1}{2}i(a^*/b^*)r^2]r^l L_N^{(n/2+l-1)}[r^2\,\mathrm{Im}(a/b)], \quad (3.3a)$$

with

$$\vartheta_N = \exp[i(2N+\tfrac{1}{2}n+l)(\arg b + \tfrac{1}{2}\pi)]. \quad (3.3b)$$

We can now verify that[16]

$$A_{nl}(\rho,r) = \sum_{N=0}^{\infty}\bar{\phi}_N(\rho)\phi_N(r)^*. \quad (3.4)$$

At this point it is apparent that a second pair of transform orthonormal bases for $\mathcal{H}^+_n$ and $\mathcal{J}^+_{nl}$ is useful, since the limit $v\to 0$ of real transformations of (2.9)—(3.3) is not manifest. As in I, we choose the basis functions $\psi_{Nl}(r)$ for $\mathcal{H}^+_n$ to be the radial part of the solutions of a harmonic oscillator with centrifugal force Hamiltonian in n dimensions given by

$$2I_3\psi_{Nl}(r)Y_L^M(\omega) \equiv \tfrac{1}{2}[\hat{\mathbf{p}} + g\hat{\mathbf{x}}^{-2} + \hat{\mathbf{x}}^2]\psi_{Nl}(r)Y_L^M(\omega) = \frac{1}{2}\left[-\frac{\partial^2}{\partial r^2} - \frac{n-1}{r}\frac{\partial}{\partial r^2} + \frac{g+L(L+n-2)}{r^2} + r^2\right] \times \psi_{Nl}(r)Y_L^M(\omega) = [2N+\tfrac{1}{2}n+l]\psi_{Nl}(r)Y_L^M(\omega), \quad (3.5a)$$

where $Y_L^M(\omega)$ is the n-dimensional normalized spherical harmonic, the collective label M standing for the transformation properties under $SO(n-1)\supset\cdots\supset SO(2)$, while the $SO(n)$ label L enters into the differential operator and relates to l through

$$l(l+n-2) \equiv -\lambda \equiv g + L(L+n-2), \quad (3.5b)$$

giving two values of l for each g and L, in general. The solution of the radial equation is

$$\psi_{Nl}(r) = [2N!/\Gamma(N+\tfrac{1}{2}n+l)]^{1/2}\exp(-r^2/2)r^l L_N^{(n/2+l-1)}(r^2), \quad (3.6)$$

whose corresponding raising and lowering operators can be obtained from (3.5a) and

$$I_1 \equiv \tfrac{1}{4}[\hat{\mathbf{p}}^2 + g\hat{\mathbf{x}}^{-2} - \hat{\mathbf{x}}^2], \tag{3.7a}$$

$$I_2 \equiv \tfrac{1}{4}[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}} + \hat{\mathbf{p}}\cdot\hat{\mathbf{x}}], \tag{3.7b}$$

which can be verified to close into an $so(2,1)$ algebra. The transform basis functions can be calculated directly using the transform (1.7a), (3.1), (3.6), yielding[17]

$$\overline{\psi}_{Nl}(\rho) = [2N!/\Gamma(N+\tfrac{1}{2}n+l)]^{1/2}(a+ib)^{-n/2-l}[(a-ib)/(a+ib)]^N \times \exp\left(-\frac{1}{2}\,\frac{d-ic}{a+ib}\,\rho^2\right)\rho^l L_N^{(n/2+l-1)}(\rho^2/[a^2+b^2]). \tag{3.8}$$

In particular, notice that when we have the Bargmann case (1.2), (2.5) gives $u=0$, $v=1$, only the leading term of the Laguerre function remains, and both bases coincide as (3.8) becomes proportional to ρ^{2N+l} and equal to $\overline{\phi}_N(\rho)$. This determined our choice of phase for the latter.

The unitarity of the transform pair (1.7) with the kernel (3.1) between $\mathcal{H}_n^+$ and $\mathcal{J}_{nl}^+$ can be established following the same steps as in Bargmann's original work.[18] That it transforms the orthonormal basis $\{\phi_N(r)\}$ to the orthonormal basis $\{\overline{\phi}_N(\rho)\}$ shows that the mapping is isometric. The completeness of the basis $\{\overline{\phi}_N(\rho)\}$ in $\mathcal{J}_{nl}^+$ was found in (2.11)–(2.12) and, moreover, we can perform directly[19]

$$\int_0^\infty r^{n-1}\,dr\,A_{nl}(\rho,r)A_{nl}(\rho',r)^* = K_{nl}(\rho,\rho'), \tag{3.9}$$

when (1.7a) can be performed, i.e., when the kernel (3.1) is bounded, namely for $\mathrm{Im}(a/b) \geq 0$ $(v\geq 0)$ or, when $a=0$, b should be real. As $(\bar{f},\bar{g})_{nl} = (f,g)_0$ for any f,g in $\mathcal{H}_n^+$, the mapping is unitary and the existence conditions are identical with those found in I for the linear complex transforms.

4. LIMITS AND PARTICULAR CASES

Real transformations: We want to show that, as in I, when the transformation parameters a,b,c,d in (1.1) become real, the space $\mathcal{J}_{nl}^+$ with a scalar product (1.10) over $\mathbb{C}^+$ collapses to $\mathcal{H}_n^+$ with a scalar product (1.7) over $\mathbb{R}^+$. The said limit involves first determining the behavior of the weight function in (2.6) as, in (2.5), $v\to 0$ and, since $|u|^2+vw=1$ with $w\equiv 2\,\mathrm{Im}\,c^*d$, for $u=\omega\exp(i\varphi)$, $\omega\to 1$. Recalling that[20] $K_\mu(z)\sim[\pi/2z]^{1/2}e^{-z}$ as $|z|\to\infty$, $\omega=|1-vw|^{1/2}\sim 1-\tfrac{1}{2}vw$, $\mathrm{l.i.m.}\,\epsilon^{-1/2}\times\exp[-z^2/\epsilon]=\pi^{1/2}\delta(z)$ for real positive $\epsilon\to 0$ and the fact that $\nu_{nl}(\rho,\rho^*)$ is under the integral $\int_{\mathbb{C}^+} d\,\mathrm{Re}\rho\, d\,\mathrm{Im}\rho = \int_0^\infty |\rho|\,d|\rho|\int_{-\pi/2}^{\pi/2} d\theta$,

$$\begin{aligned}
\underset{v\to 0}{\mathrm{l.i.m.}}\,\nu_{nl}(\rho,\rho^*) &= \underset{v\to 0}{\mathrm{l.i.m.}}\,[2/\pi v]^{1/2}\exp[-(|\rho|^2/v)(1-\cos\{\varphi+2\theta\}) \\
&\quad -\tfrac{1}{2}w|\rho|^2\cos\{\varphi+2\theta\}] \\
&= |\rho|^{n-1}\delta(|\rho|\sin(\tfrac{1}{2}\varphi+\theta))\exp[-\tfrac{1}{2}w|\rho|^2\cos(\varphi+2\theta)] \\
&= |\rho|^{n-1}[\delta(\tfrac{1}{2}\varphi+\theta)+\delta(\tfrac{1}{2}\varphi+\theta-\pi)]\exp(-\tfrac{1}{2}w|\rho|^2).
\end{aligned} \tag{4.1}$$

Now, as $\theta\in[-\tfrac{1}{2}\pi,\tfrac{1}{2}\pi)$, only the first δ contributes to pick out the value $\theta=\tfrac{1}{2}\varphi$ in the integral, so that for $r'=|\rho|$,

$$\lim_{v\to 0}\int_{\mathbb{C}^+} d\mu_{nl}(\rho)\bar{f}(\rho)^*\bar{g}(\rho) = \int_{\mathbb{R}^+e^{-i\varphi/2}} r'^{n-1}\,dr'\exp(-wr'^2/2) \times\bar{f}(r')^*\bar{g}(r'), \tag{4.2}$$

and the normalization coefficient for ν_{nl} is thus seen to be the appropriate one and the parameter l has disappeared from the right-hand side of (4.2). Since ur'^2 is real, from the discussion below Eq. (2.9) we can see that the functions $\bar{f},\bar{g}$ must be of growth $(2,1/2v-\omega/2v)\sim(2,\tfrac{1}{4}w)$ in r'. In the limit when the transformation parameters become real, $w\to 0$ and $\varphi\to 0$, the integral in the right-hand side of (4.2) is over $\mathbb{R}^+$ and $\mathcal{J}_{nl}^+$ has become identical with $\mathcal{H}_n^+$. The transform kernel $A_{nl}(\rho,r)$ in (3.1) is uneventful in this limit and now becomes a transform in $\mathcal{H}_n^+$ which coincides with the unitary representations of $SO(2,1)$ in "radial" space.[21]

Transformations where $b\to 0$ can be obtained out of the development above since $b\to 0$ implies $v\to 0$, plus the analysis of the behavior of $A_{nl}(\rho,r)$ in (3.1). It can be shown[22] with due care to the phases involved for $r\geq 0$, $\arg r'\in[-\tfrac{1}{2}\pi,\tfrac{1}{2}\pi)$,

$$\underset{b\to 0}{\mathrm{l.i.m.}}\,A_{nl}(r',r) = r'^{1-n}a^{n/2-1}\delta(r-a^{-1}r')\exp[ic/2a)r'^2]. \tag{4.3}$$

Since $\arg a = -\tfrac{1}{2}\arg u = -\tfrac{1}{2}\varphi = \theta = \arg r'$, (4.2) acts under the line integral over $\mathbb{R}^+\exp(-i\varphi/2)$ with the appropriate phase relation between r and r'. The case $a=1$, $c=iq$, q real >0 was used in Ref. 7 to reproduce the matrix elements of a Gaussian potential. The identity transformation is now obtained by simply setting $a=1$, $c=0$ in (4.2), and $A_{nl}(r',r)$ is seen to become the reproducing kernel under the scalar product (1.7). It is thus seen that our choice of the phase factor (3.1b) is appropriate.

The Hankel transform is obtained when, as for the ordinary Fourier transform in (1.1), we set $a=0=d$, $b=1=-c$. The transform kernel becomes[23]

$$A_{nl}^H(r',r) = \vartheta_{nl}(r'r)^{1-n/2}J_{n/2+l-1}(r'r). \tag{4.4}$$

The Barut-Girardello transform[11] was introduced in developing the formalism for coherent states associated with noncompact groups, these being eigenstates of the lowering operator of an $so(2,1)$ algebra in the ("discrete") $D^\pm(\Phi)$ representations $(\Phi=-\tfrac{1}{2},-1,-\tfrac{3}{2},\cdots)$. It can be obtained as a particular case of complex radial transforms for the values (1.2) of the parameters. The scalar product in the $\mathcal{J}_{nl}^+$ space has the weight function

$$\nu_{nl}^B(\rho,\rho^*) = 2\pi^{-1}|\rho|^n K_{n/2+l-1}(|\rho|^2). \tag{4.5a}$$

Similarly, the transform kernel becomes

$$A_{nl}^B(\rho,r) = 2^{1/2}(\rho r)^{1-n/2}\exp[-\tfrac{1}{2}(r^2+\rho^2)]J_{n/2+l-1}(2^{1/2}\rho r), \tag{4.5b}$$

and the orthonormal basis

$$\overline{\phi}_N^B(\rho) = (-1)^N[2^{n/2-1}N!\,\Gamma(N+\tfrac{1}{2}n+l)]^{-1/2}(2^{-1/2}\rho)^{2N+l}, \tag{4.5c}$$

with the reproducing kernel

$$K_{nl}^B(\rho,\rho') = (\rho\rho'^*)^{1-n/2}I_{n/2+l-1}(\rho\rho'^*). \tag{4.5d}$$

When $l=0$, this agrees with the scalar product in the Barut–Girardello[24] space $z\equiv\tfrac{1}{2}\rho^2\in\mathbb{C}$ for $D^+(\Phi)$ when the latter is multiplied by a factor of $2^{n/2-2}\Gamma(\tfrac{1}{2}n)$ and we set $\Phi=-\tfrac{1}{4}n$. The results of Ref. 7 are regained when we multiply our weight function by a factor $2^{n/2-2}$ and set $|q|=\tfrac{1}{2}n-1$, integer.[25] It should be noticed that the basis functions (3.6) are bases for an $so(2,1)$ representation given by the eigenvalue of $I^2=I_3^2-I_1^2-I_2^2$ obtained from (3.5)–(3.7) to be $Q\equiv\frac{1}{4}[(\frac{1}{2}n+l-1)^2-1]=\Phi(\Phi+1)$ i.e., labelled by $\Phi=-\frac{1}{2}\pm\frac{1}{2}(\frac{1}{2}n+l-1)$. Multivalued "discrete series" representations of the $SO(2,1)$ group are important as can be seen from the fact that for the ordinary one-dimensional harmonic oscillator ($n=1$, $\lambda=0$) we have the $\Phi=-\frac{1}{4}$ and $-\frac{3}{4}$ representations of $SO(2,1)$.[26]

The one-dimensional "radial" spaces are the cases when $n=1$. As no angular momentum operators exist, in (1.8), $0=\lambda=-l(l-1)$. There are two solutions for this: $l=0$ and $l=1$, i.e., $\frac{1}{2}n+l-1=\mp\frac{1}{2}$, and correspondingly two spaces, $\mathcal{F}_{10}^{+}$ and $\mathcal{F}_{11}^{+}$ are transforms of $\mathcal{H}_1^{+}$. The weight function in both spaces is, recalling $K_{\pm 1/2}(z)=[\pi/2z]^{1/2}e^{-z}$,

$$\nu_1(\rho,\rho^*)=2(2\pi v)^{-1/2}\exp[(1/2v)(u\rho^2-2\rho\rho^*+u^*\rho^{*2})]$$
$$\equiv\nu^1(\rho,\rho^*), \tag{4.6}$$

which is formally identical to the weight for the complex *linear* transform spaces in I. It has to be recalled, however, that, there,[27] the scalar product involves integration over all of $\mathbb{C}$. We shall explain this below. The two transform kernels are, using the particular expressions for $J_{\pm 1/2}$,

$$A_{10}(\rho,r)=\exp(-i\pi/4)(2/\pi b)^{1/2}$$
$$\times\exp[(i/2b)(ar^2+d\rho^2)]\cos(\rho r/b), \tag{4.7a}$$
$$A_{11}(\rho,r)=-i\exp(-i\pi/4)(2/\pi b)^{1/2}$$
$$\times\exp[(i/2b)(ar^2+d\rho^2)]\sin(\rho r/b). \tag{4.7b}$$

Hence in $\mathcal{F}_{10}^{+}$, the transform functions have the property $\bar{f}_0(\rho)=\bar{f}_0(-\rho)$ under inversion of the space, while in $\mathcal{F}_{11}^{+}$ $\bar{f}_0(\rho)=-\bar{f}_0(-\rho)$, as can be seen from the bases (2.9). Now if for a given function $f(r)$ on $r\in\mathbb{R}^+$ we extend the domain to the whole of $\mathbb{R}$ and write $f(r)=f_+(r)+f_-(r)$, $f_\pm(r)=\frac{1}{2}[f(r)\pm f(-r)]$, expanding f into its odd and even components and further demand that a transform $\bar{f}(\rho)$ have the same parity under inversion of the argument as the original function [this corresponds to having L^2 with the same eigenvalue λ in both spaces, the transformation properties under $O(n)$ now collapsing to C_2], we can write $\bar{f}_0$ as the transform of f_+ and $\bar{f}_1$ as that of f_-. Suppressing arguments,

$$\bar{f}\equiv\bar{f}_0+\bar{f}_1=\int_{\mathbb{R}^+}dr\,A_{10}f_+ +\int_{\mathbb{R}^+}dr\,A_{11}f_-=\int_{\mathbb{R}^+}dr\,A^1 f \tag{4.8}$$

with

$$A^1(\rho,r)\equiv\tfrac{1}{2}(A_{10}+A_{11})(\rho,r)$$
$$=(2\pi b)^{-1/2}\exp(-i\pi/4)\exp[(i/2b)(ar^2-2r\rho+d\rho^2)], \tag{4.9}$$

regaining the complex linear transform in I between $\mathcal{H}\equiv\mathcal{L}^2(-\infty,\infty)$ and $\mathcal{F}$ with the scalar product

$$(\bar{f},\bar{g})^1=2(\bar{f}_0,\bar{g}_0)_{10}+2(\bar{f}_1,\bar{g}_1)_{11}$$
$$=\int_{\mathbb{C}}d\,\mathrm{Re}\rho\,d\,\mathrm{Im}\rho\,\nu^1(\rho,\rho^*)\bar{f}(\rho)^*\bar{g}(\rho). \tag{4.10}$$

For the values (1.2) of the parameters, this is the Bargmann transform.[2]

Another, quite different, way of obtaining back the complex linear transforms is to follow the procedure of Barut and Girardello[11] of considering functions of $z=\epsilon^{-1/2}$ with $z=\frac{1}{2}\rho^2$ and letting $n\to\infty$ such that ϵn remain a finite number. This effects the contraction of the representations of the $so(2,1)$ algebra in (2.14) in the orthonormal basis (2.13) to that of the Heisenberg algebra. The limiting procedure is a delicate one, and we shall not pursue this point further.

5. COHERENT STATES FOR THE RADIAL HARMONIC OSCILLATOR WITH A CENTRIFUGAL FORCE

The Bargmann transform has proven to be the natural tool for the construction of coherent states for the harmonic oscillator since they map the eigenstates $\psi_N(x)$ of the one-dimensional system on functions of the complex variable $z\in\mathbb{C}$, $\bar{\psi}_N(z)=[(2\pi)^{1/2}N!]^{-1/2}z^N$ (using the normalization of I). The coherent states, defined[28] as $|z)=\sum|N\rangle\bar{\psi}_N(z)$ are eigenstates of the lowering operator $\hat{z}=2^{-1/2}(\hat{x}+i\hat{p})$ with eigenvalue z. They resolve the identity as $\mathbb{1}=\int|z)\,d\mu^1(z)(z|$ [using the measure $d\mu^1(z)$ of I] and are overcomplete[29] as $(z|z')=K^1(z,z')$, the reproducing kernel in the scalar product with measure $d\mu^1(z)$.

A similar construction for the radial functions of an n-dimensional harmonic oscillator with centrifugal force can now be made. The angular part of the wavefunctions continues to be the n-dimensional spherical harmonic in the $n-1$ angles of real or complex space as in (3.5a) (see Appendix B). We shall now examine the proper quantum-mechanical solutions of the radial part of the operator (3.5a). These are (3.6) plus the conditions that I_3 be self-adjoint between them, which means that the constant terms in the partial integrations be zero (which imposes conditions on the behavior of the functions at $r=0$) and that ψ_{Nl}, $r^{-1}\psi_{Nl}$, and $(d/dr)\psi_{Nl}$ be square-integrable.[30] From (3.5b) we see that for each n, L, and g, the two solutions

$$l_\pm=(1-\tfrac{1}{2}n)\pm[(1-\tfrac{1}{2}n)^2+L(L+n-2)+g]^{1/2} \tag{5.1}$$

are real for centrifugal forces which include attractive ones but which are not more attractive than those allowed by the zero of the discriminant for the lowest angular momentum $L=0$ namely

$$g\geqslant-(1-\tfrac{1}{2}n)^2. \tag{5.2}$$

Given this condition is fulfilled, square-integrability of ψ_{Nl} under the scalar product in $\mathcal{H}_n^{+}$ (since it is assured that the behavior at infinity is adequate), places restrictions on the behavior at the origin: $l>-\frac{1}{2}n$. The same conditions on $r^{-1}\psi_{Nl}$ and $(d/dr)\psi_{Nl}$ narrows the choice to $l>1-\frac{1}{2}n$. Hence, only l_+ of the two choices in (5.1) is possible for general g and n satisfying (5.2). Only in the case when the latter two conditions are absent (i.e., $g=0$, $n=1$, $L\equiv 0$, and $l_-=0$), do we need the two solutions of (5.1). This is convenient since for all cases, except the one-dimensional oscillator with no centrifugal force, the space $\mathcal{F}_{nl_+}^{+}$ contains all the states of the system for a given angular momentum[26] L. Henceforth denote $l_+\equiv l(L,n,g)$. Recalling (4.4) define now the kets

$$|\rho)_{nL}\equiv\sum_{N=0}^{\infty}|N\rangle_{nl}\bar{\phi}_N^B(\rho)$$

$$=2^{-(n/2+l-1)/2}\rho^l \sum_{N=0}^{\infty} |N\rangle_{nl}[N!\Gamma(N+\tfrac{1}{2}n+l)]^{-1/2}(-1)^N\rho^{2N}, \quad \rho\in\mathbb{C}^+, \tag{5.3}$$

where $|N\rangle_{nl}$ stands for the state (3.6). The ket (5.3) can be considered as a *coherent* state for the system since it is an eigenket of the lowering operator defined, parallel to (2.14), with (3.7) as

$$L\equiv I_1-iI_2=-\tfrac{1}{2}[2^{-1/2}(\hat{\mathrm{x}}+i\hat{\mathrm{p}})]^2+\tfrac{1}{4}gr^{-2} \tag{5.4a}$$

with eigenvalue $-\frac{1}{2}\rho^2$, as the bracketing suggests for $g\to 0$. This can be proven immediately using the $so(2,1)$ raising and lowering operator matrix elements (2.13):

$$\begin{aligned}L|\rho)_{nL}&=\sum_{N=0}^{\infty}[N(N+\tfrac{1}{2}n+l-1)]^{1/2}|N-1\rangle_{nl}\overline{\phi}_N^B\\&=\sum_{N'=0}^{\infty}|N'\rangle_{nl}[(N'+1)(N'+\tfrac{1}{2}n+l)]^{1/2}\overline{\phi}_{N'+1}^B\\&=\sum_{N'=0}^{\infty}|N'\rangle_{nl}\mathcal{R}\,\overline{\phi}_{N'}^B=-\tfrac{1}{2}\rho^2|\rho)_{nL}.\end{aligned} \tag{5.5}$$

The usual coherent-state properties follow,[28] as ${}_{nL}(\rho|\rho')_{nL}=K_{nl}^B(\rho,\rho')$ and $\int_{\mathbb{C}^+}|\rho)_{nL}\,d\mu^B(\rho)_{nL}(\rho|=\mathbb{1}$. It would seem desirable to change the labels $z=\frac{1}{2}\rho^2\in\mathbb{C}$ so as to coincide with the treatment in Ref. 11 with $l=0$ and $n=-4\Phi$. There is the problem, however, that for $l\neq 2\times$ integer, an $f(z)=(\rho|f)$ would not be an entire function of z, but one with a branch cut from 0 to ∞. A completeness statement[29] on the coherent states (5.3) is also wanting. Since a connection exists between the radial differential equations of the harmonic oscillator and Coulomb systems,[13,31] one expects that similar coherent states can be defined for the latter. This will be taken up elsewhere.

6. COMPOSITION OF TRANSFORMS AND REPRESENTATIONS OF $HSL(2,\mathbb{C})$

Two related topics which are virtually identical with their counterparts for complex linear transforms will now be presented in the briefest manner. The first one pertains the possibility of composition of transforms, seen as *active* transformations $\mathcal{A}_1:\mathcal{H}^+=\mathcal{J}_1^+$ and $\mathcal{A}_2:\mathcal{H}^+=\mathcal{J}_2^+$ into one transform $\mathcal{J}_2^+=\mathcal{A}_2\mathcal{A}_1^{-1}\mathcal{J}_1^+\equiv\mathcal{A}_{21}\mathcal{J}_1^+$ between $\mathcal{J}_1^+$ and $\mathcal{J}_2^+$ with the same n,l but differing in the parameters a,b,c,d, as

$$\bar{f}^{(2)}(\rho)=\int_{\mathbb{C}^+}d\mu_1(\rho')A_{(2,1)}(\rho,\rho')\bar{f}^{(1)}(\rho'), \tag{6.1a}$$

$$\bar{f}^{(1)}(\rho')=\int_{\mathbb{C}^+}d\mu_2(\rho)A_{(2,1)}(\rho,\rho')^*\bar{f}^{(2)}(\rho), \tag{6.1b}$$

where $d\mu_1(\rho')$ and $d\mu_2(\rho)$ are the corresponding measures and the transform kernel is

$$\begin{aligned}A_{(2,1)}(\rho,\rho')&=\int_{\mathbb{R}^+}r^{n-1}\,dr\,A_{(2)}(\rho,r)A_{(1)}(\rho',r)^*\\&=\Phi(b_2,-b_1^*;b)^n\vartheta_{nl}b^{-1}\exp[(i/2b)(a\rho'^{*2}+d\rho^2)]\\&\quad\times J_{n/2+l-1}(\rho\rho'^*/b)\\&=A_{(1,2)}(\rho',\rho)^*\end{aligned} \tag{6.2a}$$

where

$$M\equiv\begin{pmatrix}a&b\\c&d\end{pmatrix}=\begin{pmatrix}a_2&b_2\\c_2&d_2\end{pmatrix}\begin{pmatrix}a_1&b_1\\c_1&d_1\end{pmatrix}^{*-1} \tag{6.2b}$$

and

$$\Phi(b',b'';b)=\exp\{-\tfrac{1}{2}i[\arg b'+\arg b''-\arg b-\arg(b'b''/b)]\}=\pm1, \tag{6.2c}$$

when the conditions for existence of $\mathcal{A}_1$ and $\mathcal{A}_2$ are fulfilled [i.e., $\mathrm{Im}(a_1/b_1)\geq 0$ and $\mathrm{Im}(a_2/b_2)\geq 0$, etc].

The second point is that the composition of transforms can also be seen as that of *passive* transformations of the space $\mathcal{H}_n^+$ onto itself through a set of operators (6.2) and such that with each matrix M as defined in (6.2b) we associate a "D function"

$$D_{rr'}^{(0)}(M)=A_{(2,1)}(r,r') \tag{6.3}$$

which satisfies[32]

$$\int_{\mathbb{R}^+}r'^{n-1}\,dr'\,D_{rr'}^{(0);n,l}(M_1)\,D_{r'r''}^{(0);n,l}(M_2)=\Phi(b_1,b_2;b_{12})^nD_{rr''}^{(0);n,l}(M_1M_2). \tag{6.4}$$

We have thus a ray representation of that subset of $M\in SL(2,\mathbb{C})$ for which integration is possible. The conditions for the kernels to be bounded (or Hilbert–Schmidt) were examined in I. This forms a subsemigroup of $SL(2,\mathbb{C})$ called $HSL(2,\mathbb{C})$ in Ref. 7 and (6.3) is a representation of $HSL(2,\mathbb{C})$ labeled by the indices n,l. A continuum of such representations can be built as

$$D_{\rho,\rho'}^{(k);n,l}(M)\equiv D_{\rho,\rho'}^{(0);n,l}(M_kMM_k^{*-1})=D_{\rho',\rho}^{(k);n,l}(M^{*-1})^*, \tag{6.5}$$

for $M_k\in HSL(2,\mathbb{C})$, with a composition law which replaces the integration over $\mathbb{R}^+$ with $\int_{\mathbb{C}^+}d\mu_k(\rho)$. From (6.5) we see that for $M\in SL(2,\mathbb{R})\subset HSL(2,\mathbb{C})$, the representation is unitary.

7. CANONICAL TRANSFORMATIONS IN QUANTUM MECHANICS, EXTENDED

In the way of conclusion, the results of I and this paper seem to indicate that the definition of a canonical transformation in quantum mechanics as that which preserves the Heisenberg algebra[9] in (1.1c) can be extended. Equation (1.1c) is the quantum analog of the classical concept of a canonical transformation to that which preserves the Poisson bracket between canonically conjugate variables. The validity of (1.1c) is thus restricted to those transformations where the new operators $\hat{\eta}$ and $\hat{\xi}$ exist and have the same domain and spectrum as the original, usual $\hat{x}$ and $\hat{p}$. Classical mechanics can work with the radial coordinate r and it conjugate momentum p_r and establish that (1.5) is a proper canonical transformation and, being a local theory, avoid specifying what happens at $r=0$. The translation of (1.5) to quantum mechanics appears difficult, as operators "$\hat{\rho}$" and "$\hat{p}_\rho$" are not of the usual kind as they have no self-adjoint extension.[33]

The picture we seem to be arriving at overcomes this limitation on two accounts: First, we make use of operators which *are* properly defined [as the $so(2,1)$ generators (3.5a)–(3.7) or their linear combinations $\hat{\mathrm{x}}^2$, $\frac{1}{2}(\hat{\mathrm{x}}\cdot\hat{\mathrm{p}}+\hat{\mathrm{p}}\cdot\hat{\mathrm{x}})$ and $\hat{\mathrm{p}}^2$ with the extra centrifugal force term added to the angular momentum one] and say that the transformation

$$\begin{pmatrix} I_1' \\ I_2' \\ I_3' \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[a^2-b^2-c^2+d^2] & -ab+cd & \frac{1}{2}[-a^2-b^2+c^2+d^2] \\ -ac+bd & -ac-bd & ac+bd \\ \frac{1}{2}[-a^2+b^2-c^2+d^2] & ab+cd & \frac{1}{2}[a^2+b^2+c^2+d^2] \end{pmatrix} \times \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \tag{7.1}$$

obtained from (1.1a), (1.3) and (3.5)–(3.7) is canonical in this extended context since, as can be verified

$$[I_j, I_k] = i\epsilon_l I_l \Leftrightarrow [I_j', I_k'] = i\epsilon_l I_l' \tag{7.2}$$

with (j,k,l) cyclic permutations of $(1,2,3)$ and $\epsilon_1=\epsilon_2=-\epsilon_3=1$. The $so(2,1)$ algebra is thus conserved and we can turn the procedure of finding the weight function ν_{nl} and transform kernel A_{nl} to stem from (6.1) and the hermiticity conditions on the $\{I_j'\}$ implied by the $\{I_j\}$ being self-adjoint. Although a Heisenberg algebra is undefined here, ρ itself retains the meaning of an underlying space variable. The classical limit of (6.1) is (1.5).

Second, we have permitted the transformation parameters a,b,c,d to be complex. This is in line with the fact that quantum mechanics allows—indeed needs—the complex field as the domain of definition of its functions. The consequence of the second extension is to require Hilbert spaces of functions which include the usual Dirac[10] and Bargmann[2,3] spaces. The transformation (6.1) is the most general one allowed by (6.2), since the group of linear real automorphisms of the algebra $so(2,1)$ is $O(2,1)$ and its complexification is $SL(2,\mathbb{C})$.

Among the canonical transformations which have been useful in classical mechanics is the one mapping the phase-space coordinates on a conserved quantity—angular momentum or the Hamiltonian—and its conjugate—angle or time. One of the aims of this program[26] is to give an extended quantum mechanical meaning to these mappings.

ACKNOWLEDGMENT

I would like to thank Dr. Charles P. Boyer for several useful discussions.

APPENDIX A: REALIZATION THROUGH HYPERDIFFERENTIAL OPERATORS

As in I, we introduce a Lie algebra structure for the $SL(2,\mathbb{C})$ set of canonical transforms, disregarding the Hilbert-space structure of the functions involved, as

$$\bar{f}(r) = \int_{\mathbb{R}^+} r'^{n-1}\, dr' A_{nl\,(\tau)}(r,r') f(r') = \exp\left[i\tau H\left(r, \frac{d}{dr}\right)\right] f(r) \tag{A1}$$

where τ labels one-parameter subgroups and asking only the integrals involved to exist. The operator $H(r, d/dr)$ need not be bounded.[34] The differential operator $H(r,d/dr)$ can be found by inspection from

$$H\left(r, \frac{d}{dr}\right) f(r) = -i\int_{\mathbb{R}^+} r'^{n-1}\, dr' \left(\frac{\partial}{\partial\tau} A_{nl\,(\tau)}(r,r')\Big|_{\tau=0}\right) f(r') \tag{A2}$$

and by using the differential equations satisfied by the integration kernel, to pass the partial derivatives to act on f through partial integration, assuming the constant terms to vanish.

In agreement with what we expect from I, we find

$$\exp[ic\tfrac{1}{2}(r^2)] = \exp[ic\tfrac{1}{2}\hat{\mathbf{x}}^2]: \quad \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}, \tag{A3a}$$

$$\exp(ib\tfrac{1}{2}\{\partial_r^2 + [(n-1)/r]\partial_r + \lambda/r^2\}) = \exp(-ib\tfrac{1}{2}\hat{\mathbf{p}}^2): \quad \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \tag{A3b}$$

$$\exp(i\alpha\tfrac{1}{4}\{\partial_r^2 + [(n-1)/r]\partial_r + \lambda/r^2 + r^2\}) = \exp[-i\alpha\tfrac{1}{4}(\hat{\mathbf{p}}^2 - \hat{\mathbf{x}}^2)]: \quad \begin{pmatrix} \cosh\frac{1}{2}\alpha & \sinh\frac{1}{2}\alpha \\ \sinh\frac{1}{2}\alpha & \cosh\frac{1}{2}\alpha \end{pmatrix}, \tag{A3c}$$

$$\exp[-\beta(r\partial_r + \tfrac{1}{2}n)] = \exp[-i\beta\tfrac{1}{4}(\hat{\mathbf{x}}\cdot\hat{\mathbf{p}} + \hat{\mathbf{p}}\cdot\hat{\mathbf{x}})]: \quad \begin{pmatrix} e^{\beta/2} & 0 \\ 0 & e^{-\beta/2} \end{pmatrix}, \tag{A3d}$$

$$\exp(i\gamma\tfrac{1}{4}\{\partial_r^2 + [(n-1)/r]\partial_r + \lambda/r^2 - r^2\}) = \exp[-i\gamma\tfrac{1}{4}(\hat{\mathbf{p}}^2 + \hat{\mathbf{x}}^2)]: \quad \begin{pmatrix} \cos\frac{1}{2}\gamma & \sin\frac{1}{2}\gamma \\ -\sin\frac{1}{2}\gamma & \cos\frac{1}{2}\gamma \end{pmatrix}. \tag{A3e}$$

The generators of the last three transforms constitute the $so(2,1)$ dynamical algebra for the radial oscillator with centrifugal force. Associating thus products of 2×2 complex matrices to hyperdifferential operators yields Baker–Campbell–Hausdorff relations[35] including ∂_r^2, $(1/r)\partial_r$, $r\partial_r$, r^2, and r^{-2} terms. A particular composition used in I is

$$\begin{pmatrix} \cosh\theta & -\sinh\theta \\ -\sinh\theta & \cosh\theta \end{pmatrix} = \begin{pmatrix} 1 & -\tanh\theta \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1/\cosh\theta & 0 \\ 0 & \cosh\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\tanh\theta & 1 \end{pmatrix} \tag{A4}$$

and involves the use of (A3) for $b=-\tanh\theta=c$, $\beta=-2\ln\cosh\theta$. Rather than write the lengthy resulting relation, we take $\theta=i\frac{1}{4}\pi$. This gives the Bargmann (i.e., Barut–Girardello, for arbitrary l) transform (4.4) as

$$\bar{f}(r) = \exp\left[\tfrac{1}{8}\pi\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{\lambda}{dr} + \frac{\lambda}{r^2} + r^2\right) f(r)\right] = 2^{-n/4}\exp\left[\frac{1}{2}\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{d}{dr} + \frac{\lambda}{r^2}\right)\right] e^{r^2/4} f(2^{-1/2}r). \tag{A5}$$

Writing for f the radial wavefunction (3.6) and for $\bar{f}$ the corresponding (3.8) [i.e., (2.9) for $u=0$, $v=1$] and recalling (1.8), we obtain

$$\exp\left[\frac{1}{2}\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{d}{dr} - \frac{l(l+n-2)}{r^2}\right)\right](2^{-1/2}r)^l L_N^{(n/2+l-1)}(\tfrac{1}{2}r^2) = \frac{(-1)^N}{N!}(2^{-1/2}r)^{2N+l}. \tag{A6}$$

A special function relation which seems to be new is obtained setting $z=\frac{1}{2}r^2$ and inverting (A6) as

$$\exp\left[-\left(z\frac{d^2}{dz^2}+\tfrac{1}{2}n\frac{d}{dz}-\frac{l(l+n-2)}{4z}\right)\right]z^{N+l/2}$$
$$=(-1)^N N!\, z^{l/2}L_N^{(n/2+l-1)}(z) \tag{A7}$$

and can be verified to hold independently by expanding in series.

APPENDIX B: THE PASSAGE FROM n-DIMENSIONAL TO RADIAL TRANSFORMS

In I, Appendix B, we gave results concerning the extension to n dimensions of the complex linear transforms. For the case when the canonical transform is of the type (1.1), that is, when the transformation submatrices A,B,C,D of $n\times n$, are multiples a,b,c,d of $\mathbb{1}$, these take the form

$$A^n(\eta,\mathbf{x})=\{(2\pi|b|)^{-1/2}\exp[-\tfrac{1}{2}i(\tfrac{1}{2}\pi+\arg b)]\}^n$$
$$\times\exp[(i/2b)(a\mathbf{x}^2-2\mathbf{x}\cdot\eta+d\eta^2)]. \tag{B1}$$

the integration over $\mathbf{x}$-space being over $\mathbb{R}^n$, with measure $d^n\mathbf{x}$, and the scalar product $(\bar{f},\bar{g})^{(n)}$ involving an integration over η-space, over $\mathbb{C}^n$ with measure $\nu^n(\eta,\eta^*)\,d^n\,\mathrm{Re}\eta\, d^n\,\mathrm{Im}\eta$,

$$\nu^n(\eta,\eta')=(\tfrac{1}{2}\pi v)^{-n/2}\exp[(1/2v)(u\eta^2-2\eta\cdot\eta^*+u^*\eta^{*2})]. \tag{B2}$$

We want to show here how expressions (B1) and (B2) relate to the corresponding radial kernel (3.1) and measure (1.10)–(2.6). Consider first the two-dimensional case ($n=2$). Parametrize $\mathbb{R}^2$ as $x_1=r\sin\theta$, $x_2=r\cos\theta$ with $r\in[0,\infty)$, $\theta\in[0,2\pi)$, and $d^2\mathbf{x}=r\,dr\,d\theta$. Now parametrize $\mathbb{C}^2$ as $\eta_1=\rho\sin\Theta$, $\eta_2=\rho\cos\Theta$ with $\rho\in\mathbb{C}^+$ {i.e., $\arg\rho\in[-\frac{1}{2}\pi,\frac{1}{2}\pi)$}, $\mathrm{Re}\Theta\in[0,2\pi)$, $\mathrm{Im}\Theta\in(-\infty,\infty)$. Noticing that if $y=f(z)$ and $dy=f'(z)\,dz$, then $d\,\mathrm{Re}y\,d\,\mathrm{Im}y=|f'(z)|^2\,d\,\mathrm{Re}z\,d\,\mathrm{Im}z$, we have that the measure in $\mathbb{C}^2$ is

$$d^2\mathrm{Re}\eta\, d^2\mathrm{Im}\eta=|\rho|^2\,d\mathrm{Re}\rho\, d\mathrm{Im}\rho\, d\mathrm{Re}\Theta\, d\mathrm{Im}\Theta.$$

Now, using $\mathbf{x}\cdot\eta=r\rho\cos(\theta-\Theta)$ and the Bessel generating function, we have

$$A^2(\eta,\mathbf{x})=(2\pi b)^{-1}\exp(-i\tfrac{1}{2}\pi)\exp[(i/2b)(ar^2+d\rho^2)]$$
$$\times\sum_{m=-\infty}^{\infty}(-i\exp[-i(\theta-\Theta)])^m J_m(\rho r/b)$$
$$=\sum_{m=-\infty}^{\infty}A_{2,m}(\rho,r)[(2\pi)^{-1/2}\exp(im\theta)]^*$$
$$\times[(2\pi)^{-1/2}\exp(im\Theta)], \tag{B3}$$

where $A_{2,m}(\rho,r)$ is given, with correct phase and normalization, by (3.1). This means that if we have a function $f(\mathbf{x})$ of definite eigenvalue m under L_{12} in the form $f_m(r)[(2\pi)^{-1/2}\exp(im\theta)]$ (so that the angular part be normalized), then

$$\bar{f}(\eta)=\int_{\mathbb{R}^2}d^2\mathbf{x}\,A^2(\eta,\mathbf{x})f(\mathbf{x})$$
$$=\int_{\mathbb{R}^+}r\,dr\,A_{2,m}(\rho,r)f_m(r)[(2\pi)^{-1/2}\exp(im\Theta)]$$
$$=\bar{f}_m(\rho)[(2\pi)^{-1/2}\exp(im\Theta)]. \tag{B4}$$

and the dependence of $\bar{f}$ on Θ is the same as that of f on θ (the range of the former being now over a strip in the complex plane), and only a transform of the *radial* part has taken place. The scalar product in the transform space of two such functions can now be calculated using (B2) and $\eta\cdot\eta^*=|\rho|^2\cosh(2\,\mathrm{Im}\Theta)$, and an integral representation of the Macdonald function[36]

$$(\bar{f},\bar{g})^{(2)}=\int_{\mathbb{C}^2}d^2\mathrm{Re}\eta\, d^2\mathrm{Im}\eta\,\nu^2(\eta,\eta^*)$$
$$\times\{\bar{f}_m(\rho)(2\pi)^{-1/2}\exp[im(\mathrm{Re}\Theta+i\,\mathrm{Im}\Theta)]\}^*$$
$$\times\{\bar{g}_m(\rho)(2\pi)^{-1/2}\exp[im(\mathrm{Re}\Theta+i\,\mathrm{Im}\Theta)]\}$$
$$=(2/\pi v)\int_{\mathbb{C}^+}|\rho|^2\,d\mathrm{Re}\rho\, d\mathrm{Im}\rho$$
$$\times\exp[(1/2v)(u\rho^2+u^*\rho^{*2})]\bar{f}_m(\rho)^*\bar{g}_m(\rho)$$
$$\times\int_{-\infty}^{\infty}d\mathrm{Im}\Theta\exp[-(1/v)|\rho|^2\cosh(2\,\mathrm{Im}\Theta)]$$
$$\times\exp(-2m\,\mathrm{Im}\Theta)$$
$$=\int_{\mathbb{C}^+}d\mathrm{Re}\rho\, d\mathrm{Im}\rho\,\nu_{2,m}(\rho,\rho^*)\bar{f}_m(\rho)^*\bar{g}_m(\rho)$$
$$=(\bar{f}_m,\bar{g}_m)_{2,m}, \tag{B5}$$

where $\nu_{2m}(\rho,\rho^*)$ is given correctly by (2.6). Indeed, had we used different angle dependence for $\bar{f}$ and $\bar{g}$, a Kronecker δ in their eigenvalue under L_{12} would appear.

The problem for the n-dimensional case can be formulated similarly: Parametrize the real n-space $\mathbb{R}^n$ in the usual hyperspherical coordinates where the jth component reads $x_j=r\sin\theta_{n-1}\cdots\sin\theta_j\cos\theta_{j-1}$ for $1\le j\le n-1$ ($\theta_0\equiv 0$) and $x_n=r\cos\theta_{n-1}$. The ranges are $r\in[0,\infty)$, $\theta_1\in[0,2\pi)$, and $\theta_k\in[0,\pi]$ for $2\le k\le n-1$. Now parametrize the complex n-space $\mathbb{C}^n$ replacing r by ρ and θ_k by Θ_k with $\rho\in\mathbb{C}^+$. $\mathrm{Re}\Theta_k$ having the same ranges as θ_k and[37] $\mathrm{Im}\Theta_k\in(-\infty,\infty)$. The measure in $\mathbb{R}^n$ is $d^n\mathbf{x}=r^{n-1}dr\,d^{n-1}\omega_{n-1}$ with $d^{n-1}\omega_{n-1}=\sin^{n-2}\theta_{n-1}\,d\theta_{n-1}\,d^{n-2}\omega_{n-2}$ and $d\omega_1=d\theta_1$ while, in $\mathbb{C}^n$, $d^n\,\mathrm{Re}\eta\, d^n\,\mathrm{Im}\eta$ is found from the former with the weight function given by the absolute square of the weight function in $\mathbb{R}^n$. In order to express the n-dimensional transform kernel (B1) in a suitable way, expand the factor $\exp(-i\mathbf{x}\cdot\eta/b)$ in a series of Bessel times Gegenbauer polynomials,[38] the former in $r\rho/b$ and the latter in

$$\cos\theta_{n-1}\cos\Theta_{n-1}+\sin\theta_{n-1}\sin\Theta_{n-1}[\cos\theta_{n-2}\cos\Theta_{n-2}$$
$$+\sin\theta_{n-2}\sin\Theta_{n-2}(\cdots)]$$

which can be identified with a degenerate $SO(n)$ d^2_{000} function[39] and turned into a sum of products of hyperspherical harmonics in $\omega\equiv\{\theta_j\}$ and $\Omega\equiv\{\Theta_j\}$ as

$$\exp(-i\mathbf{x}\cdot\eta/b)=(2\pi)^{n/2}(r\rho/b)^{1-n/2}\sum_{l=0}^{\infty}\exp(-i\pi l/2)$$
$$\times J_{n/2+l-1}(r\rho/b)\sum_M Y_l^M(\omega)^*Y_l^M(\Omega), \tag{B6}$$

where the sum over the collective index M runs over the allowed $SO(n-1)\supset\cdots\supset SO(2)$ irreducible representation labels. Replacement of (B6) in (B1) and comparison with (3.1) gives

$$A^n(\eta,\mathbf{x})=\sum_{l=0}^{\infty}A_{nl}(\rho,r)\sum_M Y_l^M(\omega)^*Y_l^M(\Omega), \tag{B7}$$

which is the n-dimensional version of (B3) and which tells us, performing the integrations parallel to (B4) that the angular dependence of $\bar{f}$ is the same as that of f, with only the additional domain of the angles in the complex plane. Finally, in order to show the n-dimensional analog of (B5),

$$(\bar{f},\bar{g})^{(n)}=\int_{\mathbb{C}^n}d^n\,\mathrm{Re}\eta\, d^n\,\mathrm{Im}\eta\,\nu^n(\eta,\eta^*)$$
$$\times[\bar{f}_l(\rho)Y_l^M(\Omega)]^*[\bar{g}_l(\rho)Y_l^M(\Omega)]$$

$$= \int_{\mathbb{C}^+} d\,\mathrm{Re}\rho\, d\,\mathrm{Im}\rho\, v_{nl}(\rho, \rho^*) \bar{f}_l(\rho)^* \bar{g}_l(\rho)$$
$$= (\bar{f}_l, \bar{g}_l)_{nl}, \tag{B8}$$

we must prove

$$\int d^{n-1}\mathrm{Re}\Omega\, d^{n-1}\mathrm{Im}\Omega\, Y_l^M(\Omega)^* Y_l^M(\Omega) \exp[-(1/v)\eta\cdot\eta^*]$$
$$= (2\rho\rho^*/\pi v)^{1-n/2} K_{n/2+l-1}(\rho\rho^*/v), \tag{B9}$$

where the integration ranges over the strips in the complex plane of each of the angles as indicated above. The direct proof of Eq. (B9) is difficult. Differential or recursion-relation manipulations run into hopeless multiple integrals or combinatorics. A procedure which has allowed the verification of a fair number of individual cases for low l is that which uses the fact that (B9) is independent of M and shows that the Nth moment of the two sides of Eq. (B9) in $|\rho|^2$ are equal. For this, multiply Eq. (B9) by $(\rho\rho^*)^{2N+n+2l-1}$ and integrate over $\rho \in \mathbb{C}^+$. By using (2.6) and (2.9), the right-hand side has the value

$$\tfrac{1}{2}\pi^{n/2} v^n (2v)^{l+2N} N!\,\Gamma(N+\tfrac{1}{2}n+l)$$

while the left-hand side has become, for $\xi = v^{-1/2}\eta$ the Bargmann integral over $\mathbb{C}^n$ of the absolute square of $(\xi^2)^N \mathcal{Y}_l^l(\xi)$, where

$$\mathcal{Y}_l^l(\xi) = [\Gamma(\tfrac{1}{2}n+l)/2\pi^{n/2}\Gamma(l+1)]^{1/2}(\xi_1 + i\xi_2)^l$$

is the extreme, normalized, solid spherical harmonic. This seems to point out that no true Bargmann-type integral tables exist. The separation of n-dimensional integrals into radial and angular[40] parts can be seen as a step in that direction.

Note added in proof: It has been pointed out by Professor M. Toller that the semigroup $HSL(2,\mathbb{C})$ used here and in Ref. 1 has also been exploited in the harmonic analysis approach to multiperipheral dynamics. See G. Soliani and M. Toller, Nuovo Cimento **15**, 430 (1973) and S. Ferrara, G. Mattioli, G. Rossi, and M. Toller, Nucl. Phys. **B53**, 366 (1974). A particular case of Eq. (A7), for $l=0$, appears in C.M. King, M. Sc. Thesis, Auburn University (1963), unpublished.

[1]K. B. Wolf, J. Math. Phys. 15, 1295 (1974).
[2]V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).
[3]V. Bargmann, Commun. Pure Appl. Math. 20, 1 (1967); "Group Representations in Hilbert Spaces of Analytic Functions" in *Analytical Methods in Mathematical Physics*, edited by R. P. Gilbert and R. G. Newton (Gordon and Breach, New York, 1970).
[4]R.J. Glauber, Phys. Rev. 130, 2529 (1963); 131, 2766 (1963); P. Carruthers and M.M. Nieto, Am. J. Phys. 33, 537 (1965).
[5]M. Moshinsky and C. Quesne, J. Math. Phys. 12, 1772, 1780 (1971).
[6]M. Moshinsky, SIAM J. App. Math. 25, 193 (1973).
[7]P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex Extensions of Canonical Transformations and Quantum Mechanics" in *Group Theory and Its Applications, Vol. III*, edited by E.M. Loebl (Academic, New York, to be published).
[8]Ref. 1, Eq. (6.8).
[9]K. B. Wolf, "The Heisenberg—Weyl Ring in Quantum Mechanics," in *Group Theory and Its Applications, Vol. III*, edited by E. M. Loebl (Academic, New York, to be published).
[10]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford Clarendon Press, Oxford, 1958), 4th ed.
[11]A. O. Barut and L. Girardello, Commun. Math. Phys. 21, 41 (1971).
[12]Ref. 7, Sec. VI.
[13]M. Moshinsky, T.H. Seligman, and K.B. Wolf, J. Math. Phys. 13, 901 (1972).
[14]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Eq. 6.561.16.
[15]See Ref. 2, Sec. 1b, and Ref. 11, Sec. VI.
[16]Ref. 14, Eq. 8.975.3.
[17]Ref. 14, Eq. 7.421.4.
[18]Ref. 2, Sec. 2.
[19]Ref. 14, Eq. 6.633.2.
[20]Ref. 14, Eq. 8.451.6.
[21]Ref. 13, Eq. (2.17), which does not consider our phase factor for (3.1b), and Ref. 6, Eq. (6.14), which does.
[22]Ref. 14, Eq. 8.451.7.
[23]A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[24]Ref. 11, Eqs. (6.2), (6.3). As pointed out in Ref. 7, the index of the Macdonald function is off by a factor of 1/2 from its correct value.
[25]Ref. 7, Eqs. (6.17) and (6.23).
[26]C. P. Boyer and K. B. Wolf, "Canonical transforms, III. Configuration and phase descriptions of quantum systems posessing an $sl(2,R)$ Dynamical Algebra" J. Math Phys. 16, (to be published).
[27]Ref. 1, Eqs. (1.5) and (2.7).
[28]Here we follow the use given in Ref. 7, rather than the original work in Refs. 4, because of some convenient features such as integration over z always appearing in the form $\int_{\mathbb{C}} d\mu(z)$, etc. Our notation relates as $|z\rangle = (2\pi)^{1/4} \times \exp(-|z|^2/2)\,|z)$.
[29]V. Bargmann, P. Butera, L. Girardello, and J.R. Klauder, Rep. Math. Phys. 2, 221 (1971).
[30]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Sec. IV-4.
[31]D. Bergmann and Y. Frischman, J. Math. Phys. 6, 1855 (1965).
[32]The direct calculation of the phase factors in Eqs. (6.2) and (6.4) poses difficult complex integration problems: Eq. 6.633.2 in Ref. 14 is valid only for real values of the Bessel function. The calculation can be performed using the corresponding results of I, where direct integration is possible, together with those of Appendix B in this article.
[33]N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1966), Vol. I, Sec. 49.
[34]F. Trèves, Bull. Soc. Math. France 97, 193 (1969); M. Miller and S. Steinberg, Commun. Math. Phys. 24, 40 (1971).
[35]R. M. Wilcox, J. Math. Phys. 8, 962 (1967); E. Eriksen, J. Math. Phys. 9, 790 (1968).
[36]Ref. 14, Eq. 8.432.1.
[37]Notice that this is *different* from the more usual parametrization of the complex sphere such as that presented in, for instance, K.B. Wolf, J. Math. Phys. 13, 1634 (1972).
[38]Ref. 14, Eq. 8.534.
[39]J. Fischer, J. Niederle, and R. Rączka, J. Math. Phys. 7, 816 (1966); R. Rączka, N. Limić, and J. Niederle, J. Math. Phys. 7, 1861 (1966); N. Limić, J. Niederle, and R. Rączka, J. Math. Phys. 7, 2026 (1966); 8, 1079 (1967); C.P. Boyer and K.B. Wolf, J. Math. Phys. 15, 560 (1974).
[40]The angular part of Bargmann space has been shown to be relevant to the separation of the harmonic oscillator system in elliptic coordinates, C.P. Boyer, E. Kalnins, and W. Miller Jr., Preprint CRM-430 (1974) (Centre de Recherches Mathématiques, Montréal) J. Math. Phys. (to be published).

Green's function for Laplace's equation in an infinite cylindrical cell

A. Peskoff

Department of Physiology, University of California at Los Angeles, Los Angeles, California 90024
(Received 1 August 1974)

The Green's function for Laplace's equation in an infinite-length cylinder with a homogeneous mixed boundary condition is considered. Its eigenfunction expansion converges slowly when the axial separation between the source and observation points is small compared to the cylinder radius, and diverges when the axial separation is zero. Applying a modified form of a contour integral method of Watson to an integral representation of the Green's function, a more general expansion of the Green's function is derived. Watson's original method had previously been applied to the case when the source and observation points were both on the axis of the cylinder. The expansion contains a free parameter which may be adjusted to give rapid convergence for any axial separation. It fails, however, when the source and observation points are both near the surface of the cylinder. For two special values of the parameter, the general expansion reduces to the eigenfunction expansion or to the integral representation. The derivation is somewhat obscure, but the resulting formula has a simple interpretation as the superposition of the potential of two related boundary value problems in finite-length cylinders. Some numerical results are given in the spatial region which previously could not be calculated, for a boundary condition approaching a homogeneous Neumann condition, and for a homogeneous Dirichlet condition.

1. INTRODUCTION

In this paper, the Green's function for Laplace's equation in an infinite length cylinder is considered. The Green's function satisfies a homogeneous mixed boundary condition (a linear combination of the potential and its normal derivative vanishes) on the cylinder surface. The Green's function for this problem can be represented by an eigenfunction expansion.[1] A difficulty with the eigenfunction representation, however, is that when the axial coordinates of the source and observation points are almost equal, the expansion converges slowly, and when they are equal, it diverges. The present problem is to find an alternative representation which does not have this shortcoming, and can be used to compute the Green's function when the axial separation between the source and observation points is small.

When the two points are both on the cylinder axis, Bouwkamp and de Bruijn[2] have obtained a rapidly converging summation representation in the case of the homogeneous Dirichlet boundary condition, using a contour integration method of Watson.[3] By a variation of this method, it will be shown that their result can be generalized to the case of eccentric source and observation points, as well as to the more general boundary condition. The representation obtained contains a free parameter which can be adjusted to vary the rate of convergence without affecting the value to which the representation converges.

The mixed boundary condition for Laplace's equation occurs in biological problems. The interior of a cell is a medium of moderate electrical conductivity, a salt solution, surrounded by a thin, highly resistive membrane. The present analysis was initiated to describe the potential induced by a microelectrode current source inserted in a nerve axon. The tip of the microelectrode can be represented by a point source, so in this case the Green's function is itself the potential that one would observe in the cell. If it is assumed that the medium outside the cell is a perfect conductor maintained at zero potential, then the homogeneous mixed boundary condition applies, that is, for sufficiently small membrane current density so that the current-voltage relation of the membrane is in its linear range, the normal derivative of the potential just inside the membrane is proportional to the potential drop across the membrane.[4]

When this membrane boundary condition is written in dimensionless variables, in the biological problem the proportionality constant ϵ is a small quantity, of order 10^{-4} or smaller. Thus the boundary condition is almost a homogeneous Neumann boundary condition. Since the Neumann problem is not self-consistent (a current source inside the cell with no current crossing the membrane requires infinite potential inside the cell), it is convenient to solve the mixed boundary value problem first and then take the limit as the boundary condition approaches the Neumann boundary condition. This also has the advantage of making the results more generally applicable.

The mixed boundary condition for Laplace's equation also occurs in many physical problems. It occurs in the description of steady heat flow in a body with heat radiation and convection to the surrounding medium[5]; diffusion with evaporation at a surface[6]; current flow across a thin layer of gas separating a metallic electrode and an electrolyte[7]; etc.

In Sec. 2, the mathematical problem is stated and solved by Fourier transformation. The potential is represented as a Fourier cosine series in the polar angle and a Fourier cosine integral in the axial coordinate. This integral representation is converted to an eigenfunction expansion by closing the contour and applying the residue theorem. In Sec. 3 a rapidly converging summation representation of the potential is obtained. This is done using a modified form of Watson's method, starting with the Fourier cosine integral representation of the potential. The modification is necessary because of a branch point of the integrand. The derivation appears obscure, but the resulting formula has a physical interpretation. It is shown that the representation obtained, which is valid at the source and in a variable finite length of the infinite cylinder, is the superposition of the potential of two related boundary value problems in a finite length cylinder. In Sec. 4, by expanding the potential of Sec. 3 in powers of ϵ, the limiting case when the boundary condition approaches the Neumann boundary condition is considered, and graphs of numerical results given. In Sec. 5, by expanding the potential of Sec. 3 in powers of ϵ^{-1}, a Dirichlet boundary condition is considered and graphs of numerical results are presented. Although the rapidly converging expansion has been obtained for a specific boundary value problem of current interest, it is suspected that the technique is of wider applicability.

2. FORMULAS FOR THE POTENTIAL BY FOURIER TRANSFORMS

In this section we derive an exact representation, as a Fourier integral in the longitudinal coordinate and a Fourier series in the angular coordinate, for the electrostatic potential produced by a point source of current inside a right-circular cylindrical cell. The cell is a homogeneous, isotropic conductor bounded by a resistive membrane which is surrounded by a perfectly conducting medium at zero potential. The problem is formulated in cylindrical coordinates (r, θ, z). The source is located at the point $(R, 0, 0)$; the potential satisfies a mixed boundary condition at the inside surface of the membrane, $r=1$, and approaches the zero potential of the outer surface of the membrane as the axial coordinate z approaches plus or minus infinity. The problem for determining the potential may be written, in dimensionless variables,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\partial^2 V}{\partial z^2}=-\frac{1}{r}\delta(r-R)\delta(\theta)\delta(z), \tag{1}$$

$$\frac{\partial V}{\partial r}(1,\theta,z)+\epsilon V(1,\theta,z)=0, \tag{2}$$

$$V(r,\theta,\pm\infty)=0. \tag{3}$$

Exploiting the symmetry of the problem, the potential for a point source at (R,Θ,Z), the Green's function, is related to the potential $V(r,\theta,z)$ by $G(r,\theta,z\,|\,R,\Theta,Z) = V(r,\theta-\Theta,z-Z)$.

The problem posed by (1)—(3) can be solved by Fourier transforming in the θ and z coordinates. Defining the double Fourier cosine transform of V by

$$\psi_n(r,k)=\int_{-\pi}^{\pi}d\theta\cos(n\theta)\int_{-\infty}^{\infty}dz\cos(kz)\,V(r,\theta,z),$$

$$V(r,\theta,z)=\frac{1}{2\pi^2}\int_0^\infty dk\cos(kz)\sum_{n=0}^\infty\epsilon_n\cos(n\theta)\,\psi_n(r,k), \tag{4}$$

where $\epsilon_0=1$ and $\epsilon_1=\epsilon_2\cdots=2$, noting that V is an even function of θ and z, we obtain for the transform of (1),

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi_n}{\partial r}\right)-\left(k^2+\frac{n^2}{r^2}\right)\psi_n=-\frac{1}{r}\delta(r-R). \tag{5}$$

Transforming the boundary condition (2) by (4) gives the corresponding condition on ψ_n at $r=1$,

$$\frac{\partial\psi_n}{\partial r}(1,k)+\epsilon\psi_n(1,k)=0. \tag{6}$$

Integrating Eq. (5) across the delta function at $r=R$ shows that there is a discontinuity in the derivative of ψ_n given by

$$\frac{\partial\psi_n}{\partial r}(R+,k)-\frac{\partial\psi_n}{\partial r}(R-,k)=-\frac{1}{R}. \tag{7}$$

If the first derivative has the finite discontinuity (7), ψ_n itself must be continuous,

$$\psi_n(R+,k)=\psi_n(R-,k). \tag{8}$$

The solution to (5) which satisfies the conditions (6)—(8) is

$$\psi_n=-\frac{\epsilon K_n(k)+kK_n'(k)}{\epsilon I_n(k)+kI_n'(k)}I_n(kr)I_n(kR)+\begin{cases}I_n(kr)K_n(kR), & 0\le r\le R\\ I_n(kR)K_n(kr), & R\le r\le 1\end{cases} \tag{9}$$

where I_n and K_n are the modified Bessel functions of order n.

Taking the inverse transforms (4) of (9), using the addition theorem[8]

$$\sum_{n=0}^\infty\epsilon_n\cos(n\theta)\begin{cases}I_n(kr)K_n(kR), & 0\le r\le R\\ I_n(kR)K_n(kr), & R\le r\le\infty\end{cases}=K_0[k(r^2+R^2-2rR\cos\theta)^{1/2}]$$

and the Fourier cosine transform[9]

$$\int_0^\infty dk\cos(kz)K_0[k(r^2+R^2-2rR\cos\theta)^{1/2}]=\frac{\pi}{2}(z^2+r^2+R^2-2rR\cos\theta)^{-1/2},$$

yields the expression for the potential

$$V(r,\theta,z)=\frac{1}{4\pi}(z^2+r^2+R^2-2rR\cos\theta)^{-1/2}-\frac{1}{2\pi^2}\sum_{n=0}^\infty\epsilon_n\cos(n\theta)\int_0^\infty dk\cos(kz)\times\frac{\epsilon K_n(k)+kK_n'(k)}{\epsilon I_n(k)+kI_n'(k)}I_n(kr)I_n(kR). \tag{10}$$

As $k\to\infty$, the integrand in (10) approaches $-\cos(kz)\times\exp[k(r+R-2)]/2kr^{1/2}R^{1/2}$, and thus is rapidly convergent unless $r+R\approx 2$, i.e., unless the source and observation points are both near the boundary.

Equation (10) can be converted to an eigenfunction expansion by considering the real part of the contour integral

$$\oint dw\,e^{iw|z|}\frac{\epsilon K_n(w)+wK_n'(w)}{\epsilon I_n(w)+wI_n'(w)}I_n(wr)I_n(wR)=0 \tag{11}$$

where the contour in the $w=k+i\lambda$ plane goes out the positive real axis [the integration path in (10)], then along a circular arc in the first quadrant, and finally down the imaginary axis, with detours around the zeros of the denominator and the branch point at the origin. The zeros occur on the imaginary axis at $w=i\lambda_{n,s}$ where $\lambda_{n,s}$ is the sth positive root of

$$\epsilon J_n(\lambda_{n,s})+\lambda_{n,s}J_n'(\lambda_{n,s})=0 \tag{12}$$

arranged in order of ascending magnitudes, with $s=1,2,3\cdots$. The contour integral is zero because no singularities are enclosed. The integral along the circular portion vanishes exponentially as the radius becomes infinite, provided that $r+R<2$ (i.e., the source and observer are not both on the boundary). Hence, by the residue theorem, the integral in (10) equals the real part of the sum of πi times the residues at the poles on the imaginary axis and the principal value of the integral up the imaginary axis. The latter integral, when substituted in (10), cancels the $(z^2+r^2+R^2-2rR\cos\theta)^{-1/2}$ term. The result is the eigenfunction expansion,

$$V(r,\theta,z)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)$$
$$\times\sum_{s=1}^{\infty}\frac{\lambda_{n,s}\exp(-\lambda_{n,s}|z)}{\lambda_{n,s}^2+\epsilon^2-n^2}\;\frac{J_n(\lambda_{n,s}r)J_n(\lambda_{n,s}R)}{J_n^2(\lambda_{n,s})} \qquad (13)$$

which is equivalent to (10). The details of the derivation to obtain (13) from (10) are not given, since (13) is a special case of (23), which will be derived below, and because (13) has been given elsewhere.[1]

The double sum in (13) converges rapidly because of the exponential factor $\exp(-\lambda_{n,s}|z|)$ at positions for which the longitudinal coordinate satisfies $|z|\gtrsim 1$. In this case (13) is a convenient representation for numerical computation of the potential.[1,10] On the other hand, (13) converges slowly when $|z|\ll 1$, and diverges on the whole disc $z=0$, and hence is not a useful representation for computing the potential under these conditions. The integral representation of the potential (10) does not have this pathological behavior at $z=0$ because the singularity at $(R,0,0)$ is isolated in the $(z^2+r^2+R^2-2rR\cos\theta)^{-1/2}$ term. In the next section, a summation formula will be derived in which the $(z^2+r^2+R^2-2rR\cos\theta)^{-1/2}$ singularity is isolated.

3. CONVERGENT EXPANSION AT $z=0$

In this section we consider the problem of computing the potential near $z=0$, where the eigenfunction expansion (13) converges slowly. A more general summation representation for the potential is developed in which the singularity at $(R,0,0)$ is explicitly isolated, as in the integral representation (13). The general representation contains an adjustable parameter to vary its convergence rate without changing the value to which it converges. It reduces to (10) or (13) for two special values of the parameter. The representation can be made rapidly convergent on the disc $z=0$, except when both source and observation points are near the boundary (i.e., except when $r+R\approx 2$). This limitation is a consequence of the poor convergence of (10) when $r+R\approx 2$. The method used to obtain this representation of the potential is a variation of one used by Watson to develop rapidly convergent sums for computing integrals of the form $\int_0^\infty dt\, t^{2\nu+2n}/I_\nu^2(t)$.[3]

Watson's derivation begins by considering the judiciously chosen contour integral $\oint dw\, w^{2\nu+2n}/(w-t)I_\nu^2(w)\cos(\pi w/\beta)$ taken around the circular contour $|w|=\infty$. Replacing the contour integral, which equals zero, by $2\pi i$ times the sum of the residues at the poles yields a partial fraction expansion for $t^{2\nu+2n}/I_\nu^2(t)\cos(\pi t/\beta)$. A multiplication by $\cos(\pi t/\beta)$ recovers the original integrand and a subsequent integration over t from zero to infinity gives an infinite sum representation of the desired integral, containing the parameter β for adjusting the convergence rate. Bouwkamp and de Bruijn[2] used the same contour and the same factor $[(w-t)\cos(\pi w/\beta)]^{-1}$ to evaluate $\int_0^\infty dt\sin(tz)/tI_0^2(t)$, which appears in their formula for the electrostatic potential on the axis of a dielectric cylinder bounded by a perfect conductor, with a point source also on the axis.

For the more general problem being considered here, where the source and observation points are allowed to be off-axis, a modification of the contour and multiplicative factor in the integrand is necessary before the technique can be applied. The modifications are somewhat arbitrary, their ultimate justification being that they lead to a useful formula for calculating the potential.

Extending the integrand in (10) into the complex w-plane, with $w=k+i\lambda$, the integrand is seen to approach infinity exponentially for $|w|\to\infty$ in the left half-plane. Therefore, the contour must be restricted to the right half-plane. Furthermore, Watson's circular contour would not be permissible because of the branch point of $K_n(w)$ at $w=0$. Since we are restricted to the right half-plane, we try the contour shown in Fig. 1, with the hope that we will be able to deal with the resulting principal value integral along the imaginary axis, which did not appear in Watson's analysis.

We are still free to select an extra multiplicative factor in the integrand of (10), and need not be limited to Watson's choice. Before making this selection, we study the form which the integrand in (10) takes on the imaginary axis. Expressing the modified Bessel functions in (10), in terms of ordinary Bessel functions, we obtain for the integrand on the positive and negative imaginary axes, respectively,

$$\cosh(\lambda z)$$
$$\times\frac{\epsilon K_n[\exp(\pm i\pi/2)|\lambda|]+\exp(\pm i\pi/2)|\lambda|K_n'[\exp(\pm i\pi/2)|\lambda|]}{\epsilon I_n[\exp(\pm i\pi/2)|\lambda|]+\exp(\pm i\pi/2)|\lambda|I_n'[\exp(\pm i\pi/2)|\lambda|]}$$
$$\times I_n[\exp(\pm i\pi/2)|\lambda|r]I_n[\exp(\pm i\pi/2)|\lambda|R]$$
$$=-\frac{\pi}{2}\cosh(\lambda z)\left(\frac{\epsilon Y_n(|\lambda|)+|\lambda|Y_n'(|\lambda|)}{\epsilon J_n(|\lambda|)+|\lambda|J_n'(|\lambda|)}\pm i\right)$$
$$\times J_n(|\lambda|r)J_n(|\lambda|R). \qquad (14)$$

The real part of (14) is an even function of λ, whereas

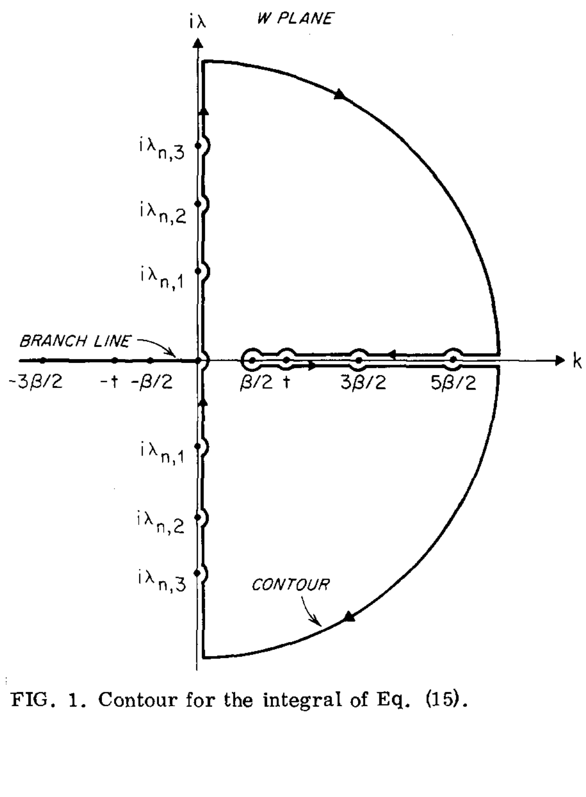

FIG. 1. Contour for the integral of Eq. (15).

the imaginary part is odd. If we choose our extra multiplicative factor to be an odd function of w, we can eliminate the real part of (14), and hence the poles, from the integral along the imaginary axis and it is no longer a principal value integral. The simplest such factor which has all the necessary properties of Watson's factor (what is necessary will be more apparent later), but yet is an odd function of w, is the function $w/(w^2-t^2)\cos(\pi w/\beta)$.

Thus we are led to consider the contour integral

$$\frac{1}{2\pi i}\oint dw \cos(wz)\frac{\epsilon K_n(w)+wK_n'(w)}{\epsilon I_n(w)+wI_n'(w)}I_n(wr)I_n(wR) \times\frac{w}{(w^2-t^2)\cos(\pi w/\beta)} \tag{15}$$

where the integration path is given in Fig. 1. There is a branch cut along the negative real axis. β and t are positive real parameters, and $\cos(\pi t/\beta)\neq 0$, so that the poles introduced are simple and distinct. If $|z|<\pi/\beta$, and $r+R<2$, the integrand along the semicircular arc tends to zero exponentially as the radius of the circle tends to infinity. There are no singularities enclosed by the contour so the integral (15) is zero. Therefore, the sum of $1/2\pi i$ times the integral up the imaginary axis, one-half times the residues on the imaginary axis, and the residues on the real axis is zero.

The *residue at* $w=t$ is

$$\frac{\epsilon K_n(t)+tK_n'(t)}{\epsilon I_n(t)+tI_n'(t)}I_n(tr)I_n(tR)\frac{\cos(tz)}{2\cos(\pi t/\beta)} \tag{16}$$

Note that $2\cos(\pi t/\beta)$ times (16) is the integrand in (10).

Substituting (14) in (15), we find that $1/2\pi i$ *times the integral up the imaginary axis* is

$$-\frac{1}{2}\int_0^\infty d\lambda\frac{\cosh(\lambda z)}{\cosh(\pi\lambda/\beta)}\frac{\lambda}{\lambda^2+t^2}J_n(\lambda r)J_n(\lambda R). \tag{17}$$

The *residues at* $w=(\nu+\frac{1}{2})\beta$, $\nu=0,1,2,\cdots$ are

$$-\frac{\beta}{\pi}(-)^\nu\frac{(\nu+\frac{1}{2})\beta\cos[(\nu+\frac{1}{2})\beta z]}{(\nu+\frac{1}{2})^2\beta^2-t^2} \times\frac{\epsilon K_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta K_n'[(\nu+\frac{1}{2})\beta]}{\epsilon I_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta I_n'[(\nu+\frac{1}{2})\beta]}I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R] \tag{18}$$

and the *residues at* $w=\pm i\lambda_{n,s}$ are

$$\frac{\lambda_{n,s}^2}{(\lambda_{n,s}^2+\epsilon^2-n^2)(\lambda_{n,s}^2+t^2)}\frac{\cosh(\lambda_{n,s}z)}{\cosh(\pi\lambda_{n,s}/\beta)} \times\frac{J_n(\lambda_{n,s}r)J_n(\lambda_{n,s}R)}{J_n^2(\lambda_{n,s})}. \tag{19}$$

Since the contour integral (15) vanishes, the residue at $w=t$, (16), is equal to minus $1/2\pi i$ times the integral up the imaginary axis, (17), minus the residues at $w=(\nu+\frac{1}{2})\beta$, (18), minus one-half times the residues at $w=\pm i\lambda_{n,s}$, (19), or,

$$\begin{aligned}\frac{\epsilon K_n(t)+tK_n'(t)}{\epsilon I_n(t)+tI_n'(t)}&\frac{\cos(tz)}{2\cos(\pi t/\beta)}I_n(tr)I_n(tR)\\ &=\frac{1}{2}\int_0^\infty d\lambda\cosh(\lambda z)\frac{\lambda J_n(\lambda r)J_n(\lambda R)}{(\lambda^2+t^2)\cosh(\pi\lambda/\beta)}\\ &+\frac{\beta}{\pi}\sum_{\nu=0}^\infty(-)^\nu\frac{\epsilon K_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta K_n'[(\nu+\frac{1}{2})\beta]}{\epsilon I_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta I_n'[(\nu+\frac{1}{2})\beta]}\\ &\times\frac{(\nu+\frac{1}{2})\beta\cos[(\nu+\frac{1}{2})\beta z]}{(\nu+\frac{1}{2})^2\beta^2-t^2}I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R]\\ &-\sum_{s=1}^\infty\frac{\lambda_{n,s}^2\cosh(\lambda_{n,s}z)J_n(\lambda_{n,s}r)J_n(\lambda_{n,s}R)}{(\lambda_{n,s}^2+\epsilon^2-n^2)(\lambda_{n,s}^2+t^2)\cosh(\pi\lambda_{n,s}/\beta)J_n^2(\lambda_{n,s})}.\end{aligned} \tag{20}$$

We now multiply (20) by $2\cos(\pi t/\beta)$, and integrate over t from zero to infinity. The left side is then the integral to be evaluated in (10). The three integrals on the right side can be evaluated using two tabulated integrals[11] so that the desired integral in (10) may be expressed as

$$\begin{aligned}\int_0^\infty dt\cos(tz)&\frac{\epsilon K_n(t)+tK_n'(t)}{\epsilon I_n(t)+tI_n'(t)}I_n(tr)I_n(tR)\\ &=\frac{\pi}{2}\int_0^\infty d\lambda\cosh(\lambda z)\frac{\exp(-\pi\lambda/\beta)}{\cosh(\pi\lambda/\beta)}J_n(\lambda r)J_n(\lambda R)\\ &+\beta\sum_{\nu=0}^\infty\cos[(\nu+\tfrac{1}{2})\beta z]\frac{\epsilon K_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta K_n'[(\nu+\frac{1}{2})\beta]}{\epsilon I_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta I_n'[(\nu+\frac{1}{2})\beta]}\\ &\times I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R]\\ &-\pi\sum_{s=1}^\infty\frac{\lambda_{n,s}\cosh(\lambda_{n,s}z)\exp(-\pi\lambda_{n,s}/\beta)J_n(\lambda_{n,s}r)J_n(\lambda_{n,s}R)}{(\lambda_{n,s}^2+\epsilon^2-n^2)\cosh(\pi\lambda_{n,s}/\beta)J_n^2(\lambda_{n,s})}\end{aligned} \tag{21}$$

To obtain the potential we must multiply (21) by $-\epsilon_n\cos(n\theta)/2\pi^2$, sum over n, and substitute the result in (10). Performing these operations on the integral on the right side of (21), noting that

$$\frac{\exp(-\pi\lambda/\beta)}{2\cosh(\pi\lambda/\beta)}=\sum_{m=1}^\infty(-)^{m+1}\exp(-2\pi m\lambda/\beta),$$

and using the addition formula for J_0,[12] we obtain

$$\begin{aligned}\frac{1}{4\pi}\sum_{m=1}^\infty(-)^m&\int_0^\infty d\lambda\{\exp[-\lambda(z+2\pi m/\beta)]\\ &+\exp[\lambda(z-2\pi m/\beta)]\}J_0[\lambda(r^2+R^2-2rR\cos\theta)^{1/2}]\\ &=\frac{1}{4\pi}\left(\sum_{m=-\infty}^{-1}+\sum_{m=1}^\infty\right)(-)^m\left[\left(z-\frac{2\pi m}{\beta}\right)^2+r^2+R^2\right.\\ &\left.-2rR\cos\theta\right]^{-1/2},\quad |z|<2\pi/\beta\end{aligned} \tag{22}$$

where a Bessel function integral[13] has been used to obtain the right side of (22). We see that the free-space potential of the point source, which appears in (10), is just the missing $m=0$ term of the series in (22).

Substituting (22) in (21) and the result in (10) yields a summation representation for the potential:

$$\begin{aligned}V(r,\theta,z)=&\frac{1}{4\pi}\sum_{m=-\infty}^\infty(-)^m\left[\left(z-\frac{2\pi m}{\beta}\right)^2+r^2+R^2-2rR\cos\theta\right]^{-1/2}\\ &-\frac{\beta}{2\pi^2}\sum_{n=0}^\infty\epsilon_n\cos(n\theta)\sum_{\nu=0}^\infty\cos[(\nu+\tfrac{1}{2})\beta z]\\ &\times\frac{\epsilon K_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta K_n'[(\nu+\frac{1}{2})\beta]}{\epsilon I_n[(\nu+\frac{1}{2})\beta]+(\nu+\frac{1}{2})\beta I_n'[(\nu+\frac{1}{2})\beta]}\\ &\times I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R]\\ &+\frac{1}{\pi}\sum_{n=0}^\infty\epsilon_n\cos(n\theta)\sum_{s=1}^\infty\frac{\cosh(\lambda_{n,s}z)}{1+\exp(2\pi\lambda_{n,s}/\beta)}\end{aligned}$$

$$\times \frac{\lambda_{n,s} J_n(\lambda_{n,s} r) J_n(\lambda_{n,s} R)}{(\lambda_{n,s}^2 - n^2 + \epsilon^2) J_n^2(\lambda_{n,s})}. \tag{23}$$

Equation (23) is an alternative to our original formulas (10) and (13), which is much more suitable for calculating the potential near and at $z=0$. We will now discuss the numerical characteristics and physical interpretations of the three terms in (23).

The first term in (23) is the sum of the free-space potentials of the source at $(R, 0, 0)$ and an infinite number of images located at $(R, 0, \pm 2\pi m/\beta)$, $m=1,2,3,\cdots$, with alternating signs. These are the images obtained by reflecting the source in the planes $z=\pm\pi/\beta$, then reflecting the two images in these two planes, and so on. It is interesting to notice that the two planes $z=\pm\pi/\beta$ are part of the boundary of the region of convergence of the contour integral (15). The images are located *inside* the cylinder, but outside the region of validity of (23). The sum can be recognized as the potential of a point source at $(R, 0, 0)$ with two perfectly conducting planes (homogeneous Dirichlet boundary condition) at $z=\pm\pi/\beta$. An equivalent eigenfunction expansion can be obtained, either directly, or from the image expansion in (23). Adding the $m=0$ term to (22), we find

$$\sum_{m=-\infty}^{\infty} (-)^m [(z - 2\pi m/\beta)^2 + r^2 + R^2 - 2rR\cos\theta]^{-1/2}$$
$$= \int_0^{\infty} d\lambda \left(\exp(-\lambda|z|) - \frac{\cosh(\lambda z)\exp(-\pi\lambda/\beta)}{\cosh(\pi\lambda/\beta)}\right) \times J_0[\lambda(r^2 + R^2 - 2rR\cos\theta)^{1/2}]$$
$$= \int_0^{\infty} d\lambda \left(\frac{\exp(-\lambda|z|)}{1+\exp(-2\pi\lambda/\beta)} - \frac{\exp(\lambda|z|)}{1+\exp(2\pi\lambda/\beta)}\right) \times J_0[\lambda(r^2 + R^2 - 2rR\cos)^{1/2}].$$

On the last line we have rewritten the integrand in a form which clearly displays that it is an odd function of λ. This allows us to replace the integral by[14]

$$\frac{1}{2}\int_{-\infty}^{\infty} d\zeta \left[\frac{\exp(-\zeta|z|)}{1+\exp(-2\pi\zeta/\beta)} - \frac{\exp(\zeta|z|)}{1+\exp(2\pi\zeta/\beta)}\right] \times H_0^{(1)}[\zeta(r^2 + R^2 - 2rR\cos\theta)^{1/2}]$$

which may be evaluated by closing the contour with a large counterclockwise semicircular path in the upper half ζ-plane. Calculating $2\pi i$ times the sum of the residues at the poles at $\zeta = \pm i(\nu+\frac{1}{2})\beta$, $\nu=0,1,2,\cdots$, we obtain an equivalent eigenfunction expansion,

$$\frac{1}{4\pi}\sum_{m=-\infty}^{\infty} (-)^m [(z - 2\pi m/\beta)^2 + r^2 + R^2 - 2rR\cos\theta]^{-1/2}$$
$$= \frac{\beta}{2\pi^2}\sum_{\nu=0}^{\infty} \cos[(\nu+\tfrac{1}{2})\beta z] K_0[(\nu+\tfrac{1}{2})\beta(r^2 + R^2 - 2rR\cos\theta)^{1/2}]. \tag{24}$$

The eigenfunction expansion on the right side of (24) is rapidly convergent if $\beta(r^2 + R^2 - 2rR\cos\theta)^{1/2}$ is not too small, in which case the terms decrease exponentially in magnitude. When this is not the case, the image expansion should be used. (In the computations for the curves given in Secs. 4 and 5, for convenience, we always use the image expansion, but we use a nonlinear sequence-to-sequence transformation, the diagonal of the Padé table,[15] to improve its convergence.)

Since the singularity in the potential has been isolated in the $m=0$ term of the image expansion, we do not expect the remaining terms in (23) to have pathological behavior at $z=0$ as did the eigenfunction expansion (13).

In the second term of (23), the sum over ν has the numerical interpretation that it is a discrete rectangular approximation, with spacing β, for the integral over k in (10). The sum of the image potentials and the last term in (23) (the double sum over n and s) then represents the correction to this rectangular approximation.

In addition to this interpretation, which is interesting from the point of view of numerical analysis, a physical interpretation of the double sum over n and ν is also possible. Using the addition theorem for K_0,[8] the image expansion in (24) can be written as the double sum

$$\frac{\beta}{2\pi^2}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{\nu=0}^{\infty}\cos[(\nu+\tfrac{1}{2})\beta z] K_n[(\nu+\tfrac{1}{2})\beta r] I_n[(\nu+\tfrac{1}{2})\beta R] \tag{25}$$

for $r > R$, or the same expression with r and R interchanged, for $r < R$. As a consequence, we see that the sum of the image expansion (24) and the "rectangular approximation" [the sum over n and ν in (23)] satisfies the homogeneous mixed boundary condition (2) at $r=1$. Furthermore, they each satisfy a homogeneous Dirichlet boundary condition at $z=\pm\pi/\beta$. Therefore, together, these first two terms in (23) represent the solution to the problem of a point source at $(R, 0, 0)$ inside a finite length cylinder with homogeneous Dirichlet boundary conditions at both ends, $z=\pm\pi/\beta$, and the homogeneous mixed boundary condition (2) on the cylindrical boundary, $r=1$. It should be mentioned that another eigenfunction expansion can be obtained for this problem. It is

$$\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\operatorname{sech}(\pi\lambda_{n,s}/\beta)\sinh\left[\lambda_{n,s}\left(\frac{\pi}{\beta}-|z|\right)\right] \times \frac{\lambda_{n,s} J_n(\lambda_{n,s} r) J_n(\lambda_{n,s} R)}{(\lambda_{n,s}^2 - n^2 + \epsilon^2) J_n^2(\lambda_{n,s})} \tag{26}$$

in which the eigenfunctions are the same as those in (13). The expansion (26) may be obtained starting with the eigenfunction expansion (13) for the infinite cylinder and using the method of images as has been applied to the corresponding Dirichlet problem.[16] It may be observed that (26) has the same divergence problems as (13) at $z=0$.

Having two interpretations of the rectangular approximation sum, we now look at its convergence properties. For large ν, the Fourier coefficients in the sum over ν approach $-\exp[-(\nu+\frac{1}{2})\beta(2-r-R)]/(2\nu+1)\beta r^{1/2}R^{1/2}$, so that the terms decrease exponentially with increasing ν, and the sum converges rapidly, except near $r=R=1$. For fixed ν, the terms in the sum over n approach $r^n R^n/2n$ for large n. Consequently, this sum over n is also nicely convergent, except near $r=R=1$.

Finally, the last term in (23) (the double sum over n and s) looks like the eigenfunction expansion in (13), except for the important difference that the $\exp(-\lambda_{n,s}|z|)$ factor is replaced by $\cosh(\lambda_{n,s} z)/[1+\exp(2\pi\lambda_{n,s}/\beta)]$. With an appropriate choice of β, this makes the sum rapidly convergent for any z, including $z=0$. The sum converges for $|z|<2\pi/\beta$, which is twice the range for which (15) converges.

On the cylindrical boundary $r=1$ the double sum satisfies the mixed boundary condition (2). If we let $z=\pm\pi/\beta$, the double sum reduces to (13). Therefore the sum represents the source-free potential which satisfies (2) on $r=1$ and equals the true potential (10) or (13) on the planes $z=\pm\pi/\beta$.

The derivation of (23) implies that it is valid for $|z|<\pi/\beta$, but since all terms in (23) are convergent for $|z|<2\pi/\beta$, by analytic continuation (23) must be valid for this larger range. Our physical interpretation of the terms in (23), however, does not have meaning in the range $\pi/\beta\leq|z|\leq 2\pi/\beta$. Because the parameter β is arbitrary, it can be chosen to make the formula converge rapidly in the vicinity of $z=0$. The rate of convergence and the value of each sum in (23) depends on β, but the total value of (23) is, of course, independent of β. Near $r+R=2$, however, the rectangular approximation term converges slowly, for any reasonable choice of β.

As mentioned previously, there are two limiting cases for which (23) reduces to (10) or (13). In the limit $\beta\to 0$, all but the $m=0$ term in the first sum vanish (the images move off toward $z=\pm\infty$), the second sum vanishes, and the sum over s approaches the integral in (10). Hence, when $\beta=0$, (23) reduces to the integral form (10) for the potential. When $\beta=\pi/|z|$, the sum over m vanishes because of a cancellation of the $m=1,2,3,\cdots$ terms with the $m=0,-1,-2,\cdots$ terms, respectively (the infinite array of point sources are in an antisymmetric arrangement around the planes $z=\pm\pi/\beta$); the sum over ν vanishes because of the $\cos[(\nu+\frac{1}{2})\pi]$ factor and the sum over n and s reduces to the double sum in (13).

4. BOUNDARY CONDITION APPROACHING NEUMANN CONDITION

We would now like to study the potential for small ϵ, that is, as the boundary condition (2) approaches a Neumann boundary condition. This is the case of biological interest in which the interior of a cylindrical cell is enclosed by a highly resistive membrane. It is not possible to allow ϵ to equal zero, because the problem becomes singular. This singularity occurs because if $\epsilon=0$, no current can cross the membrane; all the current injected at $(R,0,0)$ would travel inside the cell toward $z=\pm\infty$. Consequently, the potential would contain a term decreasing linearly with $|z|$, so that the difference in potential between a finite z and $z=\infty$ would be infinite, in violation of the $|z|=\infty$ boundary condition. It will be seen that the potential approaches infinity as $\epsilon^{-1/2}$ when ϵ goes to zero.

We will first consider the $\epsilon\to 0$ limit of (13), then return to consider (23). We obtain this limit by expanding the double sum over n and s, appearing in (13) in powers of ϵ. This is accomplished by expressing $\lambda_{n,s}$, the roots of (12), in terms of $j'_{n,s}$, the roots of (12) with $\epsilon=0$. It should be noted that $\lambda_{0,1}$ is an exceptional case since $\lambda_{0,1}\to 0$ as $\epsilon\to 0$, whereas all other $\lambda_{n,s}$ approach positive values. Thus,

$$\begin{aligned}&\lambda_{n,s}\to j'_{n,s}, \quad n=1,2,3,\cdots, \quad s=1,2,3,\cdots,\\ &\lambda_{0,s}\to j'_{0,s-1}, \qquad\qquad s=2,3,4,\ldots\\ &\lambda_{0,1}\to 0\end{aligned} \tag{27}$$

as $\epsilon\to 0$, where the $j'_{n,s}$ are the positive roots of

$$J'_n(j'_{n,s})=0 \tag{28}$$

arranged in order of ascending magnitudes, with $s=1,2,3,\cdots$. This is the notation used by Watson.[17]

Letting $n=0$ in (12) and expanding around $\lambda=0$, leads to a series for ϵ in powers of $\lambda_{0,1}$, the reversion of which gives

$$\lambda_{0,1}=(2\epsilon)^{1/2}\left[1-\frac{1}{8}\epsilon+\frac{5}{384}\epsilon^2-\cdots\right] \tag{29a}$$

which expresses the smallest root of (12) in terms of ϵ.[18]

The other roots $\lambda_{n,s}$ are found in a similar way by expanding both sides of (12) around the corresponding points $j'_{n,s}$. The resulting expansion of the root $\lambda_{n,s}$, $n=1,2,3,\cdots$, $s=1,2,3,\cdots$ about the point $j'_{n,s}$ is

$$\lambda_{n,s}=j'_{n,s}\left(1+\frac{\epsilon}{j'^2_{n,s}-n^2}-\frac{(j'^2_{n,s}+n^2)\epsilon^2}{2(j'^2_{n,s}-n^2)^3}+O(\epsilon^3)\right) \tag{29b}$$

and for $n=0$, $s=2,3,4,\cdots$,

$$\lambda_{0,s}=j'_{0,s-1}+\frac{\epsilon}{j'_{0,s-1}}-\frac{\epsilon^2}{2j'^3_{0,s-1}}+O(\epsilon^3) \tag{29c}$$

Substituting (29a, b, and c) in the representation (13) for the potential, we obtain an expansion for the potential in powers of ϵ,

$$\begin{aligned}V(r,\theta,z)=&\frac{2^{1/2}}{4\pi}\exp\left[-(2\epsilon)^{1/2}|z|\left(1-\frac{\epsilon}{8}+\frac{5\epsilon^2}{384}+O(\epsilon^3)\right)\right]\\ &\times[\epsilon^{-1/2}+\tfrac{1}{2}\epsilon^{1/2}(\tfrac{5}{4}-r^2-R^2)\\ &+\tfrac{1}{16}\epsilon^{3/2}(\tfrac{25}{4}+r^4-3r^2+4r^2R^2-3R^2+R^4)+O(\epsilon^{5/2})]\\ &+\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\exp\left[-j'_{n,s}|z|\right.\\ &\left.\times\left(1+\frac{\epsilon}{j'^2_{n,s}-n^2}+O(\epsilon^2)\right)\right]\\ &\times\frac{j'_{n,s}}{j'^2_{n,s}-n^2}\left(\frac{J_n(j'_{n,s}r)J_n(j'_{n,s}R)}{J_n^2(j'_{n,s})}-\frac{\epsilon}{(j'^2_{n,s}-n^2)J_n^2(j'_{n,s})}\right.\\ &\left.\times\left\{\frac{j'_{n,s}+n^2}{j'^2_{n,s}-n^2}-j'_{n,s}\frac{d}{dj'_{n,s}}\right\}\{J_n(j'_{n,s}r)J_n(j'_{n,s}R)\}+O(\epsilon^2)\right).\end{aligned} \tag{30}$$

The expansion (30) demonstrates the singular behavior in the limit of ϵ approaching zero. The first term (the exponential times the series with algebraic terms) originates from the $n=0$, $s=1$ term of (13). It contains a factor which decays exponentially in $|z|$, with a length constant which, in the $\epsilon\to 0$ limit, approaches infinity as $\epsilon^{-1/2}$. The series multiplying the exponential starts with $\epsilon^{-1/2}$, so that the potential also approaches infinity as $\epsilon^{-1/2}$ in the $\epsilon\to 0$ limit.

The next higher order part of the potential is the $O(1)$ part of the double sum over n and s. This contains the $(z^2+r^2+R^2-2rR\cos\theta)^{-1/2}$ singularity at the source point $(R,0,0)$ so that at points close to the source (within a distance of order $\epsilon^{1/2}$) it is comparable to the $O(\epsilon^{-1/2})$ term.

For sufficiently large $|z|$ each term in the double sum over n and s is exponentially small compared to the first term in (30), since each $j'_{n,s}$ is positive. In the *far field*, which we define by taking the limit $\epsilon \to 0$ while holding $(2\epsilon)^{1/2}|z|$ fixed, the potential is given asymptotically by the first term alone:

$$V(r,\theta,z) \sim \frac{2^{1/2}}{4\pi}\exp\left[-(2\epsilon)^{1/2}|z|\left(1-\frac{\epsilon}{8}+\frac{5\epsilon^2}{384}+O(\epsilon^3)\right)\right]$$
$$\times\left(\epsilon^{-1/2}+\tfrac{1}{2}\epsilon^{1/2}(\tfrac{5}{4}-r^2-R^2)+\frac{\epsilon^{3/2}}{16}(\tfrac{25}{4}+r^4-3r^2+4r^2R^2-3R^2+R^4)+O(\epsilon^{5/2})\right).$$

Note that the far field expansion is indedependent of θ but its higher order terms do depend on radial position r and radial location of the source R. This expansion for the far field can also be obtained as the *outer expansion* in a singular perturbation analysis of the present problem by matched asymptotic expansions.[19]

Taking the $\epsilon \to 0$ limit of (30), holding z fixed, we obtain the *near field* expansion,

$$V(r,\theta,z)=\frac{(2\epsilon)^{-1/2}}{2\pi}-\frac{|z|}{2\pi}+\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\exp(-j'_{n,s}|z|)$$
$$\times\frac{j'_{n,s}J_n(j'_{n,s}r)J_n(j'_{n,s}R)}{(j'^2_{n,s}-n^2)J_n^2(j'_{n,s})}+\frac{(2\epsilon)^{1/2}}{8\pi}(\tfrac{5}{4}+2z^2-r^2-R^2)+O(\epsilon), \qquad (31)$$

where we have omitted $O(\epsilon)$ terms for brevity. The $O(\epsilon)$ term is given elsewhere,[19] where (31) is obtained as the *inner expansion* using the method of matched asymptotic expansions. From (31), it can be seen clearly that in the near field region the potential inside the cylinder consists of a large constant term of order $\epsilon^{-1/2}$, a term decreasing linearly with increasing $|z|$, an expansion in the eigenfunctions of a perfectly insulated ($\epsilon=0$) cylinder, and higher order terms.

The near field expansion (31) is valid when the linearly decreasing term is small compared to the constant term, i.e., when $|z| \ll (2\epsilon)^{-1/2}$. For typical values of ϵ found in biological cells, (31) will be applicable for values of $|z|$ between zero and many times the cell cross-sectional radius. However, it suffers from the same poor convergence rate when $|z|$ is much smaller than the radius, as did the expansion (13) for arbitrary ϵ. Therefore, the rapidly converging form (23) should be specialized to the near field, small ϵ limit to obtain a rapidly converging substitute for (31) when $|z| \ll 1$.

Using the same procedure on (23) as was used to obtain (31) from (13), i.e., substituting (29a, b, and c) in (23), and taking the limit as ϵ approaches zero with z fixed, we obtain the general form of the near field expansion (31),

$$V(r,\theta,z)=\frac{(2\epsilon)^{-1/2}}{2\pi}-\frac{1}{2\beta}+\frac{1}{4\pi}\sum_{m=-\infty}^{\infty}(-)^m$$
$$\times[(z-2\pi m/\beta)^2+r^2+R^2-2rR\cos\theta]^{-1/2}$$
$$-\frac{\beta}{2\pi^2}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{\nu=0}^{\infty}\cos[(\nu+\tfrac{1}{2})\beta z]$$
$$\times\frac{K'_n[(\nu+\frac{1}{2})\beta]}{I'_n[(\nu+\frac{1}{2})\beta]}I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R]$$
$$+\frac{1}{\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\frac{\cosh(j_{n,s}z)}{1+\exp(2\pi j'_{n,s}/\beta)}$$
$$\times\frac{j'_{n,s}J_n(j'_{n,s}r)J_n(j'_{n,s}R)}{(j'^2_{n,s}-n^2)J_n^2(j'_{n,s})}+O(\epsilon^{1/2}). \qquad (32)$$

The curves in Figure 2 are the results of numerical computations using (32) for $|z|<\pi/\beta$ and (31) for $|z|>\pi/\beta$. The $O(1)$ term in the potential $V(r,0,z)$ is plotted as a function of z for $R=1$ and $\theta=0$ for several values of r between $r=0.9$ and $r=0$, and for $\theta=\pi$ and $r=0.9$.

The value of β used in (32) and the number of terms required to attain convergence in each sum in (31) and (32) (to three decimal places for obtaining the curves) depends on the value of $r+R$. When $r+R$ approaches 2, as has already been observed, convergence of the first double sum (over n and ν) in (32) becomes increasingly slow. We are forced to take more terms in the sum over n, but the convergence rate of the sum over ν can be improved by increasing the value of β. Increasing β, however, reduces the convergence rate of the last double sum (over n and s) in (32), so some compromise value of β must be taken. As $r+R$ varies between zero and 1.9, reasonable choices of β seem to vary from about 2 to 16.

For computation of the most slowly convergent cases in Fig. 2, the two curves for which $R=1.0$, $r=0.9$, we have set $\beta=8$. To obtain three decimal place accuracy, in the sum over n and ν, we let n go from 0 to 50 and ν from 0 to 8, and in the sum over n and s, let n go from 0 to 13 and s go from 1 to 3. The latter sum converges most slowly at the largest value of $|z|$ for which it is used, at $|z|=\pi/8$. At smaller values of $|z|$, less terms are required. The Bessel functions in (31) and (32) were computed using their recursion relations, and the location of the zeros, $j'_{n,s}$, were taken from tables.[20]

At the opposite extreme, the most rapidly convergent

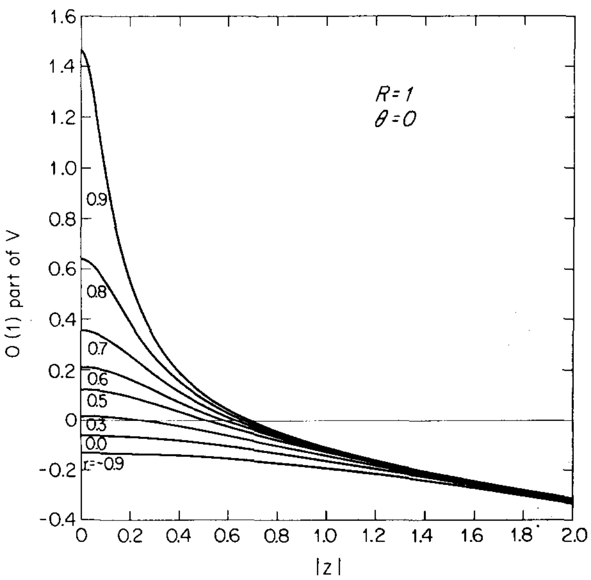

FIG. 2. The $O(1)$ term in the ϵ-expansion of the potential as a function of $|z|$, for the boundary condition approaching a homogeneous Neumann condition, with $R=1$, $\theta=0$, $r=0.0$, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, and $R=1$, $\theta=\pi$, $r=0.9$.

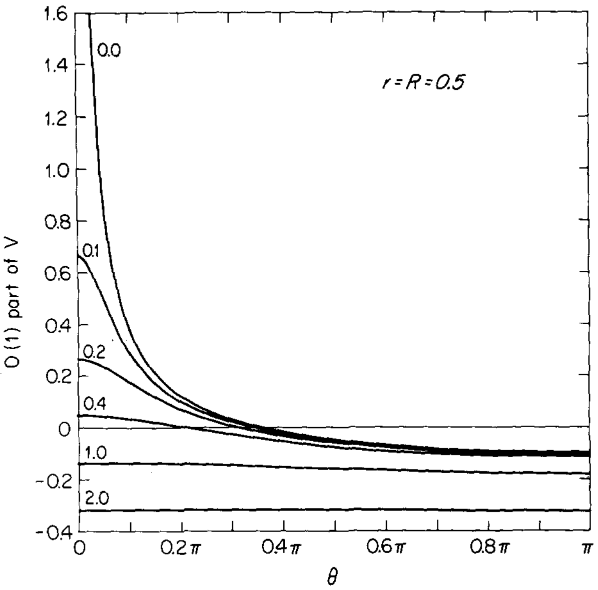

FIG. 3. The $O(1)$ term in the ϵ-expansion of the potential as a function of θ, for the boundary condition approaching a homogeneous Neumann condition with $r=R=0.5$, $|z|=0.0,\ 0.1,\ 0.2,\ 0.4,\ 1.0$, and 2.0.

case in Fig. 2 is that of $R=1$, $r=0$. In this case we set $\beta=4$. Because of the cylindrical symmetry only the $n=0$ terms are present. To obtain three decimal place accuracy we need only the $\nu=0$ and $s=1$ terms.

The sum over the source and images in (32) was computed using a nonlinear sequence-to-sequence transformation. We start with the decreasing sequence s_m with s_0 equal to the $m=0$ term in the image expansion, and $s_m(m=1,2,\cdots)$ equal to the sum of the $\pm m$ terms of the image expansion. This sequence is then transformed to the sequence of diagonal elements of its Padé table.[15] The Nth diagonal element is a ratio of two $N\times N$ determinants which utilize the first $2N+2$ terms of s_m. As an example, for the $R=1$, $r=0.9$ case, three decimal place accuracy is attained in the third diagonal element, which requires computation of two 3×3 determinants, utilizing the $m=0,\pm 1,\pm 2,\ldots,\pm 7$ terms of the image expansion. Using the equivalent eigenfunction expansion (24) in this case requires 16 terms.

In Fig. 3 the $O(1)$ term in the potential $V(0.5,\theta,z)$ is plotted as a function of θ, for $R=0.5$ and $z=0,\ 0.2,\ 0.4,\ 1.0,\ 2.0$. We set $\beta=4$, and to obtain three decimal place accuracy in the first double sum let n go from 0 to 3 and s go from 1 to 2. In this case one cannot use the eigenfunction equivalent, (24), of the image expansion because it diverges when $r=R$.

5. BOUNDARY CONDITION APPROACHING DIRICHLET CONDITION

We now turn to the opposite extreme of large ϵ, in which the boundary condition approaches a Dirichlet boundary condition. When $\epsilon=\infty$, the problem represents, for example, that of the electrostatic potential in a dielectric cylinder surrounded by a grounded perfect conductor. The $\epsilon\to\infty$ limit is not singular, in contrast to the $\epsilon\to 0$ limit. Rather than set $\epsilon=\infty$, however, we use the same procedure as in Sec. 4 to obtain the potential correct to $O(\epsilon^{-1})$ in the $\epsilon\to\infty$ limit. The solution can be applied to the description of steady heat flow in a cylinder for which the heat transfer coefficient to the surroundings is very large, or diffusion of a liquid in a porous rod, evaporating at a rapid rate into the surrounding atmosphere.

In the $\epsilon\to\infty$ limit,

$$\lambda_{n,s}\to j_{n,s},\quad n=0,1,2,\cdots,\quad s=1,2,3,\cdots, \tag{33}$$

where the $j_{n,s}$ are the positive roots of

$$J_n(j_{n,s})=0 \tag{34}$$

arranged in ascending magnitudes, again using the notation Watson.[17]

The root $\lambda_{n,s}$ can be expressed in terms of $j_{n,s}$ by a series in ascending powers of ϵ^{-1}. The series is obtained by reversion of the expansion in powers of $\lambda_{n,s}-j_{n,s}$ of (12). Substituting the series for $\lambda_{n,s}$ in the eigenfunction expansion (13), we obtain the expansion in powers of ϵ^{-1} for the potential

$$V(r,\theta,z)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\exp\{-j_{n,s}|z|[1-\epsilon^{-1}+O(\epsilon^{-2})]\}\times\frac{J_n(j_{n,s}r)J_n(j_{n,s}R)}{j_{n,s}J_{n+1}^2(j_{n,s})}\left[1-\epsilon^{-1}\left(1+\frac{j_{n,s}(d/dj_{n,s})[J_n(j_{n,s}r)J_n(j_{n,s}R)]}{J_n(j_{n,s}r)J_n(j_{n,s}R)}\right)+O(\epsilon^{-2})\right]. \tag{35}$$

This is the large-ϵ analog of (31).

Setting $\epsilon=\infty$ in (35), we obtain the potential for the Dirichlet boundary condition[16]

$$V(r,\theta,z)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{s=1}^{\infty}\frac{\exp(-j_{n,s}|z|)J_n(j_{n,s}r)J_n(j_{n,s}R)}{j_{n,s}J_{n+1}^2(j_{n,s})}. \tag{36}$$

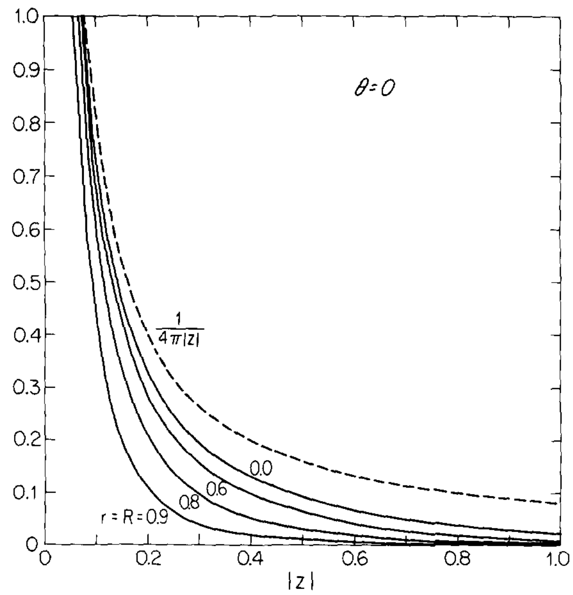

FIG. 4. The potential as a function of $|z|$ for a homogeneous Dirichlet boundary condition with $\theta=0$ and $r=R=0.0,\ 0.6,\ 0.8,\ 0.9$ (solid curves) and $1/4\pi|z|$, the free-space potential at $(R,0,z)$ of a point source at $(R,0,0)$ (broken curve).

This eigenfunction expansion is useful for computing the potential when $|z| \gtrsim 1$. When $|z| \ll 1$, we need the $\epsilon \to \infty$ limit of (23), which can be obtained by the same procedure used to obtain (36). We find, for the general representation of the potential with a homogeneous Dirichlet boundary condition,

$$V(r,\theta,z)=\frac{1}{4\pi}\sum_{m=-\infty}^{\infty}(-)^m[(z-2\pi m/\beta)^2+r^2+R^2-2rR\cos\theta]^{-1/2}$$
$$-\frac{\beta}{2\pi^2}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\sum_{\nu=0}^{\infty}\cos[(\nu+\tfrac{1}{2})\beta z]\times\frac{K_n[(\nu+\frac{1}{2})\beta]}{I_n[(\nu+\frac{1}{2})\beta]}I_n[(\nu+\tfrac{1}{2})\beta r]I_n[(\nu+\tfrac{1}{2})\beta R]$$
$$+\frac{1}{2\pi}\sum_{n=0}^{\infty}\epsilon_n\cos(n\theta)\times\sum_{s=1}^{\infty}\frac{2\cosh(j_{n,s}z)}{1+\exp(2\pi j_{n,s}/\beta)}\frac{J_n(j_{n,s}r)J_n(j_{n,s}R)}{j_{n,s}J_{n+1}^2(j_{n,s})}\,. \tag{37}$$

When $r=R=0$, (37) must be equivalent to Eq. (28) of Bouwkamp and de Bruijn.[2] A numerical comparison of the two equations shows that they agree, although there does not appear to be a simple way to show this equivalence analytically.

The curves in Fig. 4 are plots of the potential $V(r,\theta,z)$ for $\theta=0$ and $r=R=0.9$, 0.8, 0.6, 0.0. To obtain the first two curves we set $\beta=8$ and use (37) in the range $0 \leqslant |z| < \pi/8$ and (36) in the range $\pi/8 < |z| \leqslant 1$; to obtain the last two curves we set $\beta=4$ and use (37) in the range $0 \leqslant |z| < \pi/4$ and (36) in the range $\pi/4 < |z| \leqslant 1$. The dependence of the convergence rate on $r+R$ is essentially the same as in the examples cited for the Neumann boundary condition computations in Sec. 4.

The broken curve is the free-space potential of the point source $1/4\pi|z|$ plotted for comparison. It is seen that the potential approaches zero more rapidly than the free-space potential at a rate that increases as the source approaches the surface of the cylinder.

ACKNOWLEDGMENTS

I would like to thank Dr. R. S. Eisenberg for bringing to my attention the problem discussed in this paper. I am grateful to C. Clausen for writing the computer program for obtaining the curves, and to K. Aarons for typing the manuscript, especially the equations. This work was supported initially by National Science Foundation Grant No. GB 24965, and completed during the tenure of a National Institutes of Health Special Research Fellowship.

[1]R.S. Eisenberg and E.A Johnson, Prog. Biophys. Mol. Biol. 20, 1 (1970).
[2]C.J. Bouwkamp and N.G. de Bruijn, J. Appl. Phys. 18, 562 (1947).
[3]G.N. Watson, Proc. Roy. Soc. 130, 29 (1930).
[4]A.M. Weinberg, Bull. Math. Biophys. 4, 107 (1942).
[5]A.V. Luikov, *Analytical Heat Diffusion Theory* (Academic, New York, 1968), pp. 202, 205.
[6]J. Crank, *The Mathematics of Diffusion* (Oxford, 1957), p. 42.
[7]A. Gray and G.B. Mathews, *A Treatise on Bessel Functions and Their Applications to Physics* (Dover, New York, 1966), p. 148.
[8]G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge U.P., Cambridge, 1966), p. 361, Sec. 11.3, Eq. (8).
[9]Ref. 8, p. 388, Sec. 13.21, Eq. (10).
[10]R.H. Adrian, L.L. Costantin, and L.D. Peachey, J. Physiol. 204, 231 (1969).
[11]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, 4th Ed. (Academic, New York, 1965), p. 406, formula 3.723-2; p. 407, formual 3.723-9
[12]Ref. 8, p. 361, Sec. 11.3, Eq. (4).
[13]Ref. 8, p. 386, Sec. 13.2, Eq. (8).
[14]Ref. 8, p. 427.
[15]D. Shanks, J. Math. and Phys. XXXIV, 1 (1955).
[16]W.R. Smythe, *Static and Dynamic Electricity*, 2nd ed. (McGraw-Hill, New York, 1950), Secs. 5.298, 5.299.
[17]Ref. 8, Chap. XV.
[18]The first term in the series (30) was given in Ref. 1.
[19]A. Peskoff, R.S. Eisenberg, and J.D. Cole, "Potential Induced by a Point Source of Current Inside an Infinite Cylindrical Cell," UCLA-ENG-7303, UCLA School of Engineering and Applied Science, Los Angeles, 1973 (to be published).
[20]F.W.J. Olver, editor *Royal Society Mathematics Tables* (Cambridge U.P., Cambridge, 1960), Vol. 7, Part III.

Limits of the Tomimatsu–Sato gravitational field

William Kinnersley

Department of Applied Mathematics, California Institute of Technology, Pasadena, California 91109

Edward F. Kelley

Department of Physics, Montana State University, Bozeman, Montana 59715
(Received 1 May 1974)

The Tomimatsu–Sato (TS) solutions of the Einstein field equations are studied in several limiting cases. In the weak-field limit we construct two Newtonian models for the source, one consisting of a rotating disc of radius a/n, the other made up of n complex point multipoles. The "extreme" limit $q=1$ is also examined in detail, and we find there are many distinct ways of taking this limit. We are thereby led to a new two-parameter family of exact solutions which, unlike the TS metrics, are not asymptotically flat.

I. INTRODUCTION

Recently Tomimatsu and Sato[1] have found a series of exact solutions of the Einstein vacuum field equations which they claim are suitable to represent the gravitational field of a rotating body. This would only be the second time such a solution has been discovered, the previous example being the Kerr metric. Further investigations by Glass[2] and by Gibbons and Russell-Clark[3] have shown that the TS solutions contain a naked singularity outside their event horizons and are therefore not black holes. The current popularity of black holes is so great that many persons would automatically reject a solution on this basis alone. However, there are perfectly good reasons for studying solutions that contain naked singularities. One whose conscience is troubled by them on astrophysical grounds may regard them as primordial remnants of the big bang, or else imagine an appropriate interior solution covering the region that would otherwise be offensive. At any rate we feel that the TS solutions have considerable mathematical and physical interest and deserve a great deal of further study.

In this paper we first try to understand the structure of the sources necessary to produce the TS field. We examine a weak-field limit in which the source has vanishingly small rest mass but finite size. In terms of the TS parameters, this implies the limit must be taken as $p, q\to\infty$. The linearized gravitational field obtained in this manner is exhibited in terms of a complex Newtonian potential. To understand its nature at large radial distances, we analyze the field into multipole moments. Near the origin, its singularities determine the source in terms of mass and mass-current distributions. The Newtonian models we thus obtain for the TS metrics are rotating discs qualitatively similar to the model discussed by Israel[4] for the Kerr metric but with a smaller radius, a/n. We also give a simpler but more abstract model, in which the source is represented as a small number of point multipoles, all located a complex distance along the axis of symmetry.

The limit $p=0$, $q=1$ is another case of particular interest which we examine in detail. We find that this limit is not unique and that different metrics result from taking it in various ways. The metrics we obtain in this manner are new rotating solutions, considerably simpler than TS. In this limit we find that the TS parameter n need no longer be restricted to integer values. Unfortunately, the new solutions are not asymptotically flat.

II. ERNST POTENTIAL

The line elements for the TS solutions are quoted in the Weyl–Papapetrou canonical form for stationary axisymmetric fields:

$$ds^2=f(dt-\omega d\varphi)^2-f^{-1}(e^{2\gamma}[d\rho^2+dz^2]+\rho^2d\varphi^2). \tag{1}$$

Here ρ, φ, z are to be thought of as cylindrical coordinates in a flat 3-space, which we call the "Weyl space," and f, ω, γ are the field variables which depend only on ρ, z. Rather than work directly with these metric components, we find it convenient to follow other authors and focus attention on a quantity called the Ernst potential.[5]

The Ernst potential ξ is a complex scalar field. It is related to the metric by[6]

$$\mathrm{Re}\left(\frac{1-\xi}{1+\xi}\right)=f, \tag{2}$$

$$\nabla\mathrm{Im}\left(\frac{1-\xi}{1+\xi}\right)=\rho^{-1}f^2\mathbf{e}_\varphi\times\nabla\omega, \tag{3}$$

where ∇ is the gradient operator on the Weyl space and $\mathbf{e}_\varphi$ is a unit vector in the φ direction. In this formulation the Einstein equations accomplish three things for us: (i) They insure that Eq. (3) is integrable for ω, (ii) they tell us how to construct γ once f, ω are given, and (ii) they provide an equation which ξ must satisfy

$$(\xi\xi^*-1)\nabla^2\xi=2\xi^*\nabla\xi\cdot\nabla\xi. \tag{4}$$

If we write ξ in terms of its real and imaginary parts

$$\xi=\Phi+i\Omega,$$

then in the weak-field limit we have from Eqs. (2), (4),

$$f=1-2\Phi,\quad \nabla^2\Phi=0,\quad \nabla^2\Omega=0.$$

Hence the real part Φ of the Ernst potential becomes the ordinary Newtonian potential. The imaginary part Ω plays the role of a "magnetic" scalar potential in analogy with electrodynamics. This can be seen from the weak-field limit of the geodesic equations, where we find that the acceleration of a slowly moving test particle is

$$\mathbf{a}=-\nabla\Phi+\mathbf{v}\times\nabla\Omega.$$

Even in the exact theory, the Ernst potential may

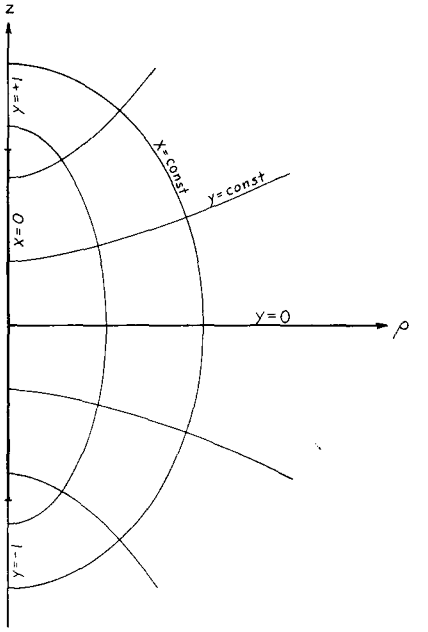

FIG. 1. Prolate spheroidal coordinates. The level surfaces $x=\text{const}$ are an orthogonal family of ellipsoids and hyperboloids. The semimajor axis of an ellipsoid is given by mpx/n, and the asymptotic angle of inclination of a hyperboloid is $\theta=\cos^{-1}y$.

still be conveniently regarded as a complexified nonlinear version of the Newtonian potential. We therefore feel it appropriate to concentrate on ξ as the quantity of direct physical and mathematical significance.

III. TS SOLUTIONS

The TS solutions to Eq. (4) contain arbitrary constant parameters m, n, p, q, where $m>0$ is the mass, n is a positive integer, and p, q are dimensionless numbers related by

$$p^2+q^2=1. \tag{5}$$

For the first two values $n=1, 2$ the solutions are[1]

$$\xi_1^{-1}=px-iqy, \tag{6}$$

$$\xi_2^{-1}=\frac{p^2(x^4-1)-2ipqxy(x^2-y^2)-q^2(1-y^4)}{2px(x^2-1)-2iqy(1-y^2)} \tag{7}$$

Here x, y are prolate spheroidal coordinates in the Weyl space (see Fig. 1) related to ρ, z by

$$\begin{aligned}\rho&=(mp/n)(x^2-1)^{1/2}(1-y^2)^{1/2},\\ z&=(mp/n)\,xy.\end{aligned} \tag{8}$$

The cases $n=3, 4$ are also given in Ref. 1. The case $n=1$ is equivalent to the Kerr metric as discussed by Ernst.[5] For $n\geq 5$ the TS solutions have not been calculated but are presumed to exist. Since the $n=4$ solution already fills half a page, there is little incentive to pursue the matter further unless a general form for all n can be discovered. Charged TS solutions could also be written down, but since the procedure for doing this is now completely automatic and understood,[7] we feel that doing so would be a definite waste of time.

If we make the natural assumption that p is real, then Eq. (5) restricts q to the range $|q|\leq 1$. As TS themselves point out,[1] the solutions may easily be extended beyond this range via a complex coordinate transformation. We let

$$\hat{p}=-ip, \quad \hat{x}=ix \tag{9}$$

and assume instead that $\hat{p}, \hat{x}$ are the quantities which are real. Then we note the following facts:

(i) Since ξ contains p, x only in quadratic combinations, none of its terms lose their reality. Thus the meaning of ξ^* is unaltered, and we still have a solution of Eq. (4).

(ii) The relation between $\hat{p}, q$ is

$$q^2=\hat{p}^2+1, \tag{10}$$

and the restriction on q is now just the opposite of what it was before, namely $|q|\geq 1$.

(iii) The new coordinates $(\hat{x}, y)$ are *oblate* spheroidal coordinates in the Weyl space (see Fig. 2) and Eq. (8) is replaced by

$$\begin{aligned}\rho&=(m\hat{p}/n)(\hat{x}^2+1)^{1/2}(1-y^2)^{1/2},\\ z&=(m\hat{p}/n)\,\hat{x}y.\end{aligned} \tag{11}$$

Finally we note that, in all of the cases discussed above, $\xi\to 0$ as $\hat{x}$ or $x\to\infty$, and hence the solutions are all asymptotically flat.

IV. WEAK FIELD LIMIT

The exact mass, angular momentum, and quadrupole moment for the TS solutions have been given by Tomimatsu and Sato[1]:

$$M=m, \quad J=m^2q, \quad Q=m^3\left(\frac{n^2-1}{3n^2}p^2+q^2\right). \tag{12}$$

In the weak field limit as $m\to 0$ we see that J will be only $O(m^2)$, too small to survive, unless $q\to\infty$ at the same time. We therefore need to use the extended TS solutions. We define a Kerr parameter a by the equations

$$q=a/m, \quad p=(a^2-m^2)^{1/2}/m \tag{13}$$

and take the weak-field limit holding a finite. Just as in the Kerr metric itself, the parameter a has the dimensions of length and serves to describe the linear extent of the source.

Carrying out the stated limit on Eqs. (6), (7), we obtain for $n=1, 2$,

$$\xi_1=(m/a)X^{-1}, \tag{14}$$

$$\xi_2=(2m/a)(X^{-1}+i(\hat{x}y-i)X^{-3}, \tag{15}$$

where

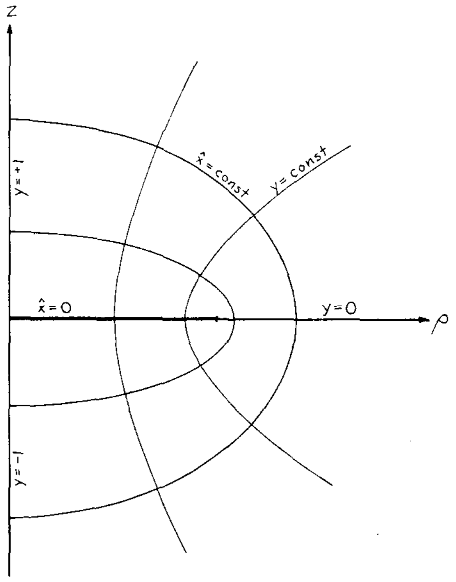

FIG. 2. Oblate spheroidal coordinates, Now $m\hat{p}\hat{x}/n$ specifies the semiminor axis of the ellipsoids.

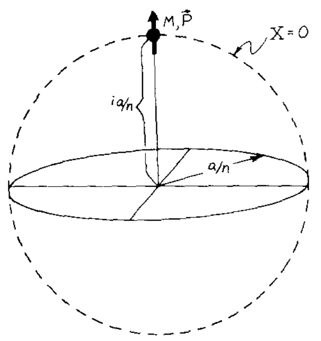

FIG. 3. Placement of point multipoles along the imaginary z-axis for the model discussed in Sec. IV.

$$X \equiv \hat{x} - iy. \tag{16}$$

Similarly from the Ernst potential for $n=3$ given in TS we obtain

$$\xi_3 = (3m/a)[X^{-1} + 2i(\hat{x}y - i)X^{-3} - \tfrac{2}{3}(3\hat{x}^2y^2 - \hat{x}^2 - 4i\hat{x}y + \hat{y}^2 - 3)X^{-5}]. \tag{17}$$

Equations (14), (15), (17) are all complex solutions of Laplace's equation, nonsingular everywhere except at $\hat{x}=y=0$. According to Eq. (11) this is the locus of a ring with radius a/n, so that whatever source is producing the field should reside there.

The results may be written in terms of spherical coordinates r, θ given by

$$\rho = r\sin\theta, \quad z = r\cos\theta$$

or alternatively in terms of coordinates $\bar{r}$, $\bar{\theta}$ defined by

$$\bar{r} = [r^2 - 2(ia/n)\cos\theta + (ia/n)^2]^{1/2} = (a/n)X, \tag{18}$$

$$\bar{r}\cos\bar{\theta} = r\cos\theta - (ia/n) = (a/n)(\hat{x}y - i). \tag{19}$$

Geometrically, $\bar{r}$, $\bar{\theta}$ are also spherical coordinates (see Fig. 3), but with their origin located at a point $z = ia/n$ on the symmetry axis, i.e., at $\hat{x}=i$, $y=1$. The multipole expansion of ξ about this point,

$$\xi_n = \sum C_{nl}\, \bar{r}^{-(l+1)} P_l(\cos\bar{\theta})$$

is quite simple, and Eqs. (14), (15), (17) show that it contains only terms up to $l=n$. Thus we have a model in which the source consists of *a finite number of point multipoles all placed at the same complex location* $z=ia/n$. It seems probable that the same type of model exists for higher values of n than $n=3$, but we have not explored this matter further.

At large radial distances one would like to have the linearized TS fields analyzed into their multipole moments. This is readily accomplished since Eqs. (14), (15), (17) can be expanded in spherical harmonics by means of a generating function. With the definition

$$\xi_n = \sum Q_{nl}\, r^{-(l+1)} P_l(\cos\theta)$$

we find that

$$Q_{1l} = m(ia)^l,$$
$$Q_{2l} = m(l+1)(ia/2)^l, \tag{20}$$
$$Q_{3l} = \tfrac{1}{3}m(2l^2 + 4l + 3)(ia/3)^l.$$

These linearized moments are in agreement with the first few exact moments quoted in Eq. (12). One of us (W.K.) has also previously calculated[8] the exact multipole moments for the Kerr metric as far as $l=8$. This was done with the aid of SYMBAL, a formula-manipulating program available for the CDC6600 computer. To that point in the calculation, the amazing result was that *no term of order m^2 had appeared,* and the exact Janis moments were still reproducing exactly the linearized moments given in Eq. (20)! With the other TS metrics this is obviously not the case.

V. DISC MODELS

The attractive simplicity of the model discussed above is offset by its somewhat symbolic use of complex coordinates, and we therefore now consider an alternative Newtonian model with real mass and mass-current distributions. We have stated that the Ernst potential is singular at $\hat{x}=y=0$ which is a ring of radius a/n. Another singularity arises from the spheroidal coordinate system itself (see Fig. 2). The coordinate y is discontinuous across the entire disc $\hat{x}=0$, being positive on one face and negative on the other. This leads to a corresponding discontinuity in ξ, and a consequent necessity for sources everywhere on this surface.

On the disc $\hat{x}=0$, the radial coordinate is

$$\rho = (a/n)(1-y^2)^{1/2}, \tag{21}$$

the element of surface area is

$$|dA| = 2\pi\rho\, d\rho = 2\pi(a/n)^2 y\, dy, \tag{22}$$

and the normal derivative is

$$\frac{\partial}{\partial z} = \frac{n}{ay}\frac{\partial}{\partial x}. \tag{23}$$

On any surface $\hat{x} = \text{const} \neq 0$,

$$\xi_1 = iq(y + i\hat{x})^{-1},$$
$$\xi_2 = 2iq(y+i\hat{x})^{-1} + 2q\hat{x}(y+i\hat{x})^{-2} - 2iq(1+\hat{x})^2(y+i\hat{x})^{-3},$$
$$\xi_3 = 3iq(y+i\hat{x})^{-1} + 6q\hat{x}(y+i\hat{x})^{-2} - 4iq(2+3\hat{x}^2)(y+i\hat{x})^{-3} - 12q(\hat{x}+\hat{x}^3)(y+i\hat{x})^{-4} + 6iq(1+\hat{x}^2)^2(y+i\hat{x})^{-5}, \tag{24}$$

where $q = m/a$. As $\hat{x} \to 0$ we have

$$\xi_1 \to iqy_+^{-1},$$
$$\xi_2 \to iq(2y_+^{-1} - 2y_+^{-3}), \tag{25}$$
$$\xi_3 \to iq(3y_+^{-1} - 8y_+^{-3} + 6y_+^{-5}),$$

where

$$y_+^{-n} \equiv \lim_{\epsilon\to 0}(y + i\epsilon)^{-n}.$$

The functions y_+^{-n} must be understood as generalized functions,[9] and in that context they have nonvanishing imaginary parts. For example,

$$y_+^{-1} = y^{-1} - i\pi\delta(y).$$

All of the other functions may be obtained from this one by repeated differentiation.

In Newtonian gravity the mass density of a sheet is given by

$$\sigma = -(2\pi)^{-1}\mathbf{n}\cdot\nabla\Phi \tag{26}$$

and the current density is

$$\mathbf{j} = (2\pi)^{-1}\,\mathbf{n}\times\nabla\Omega, \tag{27}$$

where $\mathbf{n}$ is the unit normal. Using Eqs. (24), we find that

$$\begin{aligned}\sigma_1 &= (m/2\pi a^2 y)(y^{-2}),\\ \sigma_2 &= (m/2\pi a^2 y)(2y^{-2}-3y^{-4}),\\ \sigma_3 &= (m/2\pi a^2 y)(3y^{-2}-12y^{-4}+10y^{-6}).\end{aligned} \tag{28}$$

The factor $1/y$ in all of these expressions has deliberately been isolated, for it must be eventually combined with the factor of y in the area element, Eq. (22). What remains in the parenthesis in σ_n is the generalized function. Note that σ_n is either positive definite or negative definite, depending on whether n is odd or even. Also note that as we approach the edge of the disc at $y=0$, σ_n diverges as

$$\sigma_n \sim [(a/n)^2-\rho^2]^{-(2n+1)/2}. \tag{29}$$

This would appear to imply that the total mass,

$$M_n = \tfrac{1}{2}\int_{-1}^{1} 2\pi(a/n)\,\sigma_n\, y\, dy, \tag{30}$$

would have to be infinite. However, there is a further singularity in σ_n which is concentrated on the ring $y=0$ and which is due solely to its interpretation as a generalized function. The simple (but rigorous) rule for handling a divergent integral like M_n is that the expression is integrated and then evaluated at the end points $y=\pm 1$ *just as if no singularity at* $y=0$ *were present*. For all three cases we confirm in this manner that $M_n=m$. Roughly speaking, one may say that there is an infinite mass density residing on the ring, of such a sign and strength as to make the total mass of ring plus disc finite.

Now for all three values of n Eq. (25) shows that ξ_n is purely imaginary on the disc, and hence the surface is an equipotential. One might therefore wonder what relationship these solutions have to the familiar electrostatic problem of a charged conducting disc, in which the surface is also an equipotential. In that problem the solution is[10]

$$V = (2/\pi)\cot^{-1}\hat{x} = \frac{i}{\pi}\ln\left(\frac{\hat{x}-i}{\hat{x}+i}\right). \tag{31}$$

This function is singular at $\hat{x}=\pm i$, which is an entire line segment, $\rho=0$, $z=iay$, $-1\le y\le 1$. Moreover, σ, now a charge density, is once again divergent at the disc's edge,

$$\sigma = (a^2-\rho^2)^{-1/2}.$$

To examine the behavior one would generally expect to find there, let r, ϕ be a local set of cylindrical coordinates whose axis coincides with the edge. Laplace's equation in this neighborhood will have the solution

$$V \sim \sum r^m \cos m\phi.$$

Although m would normally be an integer, the presence of the disc forces the appropriate choice to be a half-(integer instead. The charged disc picks $m=\frac{1}{2}$, while for the family of TS solutions we have $m=-n+\frac{1}{2}$. (We might therefore hope that a TS solution will someday be discovered for $n=0$!) All of these solutions have period 4π in the angle ϕ, and it is therefore a quite natural thing to consider extending them to a twofold covering of Minkowski space using the ring as a branch line. This procedure is thus not a unique feature of the Kerr metric.

Returning to the mass-current densities, we find that

$$\begin{aligned}j_1 &= -(m/2\pi a^2 y)(1-y^2)^{1/2}(y^{-2}),\\ j_2 &= -(4m/2\pi a^2 y)(1-y^2)^{1/2}(y^{-2}-3y^{-4}),\\ j_3 &= -(9m/2\pi a^2 y)(1-y^2)^{1/2}(y^{-2}-8y^{-4}+10y^{-6}).\end{aligned} \tag{32}$$

The total angular momentum is

$$J_n = \tfrac{1}{2}\int_{-1}^{1}\pi(a/n)^3 j_n\, y(1-y^2)^{1/2}\,dy, \tag{33}$$

which yields $J_n=ma$ for all three values of n. The velocity of rotation, even when special relativistic effects are included, is just

$$v_n = j_n/\sigma_n.$$

From Eqs. (28), (32) we see that the rotation is not rigid and that $v_n\to 1$ as the ring is approached.

VI. THE LIMIT $q=1$

To obtain the static (i.e., nonrotating) limit $a=0$ of the Kerr metric and the other TS metrics, we simply take the expression for the Ernst potential and set $p=1$, $q=0$. On the other hand, the so-called "extreme" Kerr limit $a=m$ *cannot* be obtained in so straightforward a manner merely by setting $p=0$, $q=1$. This is to say, the metric computed from $\xi^{-1}=-iy$ is *not* extreme Kerr. The reason that this limit needs special treatment may be seen in Eq. (8), where we observe that the transformation from (ρ,z) to (x,y) becomes singular as $p\to 0$. As a consequence there are various ways in which the limiting process might be performed, depending on whether ρ or x is required to remain finite. Possibly even some intermediate method might be attempted.

Consider, for example, the situation that arises for the Kerr metric itself. The relationship between Kerr coordinates R,Θ and the Weyl–Papapetrou coordinates is

$$\begin{aligned}\rho &= (R^2-2mR+a^2)^{1/2}\sin\Theta\\ z &= (R-m)\cos\Theta\end{aligned} \tag{34}$$

or alternatively

$$\begin{aligned}mpx &= R-m\\ y &= \cos\Theta.\end{aligned} \tag{35}$$

The Ernst potential is

$$\xi^{-1} = \frac{(R-m)-ia\cos\Theta}{m}. \tag{36}$$

Now, if the limit $p\to 0$ is taken holding either ρ or R finite, we obtain the usual extreme Kerr metric. On the other hand, if we allow our coordinates to be rescaled so that x remains finite and $\xi^{-1}=-iy$, then necessarily $\rho\to 0$ and $R\to m$. One might therefore presume that the metric we obtain in this manner would be the Kerr metric restricted to the null surface $R=m$, and hence a metric that lacks the full Lorentz signature. If ρ were strictly zero this would certainly be the case. In fact, the process only confines ρ to a neighborhood of the axis and we find that metric to be

$$\xi^{-1} = -iy = -i\cos\Theta,$$

$$ds^2 = \frac{\sin^2\Theta}{1+\cos^2\Theta}(dt-2rd\varphi)^2 \tag{37}$$

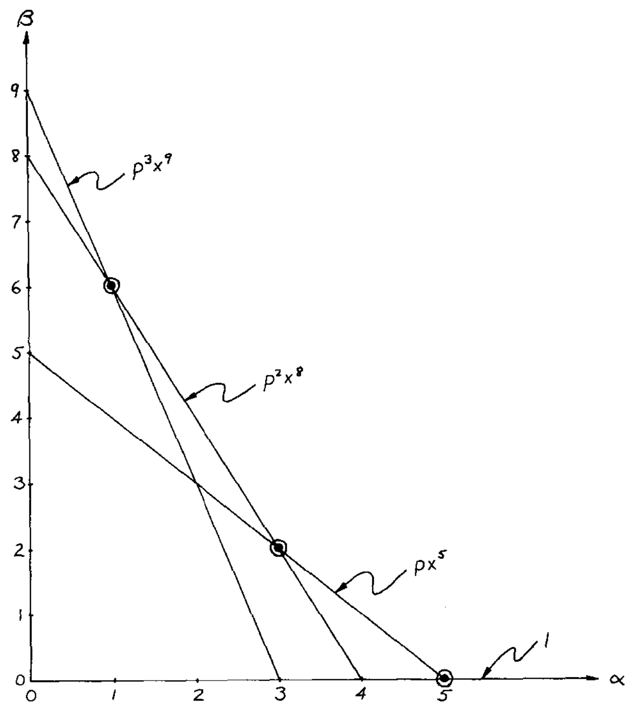

FIG. 4. Orders of magnitude of terms appearing in the Ernst potential for $n=3$. The vertical coordinate β is used to indicate that a term is $O(x^\beta)$. Bullseyes denote the distinguished limits at each odd value of α.

$$-(1+\cos^2\Theta)(r^{-2}dr^2+d\Theta^2+r^2d\varphi^2).$$

This is a type D metric and therefore a well-known one.[11,12] Its role as a limiting metric valid in the immediate vicinity of the extreme Kerr throat has been discussed previously by Bardeen.[13]

We next consider the situation that arises in the general TS metric. As Tomimatsu and Sato themselves have pointed out,[1] the limit may be taken holding the product px finite. If we then define R, Θ coordinates by

$$R-m=mpx/n, \quad \cos\Theta=y,$$

we will always obtain extreme Kerr in the limit, *regardless of which value of n we start with.*

Now we will show that there are other nontrivial ways of performing the limit that do lead to different solutions. Suppose the limit is taken in such a way that px^α remains finite, where α is an adjustable constant. The process is best illustrated using the case $n=3$. We write down from TS the Ernst potential, keeping only the leading terms in x:

$$\xi_3^{-1}=w/u,$$

$$w\approx p^3x^9-3ip^2x^8y-6px^5(1-y^4)+i(1-y^2)^3(y^3+3y), \tag{38}$$

$$u\approx 3p^2x^8-12ipx^5y(1-y^2)-(1-y^2)^3(3y^2+1).$$

The only powers of p and x that appear are p^3x^9, p^2x^8, px^5, and 1. In Fig. 4 we have plotted the order of magnitude of each term as a function of α. For almost all values of α, one term exceeds all the others in order of magnitude, and hence becomes the sole survivor as the limit is taken. For certain values, $\alpha=1, 3, 5$, the two largest terms happen to have the same order of magnitude, and we then obtain what is known as a "distinguished limit."

For $q=1$ limits which are not distinguished, the resulting Ernst potential must clearly be a function of y alone, and turns out to have the form

$$\xi^{-1}=-i\left(\frac{(1+y)^k-(1-y)^k}{(1+y)^k+(1-y)^k}\right), \tag{39}$$

where k is an integer, $k\leq n$. These solutions are not really new, since they can be quite easily obtained from the Voorhees metric[14] by making the replacement $x\longleftrightarrow y$. However, they may deserve more attention than has been previously been paid to them. The entire metric is

$$ds^2=f(dt-2kr\,d\varphi)^2$$
$$-f^{-1}[\sin^{2k^2}\Theta\; r^{-2k^2}(dr^2+r^2d\Theta^2)+r^2\sin^2\Theta d\varphi^2], \tag{40}$$

where f may be written as

$$f=2[\tan^{2k}(\theta/2)+\cot^{2k}(\theta/2)]^{-1}. \tag{41}$$

This solution is a type I generalization of Eq. (37) and describes a region of the TS metric near its ergosphere. Like Eq. (37), this metric is not asymptotically flat.

The distinguished limits lead to metrics which are apparently new ones. For example, for $\alpha=3$ we find

$$\xi^{-1}=-i\left(\frac{(1-y^4)+2ipx^3y}{2y(1-y^2)+2ipx^3}\right) \tag{42}$$

and this same solution is obtained for $\alpha=3$ from every TS metric regardless of which value of n we start with (provided only $n\geq 2$). The coordinates r, θ defined by

$$(r/m)^3=\tfrac{1}{2}px^3, \quad \cos\theta=y \tag{43}$$

will be spherical coordinates in the Weyl space, and ξ may be conveniently written in terms of them.

In general, for $\alpha=2k-1$ we define

$$\left(\frac{r}{m}\right)^{2k-1}=\frac{(k!)^2}{(2k)!}\,px^{2k-1},$$
$$\cos\theta=y \tag{44}$$

and obtain the exact solution

$$\xi^{-1}=w/u,$$

$$w=(r/m)^{2k-1}[(1+y)^{k-1}-(1-y)^{k-1}]$$
$$-i(1-y^2)^{k-1}[(1+y)^k+(1-y)^k],$$

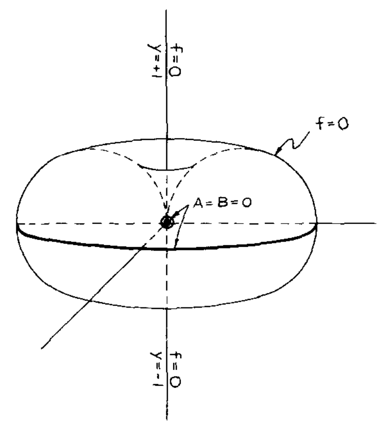

FIG. 5. Singularities of the function f in the new family of solutions, Eq. (46). Directional singularities occur at the origin and along the equator of the torus, as indicated.

$$u = (1-y^2)^{k-1}[(1+y)^k - (1-y)^k] + i(r/m)^{2k-1}[(1+y)^{k-1} + (1-y)^{k-1}]. \tag{45}$$

Furthermore, if Eq. (45) is now regarded simply as a solution in its own right (not derived from a TS solution), there is no reason to restrict k to be an integer, and we may allow it to take on any real value.

The metric has the form of Eq. (1) with

$$\begin{aligned} f &= 2(1-y^2)^{k-1}A/B, \\ \omega &= rC/mA, \\ \exp(2\gamma) &= A(1-y^2)^{(k-1)^2}(r/m)^{-2k^2} \end{aligned} \tag{46}$$

and

$$\begin{aligned} A &= (r/m)^{4k-2} - (1-y^2)^{2k-1}, \\ B &= (r/m)^{4k-2}[(1+y)^{2k-2} + (1-y)^{2k-2}] \\ &\quad - 4(r/m)^{2k-1}(1-y^2)^{2k-2} + (1-y^2)^{2k-2}[(1+y)^{2k} + (1-y)^{2k}], \\ C &= 2(k-1)(r/m)^{4k-2} + (r/m)^{2k-1}[(1+y)^{2k-1} \\ &\quad + (1-y)^{2k-1}] - 2k(1-y^2)^{2k-1}. \end{aligned} \tag{47}$$

These solutions have not yet been examined in any great detail, but we can make a few preliminary remarks about their properties. For either $r \to 0$ for $r \to \infty$ they asymptotically approach one or another of the undistinguished solutions of Eq. (42). Hence they are not asymptotically flat. The function f has zeroes on the symmetry axis $y = \pm 1$, and on a torus $(r/m)^2 = 1 - y^2$ (see Fig. 5) and one would expect these to be surfaces of infinite redshift. However, the denominator B also vanishes at the origin $r = 0$, $y = \pm 1$, and on the ring $r = m$, $y = 0$. At these points f will possess an angular singularity similar to those which have already been discussed for the Weyl and TS metrics. For example, let

$$y = \epsilon, \quad r = m + \eta,$$

where ϵ, η are assumed small. Then we find, in a neighborhood of the ring,

$$f \approx \frac{1}{2k-1}\left(\frac{\epsilon^2 + 2\eta}{\epsilon^2 + \eta^2}\right).$$

The limiting value of f will be infinite as long as we approach the ring along a straight path, $\epsilon/\eta = \text{const}$, but if we approach it along a parabola $\epsilon^2/\eta = \text{const}$, we can obtain a limit which is any finite value we please, including zero.

*Supported in part by NSF Grant GT-32157X.

[1] A. Tomimatsu and H. Sato, Prog. Theor. Phys. 50, 95 (1973).
[2] E. Glass, Phys. Rev. D 7, 3127 (1973).
[3] G. Gibbons and R. Russell-Clark, Phys. Rev. Lett. 30, 398 (1973).
[4] W. Israel, Phys. Rev. D 2, 641 (1970).
[5] F. Ernst, Phys. Rev. 168, 1415 (1968).
[6] Our Ernst potential is the reciprocal of the one introduced in Ref. 5. Equation (4) is invariant under $\xi \to \xi^{-1}$, and we have used this fact in order to have $\xi \to 0$ when $r \to \infty$.
[7] W. Kinnersley, J. Math. Phys. 14, 651 (1973).
[8] W. Kinnersley (unpublished).
[9] I. M. Gel'fand and G.E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I, p. 60. We have called $y_{\mp}^{n}$ what Gel'fand calls $(y + i0)^{-n}$.
[10] J. Jeans, *Mathematical Theory of Electricity and Magnetism* (Cambridge U.P., Cambridge, 1927), p. 249.
[11] B. Carter, Comm. Math. Phys. 10, 280 (1968).
[12] W. Kinnersley, J. Math. Phys. 10, 1195 (1969).
[13] J. Bardeen, unpublished communication to C. Misner.
[14] B. Voorhees, Phys. Rev. D 2, 2119 (1970).

Effects of long range interactions in harmonically coupled systems. I. Equilibrium fluctuations and diffusion

Paul E. Phillipson

*Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado 80302**

C.N.R. Center for Molecular Biology at the Institute of Biochemistry, University of Rome, Italy
(Received 12 July 1974)

Systems of harmonically coupled identical particles at thermal equilibrium provide dynamical models for studies of diffusion due to equilibrium fluctuations. The velocity autocorrelation function and mean square displacement of a particle selected from a given system are investigated for various models which have the common feature that the particle is directly coupled to $L > 1$ neighbors, reflecting the influence of long range interactions. Theorems are developed which indicate how the time course of diffusion is dictated by analytic properties of the vibrational frequency distribution as well as by quantum fluctuations whose presence is betrayed by the increasingly important role at progressively lower temperatures of $\tau_q = \hbar/\pi k T$, the quantum transient time. The formalism is first applied to a system for which the long range couplings are so parametrized by a range parameter z that when $z=0$ the frequency distribution is identical to that for nearest neighbor coupling only ($L=1$), while as z approaches unity ($L\rightarrow\infty$) the frequency distribution becomes identifiable with that of Ford, Kac, and Mazur which served as the starting point for their dynamical theory of Brownian motion. Consequences of this model are: (1) when $z < 0.5$, the classical velocity autocorrelation functions exhibit similar qualitative features to those computed for molecular diffusion in simple liquids; (2) as z approaches unity, the classical velocity autocorrelation function approaches the $e^{-\lambda\tau}$ Gaussian Markoffian form, and the mean square displacement in the same limit is identical to that predicted by the Langevin equation; (3) at low temperatures such that $\lambda\tau_q > 1$, quantum fluctuations tend to dominate thermal fluctuations, resulting in severe departures from Gaussian Markoffian behavior. The low temperature effects are analyzed in some detail, and it is suggested that the predicted departure of the mean square displacement from its classical behavior might be displayed by a particle of macroscopic size suspended in a superfluid. Other models are developed which yield mean square displacements which depart even at high temperature from the linear dependence upon time characteristic of classical diffusion. The reasons and possible physical implications of these behaviors are discussed, together with a brief consideration of Poincaré cycles, whose neglect is implicit in any dynamical theory of irreversible processes.

I. INTRODUCTION

The dynamics of harmonically coupled particles have for several years[1–4] provided a useful basis for analytically investigating many-body effects within the framework of a solvable physical model. The procedure involves two steps. First the motion of each particle is found in terms of a linear combination of normal modes of vibration. Then assumptions are made about the initial configuration of the system at zero time. If the initial dynamical configuration is defined in terms of a statistical (Gaussian) distribution over the normal modes,[2,3] or particle displacements and velocities,[4] and if the number of particles goes to infinity, then the time evolution of a single particle selected from the system exhibits features characteristic of a particle coupled to a heat bath. Specifically, by choosing the initial conditions on the coordinates and velocities of each normal mode to be fixed consistent with thermal equilibrium at temperature T, then the system as a whole is simply a collection of independent harmonic oscillators at thermodynamic equilibrium, and any single particle will show the effects of equilibrium fluctuations[5] in its behavior. The consequences of equilibrium fluctuations upon single particle dynamics are most naturally analyzed in terms of the single particle velocity autocorrelation function.[6] This quantity in general assumes some initial value fixed by the temperature at time $t=0$ and approaches zero as $t\rightarrow\infty$. Derivable from the velocity autocorrelation function is the particle mean square displacement whose evolution in time reflects particle diffusion due to equilibrium fluctuations. The precise behavior of the velocity autocorrelation function and the nature of the consequent diffusive motion, if it occurs at all, are dictated by the properties of the normal mode frequency distribution.

These investigations, with one important exception, assume that the heat bath oscillator assembly is constructed of particles harmonically coupled to their nearest neighbors, as in a one-dimensional solid. However, from a theoretical point of view it is possible to construct one-dimensional systems in which a given particle is harmonically coupled not only to its nearest neighbors, but to any other neighbors as well. In such a case the vibrational frequency distribution becomes parametrically a function of a force constant distribution which reflects long range interactions between "active neighbors"[7] far from a given particle. Such frequency distributions were discussed briefly many years ago by Brillouin,[8] and a particular force constant distribution is implicit in the previously mentioned exception, the investigations of Ford, Kac, and Mazur.[2] These authors showed that for a particular choice of frequency distribution they could derive a statistical mechanical theory of Brownian motion. This work has recently been extended by Zwanzig[9] to derive generalized Langevin equations for nonlinear systems interacting with suitably constructed heat baths.

The purpose of the present work is to examine the effects of long range couplings upon the diffusive behavior of a single particle embedded in a one-dimensional harmonic oscillator assembly of identical particles. Interest will be confined to the single particle velocity autocorrelation function and the consequent mean square displacement. In Sec. II the assumptions underlying the systems to be treated are stated together with the solutions of the normal mode problem. In Sec. III expressions are derived for the velocity autocorrelation function and mean square displacement within the framework of quantum statistical mechanics. Certain statements can be made concerning the latter which depend on general analytic properties of the frequency distribution function $\rho(\omega)$ defined in Sec. II. It is shown that if $\rho(\omega)$ is a slowly varying function of ω then in the high temperature limit that $T\to\infty$, the mean square displacement behaves quite generally as $2Dt$, the result of classical self-diffusion. The diffusion coefficient D is determined by $\rho(\omega)$ evaluated at zero frequency. This time dependence is true asymptotically, however, and holds only for times such that $\omega_0 t \gg 1$, where ω_0 defines the cutoff of the frequency distribution. The nearest neighbor lattice and the model of Ford *et al.*[2] are special cases of this general result. It is possible to connect smoothly the nearest neighbor model with that which includes long range couplings implied by Ford *et al.* through introduction of a geometric progression of interaction constants. This program is carried out in Sec. IV. The infinite cutoff frequency which these authors postulated is replaced by a finite cutoff frequency directly reflecting long range interactions. Its presence produces velocity autocorrelation functions which behave qualitatively as those computed for center of mass molecular motion in simple liquids.[10] The model in general permits the tracing of the progressive role of long range couplings as time goes on to produce the diffusive behavior characteristic of Brownian motion.

When the above two conditions, $T\to\infty$ and $\rho(\omega)$ slowly varying, are relaxed new features appear. The behavior of these model systems for arbitrary temperatures introduces another physical parameter, the quantum transient time $\hbar/kT$ which has been considered by Ullersma[3] and earlier by MacDonald.[11] In the classical limit it is of course zero and hence does not appear. At low temperatures, on the contrary, it assumes progressively more importance. The general considerations of Sec. III show that at absolute zero such that $\hbar/kT\to\infty$, if $\rho(\omega)$ is slowly varying the linear time dependence of the mean square displacement in the classical limit is replaced in the quantum limit by a logarithmic dependence on time for $\omega_0 t\gg 1$. The cause of this "zero point drift" is associated with the zero point vibrations of the normal modes. However, as soon as there is a departure from zero temperature, albeit small, there is a competitive asymptotic contribution which adds to the zero point term. It is of the same form as the corresponding classical form, proportional to T and linearly proportional to time. The latter thus tends to swamp out the logarithmic contribution. On the other hand, the classical Langevin theory of Brownian motion[12] predicts a mean square displacement which is the sum of two terms, a term linear in time and a decaying term which behaves as $\exp(-\lambda t)$, where λ^{-1} is a characteristic relaxation time. The effect of a non-zero quantum transient time is to severely modify this second term. Accordingly, section V will consist of an analysis of the competitive effects of the quantum transient time upon Brownian motion within the framework of the geometric progression model. As $T\to\infty$, of course, the classical result is obtained, while at $T=0$ the mean square displacement is of an elaborate functional form, reducing to $\log\lambda t$ as $\lambda t\to\infty$.

When the second condition is relaxed more exotic diffusive behavior appears. If $\rho(0)$ is infinite the mean square displacement is asymptotically proportional to $t^{\alpha}(\alpha>1)$ in the classical limit and $t^{\alpha-1}$ in the quantum limit. The value of α is determined by the nature of the singularity of $\rho(\omega)$ as $\omega\to 0$. If, on the other hand $\rho(0)$ vanishes, then $\alpha<1$, so that diffusive behavior can appear in the classical limit but not in the quantum limit. In the latter case the particle never escapes from an equilibrium position. Diffusion in the context of the present scheme, then, becomes classifiable in terms of the nature of the singular or vanishing behavior of $\rho(0)$. Examples of both cases in the classical and quantum limits are investigated for particular force constant distributions in Sec. VI.

The choices of force constant distributions are not intended to be exhaustive, but are selected to illustrate one or more aspects of the quantum statistical mechanics of irreversibility. Section VII will briefly consider the general problem of Poincaré cycles and irreversibility within the context of harmonic models, and concludes this work with a general assessment of their applicability to physical systems.

II. THE MODEL

Consider $2N+1$ identical particles each of mass M arranged to form a linear one-dimensional lattice with free ends. It is assumed there exists a stable configuration such that if all the particles were at rest, adjacent particles would be a lattice distance d apart: particle n is at nd measured to the right of the origin along the X axis. The position of the nth particle at time t is $X_n(t)=x_n(t)+nd$, so that $x_n(t)$ represents the displacement of the nth particle with respect to its rest position. Maximally there can be $2N$ interactions between nearest neighbors, $2N-1$ between next nearest neighbors ... $(2N+1)-m$ between particles md apart... one interaction between end particles, for a total of $N(2N+1)$ possible interparticle interactions. For small displacements around the rest positions the potential energy in the harmonic approximation is

$$V=\frac{1}{2}\sum_{m=1}^{L\le 2N}\sum_{n=1}^{(2N+1)-m} g_m(x_{n+m}-x_n)^2. \tag{1}$$

$L=1$ corresponds to nearest neighbor coupling only ($2N$ couplings) while if $L=2N$ the interactions are maximal [$N(2N+1)$ couplings]. Each force constant g_m is physically a measure of direct coupling of particle n with the particle a lattice distance md from it. As an example, a lattice of seven particles is illustrated in Fig. 1. The inner sum in the potential energy is constrained for each n depending upon m because the parti-

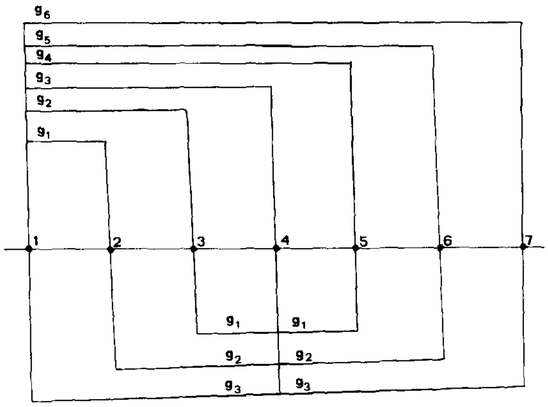

FIG. 1. Long range harmonic couplings in a linear lattice of seven particles.

cles are in different dynamical environments in pairs, with an exceptional central particle. The central particle $n=N+1$ is in a symmetric environment in which the force to its right is balanced by an identical one to its left: it can experience at most N interactions $g_1, g_2 \cdots g_N$. Particles n and $(2N+2)-n$ are identical in pairs and can experience at most $(2N+1)-n$ interactions $g_1, g_2 \cdots g_{2N+1-n}$.

For $L=1$ the solutions to the equations of motion which Eq. (1) leads to for free ends $(x_0 = x_1,\ x_{2N+1} = x_{2N+2})$ are well known.[13] When long range couplings are included, however, $(L>1)$ the boundary conditions become very complex and the dynamical problem admits of no simple analytical solutions, a point made already by Brillouin.[8] Physically these complications represent extended boundary effects which can be formally eliminated by the requirement that the system be translationally invariant. This condition implies that while $g_1 \cdots g_N$ can be formally independent quantities, the remaining constants $g_{N+1} \cdots g_{2N}$ are related to them by $g_{2N} = g_1, g_{2N-1} = g_2 \cdots g_{N+1} = g_N$ or in general

$$g_{2N-m} = g_{m+1}, \quad m = 0, 1 \cdots N-1. \tag{2}$$

For seven particles, $g_6 = g_1$, $g_5 = g_2$, $g_4 = g_3$ and the particles can now be visualized as distributed on the circumference of a ring with the interactions represented by cords connecting the particles. This is shown in Fig. 2. In the general case, traversing the ring in, say, a clockwise direction returns one to a given particle after counting $2N+1$ particles. Thus $x_n = x_{n+2N+1}$ which is the usual statement of cyclic boundary conditions.[13] Analytically, from Eqs. (1) and (2)

$$\begin{aligned}
V_{\text{cyclic}} &= V(g_{2N-m} = g_{m+1}) \\
&= \frac{1}{2}\Big\{ g_1 \sum_{n=1}^{2N} (x_{n+1} - x_n)^2 + (g_{2N} = g_1)\,[(x_{2N+1} - x_1)^2] \\
&\quad + g_2 \sum_{n=1}^{2N-1} (x_{n+2} - x_n)^2 + (g_{2N-1} = g_2)\,[(x_{2N} - x_1)^2 \\
&\quad + (x_{2N+1} - x_2)^2] \\
&\quad \cdots\cdots \\
&\quad + g_{m+1} \sum_{n=1}^{2N-m} (x_{n+m+1} - x_n)^2 + (g_{2N-m} = g_{m+1}) \\
&\quad \times \sum_{n=1}^{m+1} (x_{n+2N-m} - x_n)^2 \\
&\quad \cdots\cdots\cdots \\
&\quad + g_N \sum_{n=1}^{N+1} (x_{n+N} - x_n)^2 + (g_{N+1} = g_N)\left[\sum_{n=1}^{N} (x_{n+N+1} - x_n)^2\right]\Big\}
\end{aligned} \tag{3}$$

$$\begin{aligned}
&= \frac{1}{2}\Big\{ g_1 \sum_{n=1}^{2N+1} (x_{n+1} - x_n)^2 \ (x_1 = x_{2N+2}) \\
&\quad + g_2 \sum_{n=1}^{2N+1} (x_{n+2} - x_n)^2 \ (x_1 = x_{2N+2},\ x_2 = x_{2N+3}) \\
&\quad \cdots\cdots \\
&\quad + g_{m+1} \sum_{n=1}^{2N+1} (x_{n+m+1} - x_n)^2 \ (x_1 = x_{2N+2}, x_2 = x_{2N+3} \cdots x_{m+1} \\
&\qquad = x_{m+2N+2}) \cdots\cdots \\
&\quad + g_N \sum_{n=1}^{2N+1} (x_{n+N} - x_n)^2 \ (x_1 = x_{2N+2}, x_2 = x_{2N+3} \cdots x_n \\
&\qquad = x_{n+2N+1}) \Big\},
\end{aligned}$$

so that

$$V_{\text{cyclic}} = \frac{1}{2} \sum_{m=1}^{L \le N} \sum_{n=1}^{2N+1} g_m (x_{n+m} - x_n)^2; \quad x_n = x_{n+2N+1}. \tag{4}$$

Comparison of Eq. (4) with Eq. (1) shows that the imposition of the conditions of Eq. (2) on the force constants, plus cyclic boundary conditions, removes the dependence of the inner sum upon the range of coupling. Translational invariance restricts the maximum allowed coupling to $L=N$ and all particles behave identically, roughly as the "central particle." For nearest neighbor coupling only $(L=1)$ just the condition $x_1 = x_{2N+2}$ is necessary while the general condition $x_n = x_{n+2N+1}$ is redundant. When maximum coupling is allowed Eq. (3) indicates the more general condition is necessary.

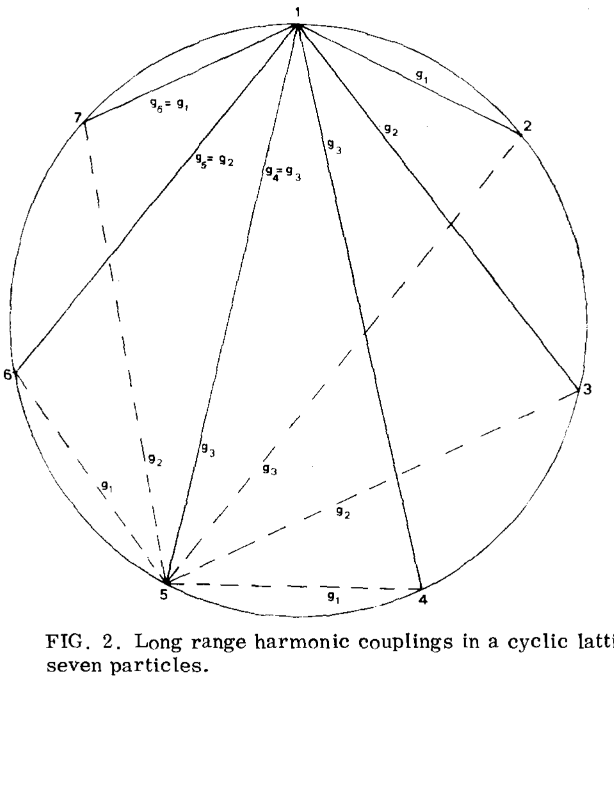

FIG. 2. Long range harmonic couplings in a cyclic lattice of seven particles.

The Hamiltonian implied by Eq. (4)

$$H=\frac{1}{2M}\sum_{n=1}^{2N+1} p_n^2+V_{\text{cyclic}};\ p_n=M\dot{x}_n \tag{5}$$

leads to the equations of motion

$$M\ddot{x}_n+\sum_{m=1}^{L} g_m[2x_n-x_{n+m}-x_{n-m}]=0;\ x_n=x_{n+2N+1}. \tag{6}$$

The solutions to Eq. (6) are given by the same linear combination of normal modes Q_j which are appropriate to $L=1$ for cyclic boundary conditions. Specifically, the solutions are effected to the following real transformation:

$$x_n(t)=\frac{1}{[M(2N+1)]^{1/2}}\sum_{j=-N}^{+N}\left(\sin\frac{2\pi nj}{2N+1}+\cos\frac{2\pi nj}{2N+1}\right)Q_j(t)$$

$$Q_j(t)=Q_j(0)\cos\omega_j t+P_j(0)(\sin\omega_j t/\omega_j);\ P_j(t)=\dot{Q}_j(t) \tag{7}$$

so that the Hamiltonian is now in diagonal form

$$H=\frac{1}{2}\sum_{j=-N}^{+N}(P_j^2+\omega_j^2Q_j^2), \tag{8}$$

where the frequencies of the normal modes of vibration are[2,8]

$$\omega_j^2=\frac{2}{M}\sum_{m=1}^{L\leq N} g_m\left(1-\cos\frac{2\pi mj}{2N+1}\right),\ j=0,\pm 1\cdots\pm N. \tag{9}$$

The frequencies are doubly degenerate, $\omega_j=\omega_{-j}$ except when $j=0$ for which $\omega_0=0$ corresponds to uniform translation of all the particles. This result generalizes the nearest neighbor case, where the frequencies now involve a sum over L terms for each $j\neq 0$.

The frequencies of Eq. (9) have an interesting mathematical property which has been stressed by Brillouin[8]: since ω_j^2 can be expressed as a polynomial of degree L, ω_j^2 is a single-valued function of j, but j is not necessarily a single valued function of ω_j. However, it is important to note that it can be provided $N\to\infty$. Furthermore, in the limit that $N\to\infty$ the frequencies lie continuously in the interval $0\leq\omega\leq\omega_0$, where ω_0 is the cutoff frequency

$$\omega_0^2=\frac{4}{M}\sum_{\substack{m=1\\(m\text{ odd})}}^{L} g_m. \tag{10}$$

Now L can be infinite as well as finite ($L\to N\to\infty$). Since it will prove possible to choose force constant distributions such that j is in fact a single-valued function of ω, whether L is finite or infinite, it is a simple matter in these cases to find $j(\omega)$ by inversion. The first derivative of $j(\omega)$ then defines the following frequency distribution:

$$\rho(\omega)=\begin{Bmatrix}\dfrac{2}{2N+1}\dfrac{dj}{d\omega}, & 0\leq\omega\leq\omega_0\\ 0 & ,\ \omega>\omega_0\end{Bmatrix} \tag{11}$$

$$\int_0^{\omega_0}\rho(\omega)\,d\omega=1.$$

$\rho(\omega)$ represents the well-known density of frequencies, or number of frequencies per unit frequency interval divided by the total number of frequencies,[14] with ω_0 functioning as the cutoff of the frequency spectrum.

Quantities which involve sums over j will now become integrals over ω. That is

$$\frac{1}{2N+1}\sum_{j=-N}^{+N}F(j)\xrightarrow{N\to\infty}\int_0^{\omega_0}F(\omega)\rho(\omega)\,d\omega. \tag{12}$$

The choices of force constants, which completes the dynamical problem, will be dictated by two criteria. The first is the mathematical requirement that $\omega_j^2>0$ for $j>0$ so that all frequencies are real. The second is the physical requirement that the interparticle forces decrease with distance between particle pairs

$$|g_m|>|g_{m+1}|;\ \lim_{L\to N\to\infty}|g_L|=0. \tag{13}$$

III. VELOCITY AUTOCORRELATION FUNCTION AND MEAN SQUARE DISPLACEMENT

Thus far, the description of the harmonic lattice is deterministic and completely within the framework of classical mechanics. Statistical mechanics enters the picture by postulating *a priori* that the initial conditions on the coordinates and velocities of the normal modes be fixed consistent with thermal equilibrium. This implies that at $t=0$ these modes, regarded as a collection of independent oscillators, have equilibrium dispersions derivable from the following density matrix W in the position representation[15,16]

$$W=\prod_j W_j\left[Q_j(0),\ Q_j'(0);\theta_j=\frac{\hbar\omega_j}{kT}\right] \tag{14}$$

$$W_j=\left[\frac{\omega_j}{\pi\hbar}\tanh\frac{\theta_j}{2}\right]^{1/2}\exp-\frac{\omega_j}{4\hbar}\left\{\tanh\frac{\theta_j}{2}[Q_j(0)+Q_j'(0)]^2\right.$$
$$\left.+\coth\frac{\theta_j}{2}[Q_j(0)-Q_j'(0)]^2\right\}.$$

The moments at equilibrium are found by integrating products of the operators

$$\overset{(\mathrm{op})}{Q_j}(0),\ \overset{(\mathrm{op})}{P_j}(0)=-i\hbar[d/dQ_j(0)]$$

over this distribution

$$\langle Q_j(0)\rangle=\langle P_j(0)\rangle=0,$$

$$\langle Q_j(0)\,Q_k(0)\rangle=\frac{\hbar}{2\omega_j}\coth\frac{\theta_j}{2}\,\delta_{jk}=\frac{\hbar}{\omega_j}\left[\frac{1}{\exp(\theta_j)-1}+\frac{1}{2}\right]\delta_{jk},$$

$$\langle P_j(0)\,P_k(0)\rangle=\frac{1}{2}\hbar\omega_j\coth\frac{\theta_j}{2}\,\delta_{jk}=\hbar\omega_j\left[\frac{1}{\exp(\theta_j)-1}+\frac{1}{2}\right]\delta_{jk}, \tag{15}$$

$$\langle Q_j(0)\,P_k(0)\rangle=-\langle P_k(0)\,Q_j(0)\rangle=(i\hbar/2)\,\delta_{jk},$$

where the angular brackets signify the dispersions are at thermal equilibrium.[17] The single particle velocity autocorrelation function $\langle v(t_1)\,v(t_2)\rangle_T$ for the nth particle is found by differentiating Eq. (7), forming the symmetrized product of $v_n(t_1)=\dot{x}_n(t_1)$ and $v_n(t_2)=\dot{x}_n(t_2)$ at two times t_1, t_2 and statistically averaging over the zero time quantities according to Eq. (15). The cross terms in the double sum vanish because of the Kronecker δ_{jk} factors in the dispersions, leaving a single sum.

$$\langle v(t_1)v(t_2)\rangle_T=\frac{1}{2}[\langle v_n(t_1)\,v_n(t_2)+v_n(t_2)\,v_n(t_1)\rangle]$$

$$= \frac{1}{M(2N+1)} \sum_{j=-N}^{+N} \frac{1}{2}\hbar\omega_j \coth\frac{\theta_j}{2} \cos\omega_j\tau \qquad (16)$$

$$[\tau = |t_1 - t_2|].$$

In the limit that $N\to\infty$, according to the prescription of Eq. (12), one obtains

$$\langle v(t_1)\, v(t_2)\rangle_T = \langle v(0)\, v(\tau)\rangle_T,$$
$$= \frac{\hbar}{2M}\int_0^{\omega_0} \omega \coth\left(\frac{\hbar\omega}{2kT}\right) \rho(\omega)\cos\omega(t_1-t_2)d\omega, \qquad (17)$$
$$= \frac{\hbar}{M}\int_0^{\omega_0} \frac{\omega\rho(\omega)}{\exp(\hbar\omega/kT)-1}\cos\omega\tau\, d\omega + \frac{\hbar}{2M}\int_0^{\omega_0} \times \omega\rho(\omega)\cos\omega\tau\, d\omega.$$

The velocity autocorrelation function has been defined using the symmetrized product because quantum operators at two different times do not commute.[15] It is independent of n as it must be due to dynamical equivalence of the particles. In general, a consequence of translational invariance is that any two-particle correlation function linking particles m and n must be a function of $|m-n|$; the present result is a special case of this fact.[2] The dependence of the velocity autocorrelation function on the time difference $\tau = |t_1 - t_2|$ reflects time invariance, so that any property of the system must be independent of time, the characteristic stationary property of thermal equilibrium. In the present case

$$\langle v^2(t)\rangle_T = \langle v^2\rangle_{\rm eq} = \frac{\hbar}{2M}\int_0^{\omega_0} \omega\coth\left(\frac{\hbar\omega}{2kT}\right)\rho(\omega)\, d\omega. \qquad (18)$$

This quantity is the equilibrium mean square velocity per particle.

Associated with stationary quantities are equilibrium fluctuations[5] in their values which are functions of time. The fluctuations in the position of any particle are measured by the change in its position in a time interval t, averaged over the equilibrium distribution. Thus,

$$\sigma(t) = X_n(t) - X_n(0) = x_n(t) - x_n(0) = \int_0^t v_n(t_1)dt_1 \qquad (19)$$
$$\sigma^2(t) = [x_n(t) - x_n(0)]^2 = \int_0^t\int_0^t v_n(t_1)v_n(t_2)\, dt_1\, dt_2.$$

Taking the thermal average $\langle\sigma|t\rangle_T = 0$ from the first line of Eq. (15), the mean square displacement $\langle\sigma^2(t)\rangle_T$ is given from Eq. (17) by

$$\langle\sigma^2(t)\rangle_T = \int_0^t\int_0^t \langle v(t_1)v(t_2)\rangle_T\, dt_1\, dt_2, \qquad (20)$$
$$= \frac{\hbar}{M}\int_0^{\omega_0}\left\{\left[\frac{1-\cos\omega t}{\omega}\right]\left[\coth\left(\frac{\hbar\omega}{2kT}\right)\rho(\omega)\right]\right\}d\omega,$$
$$= \frac{2\hbar}{M}\int_0^{\omega_0}\left\{\left[\frac{1-\cos\omega t}{\omega}\right]\left[\frac{\rho(\omega)}{\exp(\hbar\omega/kT)-1}\right]\right\}d\omega$$
$$+ \frac{\hbar}{M}\int_0^{\omega_0}\left[\frac{1-\cos\omega t}{\omega}\right]\rho(\omega)\, d\omega.$$

Comparison of Eq. (17) with Eq. (20) shows that the velocity autocorrelation function is related to the mean square displacement it produces by

$$\langle v(0)\, v(\tau)\rangle_T = \frac{1}{2}\left[\frac{d^2}{dt^2}\langle\sigma^2(t)\rangle_T\right]_{t=\tau}. \qquad (21)$$

One could follow a similar procedure to investigate velocity fluctuations in terms of a double time integral over the autocorrelation function for the force found from the second derivative of position coordinate. Velocity fluctuations tend to regress to zero with a time dependence dictated by the velocity autocorrelation function. The fluctuations in position will not usually in the present scheme go to zero, although they might. Whether they do or do not depends upon the functional properties of $\rho(\omega)$. If the position fluctuations go to zero in time, then each particle will have a position equilibrium mean square value $\langle x^2\rangle_{\rm eq}$ as does always the velocity [Eq. (18)]. If the fluctuations do not regress, their persistence results in a mean square displacement which increases in time, the familiar phenomenon of diffusion. These points will be discussed in more detail in Sec. VI.

It is to be observed that both the velocity autocorrelation function and the mean square displacement include terms independent of temperature. They appear because of the temperature independent $\frac{1}{2}$ factors in the equilibrium dispersion of Eq. (15). If one subtracts off the zero point contributions, the quantities

$$\langle v(0)\, v(\tau)\rangle_0 = \frac{\hbar}{2M}\int_0^{\omega_0}[\omega\rho(\omega)\cos\omega\tau]\, d\omega \qquad (22)$$

$$\langle\sigma^2(t)\rangle_0 = \frac{\hbar}{M}\int_0^{\omega_0}\left[\frac{1-\cos\omega t}{\omega}\right]\rho(\omega)\, d\omega \qquad (23)$$

would not appear. This was done by Ford *et al.*,[2] by defining the equilibrium dispersion in terms of the ordered product formalism[18] which removes the $\frac{1}{2}$ factors. The opinion here is that on physical grounds the zero point contributions should be retained. The reason is that at absolute zero the Uncertainty Principle demands motion of the particles due to zero point vibrations associated with the normal modes. Thus, any given particle suffers "collisions" the net effect being at $T=0$ a zero point drift, caused essentially by zero point velocity correlations. The dependence of the zero point drift upon time will be discussed below. Because of this contribution, it is necessary to make a distinction between the classical values of these quantities for which $\hbar = 0$ and the high temperature limit as $T\to\infty$. Denoting classical quantity by the subscript (c) then, when $\hbar=0$, from Eqs. (17) and (20)

$$\langle v(0)\, v(\tau)\rangle_c = \frac{kT}{M}\int_0^{\omega_0}[\rho(\omega)\cos\omega\tau]\, d\omega \qquad (24)$$

$$\langle\sigma^2(t)\rangle_c = \frac{2kT}{M}\int_0^{\omega_0}\left[\frac{1-\cos\omega t}{\omega^2}\right]\rho(\omega)\, d\omega. \qquad (25)$$

The differences between the high temperature limit and the classical expressions of Eqs. (24) and (25) are the corresponding zero temperature contributions of Eqs. (22) and (23).

The detailed analysis of the diffusive processes requires specification of $\rho(\omega)$. However, some general statements can be made as to the behavior of $\langle\sigma^2|t\rangle_T$ for times small such that $\omega_0 t \ll 1$, and times large such that $\omega_0 t \gg 1$. For short times the cosine term in Eq. (20) can be expanded, yielding a first nonvanishing contribution which goes as t^2. With the use of Eq. (18)

$$\langle\sigma^2(t)\rangle_T = \langle v^2\rangle_{eq}\, t^2,\quad \omega_0 t \ll 1. \tag{26}$$

For times short compared to the inverse of the cutoff frequency, the mean square displacement is like that of a free particle and characterized by the equilibrium mean square velocity. During this period the particle has yet to suffer a collision: ω_0^{-1} serves as a measure of the "microscopic interaction time" for times less than which, roughly, there is no interaction so that the particle moves undisturbed. This parallels the short time behavior of a particle executing Brownian motion and is a well known consequence of the solutions to the free particle Langevin equation.[12] The identification is not exact because the criterion for small times in the Langevin theory is $\lambda t \ll 1$, where λ is a characteristic macroscopic relaxation time parameter. The relationships between the microscopic interaction time and the macroscopic relaxation time will be of paramount importance in the classical discussion of Brownian motion in the next section. A second difference is that $\langle v^2\rangle_{eq}$ is here defined for arbitrary temperature. Specifically

$$\langle v^2\rangle_{eq} \begin{cases} \xrightarrow{\hbar=0} \dfrac{kT}{M} \equiv \langle v^2\rangle_c \\ \xrightarrow{T=0} \dfrac{\hbar\bar\omega}{2M},\quad \bar\omega = \displaystyle\int_0^{\omega_0} \omega\,\rho(\omega)\,d\omega. \end{cases} \tag{27}$$

The kinetic energy per particle is similar to that of a free particle in the classical limit, and at $T=0$ similar to that of a quantized oscillator of frequency $\bar\omega$ in the ground state, where $\bar\omega$ is the first moment of the frequency distribution. Thus

$$\langle v^2\rangle_{eq} \xrightarrow{T\to\infty} \frac{kT}{M} + \frac{\hbar\bar\omega}{2M}. \tag{28}$$

These results can be put into perspective by "turning off" the harmonic couplings between the particles. The free particle frequency distribution is given by $\rho(\omega) = \delta(\omega)$. As a consequence the velocity autocorrelation function is kT/M and $\langle\sigma^2|t)\rangle_T$ is $(kT/M)t^2$ for all times and independent of quantum effects ($\hbar$). Conversely, the very existence of particle interactions in the present scheme necessarily brings in quantum effects.

Some relationships between quantum effects and particle interactions can be seen in a more transparent light by consideration of the mean square displacement in the asymptotic limit that $\omega_0 t\to\infty$. If $\rho(\omega)$ is a slowly varying function, then from Eqs. (25) and (23) after a change of integration variable

$$\langle\sigma^2(t)\rangle_c = \frac{2kT}{M}\, t\int_0^{\omega_0 t}\left(\frac{1-\cos x}{x^2}\right)\rho\!\left(\frac{x}{t}\right)dx \to \frac{2kT}{M}\rho(0)t\left(\lim_{\omega_0 t\to\infty}\int_0^{\omega_0 t}\frac{1-\cos x}{x^2}\,dx\right), \tag{29}$$

$$\langle\sigma^2(t)\rangle_0 = \frac{\hbar}{M}\int_0^{\omega_0 t}\left(\frac{1-\cos x}{x}\right)\rho\!\left(\frac{x}{t}\right)dx \to \frac{\hbar}{M}\rho(0)\left(\lim_{\omega_0 t\to\infty}\int_0^{\omega_0 t}\frac{1-\cos x}{x}\,dx\right).$$

For $\omega_0 t$ large, the first integral is $(\pi/2)+O(\omega_0 t)^{-1}$, while the second integral is $\log\omega_0 t + \gamma - \mathrm{Ci}(\omega_0 t)$, where γ is Euler's constant and $\mathrm{Ci}(\omega_0 t)$ is the cosine integral which also decays as $(\omega_0 t)^{-1}$.[19] Thus, retaining only the terms which grow in time

$$\langle\sigma^2(t)\rangle_c \to 2Dt \tag{30}$$
$$\langle\sigma^2(t)\rangle_0 \to 2D\tau_q \log\omega_0 t,$$

where

$$D = (\pi kT/2M)\rho(0);\quad \tau_q = \hbar/\pi kT. \tag{31}$$

In the classical limit, then, if $\rho(\omega)$ is slowly varying, the mean square displacement is proportional to time in the asymptotic limit that $\omega_0 t\to\infty$, and characterized by a diffusion coefficient which is essentially given by the density of frequencies, evaluated at zero frequency. This of course is just the dependence predicted by the classical theory of diffusion. On the other hand, at absolute zero, $D=0$ but the mean square displacement increases as $\log(\omega_0 t)$ parameterized by a "diffusion coefficient" (of different dimensions) $D_q = D\tau_q = (\hbar/2M)\rho(0)$. In this approximation the zero point drift is parameterized simply by the product of the classical diffusion coefficient and a quantity τ_q which, following Ullersma, will be referred to as the quantum transient time.[3,20] This derivation as it stands, however, is apt to be misleading: it would seem to imply that for high temperatures and long times the mean square displacement evolves as the sum of the two terms of Eq. (30). But this is not true. What is true is that as long as $T\neq 0$ the mean square displacement eventually goes as $2Dt$ and the logarithmic term is cancelled out as $\omega_0 t\to\infty$. At $T=0$ this cancellation cannot occur and the logarithmic drift persists. To see this it is necessary to return to the complete expression of Eq. (20). Let $x=\omega\tau'$ in the first term, where τ' is some parameter. Then

$$\langle\sigma^2(t)\rangle_T = \frac{2\hbar}{M}\int_0^{\omega_0\tau'}\left\{\left[\frac{1-\cos(t/\tau')x}{\exp(\hbar x/kT\tau')-1}\right]\left[\frac{\rho(x/\tau')}{x}\right]\right\}dx + \langle\sigma^2(t)\rangle_0. \tag{32}$$

In the same approximation as the preceding, $\rho(\omega)$ is evaluated at $\omega=0$ and the integration limit is extended to infinity. The resultant integral yields[21]

$$\langle\sigma^2(t)\rangle_T \xrightarrow{\omega_0\tau'\to\infty} 2D\tau_q\log\left[\frac{\sinh(t/\tau_q)}{(t/\tau_q)}\right] + \langle\sigma^2(t)\rangle_0. \tag{33}$$

It is to be observed that this expression is independent of τ'. If $\tau'=\tau_q$ then its validity is restricted by the condition $\hbar\omega_0 \gg kT$; if $\tau'=t$ its validity is restricted by the condition $\omega_0 t\gg 1$. Thus, Eq. (33) holds for all times and asymptotically low temperatures or all temperatures and asymptotically long times. At $T=0$, of course, τ_q is infinite and the first term is zero for all times. This expression is not valid at all in the free particle case discussed above since this implies that $\omega_0=0$ [Eq. (10)] so that neither asymptotic condition can be satisfied. For $t\ll\tau_q$ the log term can be expanded for small values of t/τ_q. The first nonvanishing term goes as t^2. Combining this with Eq. (27) for the zero temperature contribution leads to the result

$$\langle\sigma^2(t)\rangle_T=\left[\frac{1}{6}\frac{(\pi kT)^2}{M\hbar}\rho(0)+\frac{\hbar\bar{\omega}}{2M}\right]t^2,\quad \omega_0 t\ll 1\ll\omega_0\tau_q. \tag{34}$$

For low temperatures the first temperature contribution to the equilibrium mean square velocity is proportional to T^2. In the opposite extreme that $t\gg\tau_q$, ignoring terms of the order of $\exp-(t/\tau_q)$,

$$\langle\sigma^2(t)\rangle_T=2Dt+2D\tau_q\left[\log\left(\frac{\omega_0\tau_q}{2}\right)-\log(\omega_0 t)\right] \tag{35}$$
$$+\left[\langle\sigma^2(t)\rangle_0-2D\tau_q\log(\omega_0 t)\right]$$
$$1\ll\omega_0\tau_q\ll\omega_0 t.$$

Neglecting the constant term, the zero temperature logarithmic time dependence vanishes in this approximation, leaving only the classical contribution for arbitrarily low but finite temperature. This assumes, however, that $t\gg\hbar/kT$ (independent of ω_0) and hence this asymptotic result takes longer to achieve as $T\to 0$; it is never reached at $T=0$.

The preceding can be summarized by saying that for any nonzero temperature the mean square displacement will after a sufficiently long time approach the classical linear dependence upon time. The rate of approach is dictated by the quantum transient time: the classical form is reached when $t\gg\tau_q$ for any finite value of $\omega_0\tau_q$ provided $\omega_0 t$ is sufficiently large, and hence takes longer to achieve at progressively lower temperatures. Finally, at absolute zero the linear time dependence gives way to a logarithmic behavior. It is to be reemphasized, however, that these results are valid only provided $\rho(\omega)$ is sufficiently slowly varying that it can be approximated by its value at zero frequency over the entire temperature range. Even if $\rho(0)$ is finite, it is possible for this approximation to be valid in the classical limit but not in the quantum limit. It is just this circumstance which will characterize the quantum modifications of the classical theory of Brownian motion of Sec. V within the framework of the model to be developed in the next section.

IV. GEOMETRIC PROGRESSION OF LONG RANGE INTERACTIONS AND THE CLASSICAL THEORY OF BROWNIAN MOTION

We will consider here the classical velocity autocorrelation function and mean square displacement for a force constant model in which the interaction constants are related to each other as terms in a geometric progression with alternating sign: $g_1=g$, $g_2=-zg$, $g_3=z^2g\cdots$, and in general

$$g_m=(-z)^{m-1}g;\ m=1,2\cdots\infty,\ |z|<1. \tag{36}$$

Substitution into Eq. (9) gives, after trigonometrical summation[21]

$$\omega_j^2=\frac{2g}{M}\left[\frac{1-z}{1+z}\right]\left[\frac{1-\cos[2\pi j/(2N+1)]}{1+2z\cos[2\pi j/(2N+1)]+z^2}\right] \tag{37}$$

which, by inversion, yields the following frequency distribution:

$$\rho(\omega)=\frac{2}{\pi}\frac{\omega_0^2(1-z^2)}{[\omega_0^2(1-z)^2+4z\omega^2](\omega_0^2-\omega^2)^{1/2}},$$
$$=\frac{(1+z)}{\pi(z)^{1/2}}\frac{\omega_0\lambda}{(\lambda^2+\omega^2)(\omega_0^2-\omega^2)^{1/2}}, \tag{38}$$
$$=\frac{2}{\pi}\left[\frac{1-z}{1+z}\right]\left[\frac{1}{(\omega_0^2-\omega^2)^{1/2}}+\frac{(\omega_0^2-\omega^2)^{1/2}}{\lambda^2+\omega^2}\right].$$

This frequency distribution can be thought of as characterized by two parameters: the cutoff frequency ω_0 given by Eq. (10) with $L=\infty$, and a second parameter λ related by ω_0 through z. They are given by

$$\omega_0=\left[\frac{4g}{M(1-z^2)}\right]^{1/2};\ \lambda=\frac{1-z}{2\sqrt{z}}\ \omega_0=\left[\frac{g}{Mz}\ \frac{1-z}{1+z}\right]^{1/2}. \tag{39}$$

These quantities act reciprocally as functions of z. At $z=0$, $\omega_0=(4g/M)^{1/2}$ and λ is infinite; as z approaches one, ω_0 approaches infinity while λ approaches zero. In parallel to the mathematical behavior, Eq. (36) indicates that the physical effect of progressively increasing z is to progressively increase the importance of long range interactions. By regarding z as the dependent variable, it follows from Eq. (39) that

$$z=\frac{(\omega_0^2+\lambda^2)^{1/2}-\lambda}{(\omega_0^2+\lambda^2)^{1/2}+\lambda}\to\begin{cases}\dfrac{\omega_0^2}{4\lambda^2}, & \omega_0\ll\lambda\\ \cdot 172, & \omega_0=\lambda\\ 1-\dfrac{2\lambda}{\omega_0}, & \omega_0\gg\lambda\end{cases}. \tag{40}$$

Introducing the following polar coordinates, $\omega_0=r\sin\phi$ and $\lambda=r\cos\phi$, then

$$z=\tan^2(\phi/2),\quad \tan\phi=\omega_0/\lambda. \tag{41}$$

Increasing the strength of the long range couplings from $z=0$ towards $z=1$ can be thought of as a rotation in an interaction space from $\phi=0$ towards $\phi=\pi/2$. z cannot equal one, since this would imply that all interactions are of the same magnitude in violation of the physical requirement of Eq. (13), as well as invalidating the summation leading to Eq. (37). Nevertheless, asymptotically

$$\rho(\omega)\xrightarrow{z=0}\frac{2}{\pi(\omega_0^2-\omega^2)^{1/2}} \tag{42}$$

$$\xrightarrow{z\to 1}\frac{2\lambda}{\pi(\lambda^2+\omega^2)},\quad\begin{pmatrix}\omega_0\to\infty\\ \lambda\to 0\end{pmatrix}.$$

The frequency distribution for $z=0$ corresponds to the usual case of nearest neighbor coupling only, while as z approaches unity $\rho(\omega)$ corresponds to the frequency distribution of Ford *et al.* which served as the starting point for their dynamical theory of Brownian motion. However, there is a difference in principle between their procedure and the one adopted here. They considered λ and ω_0 as independent quantities, and by choosing the latter to be infinite immediately obtained the classical velocity autocorrelation function as the complete Fourier transform of $\rho(\omega)$ to produce the time dependence $\exp(-\lambda\tau)$ characteristic of Brownian motion. In the present scheme λ and ω_0 are so related [Eq. (39)] that the former approaches zero at the same rate that the latter tends to infinity. Thus it is necessary to investigate the implications and range of validity of their approximation of infinite cutoff frequency.

As a first step in tracing the dynamical role of long range interactions embodied by the model of Eq. (36), we will first show that the classical velocity autocorrelation function can be expressed as an infinite series of Bessel functions of even order J_{2n}. From the first equality of Eq. (38), the frequency distribution can be expanded in Chebyshev polynomials T_{2n}. Setting $x=\omega/\omega_0$[22]

$$\rho(x)=\frac{1}{\omega_0}\frac{1}{(1-x^2)^{1/2}}\frac{(1-z^2)}{(1-z)^2+4zx^2}$$
$$=\frac{1}{\omega_0}\frac{1}{(1-x^2)^{1/2}}\left[1+2\sum_{n=1}^{\infty}(-z)^n T_{2n}(x)\right] \tag{43}$$

so that from Eq. (24)

$$\langle v(0)\,v(\tau)\rangle_c$$
$$=\frac{kT}{M}\left\{\frac{1}{\pi}\int_{-1}^{+1}\frac{\cos(x\omega_0\tau)}{(1-x^2)^{1/2}}\,dx+\frac{2}{\pi}\sum_{n=1}^{\infty}(-z)^n\int_{-1}^{+1}\frac{T_{2n}(x)\cos(x\omega_0\tau)}{(1-x^2)^{1/2}}\,dx\right\}$$
$$=\frac{kT}{M}\left[J_0(\omega_0\tau)+2\sum_{n=1}^{\infty}z^nJ_{2n}(\omega_0\tau)\right]$$
$$=\frac{kT}{M}\left[J_0(\omega_0\tau)+2\sum_{n=1}^{\infty}\tan^{2n}\left(\frac{\phi}{2}\right)J_{2n}(\omega_0\tau)\right], \tag{44}$$

where the second equality follows from the fact that the integrals over T_{2n} are $\pi(-1)^nJ_{2n}(\omega_0\tau)$.[21] For $\omega_0\tau\ll 1$, J_{2n} can be expanded in powers of $\omega_0\tau$ followed by summation over n. Through $(\omega_0\tau)^4$ only J_0, J_2, and J_4 contribute, giving the result

$$\langle v(0)\,v(\tau)\rangle_c$$
$$=\frac{kT}{M}\left\{1-\left[\frac{1-z}{4}\right](\omega_0\tau)^2+\left[\frac{(1-z)(3-z)}{192}\right](\omega_0\tau)^4-\cdots\right\}. \tag{45}$$

This illustrates that as $\omega_0\tau$ increases higher order Bessel functions contribute progressively greater contributions to the velocity autocorrelation function, implying the role of progressively higher powers of z: the longer range couplings become more important as $\omega_0\tau$ gets large. Oppositely, when $z=0$ and there are no long range couplings, Eq. (44) degenerates into the well known expression for nearest neighbor coupling as a zeroth order Bessel function.[3,4]

Of greater importance is the velocity autocorrelation function for values of τ large compared to the microscopic interaction time, ω_0^{-1}. This asymptotic behavior is obscured in Eq. (44) by the fact that as long as $\omega_0\tau$ is finite there is no single large argument approximation to J_{2n} which is valid for all n. It is necessary, therefore, to consider the third equality of Eq. (38) for $\rho(\omega)$. Insertion into Eq. (24) gives

$$\langle v(0)\,v(\tau)\rangle_c$$
$$=\frac{kT}{M}\left\{\left[\frac{1-z}{1+z}\right]J_0(\omega_0\tau)+\frac{2}{\pi}\left[\frac{1-z}{1+z}\right]\int_0^{\omega_0}\frac{(\omega_0^2-\omega^2)^{1/2}}{\omega^2+\lambda^2}\cos\omega\tau\,d\omega\right\}. \tag{46}$$

For $\omega_0\tau$ large the Bessel function assumes its asymptotic form, retaining only the first term which behaves as $(\omega_0\tau)^{-1/2}$. Since the integral will appear repeatedly in subsequent calculations it is considered separately in Appendix A, denoted there by $I(c)$. The asymptotic value for $I(c)$ is given by Eq. (A4). Setting $c=\lambda$, $t=\tau$ in that expression and ignoring the oscillatory term, then

$$\langle v(0)\,v(\tau)\rangle_c$$
$$=\frac{kT}{M}\left\{\exp(-\lambda\tau)+\left[\frac{1-z}{1+z}\right]\left(\frac{2}{\pi\omega_0\tau}\right)^{1/2}\cos\left(\omega_0\tau-\frac{\pi}{4}\right)\right\} \tag{47}$$

$$\left[\lambda=\frac{1-z}{2\sqrt{z}}\,\omega_0\right]$$

with the use of Eq. (39). The next term behaves as $(\omega_0\tau)^{-3/2}$ which, if included, would be the sum of the next correction to J_0 and the neglected oscillatory term of Eq. (A4). For $z=0$, λ is infinite and Eq. (47) reduces to the asymptotic result for the nearest neighbor coupling. In the present model the necessary and sufficient condition for the existence of the exponential term is the presence of long range interactions.

Brownian motion is characterized as a Gaussian Markoffian process, implying that for all τ the velocity autocorrelation function be of an exponential form. If one formally sets ω_0 infinite and keeps λ finite this is achieved in Eq. (47). This was the assumption of Ford *et al.*, although recognized by them as being an approximation. The second term of Eq. (47) represents the first correction to that approximation, that is, the first departure from Gaussian Markoffian behavior. The question is, how valid is the approximation in which, rather than setting ω_0 equal to infinity, one instead merely neglects the second term? The answer is determined by time scale. Consider first values of τ for which $0\leqslant\omega_0\tau\leqslant 10$. The velocity autocorrelation function must be determined from the exact expression of Eq. (44), and illustrated in Fig. 3 for representative values of z. Non-Markoffian oscillatory behavior is very manifest for small values of z while for values close to unity the curves are starting a predominently exponential decay, although the time scale does not yet betray their over-all behavior. It is of interest to compare these results with computer computations of velocity autocorrelation function for liquids.[10] As here the curves show a nonexponential dependence upon τ. There is typically an interval for which the velocity autocorrelation function is negative, and it is just this feature exhibited by the curves for small z of Fig. 3. For ω_0 of the order $10^{13}\,\text{sec}^{-1}$ the curves for $z=0.2$, 0.4 reproduce qualitatively the computer experiment curves. Berne has argued[10] that the negative regions indicate a particle's displacement towards its neighbors followed by a return back to its original position. For low values of z this occurs in the short time region but not yet for high values. Generally, the period for which the curve is negative is roughly a measure of the duration during which the particle retains memory of its interaction. In terms of the present scheme, the curves go negative

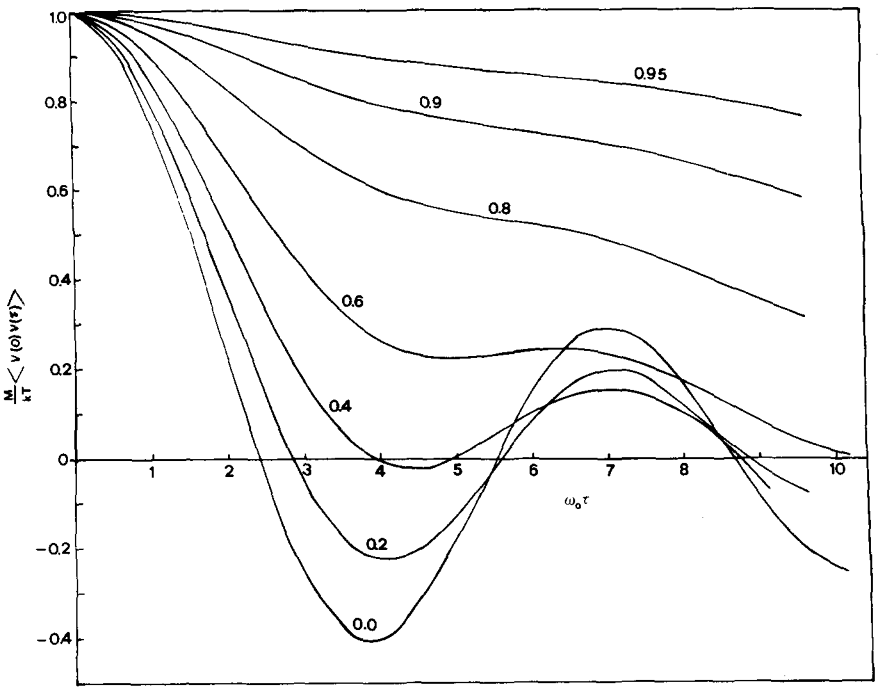

FIG. 3. Normalized classical velocity autocorrelation functions predicted by Eq. (44) for small times $\omega_0\tau \leq 10$. Curves are labelled by the corresponding value of z.

because of a finite microscopic interaction time ω_0^{-1}. It is to be emphasized that the present results show only a qualitative similarity with the much more refined computer experiments. In some latter cases the curves may go negative twice, but having gone negative never go positive. In other cases, the curves go negative once and thereafter remain positive. The present curves go both positive and negative more than once, but with diminishing amplitude as τ goes on. On the other hand, these results represent only a one parameter fit: ω_0 fixes the time scale with z (or λ) adjusted. This suggests a formal Bessel function expansion with more adjustable parameters may provide a useful device for computer simulation of liquid velocity autocorrelation functions. Continuing on, the next time scale for $10 \leq \omega_0\tau \leq 100$ is shown in Fig. 4. For z less than 0.8 the velocity autocorrelation functions have essentially decayed, while for $z \geq 0.8$ the oscillations tend to be cancelled by destructive interference of many appreciable Bessel function contributions. The lowest $z = 0.6$ "survivor" still shows many oscillations but with successively diminishing amplitude characteristic of Bessel function behavior. This is considerably less so for $z = 0.8$ and above, where the exponential behavior is very apparent. Figure 3 depicts behavior for microscopic τ of the order of ω_0^{-1}; Fig. 4 is an intermediate

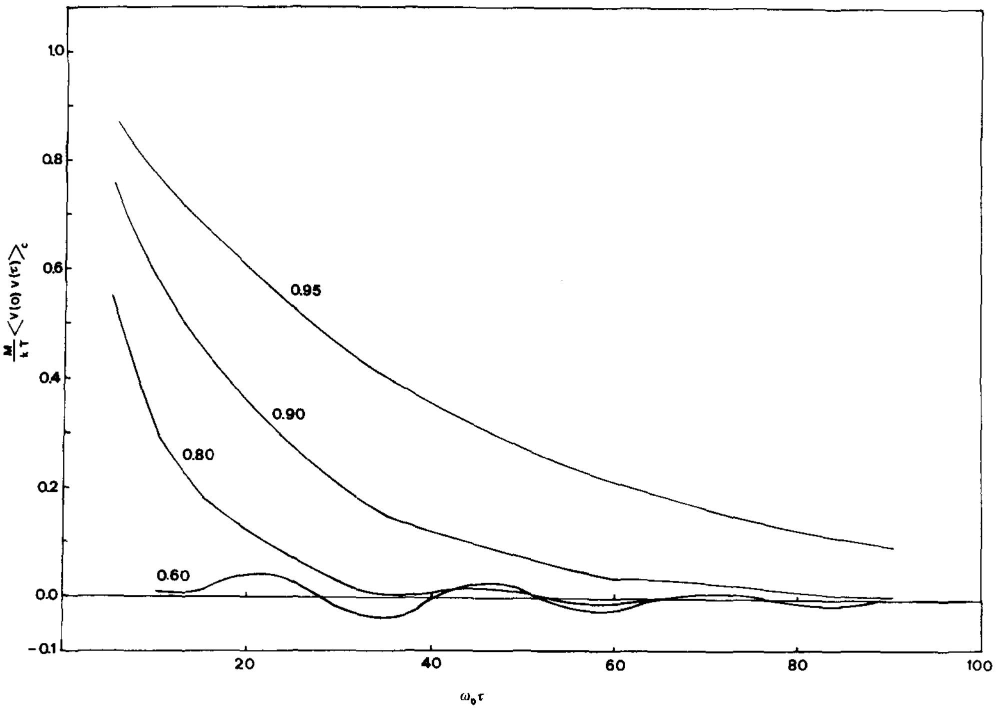

FIG. 4. Normalized classical velocity autocorrelation functions predicted by Eq. (47) for intermediate times $10 < \omega_0\tau < 100$ (see text).

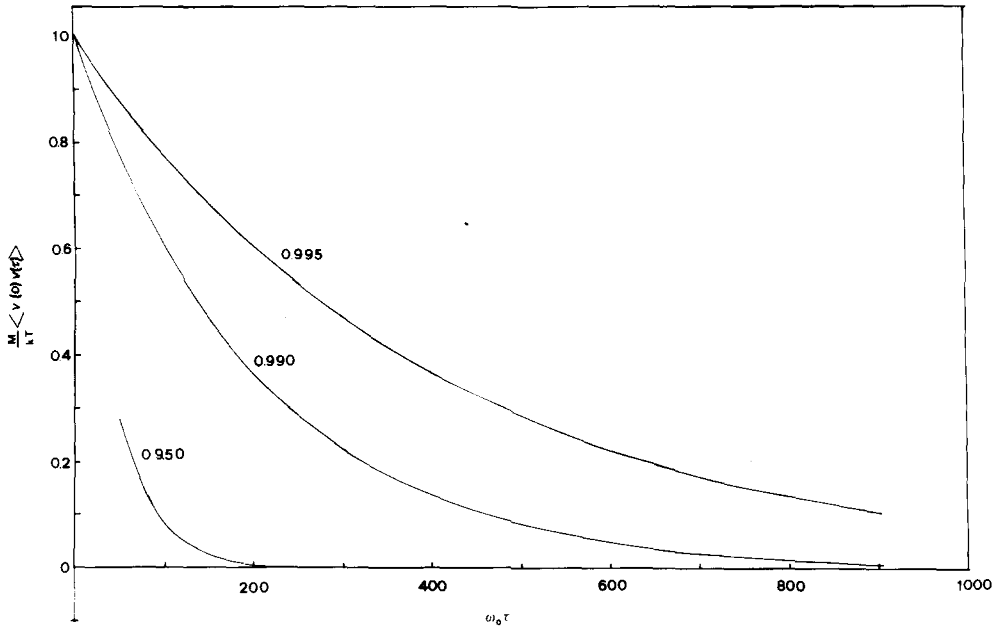

FIG. 5. Normalized classical velocity autocorrelation functions from Eq. (47) for long times $10^2 < \omega_0\tau < 10^3$.

region where $\omega_0\tau$ is large enough for the velocity autocorrelation function to exhibit Gaussian Markoffian behavior, but small enough to also exhibit residual oscillatory evolution, or memory. It should be added that for $\omega_0\tau = 10$ the exact expression of Eq. (44) with 10 Bessel function terms (through J_{20}) gave results essentially identical with the approximate asymptotic formula of Eq. (47). The latter, then, was used to compute the curves of Figs. 4–6. Figures 5 and 6 show results for the time scales $10^2 \leq \omega_0\tau \leq 10^3$, $10^3 \leq \omega_0\tau \leq 10^4$, respectively. We are clearly here in macroscopic time regions for which no oscillations are sensibly detected. The behavior appears Gaussian Markoffian, and only velocity correlations reflecting long range interactions are significant. In principle there are still oscillations, but on the macroscopic time scale they are too small to be "observed"; indeed their absence defines the time scale as macroscopic. In summary, the effect of long range interactions is to push the non-Markoffian oscillations to large time scales. Consequently, their contribution to the velocity autocorrelation function tends to diminish, since in any event they fall off roughly as $(\omega_0\tau)^{-1/2}$.

The preceding graphical discussion is supplemented to advantage by investigating the approximate duration during which the exponential term dominates the oscillatory contribution in the asymptotic formula Eq. (47). The condition for dominance is determined by the inequality

$$\exp(-\lambda\tau) > \left[\frac{1-z}{1+z}\right]\left(\frac{2}{\pi\omega_0\tau}\right)^{1/2} \tag{48}$$

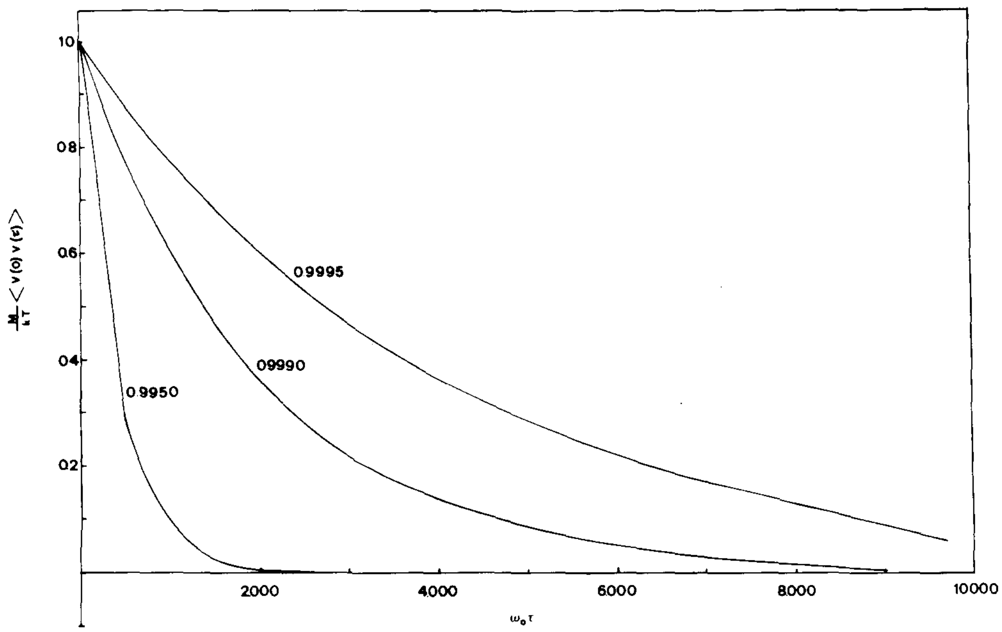

FIG. 6. Normalized classical velocity autocorrelation functions from Eq. (47) for macroscopic times $\omega_0\tau \gg 10^3$.

$$\frac{\omega_0\tau}{2\sqrt{z}}(1-z)-\frac{1}{2}\log\left[\frac{\omega_0\tau}{2\sqrt{z}}(1-z)\right]<\log\left[\left(\frac{1+z}{1-z}\right)\left(\frac{\pi\sqrt{z}}{1-z}\right)^{1/2}\right].$$

Since Eq. (47) holds only for $\omega_0\tau$ large ($\omega_0\tau>10$), the log term on the left can be neglected compared to the linear term, so that approximately

$$\omega_0\tau<\frac{2\sqrt{z}}{1-z}\log\left[\frac{\pi^{1/2}z^{1/4}(1+z)}{(1-z)^{3/2}}\right]. \tag{49}$$

For z small this condition cannot be fulfilled; the dominance of the decaying exponential term implies the necessity of appreciable long range interactions. Accordingly, let $z=1-2\mu$; then for μ small, from Eq. (39) and (40),

$$\omega_0=\left(\frac{g}{M\mu}\right)^{1/2},\quad \lambda=\left(\frac{g\mu}{M}\right)^{1/2},\quad \mu=\frac{\lambda}{\omega_0}\ll 1 \tag{50}$$

corresponding to a rotation in the interaction space of Eq. (41) by an angle $\phi=(\pi/2)-(\lambda/\omega_0)$. Equation (49) becomes

$$10<\omega_0\tau<\frac{3}{2\mu}\log\frac{1}{\mu};\quad 10\frac{\lambda}{\omega_0}<\lambda\tau<\frac{3}{2}\log\frac{\omega_0}{\lambda}, \tag{51}$$

where the lower limits are fixed by the limit of applicability of the asymptotic representation. Consider $\mu=0.1$ corresponding to $z=0.8$ or $\omega_0=10\lambda$. Then $10<\omega_0\tau<35$: for $\omega_0\tau>35$ the oscillatory behavior dominates. The curve for $z=0.8$ of Fig. 4 indeed exhibits oscillations for $\omega_0\tau>35$, but by the time they appear the velocity autocorrelation function is almost zero. The second form of the inequality of Eq. (51) sets the limits upon $\lambda\tau$ as $1<\lambda\tau<3.3$. Of course for $\omega_0\tau$ small enough the curves for $z<0.8$ will exhibit features of exponential decay, as can be seen from Fig. 3. However, they are in a time region where the asymptotic representation does not apply. Because of this, for example, the $z=0.6$ curve shows local exponential behavior to $\omega_0\tau\approx 4$, followed subsequently by large oscillatory character for $\omega_0\tau>4$. The latter is large because it starts while the velocity autocorrelation function is still large. The situation gets worse as z is smaller.

This analysis, then, leads to the following conclusion: the classical velocity autocorrelation function describes essentially exponential decay provided $z\geq 0.8$ and $\omega_0\tau\geq 10$. Under these conditions it is meaningful to regard λ as a macroscopic relaxation time parameter, which must be only one order of magnitude smaller than the microscopic cutoff frequency ω_0. The autocorrelation function, now given by

$$\langle v(0)\,v(\tau)\rangle_c\approx\frac{kT}{M}\left\{\exp(-\lambda\tau)+\mu\left(\frac{2}{\pi\omega_0\tau}\right)^{1/2}\cos\left(\omega_0\tau-\frac{\pi}{4}\right)\right\}$$

$$[\lambda=\mu\omega_0] \tag{52}$$

can be approximated solely by the exponential term with the second term neglected for $\mu\leq 0.2$ and $\omega_0\tau\geq 10$. As a consequence, if shorter times are neglected, the frequency distribution of Ford *et al.*, adequately predicts the process as Gaussian Markoffian for times $\tau\geq 10\omega_0^{-1}$. For short times, the process is non-Markoffian and only the exact frequency distribution of Eq. (38) and the consequent correlation function Eq. (44) hold. This argument is independent of the value of ω_0 which fixes the absolute time scale, since its numerical value is determined by g/M as well as μ. What is crucial is the ratio λ/ω_0. The asymptotic integrations involve setting $\omega_0\tau$ equal to infinity in the limit of integration, not ω_0, and for $\omega_0\tau\geq 10$ this is a reasonable approximation.

A further implication of the model is brought out upon comparison with that of Turner[22] and Ullersma,[3] who calculated the classical velocity autocorrelation function of a heavy particle of mass m_H substituted into a one-dimensional lattice of harmonically coupled light particles of mass m_L. Turner, using a procedure of Rubin,[23] showed that the velocity autocorrelation function for the heavy mass is given by

$$\langle v(0)\,v(\tau)\rangle_c^{(T-U)}=\frac{kT}{m_H}\frac{\mu}{\pi}\int_{-1}^{+1}\frac{(1-x^2)^{1/2}}{(1-2\mu)x^2+\mu^2}\cos(x\omega_c\tau)\,dx$$

$$\to\frac{kT}{m_H}\left\{\exp(-\lambda\tau)+\mu\left[\frac{2}{\pi(\omega_c\tau)^3}\right]^{1/2}\times\sin\left(\omega_c\tau-\frac{\pi}{4}\right)\right\} \tag{53}$$

$$\left[\mu=\frac{m_L}{m_H}\ll 1,\ \lambda=\mu\omega_c,\ \omega_c=\left(\frac{4g}{m_L}\right)^{1/2}\right],$$

where the asymptotic limit was found by Ullersma. Comparison of Eq. (46) with Eq. (53) shows, first, that the former for any τ has an added contribution $J_0(\omega_0\tau)$ which is absent in the latter. As a result both models predict asymptotically an exponential decay, but the oscillatory part behaves as $(\omega_0\tau)^{-1/2}$ in the present model while it behaves as $(\omega_0\tau)^{-3/2}$ in the Turner–Ullersma $(T-U)$ model. Secondly, the cutoff frequency parameterizing the $T-U$ model is just that for nearest neighbor interactions while in the present case is itself a function of the range of interaction. However, neglecting the asymptotic terms, the two models predict exponential decay asymptotically. If we define $m_H=M$, $m_L=M\mu/4$ the exponential factors of the two models, Eqs. (52) and (53) are formally identical. This suggests that while here all the particles have the same mass, the classical velocity autocorrelation function behaves as if a particle of mass M were embedded in a sea of light particles of effective mass $M\mu/4$. The parameter of smallness in the $T-U$ model is the ratio of light to heavy particles masses, which finds its parallel in the present scheme as essentially the ratio of a macroscopic to microscopic parameter, λ/ω_0.

The mean square displacement implied by Eq. (38) follows by substitution of the third equality into Eq. (25)

$$\langle\sigma^2(t)\rangle_c=\frac{2kT}{M}\left[\frac{1-z}{1+z}\right][I_1+I_2], \tag{54}$$

where

$$I_1=\frac{2}{\pi}\int_0^{\omega_0}\frac{1-\cos\omega t}{\omega^2(\omega_0^2-\omega^2)^{1/2}}\,d\omega=\frac{t}{\omega_0}\left[\int_0^{\omega_0 t}J_0(x)\,dx-J_1(\omega_0 t)\right],$$

$$=\frac{t}{\omega_0}\left\{[(\omega_0 t)J_0(\omega_0 t)-J_1(\omega_0 t)]+\frac{\pi}{2}(\omega_0 t)\,[J_1(\omega_0 t)\,\mathbf{H}_0(\omega_0 t)-J_0(\omega_0 t)\,\mathbf{H}_1(\omega_0 t)]\right\}, \tag{55}$$

$$I_2 = \frac{2}{\pi}\int_0^{\omega_0}\left[\frac{(\omega_0^2-\omega^2)^{1/2}}{\lambda^2+\omega^2}\right]\left[\frac{1-\cos\omega t}{\omega^2}\right]d\omega,$$

$$= \frac{1}{\lambda^2}\left\{\omega_0^2 I_1 + J_0(\omega_0 t) - \left[\frac{1+z}{1-z}\right]\left[1-\frac{2}{\pi}\left(\frac{1-z}{1+z}\right)I(\lambda)\right]\right\}.$$

The first integral is given in terms[21] of the Struve functions $\mathbf{H}_n$ and the last term of I_2 is the integral discussed in Appendix A. With the use of the asymptotic forms of the Bessel and Struve functions

$$I_1 \approx \frac{t}{\omega_0}\left\{1-\left[\frac{2}{\pi(\omega_0 t)^3}\right]^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right\} \tag{56}$$

so that through $(\omega_0 t)^{-1/2}$

$$\begin{aligned}\langle\sigma^2(t)\rangle_c &= \frac{2kT}{M}\left\{\left[\frac{1+z}{1-z}\right]\frac{t}{\omega_0} - \frac{1}{\lambda^2}[1-\exp(-\lambda t)]\right.\\ &\quad\left.-\frac{1}{\omega_0^2}\left[\frac{1-z}{1+z}\right]\left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right\}\\ &\to 2Dt,\end{aligned} \tag{57}$$

where, consistent with Eq. (31),

$$D = \frac{kT}{M\omega_0}\left[\frac{1+z}{1-z}\right] = \frac{kT}{M\lambda}\left[\frac{1+z}{2\sqrt{z}}\right] \begin{cases} \xrightarrow{z=0} \dfrac{kT}{2(Mg)^{1/2}} \\[2ex] \xrightarrow{z\to 1} \dfrac{kT}{M\lambda} = \dfrac{kT}{2[(Mg\mu/4)^{1/2}]} \end{cases} \tag{58}$$

As a consequence, for the nearest neighbor case[24]

$$\langle\sigma^2(t)\rangle_c = \frac{2kT}{M\omega_0^2}\left[\omega_0 t - \left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right] \tag{59}$$

$$[z=0,\ \omega_0 t \gg 1]$$

and for long range couplings, expressing ω_0 in terms of λ and z $(=1-2\mu)$, then for μ small

$$\begin{aligned}\langle\sigma^2(t)\rangle_c &= \frac{2kT}{M\lambda^2}[\lambda t-(1-\exp(-\lambda t))]\\ &\quad+\frac{kT}{M\lambda^2}\mu^2\left[\lambda t - 2\mu\left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right]\end{aligned} \tag{60}$$

$$[\lambda=\mu\omega_0,\ \mu\ll 1,\ \omega_0 t\gg 1].$$

The term multiplying μ^2 can be neglected compared to the first term since (1) the linear time factor adds only a second order correction to the diffusion constant, (2) the oscillatory function is dominated by the exponential for the same reasons and approximately under the same mathematical conditions as for the velocity autocorrelation function. The first term is the mean square displacement predicted by the Langevin theory of Brownian motion.[6,12]

We close this section with a speculation. In 1946 Kirkwood[25] derived a molecular theory of Brownian motion applicable to self-diffusion of an atom or molecule in a medium of similar molecules, such as liquid argon. Lebowitz and Rubin[26] demonstrated that the validity of Kirkwood's results was restricted by the condition $m_H \gg m_L$ consistent with the $T-U$ model. As pointed out by Zwanzig[27] this appears to preclude the very application which motivated Kirkwood's theory in the first place. It is suggested here that the model of Ford *et al.* implies in fact, that the Kirkwood theory holds for equal masses provided one admits the existence of effective light host particles, parameterized by the range of interaction between the particle of interest and the other particles (since all the particles are identical they serve as hosts for any given particle). This is not to say that a given atom feels direct coupling with an enormous number of host atoms at any instant. Rather, since Brownian motion involves many collisions over any macroscopic time interval—corresponding to $t\gg\omega_0^{-1}$—then in that time interval the collective effect of these collisions is as if the particle is experiencing long range interactions with particles of equal mass.

V. DYNAMICAL EFFECTS OF A FINITE QUANTUM TRANSIENT TIME

The model of the previous section ascribed all departure of the velocity autocorrelation function from its simple exponential form to the existence of a finite frequency distribution cutoff. With the inclusion of long range couplings in accordance with Eq. (36), the exponential form is essentially realized for times long compared to the inverse of this quantity, within the limits discussed graphically. Within the same limits the time course of diffusion evolves as the classical Langevin theory of Brownian motion. When one includes quantum effects embodied by a finite quantum transient time, the velocity autocorrelation function departs drastically from $\exp(-\lambda\tau)$ even within the framework of the frequency distribution of Ford *et al.* (which neglects the effects of a finite microscopic interaction time). As a consequence, in any event, all vestiges of Gaussian Markoffian behavior become lost as $\tau_q\to\infty$, with corresponding modifications of the time course of diffusion. These modifications form the subject of the present section.

In general the evaluation of the mean square displacement for any temperature is facilitated by expanding the hyperbolic cotangent of Eq. (20) in simple fractions

$$\coth x = \frac{1}{x} + 2x\sum_{n=1}^{\infty}\frac{1}{x^2+n^2\pi^2},$$

so that

$$\langle\sigma^2(t)\rangle_T = \frac{2kT}{M}\left\{\int_0^{\omega_0}\frac{1-\cos\omega t}{\omega^2}\rho(\omega)\,d\omega + 2\sum_{n=1}^{\infty}I_n\right\} \tag{61}$$

$$I_n \equiv \int_0^{\omega_0}\frac{1-\cos\omega t}{a_n^2+\omega^2}\rho(\omega)\,d\omega;\quad a_n = \frac{2n}{\tau_q} = \left(\frac{2\pi kT}{\hbar}\right)n.$$

The first term is $\langle\sigma^2(t)\rangle_c$ and the sum reflects the influence of a finite quantum transient time. In preparation for the more elaborate case associated with long range couplings, it is instructive to consider first the case of nearest neighbor coupling only. Insertion of Eq. (38) with $z=0$ into Eq. (61) gives for $\omega_0 t\gg 1$

$$\langle\sigma^2(t)\rangle_T = \frac{2kT}{M\omega_0}t + \frac{\hbar}{M\omega_0}\left\{\frac{2\nu}{\pi}\sum_{n=1}^{\infty}\frac{1-\exp-n(\omega_0 t/\nu)}{n(n^2+\nu^2)^{1/2}}\right. \tag{62}$$

$$-\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\Big\}$$

$$[\nu=\tfrac{1}{2}\omega_0\tau_q=\hbar\omega_0/2\pi kT],$$

with the use of Eq. (A5) of Appendix A, while the integrals independent of time are elementary. The sum cannot be expressed in closed form, but a simplification is achieved by observing that for values of n of the order of ν, the corresponding exponential terms in this sum are insignificant compared to the oscillatory factor which behaves as $(\omega_0 t)^{-1/2}$ for any temperature. The only terms which compete are those for which $n\ll\nu$. Accordingly the sum is split into two parts

$$S=\nu\sum_{n=1}^{\infty}\frac{1-\exp-n(\omega_0 t/\nu)}{n(n^2+\nu^2)^{1/2}}=\nu\sum_{n=1}^{N}\Big(\quad\Big)+\nu\sum_{n=N+1}^{\infty}\Big(\quad\Big),\qquad(63)$$

where N is some integer less than ν but sufficiently large that terms of the order of $\exp-N(\omega_0 t/\nu)$ are neglected. Then approximately

$$S=\sum_{n=1}^{N<\nu}\frac{1-\exp-n(\omega_0 t/\nu)}{n}\left[1-\frac{n^2}{2\nu^2}\right]+\left[\nu\int_N^\infty\frac{dn}{n(n^2+\nu^2)^{1/2}}-\frac{\nu}{2N(N^2+\nu^2)^{1/2}}\right]$$

$$\approx\Big\{\log\nu\,[1-\exp(-\omega_0 t/\nu)]+(\log 2+\gamma)+\frac{1}{2\nu^2}\,\frac{\exp-(\omega_0 t/\nu)}{[1-\exp-(\omega_0 t/\nu)]^2}\Big\}.\qquad(64)$$

The second expression follows from the first in Eq. (64) by approximating the first sum by $\log N+\gamma+(1/2N)$, extending the sums over the exponential terms to infinity, and approximating the second sum over the constant terms of Eq. (63) by the Euler–Maclaurin summation formula.[28] The last factor in the second expression can now be neglected; even at $T=0$, where it is largest, it behaves as $(\omega_0 t)^{-2}$. Then substitution of Eq. (64) into Eq. (62) gives

$$\langle\sigma^2(t)\rangle_T=2D\tau_q\left\{\log\left[\omega_0 t\left(\frac{\sinh(t/\tau_q)}{t/\tau_q}\right)\right]-\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{\pi}{2\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)+\text{const.}\right\}\qquad(65)$$

$$[D=kT/M\omega_0,\ \text{const}=\log 2+\gamma],$$

for the mean square displacement through $(\omega_0 t)^{-1/2}$. The first term is just another way of expressing Eq. (35), so that the error in approximating the frequency distribution for nearest neighbor coupling as constant over the whole temperature range is the oscillatory term diminishing as $(\omega_0 t)^{-1/2}$ also over the whole temperature range. It follows from Eq. (21) that the asymptotic functional form of the velocity autocorrelation function is insensitive to the temperature

$$\langle v(0)\,v(\tau)\rangle_T=\frac{\hbar\omega_0}{2M}\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{2}{\pi\omega_0\tau}\right)^{1/2}\cos\left(\omega_0\tau-\frac{\pi}{4}\right).\qquad(66)$$

With this caveat we are now in a position to consider the inclusion of long range couplings embodied by the model of Eq. (38). Its substitution into Eq. (61) results in

$$\langle\sigma^2(t)\rangle_T=\frac{2kT}{M}\left\{\left[\frac{1-\mu}{(1-2\mu)^{1/2}}\right]\frac{t}{\lambda}-\frac{1}{\lambda^2}\,(1-\exp(-\lambda t))\right\}+\left[\frac{\hbar^2}{M\pi^2 kT}\right]\left\{\sum_{n=1}^{\infty}\left[\frac{1}{p^2-n^2}\right]\times\left[\frac{p[1-\exp-n(\omega_0 t/\nu)]}{n\{1-[\mu/p(1-\mu)]^2(p^2-n^2)\}^{1/2}}-(1-\exp(-\lambda t))\right]\right\}$$

$$-\left[\frac{\hbar\mu^2}{M\lambda(1-\mu)(1-2\mu)^{1/2}}\right]\coth\left(\frac{\hbar\omega_0}{2kT}\right)\times\left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)$$

$$[z=1-2\mu;\ p=\hbar\lambda/2\pi kT=\lambda\tau_q/2],\qquad(67)$$

with the use of Eqs. (39), (57), (A5), and (A6). Correspondence to the classical Langevin theory is achieved by the approximation of setting $\mu=0$, in which case, with $kT=\hbar\lambda/2\pi p$

$$\langle\sigma^2(t)\rangle_T^{(L)}=\frac{\hbar}{M\pi\lambda}\left\{\frac{1}{p}\,[\lambda t+\exp(-\lambda t)+1]+2[\psi(p)+\gamma]+2p\sum_{n=1}^{\infty}\frac{n\exp(-\lambda t)-p\exp-n(\lambda t/p)}{n[p^2-n^2]}\right\}\qquad(68)$$

$$\left[\psi(p)+\gamma=-\frac{1}{p}+p\sum_{n=1}^{\infty}\frac{1}{n(n+p)}\right],$$

where constant factors have been expressed in terms[29] of the Euler psi function $\psi(p)$. The superscript (L) denotes the present $\mu=0$ approximation corresponds to the classical (Langevin) result for Brownian motion when $\hbar=0$ [Eq. (60)]. It implies that for all n, the square root factor in Eq. (67) is set equal to unity, which means $\mu^2/p^2=0$, or $\lambda=\mu\omega_0$ and $\hbar\omega_0/2\pi kT=\infty$. This is strictly true for all temperatures if ω_0 is considered infinite: the approximation of the Ford model [Eq. (68)] is equally obtained by direct use of $\rho(\omega)$ for $z\to 1$ of Eq. (42). The next approximation is to expand the square root for μ^2 small, a procedure valid for all n such that $(\mu n/p)^2\ll 1$, or $n\ll\nu$. This is the same condition which motivated the approximate evaluation of the sum appearing in the nearest neighbor coupling model. The first correction to $\langle\sigma^2(t)\rangle_T^{(L)}$ (proportional to μ^2) is calculated in Appendix B using a similar procedure. To facilitate computation of $\langle\sigma^2(t)\rangle_T^{(L)}$ itself there is little loss in generality by first restricting p to positive integers $1, 2\cdots$. This implies investigation of cases for temperatures sufficiently low that the quantum transient time is greater than the classical relaxation time. The zero temperature limit for fixed λ corresponds formally to $p=\infty$. The $n=p$ term in the sum of Eq. (68) is $-(2p^2)^{-1}(1+\lambda t)\exp(-\lambda t)$ so that after some manipulations

$$\langle\sigma^2(t)\rangle_T^{(L)}=\frac{2D_0}{\lambda}\left\{\frac{\lambda t}{p}\,(1-\exp(-\lambda t))+2[\psi(p)+\gamma]\right.\qquad(69)$$

$$+(2-\exp(-\lambda t)-\exp(\lambda t))\log[1-\exp(-\lambda t/p)]$$
$$-\left[\exp(-\lambda t)\sum_{n=1}^{p-1}\frac{\exp+n(\lambda t/p)}{n}+\exp(\lambda t)\sum_{n=1}^{p-1}\frac{\exp-n(\lambda t/p)}{n}\right]\Bigg\}$$
$$\left[\psi(p)+\gamma=\sum_{n=1}^{p-1}\frac{1}{n},\quad p=1,2\cdots;\ D_0=\frac{\hbar}{2\pi M}\right].$$

Thus, for example,

$$\langle\sigma^2(t)\rangle_T^{(L)}=\frac{2D_0}{\lambda}\Big\{\lambda t(1-\exp(-\lambda t))+(2-\exp(-\lambda t)-\exp(\lambda t))\log(1-\exp(-\lambda t))\Big\}\left[\begin{matrix}p=1\\ \tau_q=2\lambda^{-1}\end{matrix}\right]$$
$$=\frac{2D_0}{\lambda}\Big\{\frac{\lambda t}{2}(1-\exp(-\lambda t))+2\Big(1-\cosh\frac{\lambda t}{2}\Big)+(2-\exp(-\lambda t)-\exp(\lambda t))\log(1-\exp(-\lambda t/2))\Big\}\left[\begin{matrix}p=2\\ \tau_q=4\lambda^{-1}\end{matrix}\right]$$
$$=\frac{2D_0}{\lambda}\Big\{\frac{\lambda t}{3}(1-\exp(-\lambda t))+\Big(3-2\cosh\frac{\lambda t}{3}-\cosh\frac{2\lambda t}{3}\Big)+(2-\exp(-\lambda t)-\exp(\lambda t))\log(1-\exp(-\lambda t/3))\Big\}\left[\begin{matrix}p=3\\ \tau_q=6\lambda^{-1}\end{matrix}\right]\tag{70}$$

and so on. In the general case, for long times $\lambda t\gg 1$, expanding the logarithm in Eq. (69) and collecting terms

$$\langle\sigma^2(t)\rangle_T^{(L)}=\frac{2D_0}{\lambda}\Big\{\frac{\lambda t}{p}(1-\exp(-\lambda t))+2\Big[\psi(p)+\gamma+\frac{1}{2p}\Big]-\Big[2p^2\sum_{n=1}^{p-1}\frac{\exp-n(\lambda t/p)}{n(p^2-n^2)}+\frac{3}{2p}\exp(-\lambda t)\Big]+O[\exp-(1+p^{-1})\lambda t]\Big\}\tag{71}$$

while for small times $\lambda t\ll 1$, expanding the exponentials through t^2 and performing the sums

$$\langle\sigma^2(t)\rangle_T^{(L)}=(\hbar\lambda/M\pi)\,t^2(A_p-\log\lambda t)$$
$$A_p=\{(\tfrac{3}{2}+\log p)-[\psi(p)+\gamma+(2p)^{-1}]\}.\tag{72}$$

These results differ from the classical expression of Eq. (60) (with the second term neglected)

$$\langle\sigma^2(t)\rangle_c^{(L)}=\frac{2D}{\lambda}[\lambda t-(1-\exp(-\lambda t))]=\left\{\begin{matrix}2Dt, & \lambda t\gg 1\\ (kT/M)t^2, & \lambda t\ll 1\end{matrix}\right\}$$
$$[D=kT/M\lambda]\tag{73}$$

primarily in that there are $p-1$ exponential contributions which decay slower than the single exponential factor of the Langevin theory, accompanied by a corresponding reduction of the linear time term by a factor of $1/p$. For small times the mean square displacement is similarly proportional to t^2, but modulated by a logarithmic factor. It is also to be observed that the exponential terms enter with opposite sign than does the single one of Eq. (73). As a consequence, the corresponding velocity autocorrelation functions are *negative*. From Eqs. (21) and (69)

$$\langle v(0)v(\tau)\rangle_T^{(L)}=-\frac{\hbar\lambda}{2\pi M}\Big\{\frac{1}{p}(1+\lambda\tau\exp(-\lambda\tau))+(\exp(-\lambda\tau)+\exp(\lambda\tau))\times\log(1-\exp(-\lambda\tau/p))+\Big[\exp(-\lambda\tau)\sum_{n=1}^{p-1}\frac{\exp+n(\lambda\tau/p)}{n}+\exp(\lambda\tau)\sum_{n=1}^{p-1}\frac{\exp-n(\lambda\tau/p)}{n}\Big]\Big\},$$
$$=-\frac{\hbar\lambda}{2\pi M}\Big\{2\sum_{n=1}^{p-1}\frac{n\exp-n(\lambda\tau/p)}{p^2-n}+\frac{1}{p}\Big(\lambda\tau-\frac{1}{2}\Big)\times\exp(-\lambda\tau)\Big\},\quad \lambda\tau\gg 1,\tag{74}$$
$$=-\frac{\hbar\lambda}{\pi M}[B_p+\log\lambda\tau],\quad B_p=\Big[\psi(p)+\gamma+\frac{1}{2p}-\log p\Big],\quad \lambda\tau\ll 1,$$

so that for example

$$\langle v(0)v(\tau)\rangle_T^{(L)}=-\frac{\hbar\lambda}{2\pi M}\Big\{(1+\lambda\tau\exp(-\lambda\tau))+2\cosh\lambda\tau\log(1-\exp(-\lambda\tau))\Big\}\ [p=1]$$
$$=-\frac{\hbar\lambda}{2\pi M}\Big\{\frac{1}{2}(1+\lambda\tau\exp(-\lambda\tau))+2\cosh\frac{\lambda\tau}{2}+2\cosh\lambda\tau\log(1-\exp(-\lambda\tau/2))\Big\}\ [p=2]\tag{75}$$
$$=-\frac{\hbar\lambda}{2\pi M}\Big\{\frac{1}{3}(1+\lambda\tau\exp(-\lambda\tau))+\Big[2\cosh\frac{\lambda\tau}{3}+\cosh\frac{2\lambda\tau}{3}\Big]+2\cosh\lambda\tau\log(1-\exp(-\lambda\tau/3))\Big\}.\ [p=3]$$

Within the framework of the model, departure of the velocity autocorrelation function from its simple exponential (Gaussian Markoffian) behavior at any finite temperature exhibits the added feature of "persistent memory" for temperatures sufficiently low that $\lambda\tau_q>1$.

The mean square displacement and velocity autocorrelation function at absolute zero are characterized by complete disappearance of an asymptotic linear dependence upon time of the former quantity and a parallel disappearance of an exponential dependence of the latter. The results are most conveniently expressed in terms of exponential integral functions defined by[19]

$$\mathrm{Ei}(-x)=-\int_x^\infty\frac{\exp(-t)}{t}\,dt=\gamma+\log x+\sum_{n=1}^{\infty}\frac{(-x)^n}{n!n},$$
$$E^*(x)=-⨍_{-x}^{\infty}\frac{\exp(-t)}{t}\,dt=\gamma+\log x+\sum_{n=1}^{\infty}\frac{x^n}{n!n}.\tag{76}$$

Then

$$\lim_{p\to\infty}\sum_{n=1}^{p-1}\frac{\exp\pm n(\lambda t/p)}{n}=\lim_{p\to\infty}\Big\{\sum_{n=1}^{p-1}\frac{1}{n}+\sum_{n=1}^{\infty}\frac{1}{n!}\Big(\frac{\pm\lambda t}{p}\Big)^n$$

$$\times\left[\frac{(p-1)^n}{n}+\frac{(p-1)^{n-1}}{2}\cdots\right]\Bigg\}$$

$$=\log p+\gamma+\sum_{n=1}^{\infty}\frac{(\pm\lambda t)^n}{n!\,n}$$

$$=\begin{bmatrix}E^*(\lambda t)\\ \mathrm{Ei}(-\lambda t)\end{bmatrix}+\log p-\log\lambda t \qquad (77)$$

$$\lim_{p\to\infty}\begin{bmatrix}\log(1-\exp(-\lambda t/p))\\ \psi(p)\end{bmatrix}=\begin{bmatrix}\log\lambda t-\log p\\ \log p\end{bmatrix}.$$

Consequently, Eqs. (69) and (74) reduce to

$$\langle v(0)\,v(\tau)\rangle_0^{(L)}=-(\hbar\lambda/2\pi M)\left[\exp(-\lambda\tau)E^*(\lambda\tau)+\exp(\lambda\tau)\mathrm{Ei}(-\lambda\tau)\right], \qquad (78)$$

$$\langle\sigma^2(t)\rangle_0^{(L)}=2D\tau_q\{(\gamma+\log\lambda t)-\tfrac{1}{2}\left[\exp(-\lambda t)E^*(\lambda t)+\exp(\lambda t)\mathrm{Ei}(-\lambda t)\right]\}\ \left[D\tau_q=\hbar/\pi M\lambda\right],$$

the $\log p$ factors cancelling. These quantities represent the effects of quantum fluctuations due to zero temperature vibrational energy of the normal modes. From the asymptotic properties of the exponential integral functions

$$\langle v(0)\,v(\tau)\rangle_0^{(L)}=\begin{bmatrix}-\hbar/\pi M\lambda\tau^2, & \lambda\tau\gg 1\\ -(\hbar\lambda/\pi M)\left[\gamma+\log\lambda\tau\right], & \lambda\tau\ll 1\end{bmatrix}, \qquad (79)$$

$$\langle\sigma^2(t)\rangle_0^{(L)}=\begin{bmatrix}2D\tau_q\log\lambda t, & \lambda t\gg 1\\ (\hbar\lambda/\pi M)\,t^2\left[(\tfrac{3}{2}-\gamma)-\log\lambda t\right], & \lambda t\ll 1\end{bmatrix}.$$

Replacing the classical single exponential dependence is an asymptotic τ^{-2} dependence at zero temperature for the velocity autocorrelation function; the linear time dependence is correspondingly replaced by a logarithmic dependence for the mean square displacement. In the latter case, the asymptotic time evolution is just as for slowly varying frequency distributions with ω_0 formally replaced by λ. However, ω_0 is essentially infinite in the present approximation, so that the approximate method of Sec. III would lead to divergent results. The absence of the divergence is because that method is inapplicable here, since the frequency distribution behaves as ω^{-2}, which is fast varying compared to the linear denominator of Eq. (23). It is to be noted, however, that the logarithmic drift is via a rather torturous route through exponential integrals.

Approach to the high temperature classical limit can be studied by substituting $p=1/q$ in Eq. (68). The sum over constant terms can be expressed as a cotangent and $\hbar$ eliminated through q, so that

$$\langle\sigma^2(t)\rangle_T^{(L)}=\frac{2D}{\lambda}\Bigg\{\lambda t-\left[1-\left(\frac{\pi}{q}\cot\frac{\pi}{q}\right)\exp(-\lambda t)\right]+\frac{2}{q}\left[\psi\left(\frac{1}{q}\right)+\gamma+q\right]+\frac{2}{q}\sum_{n=1}^{\infty}\frac{\exp(-nq\lambda t)}{n(n^2q^2-1)}\Bigg\} \qquad (80)$$

$$=\frac{2D}{\lambda}\Bigg\{\lambda t-\left[1-\left(\frac{\pi}{q}\cot\frac{\pi}{q}\right)\exp(-\lambda t)\right]+\frac{2}{q}\left[\psi(1/q)+\gamma+q\right]+\frac{2}{q}\frac{\exp(-q\lambda t)}{q^2-1}\Bigg\},\quad \lambda t\gg 1,$$

$$=\frac{kT}{M}t^2\left[A_q-\frac{2}{q}\log\lambda t\right],\qquad \lambda t\ll 1$$

$$\left[A_q=\frac{1}{q}\left[3-q-2\log q-2\gamma-2\psi\left(\frac{1}{q}\right)\right]\right]$$

$$[q=2/\lambda\tau_q>1].$$

Similarly,

$$\langle v(0)\,v(\tau)\rangle_T^{(L)}=\frac{kT}{M}\left[\left(\frac{\pi}{q}\cot\frac{\pi}{q}\right)\exp(-\lambda\tau)+2q\sum_{n=1}^{\infty}\frac{n\exp(-nq\lambda\tau)}{n^2q^2-1}\right], \qquad (81)$$

$$=\frac{kT}{M}\left[\left(\frac{\pi}{q}\cot\frac{\pi}{q}\right)\exp(-\lambda\tau)+\frac{2q}{q^2-1}\exp(-q\lambda\tau)\right],\quad \lambda\tau\gg 1,$$

$$=\frac{kT}{M}\left[B_q-\frac{2}{q}\log\lambda\tau\right],\quad \lambda\tau\ll 1$$

$$\left[B_q=-\frac{1}{q}\left[q+2\log q+2\gamma+2\psi(1/q)\right]\right].$$

The progression of q through increasing values corresponds for fixed λ to increasing the temperature such that the quantum transient time becomes progressively smaller than the classical relaxation time. As $q\to\infty$, the sum terms vanish exponentially and $\gamma+\psi(1/q)$ goes to $-q$ so that Eq. (80) reduces to the classical result of Eq. (73), and correspondingly the velocity autocorrelation function assumes its single exponential form. The velocity autocorrelation functions are now positive, and correspondingly the classical exponential factor always dominates the mean square displacement for long times, the only vestige of quantum effects being the cotangent factor. On the other hand, for short times, both quantities are characterized by a logarithm as in the low temperature case (except when q is truly infinite). The situation is summarized by saying that for low temperatures such that the quantum transient time is greater than the relaxation time, the velocity autocorrelation function is negative; it and the mean square displacement are dominated by many significant exponential terms whose numbers increase as the temperature gets lower. For temperatures such that the quantum transient time is less than the relaxation time, the quantum features are preserved for short times but are lost for long times. In this case diffusion evolves asymptotically to the classical form, differing only in the temperature-dependent cotangent coefficient. Correspondingly, the always positive velocity autocorrelation function unfolds from its low-temperature-like form to its classical Gaussian Markoffian form, multiplied by the same coefficient.

These considerations define a critical temperature T_c such that at T_c, $\tau_q=\lambda^{-1}$. What is special about T_c is that when the quantum transient time is equal to the relaxation time ($q=2$) the cotangent coefficient vanishes. As a result, the velocity autocorrelation function in the process of "turning around" from negative to positive as one proceeds from low to high temperatures decays faster asymptotically (as $\exp(-2\lambda\tau)$) than at any other temperature. The sums in Eqs. (80) and (81) can be evaluated by expressing them as integrals. Specifically

$$\frac{2}{q}\sum_{n=1}^{\infty}\frac{\exp(-nq\lambda t)}{n(n^2q^2-1)}=S_1+S_2+\frac{2}{q}\log(1-\exp(-q\lambda t)),$$

$$S_1=\sum_{n=1}^{\infty}\frac{\exp(-nq\lambda t)}{nq-1}=\exp(-\lambda t)\int_0^{\exp(-\lambda t)}\frac{x^{q-2}}{1-x^q}\,dx, \tag{82}$$

$$S_2=\sum_{n=1}^{\infty}\frac{\exp(-nq\lambda t)}{nq+1}=\exp(\lambda t)\int_0^{\exp(-\lambda t)}\frac{dx}{1-x^q}-1,$$

so that for $q=2$

$$\begin{aligned}\langle v(0)v(\tau)\rangle_T^{(L)}&=\frac{kT}{M}\left[\cosh\lambda\tau\log\left(\frac{1+\exp(-\lambda\tau)}{1-\exp(-\lambda\tau)}\right)-1\right],\\&=\tfrac{4}{3}(kT/M)\exp(-2\lambda\tau),\quad \lambda\tau\gg1,\\&=(kT/M)\left[(\log 2-1)-\log\lambda\tau\right],\quad \lambda\tau\ll1,\end{aligned} \tag{83}$$

$$\begin{aligned}\langle\sigma^2(t)\rangle_T^{(L)}&=\frac{2D}{\lambda}\left[\lambda t+\log(1-\exp(-2\lambda t))\right.\\&\quad\left.+\cosh\lambda t\log\left(\frac{1+e^{-\lambda t}}{1-e^{-\lambda t}}\right)-2\log 2\right],\\&=(2D/\lambda)\left[\lambda t+(1-2\log 2)+\tfrac{1}{3}\exp(-2\lambda t)\right],\quad \lambda t\gg1,\\&=(kT/M)t^2\left[(\log 2+\tfrac{1}{2})-\log\lambda t\right],\quad \lambda t\ll1.\end{aligned}$$

Higher finite temperatures corresponding to higher values of q can be investigated according to the prescription of Eq. (82) for the sums, although the expressions rapidly become complicated. A more complete analytical and graphical treatment of the contents of this section will be presented elsewhere.[30]

As a final note, it should be added that due to the integrals approximation that $\omega_0\tau\gg1$, the velocity autocorrelation functions computed here, which include logarithmic behavior, are not defined at $\tau=0$. The value must be given by Eq. (18) which, at zero temperature, predicts the mean square velocity for the geometric progression model to be

$$\langle v^2\rangle_{\text{eq}}=\frac{\hbar\omega_0}{2\pi M}\left[\frac{1-z}{\sqrt{z}}\right]\log\left[\frac{1+\sqrt{z}}{1-\sqrt{z}}\right]\approx\frac{\hbar\lambda}{\pi M}\log\frac{2\omega_0}{\lambda}. \tag{84}$$

In this connection, the Langevin theory, by construction, reduces the mean square displacement to that of a free particle for λt small. The $\log\lambda t$ term which modulates the t^2 behavior in the present scheme (persisting at all finite temperatures and vanishing only in the classical limit) reflects the fact that while t is much less than λ^{-1} it is still much greater than ω_0^{-1}. Thus the particle has suffered many collisions and is not free. The persistence of the logarithm indicates collisional effects due to nonthermal quantum fluctuations. At absolute zero $\log\lambda t$ is the only thing that essentially survives [note its presence for both λt large and small in Eq. (79)]. As the temperature goes up, competitive thermal fluctuations progressively reduce its importance until the quantum contribution to the diffusive process is completely swamped out at infinite temperature. This effect could never appear, of course, in any classical theory. An assessment of the present quantum modifications of classical Brownian motion will be made in Sec. VII.

VI. OTHER TYPES OF DIFFUSIVE MOTION

The geometric progression model yielded mean square displacements for large times proportional to t in the classical limit and proportional to $\log t$ in the limit of zero temperature. This was a consequence of the fact that the frequency distribution at zero frequency for this model is finite and nonzero. The aim of the present section is to investigate the consequences of examples for which, on the contrary, $\rho(0)$ is infinite or zero.

The first case is afforded by the following force constant distribution

$$g_m=(-1)^{m+1}g\left[\frac{(L-1)!(L+1)!}{(L-m)!(L+m)!}\right],\quad m=1,2\cdots L. \tag{85}$$

Insertion into Eq. (9) yields

$$\omega_j^2=\omega_0^2\sin^{2L}\left(\frac{j\pi}{2N+1}\right) \tag{86}$$

$$\omega_0^2=\frac{g}{M}\left[\frac{2^{2L}(L-1)!(L+1)!}{(2L)!}\right].$$

By inversion, the frequency distribution is

$$\rho^{(L)}(\omega)=\frac{2}{\pi L}\left\{\omega\left[\left(\frac{\omega_0}{\omega}\right)^{2/L}-1\right]^{1/2}\right\}^{-1}, \tag{87}$$

where the subscript (L) now denotes that the range of coupling extends through the Lth neighbor. For $L=1$, Eq. (87) is the same as the nearest neighbor case, but if $L>1$, $\rho^{(L)}(0)$ is infinite. If, for illustration, one formally includes only nearest and next-nearest neighbor coupling: $g_1=g$, $g_2=-g/4$, $g_m=0$ $(m>2)$ and

$$\rho^{(2)}(\omega)=\frac{1}{\pi}\frac{1}{(\omega_0\omega-\omega^2)^{1/2}},\quad \omega_0^2=\frac{4g}{M}. \tag{88}$$

From Eqs. (22) and (24)[21]

$$\begin{aligned}&\langle v(0)v(\tau)\rangle_c^{(2)}\\&=\frac{kT}{M}J_0\left(\frac{\omega_0\tau}{2}\right)\cos\frac{\omega_0\tau}{2}\to\frac{kT}{M}\frac{1}{(2\pi\omega_0\tau)^{1/2}}\\&\quad\times[1+\cos\omega_0\tau+\sin\omega_0\tau]\\&\langle v(0)v(\tau)\rangle_0^{(2)}\\&=\frac{\hbar\omega_0}{4M}\left[J_0\left(\frac{\omega_0\tau}{2}\right)\cos\frac{\omega_0\tau}{2}-J_1\left(\frac{\omega_0\tau}{2}\right)\sin\frac{\omega_0\tau}{2}\right]\\&\quad\to\frac{\hbar\omega_0}{2M(\pi\omega_0\tau)^{1/2}}\cos\left(\omega_0\tau-\frac{\pi}{4}\right).\end{aligned} \tag{89}$$

As for nearest neighbor coupling, the correlation functions decay as $(\omega_0\tau)^{-1/2}$. But the mean square displacements predicted by Eqs. (23) and (25) are very different. Setting $a=\omega_0 t$,

$$\begin{aligned}&\langle\sigma^2(t)\rangle_c^{(2)}\\&=\frac{4}{3}\frac{kT}{M\omega_0^2}\left\{a^2\left[J_0\left(\frac{a}{2}\right)\cos\frac{a}{2}+J_1\left(\frac{a}{2}\right)\sin\frac{a}{2}\right]\right.\\&\quad\left.-a\left[J_1\left(\frac{a}{2}\right)\cos\frac{a}{2}\right]\right\}\\&\quad\to\frac{kT}{M\omega_0}\left(\frac{2}{\pi}\right)^{1/2}t\left[\frac{4}{3}a^{1/2}+a^{-1/2}\right]\to\frac{kT}{M}\left(\frac{32}{9\pi\omega_0}\right)^{1/2}t^{3/2}\\&\quad+O(t^{1/2})\end{aligned} \tag{90}$$

$$\langle\sigma^2(t)\rangle_0^{(2)} = \frac{\hbar}{M}\, t\left[J_0\left(\frac{a}{2}\right)\sin\frac{a}{2} - J_1\left(\frac{a}{2}\right)\cos\frac{a}{2}\right]$$
$$\to \frac{\hbar}{M}\left(\frac{2t}{\pi\omega_0}\right)^{1/2} + O(t^{-1/2}).$$

For times sufficiently large that the $t^{1/2}$ contribution can be neglected, the classical mean square displacement is proportional to $t^{3/2}$. This result is "midway" between the free particle t^2 behavior and the usual linear time behavior characteristic of a diffusing particle with constant diffusion coefficient. In contrast to the logarithmic drift discussed previously, the present mean square displacement at zero temperature goes as $t^{1/2}$. While a constant diffusion coefficient implies that the classical displacement evolves in time exponentially compared to the zero temperature displacement, in the present model the former exceeds the latter only linearly in time. It is striking to note that roughly the same asymptotic form of the velocity autocorrelation function ($\tau^{-1/2}$) is consistent with diffusive behavior as varied as $\log t$, t, $t^{1/2}$, and $t^{3/2}$.

For the general case of coupling through the Lth neighbor the classical mean square displacement is

$$\langle\sigma^2(t)\rangle_c^{(L)} = \frac{4kT}{\pi ML}\,\frac{t^2}{a^{1/L}}\int_0^a \frac{1-\cos x}{x^{(3L-1)/L}\,[1-(x/a)^{2/L}]^{1/2}}\,dx. \tag{91}$$

As long as constant and decaying oscillatory terms are neglected, it is legitimate to expand the square root in powers of (x/a) keeping terms only through $(x/a)^{L-1}$, and subsequently integrating term-by-term with the upper limit of the integrals put at infinity. The result is an Lth order polynomial

$$\langle\sigma^2(t)\rangle_c^{(L)} = \frac{2kT}{M\omega_0^2}\,a^{2-1/L}\sum_{n=0}^{L-1} C_n^{(L)}\,\frac{1}{a^{2n/L}}, \quad a=\omega_0 t \gg 1 \tag{92}$$

$$C_n^{(L)} = (2n)!\left\{(2^n n!)^2\,[2L-(2n+1)]\right.$$
$$\left.\times\left[\Gamma\left(\frac{2L-2n-1}{L}\right)\right]\left[\sin\pi\left(\frac{2n+1}{2L}\right)\right]\right\}^{-1}.$$

For $L=1, 2$ Eq. (92) reproduces the results obtained above. For times so large that $a^{-2/L}$ can be neglected compared to unity, the $n=0$ term tends to dominate the other terms so that

$$\langle\sigma^2(t)\rangle_c^{(L)} \to \frac{2kT}{M\omega_0^2}\left[(2L-1)\,\Gamma\left(2-\frac{1}{L}\right)\sin\frac{\pi}{2L}\right]^{-1} a^{2-1/L}. \tag{93}$$

This dominance takes longer and longer to achieve, however, as L gets large. In order for the ratio of the first neglected term to the lead term to be $\epsilon \ll 1$, a must be of the order of $(\epsilon)^{-L/2}$: $a=10^5$ for $L=10$ and $\epsilon=0.1$ for example. In the limit that $L\to\infty$, $\omega_0^2 \to (g/M)\,(2\pi L)^{1/2}$ and

$$\langle\sigma^2(t)\rangle_c^{(L)} \underset{a\to\infty}{\overset{L\to\infty}{\longrightarrow}} \frac{2kT}{\pi M}\left[\frac{g}{M}\,(2\pi L)^{1/2}\right]^{-1/2L} t^{2-1/L}, \tag{94}$$

where L now is arbitrarily large. The classical mean square displacement asymptotically approaches that for a free particle, although, of course, L cannot be set equal to infinity since then the frequencies and frequency distribution are undefined. This result can be interpreted by examination of the ratio of successive force constant terms

$$\left|\frac{g_{m+1}}{g_m}\right| = \frac{L-m}{L+m+1}. \tag{95}$$

This ratio gets smaller as $1/2L$ for m of the order of L, and $g_L \to 0$ as $L^{1/2}/2^{2L}$ for L large; but for $m \ll L$ the ratios tend to approach unity. This mathematical behavior parallels formally the asymptotic behavior of the geometric progression model as $z\to 1$: the cutoff frequency tends to infinity while long range couplings tend to make equal dynamical contributions independent of range (m). In the latter model, however, the relative strengths of the couplings did not affect the finite behavior of $\rho(0)$ and served only to affect the mode of approach to a linear dependence on time (in accord with the Langevin theory of Brownian motion). In the present model increasing the relative strength of the couplings is tantamount to increasing the range of the couplings (L), directly affecting thereby the nature of the singular behavior of the frequency distribution at zero frequency. As a consequence the asymptotic time dependence of the mean square displacement becomes a sensitive function of the singularity which goes as $\omega^{-(L-1/L)}$. The zero temperature mean square displacement shows the same sort of effect. Substitution of Eq. (87) into Eq. (23) leads to a polynomial of $(L+1)/2$ terms for L odd and $L/2$ terms for L even.

$$\langle\sigma^2(t)\rangle_0^{(L)} = \frac{\hbar}{M\omega_0}\left\{a^{1-1/L}\sum_{n=0}^{P_L} C_n^L\,\frac{1}{a^{2n/L}} + \begin{bmatrix} B_L \log a, & L \text{ odd} \\ 0, & L \text{ even}\end{bmatrix}\right\},$$

$$C_n^L = (2n)!\left\{(2^n n!)^2\,[L-(2n+1)]\left[\Gamma\left(\frac{L-2n-1}{L}\right)\right]\right.$$
$$\left.\times\left[\cos\pi\left(\frac{2n+1}{2L}\right)\right]\right\}^{-1}, \tag{96}$$
$$B_L = \frac{4}{\pi L}\,\frac{1}{2^L}\,\frac{(L-1)!}{[((L-1)/2)!]^2},$$
$$P_L = \begin{bmatrix}\dfrac{L-3}{2}, & L \text{ odd}\\ \dfrac{L}{2}-1, & L \text{ even}\end{bmatrix}.$$

Again, for a large, the $n=0$ term is dominant, and when L is large as well, the coefficient of the log term for L odd is smaller than $C_0^L \sim 1/L$ by a factor of $L^{-1/2}$. Thus, in general,

$$\langle\sigma^2(t)\rangle_0^{(L)} \underset{a\to\infty}{\overset{L\to\infty}{\longrightarrow}} \frac{\hbar}{ML}\left[\frac{g}{M}\,(2\pi L)^{1/2}\right]^{-1/2L} t^{1-1/L} \tag{97}$$

and the dependence approaches t although the coefficient multiplying the time gets small as L^{-1} while it remains finite in the classical limit [Eq. (94)]. The asymptotic vanishing of the mean square displacement at zero temperature here as a linear time dependence is approached parallels that for a free particle, consistent with Eq. (94), and the conclusion $\langle\sigma^2(t)\rangle_c/\langle\sigma^2(t)\rangle_0 \to t$ which was demonstrated for the particular case of $L=2$

is a general result for any $L \geqslant 2$. It is interesting to note that the approximate linear dependence of the zero temperature mean square displacement upon time for L large is similar to classical diffusion in the limit that M is small such that

$$\lim_{\substack{M\to 0\\ L\to\infty}} ML = m_0$$

exists.

At the opposite extreme, the vanishing of $\rho(0)$ tends to depress the diffusive process or even possibly eliminate it completely. As an example, if the couplings fall off as the inverse square, so that[21]

$$g_m = g/m^2, \quad m = 1, 2 \cdots \infty \tag{98}$$

$$\omega_j^2 = \frac{4\omega_0^2}{\pi^2}\left[\pi\left(\frac{j\pi}{2N+1}\right) - \left(\frac{j\pi}{2N+1}\right)^2\right], \quad \omega_0^2 = \frac{\pi^2 g}{2M}$$

then the frequency distribution

$$\rho(\omega) = \omega/[\omega_0(\omega_0^2 - \omega^2)^{1/2}] \tag{99}$$

vanishes as ω as $\omega \to 0$. The velocity autocorrelation functions in the classical and zero temperature limits

$$\begin{Bmatrix}\langle v(0)\,v(\tau)\rangle_c\\ \langle v(0)\,v(\tau)\rangle_0\end{Bmatrix} = \begin{Bmatrix}\dfrac{\pi kT}{2M}\ \mathrm{H}_{-1}(\omega_0\tau)\\ \dfrac{\pi\hbar\omega_0}{8M}\,[J_0(\omega_0\tau) - J_2(\omega_0\tau)]\end{Bmatrix} \to \begin{Bmatrix}\dfrac{kT}{M}\\ \dfrac{\hbar\omega_0}{2M}\end{Bmatrix} \times \left(\frac{\pi}{2\omega_0\tau}\right)^{1/2}\cos\left(\omega_0\tau - \frac{\pi}{4}\right) \tag{100}$$

again decay asymptotically as $(\omega_0\tau)^{-1/2}$ over the whole temperature range. However, in this case, diffusion only occurs in the classical limit.

$$\begin{Bmatrix}\langle\sigma^2(t)\rangle_c\\ \langle\sigma^2(t)\rangle_0\end{Bmatrix} = \begin{Bmatrix}\dfrac{\pi kT}{M\omega_0^2}\displaystyle\int_0^{\omega_0 t}\mathrm{H}_0(x)\,dx\\ \dfrac{\pi\hbar}{2M\omega_0}\,[1 - J_0(\omega_0 t)]\end{Bmatrix} \to \begin{Bmatrix}\dfrac{2kT}{M\omega_0^2}\log\omega_0 t\\ \dfrac{\pi\hbar}{2M\omega_0}\end{Bmatrix}. \tag{101}$$

The classical result is parallel to the zero temperature result for the nearest neighbor lattice, since the integrals are identical. At zero temperature, rather than diffusing, the particle assumes a stationary equilibrium. It is just this circumstance which characterizes three-dimensional crystalline lattices at all temperatures. The reason is[31] that vibrational frequency distributions of solids vanish at low frequencies as ω^2. Thus, for a Debye solid,

$$\rho^{(D)}(\omega) = 3\omega^2/\omega_D^3, \quad 0 \leqslant \omega \leqslant \omega_D \tag{102}$$

the velocity autocorrelation functions are

$$\begin{matrix}\langle v(0)\,v(\tau)\rangle_c^{(D)}\\ \langle v(0)\,v(\tau)\rangle_0^{(D)}\end{matrix} \tag{103}$$

$$= \begin{Bmatrix}\dfrac{3kT}{M}\left[\dfrac{\sin x}{x} + \dfrac{2\cos x}{x^2} - \dfrac{2\sin x}{x^3}\right]\\ \dfrac{3\hbar\omega_D}{2M}\left[\dfrac{\sin x}{x} + \dfrac{3\cos x}{x^2} - \dfrac{6\sin x}{x^3} + \dfrac{6(1-\cos x)}{x^4}\right]\end{Bmatrix} \to \begin{Bmatrix}\dfrac{3kT}{M}\\ \dfrac{3\hbar\omega_D}{2M}\end{Bmatrix}\frac{\sin\omega_D\tau}{\omega_D\tau}$$

$$[x = \omega_D\tau]$$

and the associated mean square displacements evolve to equilibrium values

$$\begin{Bmatrix}\langle\sigma^2(t)\rangle_c^{(D)}\\ \langle\sigma^2(t)\rangle_0^{(D)}\end{Bmatrix} \tag{104}$$

$$= \begin{Bmatrix}\dfrac{6kT}{M\omega_D^2}\left[1 - \dfrac{\sin x}{x}\right]\\ \dfrac{3}{2}\dfrac{\hbar}{M\omega_D}\left[1 - \dfrac{2\sin x}{x} + \dfrac{2(1-\cos x)}{x^2}\right]\end{Bmatrix} \to \begin{Bmatrix}\dfrac{6kT}{M\omega_D^2}\\ \dfrac{3}{2}\dfrac{\hbar}{M\omega_D}\end{Bmatrix}$$

$$[x = \omega_D t].$$

The physical process underlying all these models can be understood in a general way by expressing the mean square displacement in terms of the position autocorrelation function. From Eqs. (7), (12), and (15)

$$\langle x(0)\,x(\tau)\rangle_T = \frac{\hbar}{2M}\int_0^{\omega_0}\frac{\cos\omega\tau}{\omega}\coth\left(\frac{\hbar\omega}{2kT}\right)\rho(\omega)\,d\omega \tag{105}$$

$$\langle\sigma^2(t)\rangle_T = 2[\langle x^2\rangle_{\mathrm{eq}} - \langle x(0)\,x(t)\rangle_T], \quad \langle x^2(0)\rangle = \langle x^2\rangle_{\mathrm{eq}}.$$

If the position autocorrelation function is a divergent integral for any time then, in particular, $\langle x^2\rangle_{\mathrm{eq}}$ is not defined. In such a case, the mean square displacement, which is the difference between two divergent quantities grows in time, characteristic of diffusion. This is the common property exhibited by all the frequency distributions considered, except for the Debye distribution for a solid. In the latter case, the equilibrium mean square displacement is finite and the position autocorrelation function decays in time similar to the velocity autocorrelation function. In the context of the present formalism, diffusion represents the fruitless search of a particle in coordinate space for an equilibrium position dispersion which does not exist. This point of view suggests that the frequency distributions considered here, and others which can be developed along the same lines, may provide analytical models to aid understanding of dynamical processes in liquids.

VII. IMPLICATIONS OF HARMONIC MODELS

The motivation of the present work, in common with that of previous investigators, has been to deduce features of transport phenomena from an underlying simple dynamics. Two questions are of importance concerning specifically harmonic models. The first is one of principle: why should an harmonically bound particle exhibit diffusive behavior in any approximation? The answer resides in the concept underlying all these treatments that irreversibility, which diffusion reflects, appears because the particle of interest (the "system") interacts with a large number of other particles (the "heat bath"). Diffusion occurs only in the limit that the number of heat bath particles is truly infinite, which is the approximation of Eq. (12). If, on the other hand, N is finite, irreversibility appears as such for times short compared to a recurrence time $T_p(N)$ characteristic of a Poincaré cycle, while for times of the order of T_p the system approaches its initial dynamical state. As N goes to infinity so does the length of the recurrence time. These points have been investigated by Mazur and Montroll[32] in connection with the classical velocity autocorrelation function for the nearest neighbor lattice. A still simpler example is afforded by the following force constant distribution, similar to Eq. (98) but with alternating sign:

$$g_m = (-1)^{m-1}(g/m^2), \quad m = 1, 2 \cdots \infty, \tag{106}$$

so that from Eqs. (9) and (16)[21]

$$\omega_j = \frac{2\omega_0 j}{2N+1} = \frac{2\pi j}{T_p(N)}\,;\quad T_p(N) = \frac{(2N+1)\pi}{\omega_0}, \quad \omega_0 = \left(\frac{g\pi^2}{2M}\right)^{1/2}$$

$$\langle v(0)\,v(\tau)\rangle_c = \frac{kT}{M}\,\frac{1}{2N+1}\sum_{j=-N}^{+N}\cos j\left(\frac{2\pi\tau}{T_p(N)}\right) \tag{107}$$

$$= \frac{kT}{M\omega_0}\,\frac{\pi}{T_p(N)}\,\frac{\sin\omega_0\tau}{\sin(\pi\tau/T_p(N))} \xrightarrow{T_p\to\infty} \frac{kT}{M}\,\frac{\sin\omega_0\tau}{\omega_0\tau}.$$

The frequency distribution is independent of frequency, $\rho(\omega) = \omega_0^{-1}$, a feature noticed by Brillouin.[8] As long as N is finite, the velocity autocorrelation function repeats its value in time T_p, and decays only in the approximation that $\tau \ll T_p \to \infty$. Similarly the mean square displacement with respect to the center of mass evolves as $2Dt$ only in the double limit $N \to \infty$ (first), then $t \to \infty$ (second). Using the same procedure as employed in Eq. (77).

$$\langle\sigma^2(t)\rangle_c = \frac{2kT}{\pi M\omega_0}\,T_p(N)\sum_{j=1}^{N}\frac{\sin^2 j(\pi t/T_p(N))}{j^2}$$

$$\underset{\text{"Thermodynamic Limit"}}{\xrightarrow{T_p\to\infty}}\ \frac{2kT}{M\omega_0^2}\left[(\omega_0 t)\sum_{n=0}^{\infty}\frac{(-1)^n(\omega_0 t)^{2n+1}}{(2n+1)(2n+1)!}\right.$$

$$\left. + \sum_{n=1}^{\infty}\frac{(-1)^n(\omega_0 t)^{2n}}{(2n)!}\right] \tag{108}$$

$$= \frac{2kT}{M\omega_0^2}\,[(\omega_0 t)\,\mathrm{Si}(\omega_0 t) - (1-\cos\omega_0 t)]$$

$$\underset{\text{"Diffusion Limit"}}{\xrightarrow{t\to\infty}}\ 2Dt, \qquad D = (\pi/2)(kT/M\omega_0),$$

where $\mathrm{Si}(x) \to \pi/2$ is the sine integral.[28] From the first equality, with N large but finite, the mean square displacement evolves from zero at $t=0$ to a maximum value of $(\pi^2 kT/2M\omega_0^2)N$ at $t = T_p/2$, returning to zero at $t = T_p$. These results are true only to an approximation however, since the sum of Eq. (9) should strictly extend to N in this case not infinity. A more rigourous and complete treatment will be presented elsewhere.[30] That diffusion implies time scales so small recurrence times are neglected is, as stressed by Berne,[10] a general dynamical principle.

The second question, of practical relevance is, how realistically applicable to liquids is a formalism whose construction is patterned after a harmonic solid? One positive connection is the fact that Eq. (24) implies $\rho(\omega)$ is the Fourier transform of the velocity autocorrelation function. This relationship is true for liquids, $\rho(\omega)$ in such cases referred to as the spectral density.[33] Egelstaff[34] has noted that the value of the spectral density at zero frequency yields the diffusion coefficient, in parallel to Eq. (31). On the negative side, of course, long range harmonic couplings do not exist. If harmonic models are to provide an adequate base upon which to investigate problems in liquid dynamics, their success would rest upon cancellation of two errors: the error of introducing harmonic potentials at the outset compensated for by including many of them through long range couplings. The geometric progression model of Eq. (36) includes the feature that successive force constant terms are of opposite sign. As a consequence each particle of the system is experiencing a superposition of repulsive as well as attractive forces: the contributions of the former tending to push the particle offset the contributions of the latter which tend to bind the particle. The resultant effective potential can be interpreted as an approximate resolution of, and interplay between, attractive and repulsive forces characterizing true molecular interactions in liquids. For low values of z, short range couplings dominate, but they are sufficient to reproduce qualitatively the velocity autocorrelation functions of computer molecular dynamics studies. As z approaches unity longer range couplings take on greater significance and the physical picture changes to that of a heavy particle embedded in a heat bath of effective light particles. In this case the particle is experiencing an enormous number of significant interactions which finds its parallel in the large number of collisions suffered by a particle of colloidal size immersed in a liquid. However, the most gratuitous acceptance of this interpretation must be tempered by the following objection to the low temperature extrapolations of Sec. V. While the Ford model and the present geometric progression model which purports to be its extension reduce in the classical limit to the prediction of the Langevin theory as far as time evolution is concerned, the origin of the relaxation time parameter is very different. In the former cases λ emerges as a purely mechanical quantity, while in the latter case it is identified as proportional to the viscosity, itself a function of temperature. One could formally introduce temperature dependence into λ by postulating, say, that z is some function of T, but in the absence of some guiding principle, making the range of coupling itself temperature dependent is a procedure devoid of foundation. On the other hand, identification of the friction coefficient to the viscosity is independent of the dynamics predicted by the Langevin equation and appears as a separate postulate. The position taken here is $\lambda = kT/MD$ should be regarded formally as a parameter whose value is fixed by the empirically determined value of the diffusion coefficient for a real liquid, and speculate that observation of a particle of macroscopic size suspended in a superfluid may exhibit mean square displacement patterns departing from the classical theory of Eq. (73) in accordance with the general prediction of Eq. (68). As the examples of Eq. (70) show the quantum effects are of a rather complicated form. While it is, of course, impossible to detect any system at absolute zero, the $T=0$ calculations are included to show the extrapolations to zero temperature which would be indicated by very low temperature measurements.

Models of diffusion considered in Sec. VI are representative of physical situations in which the diffusion coefficient varies in time. This implies that the particle mass is changing as it moves due to chemical reaction or some other mechanism. In such cases particle motion would be in response to inelastic as well as elastic encounters, and the former would tend to either promote fragmentation into lighter parts or association into a heavier species (as in coagulation). While the examples were chosen primarily to illustrate matters of principle, it would be of interest to develop more realistic frequency distributions which might reflect the essential features of such processes.

Future communications intend to deal with other force constant distributions, other transport processes and general considerations of the response of these systems to externally applied forces.

ACKNOWLEDGMENTS

The author wishes to extend his sincere appreciation to Professor Jeffries Wyman for his interest and helpful discussions concerning this work, and to Dr. Alfredo Colosimo for his aid in the graphical computations. He also wishes to thank the European Molecular Biology Organization for the support of a Research Fellowship.

APPENDIX A. INTEGRALS

The calculations of Secs. IV and V in connection with the theory of Brownian motion reduced in many cases to the following integral for $\omega_0 t$ large:

$$I(c)=\int_0^{\omega_0}\frac{(\omega_0^2-\omega^2)^{1/2}}{c^2+\omega^2}\cos\omega t\,d\omega=\frac{1}{2}\int_{-\omega_0 t}^{+\omega_0 t}\frac{[(\omega_0 t)^2-x^2]^{1/2}}{x^2+(ct)^2}\cos x\,dx. \tag{A1}$$

As a first approximation the limits in the second expression are extended to infinity and the integral evaluated by contour integration in the upper complex plane, enclosing the pole at $ict(>0)$. Then

$$I(c)=\frac{\pi}{2c}(\omega_0^2+c^2)^{1/2}\exp(-ct)+I', \tag{A2}$$

where I' is the error introduced by extending the limits to infinity. The first correction is found by integrating the first equality in Eq. (A1) by parts. After some rearrangements

$$I(c)=\frac{\pi}{2}\frac{\omega_0^2}{\omega_0^2+c^2}\frac{J_1(\omega_0 t)}{\omega_0 t}+\frac{1}{2}\left\{[(\omega_0 t)^2+(ct)^2]^{-1}\int_{-\omega_0 t}^{+\omega_0 t}\frac{x[(\omega_0 t)^2-x^2]^{1/2}}{x^2+(ct)^2}\sin x\,dx\right. \tag{A3}$$
$$\left.+2\int_{-\omega_0 t}^{+\omega_0 t}\frac{x[(\omega_0 t)^2-x^2]^{1/2}}{[x^2+(ct)^2]^2}\sin x\,dx\right\}.$$

After extending the limits to infinity, subsequent contour integration of the two integrals produces again the first term in Eq. (A2). Consequently the first correction to I' is the Bessel function factor. Retaining only the first term of its asymptotic expansion for $\omega_0 t$ large

$$I(c)=\frac{\pi}{2c}(\omega_0^2+c^2)^{1/2}\exp(-ct)+\left(\frac{\pi}{2}\right)^{1/2}\frac{\omega_0^2}{\omega_0^2+c^2}\frac{\sin(\omega_0 t-\pi/4)}{(\omega_0 t)^{3/2}}+O(\omega_0 t)^{-5/2}. \tag{A4}$$

This result was used without proof by Ullersma in connection with his velocity autocorrelation function calculations (see text). The calculations of Sec. V involved the following two types of integrals expressible in terms of $I(c)$:

$$\int_0^{\omega_0}\frac{\cos\omega t\,d\omega}{(\omega^2+a_n^2)(\omega_0^2-\omega^2)^{1/2}}=\frac{1}{\omega_0^2+a_n^2}\left[I(a_n)+\frac{\pi}{2}J_0(\omega_0 t)\right] \tag{A5}$$
$$=\frac{\pi}{2}\left[\frac{\exp(-a_n t)}{a_n(\omega_0^2+a_n^2)^{1/2}}+\frac{1}{\omega_0^2+a_n^2}\left(\frac{2}{\pi\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right]$$

$$\int_0^{\omega_0}\frac{(\omega_0^2-\omega^2)^{1/2}\cos\omega t}{(a_n^2+\omega^2)(\lambda^2+\omega^2)}\,d\omega=\frac{1}{\lambda^2-a_n^2}[I(a_n)-I(\lambda)] \tag{A6}$$
$$=\frac{\pi}{2(\lambda^2-a_n^2)}\times\left[\frac{(\omega_0^2+a_n^2)^{1/2}}{a_n}\exp(-a_n t)-\frac{(\omega_0^2+\lambda^2)^{1/2}}{\lambda}\exp(-\lambda t)\right]$$

retaining terms only through $(\omega_0 t)^{-1/2}$.

APPENDIX B. CORRECTION TO $\langle\sigma^2(t)\rangle_T^{(L)}$ FOR FINITE MICROSCOPIC INTERACTION TIME

The mean square displacement of Eq. (68) which reduces in the classical limit to the Langevin result, assumes $\mu\approx\lambda/\omega_0=0$. The first correction to this approximation is found from Eq. (67) by expanding terms in μ. To lowest order in μ

$$\langle\sigma^2(t)\rangle_T=\langle\sigma^2(t)\rangle_T^{(L)}+\mu^2C_T(t) \tag{B1}$$

where, following a procedure similar to that of Eq. (64)

$$C_T(t)=\frac{\hbar}{\pi M\lambda}\left\{\frac{\lambda t}{2p}-\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{2\pi}{\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right. \tag{B2}$$
$$\left.+\lim_{\substack{N\to\infty\\ \mu\to 0}}\left[\sum_{n=1}^{N}\frac{1-\exp-n((\omega_0/\nu)t)}{n}+\frac{2p^2}{\mu^2}\left(\int_N^\infty\frac{dn}{n(p^2-n^2)[1-(\mu^2/p^2)(p^2-n^2)]^{1/2}}-\int_N^\infty\frac{dn}{n(p^2-n^2)^{1/2}}\right)\right]\right\}$$

and with the identity $2\pi kT/\hbar=\omega_0/\nu=\lambda/p=\lambda q$,

$$C_T(t)=\frac{\hbar}{M\pi\lambda}\left\{\frac{\lambda t}{2p}+\log\left[\frac{p}{\mu}(1-\exp(-\lambda t/p))\right]-\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{2\pi}{\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right\}$$

$$+\left[\mathrm{const}=(\hbar/\pi M\lambda)(\gamma+\tfrac{1}{2}+\log 2)\right] \tag{B3}$$

$$=\frac{kT}{M\lambda^2}\left\{\lambda t+\frac{2}{q}\log\left[\frac{1-\exp(-q\lambda t)}{\mu q}\right]-\frac{2}{q}\coth\left(\frac{\hbar\omega_0}{2kT}\right)\left(\frac{2\pi}{\omega_0 t}\right)^{1/2}\cos\left(\omega_0 t-\frac{\pi}{4}\right)\right\}$$

$$+[\mathrm{const.}].$$

Thus, for example, at zero temperature ($p=\infty$), with the use of Eq. (79)

$$\langle\sigma^2(t)\rangle_0=2D\tau_q[1+(\mu^2/2)]\log\lambda t,\quad \lambda t\gg 1 \tag{B4}$$

and in the classical limit ($\hbar=0$, $q=\infty$), Eq. (B3) reduces to the second term of Eq. (60).

*Present address.

[1]P. C. Hemmer, *Dynamic and Stochastic Types of Motion in a Linear Chain*, thesis, Trondheim, Norway, 1959.
[2]G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).
[3]P. Ullersma, Physica **32**, 27, 56, 74 (1966).
[4]M. A. Huerta and H. S. Robertson, J. Stat. Phys. **1**, 373 (1969); H. S. Robertson and M. A. Huerta, Am. J. Phys. **38**, 619 (1970); M. A. Huerta and H. S. Robertson, J. Stat. Phys. **3**, 171 (1971).
[5]W. Bernard and H. B. Callen, Rev. Mod. Phys. **31**, 1017 (1959).
[6]F. Reif, *Fundamentals of Statisical and Thermal Physics* (McGraw-Hill, New York, 1965), Chap. 15.
[7]H. B. Rosenstock, Phys. Rev. **121**, 416 (1961).
[8]L. Brillouin, *Wave Propagation in Periodic Structures* (McGraw-Hill, New York, 1946), Chap. III.
[9]R. Zwanzig, J. Stat. Phys. **9**, 215 (1973).
[10]B. J. Berne in *Physical Chemistry, An Advanced Treatise, Vol. VIII B/ Liquid State*, edited by H. Eyring, D. Henderson, and W. Jost (Academic, New York, 1971), Chap. 9.
[11]D. K. C. MacDonald, Physica **28**, 409 (1962).
[12]G. E. Uhlenbeck and L. G. Ornstein, Phys. Rev. **36**, 823 (1930); S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943); M. C. Wang and G. E. Uhlenbeck, **17**, 323 (1945).
[13]R. Weinstock, Am. J. Phys. **39**, 484 (1971).
[14]L. Van Hove, Phys. Rev. **89**, 1189 (1953).
[15]L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1969), 2nd ed.
[16]U. Fano, Rev. Mod. Phys. **29**, 74 (1957).
[17]It is possible to define the equilibrium dispersions alternatively in terms of statistical conditions on position and velocities of the physical particles $[x_n(0),\dot{x}_n(0)]$ instead of upon the normal modes. This has been done by Huerta and Robertson (Ref. 4) for the $L=1$ case, assuming $x_n(0),\dot{x}_n(0)$ obey independent Gaussian distributions. The opinion here is that since the normal modes are independent quasiparticles by construction it is more physically reasonable and consistent to have them defined initially in terms of independent distributions, so that temperature appears in a natural way [Eq. 15)]. To couple the physical particles through harmonic potentials and then require their initial positions and velocities to be statistically independent requires a separate discussion of what the temperature in fact is. The procedure here, which parallels that of Ford *et al.* (Ref. 2), provides an unambiguous prescription for treating both classical and quantum behaviours. As distinct from them, however, the zero temperature contributions to the dispersions are retained for reasons discussed below.
[18]G. C. Wick, Phys. Rev. **80**, 268 (1950).
[19]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.
[20]Ullersma (Ref. 3), who refers to τ_q as the quantum statistical transient time, defines this quantity as $\hbar/kT$ which differs from Eq. (31) by a factor of π. The present definition emphasizes a formal symmetry between the classican and zero temperature expressions of Eq. (30).
[21]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[22]R. E. Turner, Physica **26**, 269 (1960).
[23]R. J. Rubin, Bull. Am. Phys. Soc. **1**, 221 (1956).
[24]The diffusion constant for $z=0$ (nearest neighbor coupling) of Eq. (58) agrees with Hemmer's result (Ref. 1) and with that of Huerta and Robertson (Ref. 4) for canonical initial conditions. However, the latter, reported that the correction term goes as $t^{1/2}$ and not $t^{-1/2}$ as computed here, yet their expression for the mean square displacement is just I_r of Eq. (55). The cause of this discrepancy is that they let $\omega_0 t$ go to infinity in the integral of I_r (which gives the value 1) and approximated J_1 by $\sim(\omega_0 t)^{-1/2}$. However, if one takes the limit for $\omega_0 t$ large of the integrated expression in terms of Struve functions as done here, the $(\omega_0 t)^{-1/2}$ terms cancel and the first correction goes as $(\omega_0 t)^{-3/2}$ [Eq. (56)].
[25]J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).
[26]J. L. Lebowitz and E. Rubin, Phys. Rev. **131**, 2381 (1963).
[27]R. Zwanzig in Ann. Rev. Phys. Chem. **16**, 67 (1965).
[28]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[29]W. Magnus, F. Oberhittinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966).
[30]P. E. Phillipson (to be published).
[31]A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation*, Supplement 3, Solid State Physics (Academic, New York, 1963).
[32]P. Mazur and E. Montroll, J. Math. Phys. **1**, 70 (1960).
[33]See, for example, the article by S.-H. Chen in Vol. VIII/A of Ref. 10, Chap. 2.
[34]P. A. Egelstaff in *Inelastic Scattering of Neutrons in Solids and Liquids* (IAEA, Vienna, 1961), p. 25.

Summation relation for $U(N)$ Racah coefficients*

K. T. Hecht

Physics Department, University of Michigan, Ann Arbor, Michigan 48104
(Received 3 April 1974)

A summation relation is given for $U(N)$ Racah coefficients which has the *form* of an orthogonality relation, or a composition of recoupling transformations, except that the summation over column indices (for fixed row indices) is over multiplicity labels *only*. In the recoupling matrix for $[f^1] \times [f^2] \times [f^3] \to [f]$, $U(N)$ irreducible representations $[f^2]$ and $[f^3]$ are limited to be elementary, $[11...10...0] \equiv [1^k]$, or totally symmetric $[k]$, or of the form $[k^{N-1}]$. Results are tabulated as functions of the axial distances in $[f]$ for $[f^2]=[1^{N-1}]$, $[1^{N-2}]$, or $[2^{N-1}]$; $[f^3]=[1]$, $[1^2]$, or $[2]$; all cases which arise in the evaluation of squares of matrix elements of one- and two-body operators averaged over irreducible representations of $U(N)$.

1. INTRODUCTION

In recent years the Wigner–Racah calculus for the unitary groups $U(N)$ has been brought to a state of development comparable to that for the angular momentum calculus for $SU(2)$. Biedenharn, Louck, and collaborators,[1-6] especially, have developed powerful methods which make it possible to calculate all Wigner and Racah coefficients for $U(N)$. For the case of multiplicity free and extremal Wigner couplings, in particular, algebraic formulas for the Wigner coefficients can be read off directly from their diagrammatic pattern calculus.[4] In more general cases an additional algorithm is needed to extract algebraic or numerical values of the Wigner coefficients from their formalism. In the case of $SU(3)$[2,3] this has been translated into a computer program,[7,8] so that both Wigner and Racah coefficients for $SU(3)$ are now available in complete generality. Biedenharn and Louck advocate the view that there is a canonical structure for the $U(N)$ Wigner–Racah algebra. This eliminates all free choices in the resolution of the multiplicity problem for the general Wigner coupling, so that all $U(N)$ Wigner and Racah coefficients are uniquely defined. For arbitrary N, explicit algebraic constructions for Wigner couplings involving the most general multiplicity structure have so far been limited to matrix elements of the Wigner operators transforming as the $U(N)$ irreducible representation $[211...10] \equiv [21^{N-2}]$ {equivalent to $[10...0-1]$ in $SU(N)$}. As a by-product of this calculation, Louck and Biedenharn[1] also give the $U(N)$ Racah coefficients for the recoupling matrix for $[f]\times[11...10]\times[10...0] \to [f]$ in elegantly compact form. Although Racah coefficients, being independent of subgroup labels, have a simpler algebraic structure than the Wigner coefficients, general expressions for $U(N)$ Racah coefficients have so far been limited to a few very special cases, usually cases in which the four Wigner couplings in the Racah recoupling transformation are all free of multiplicity such as when two or more of the irreducible representations are totally symmetric (Moshinsky and Chacón[9] and Ališauskas, Jucys, and Jucys[10]). In the applications to physical problems Racah coefficients are often more useful than Wigner coefficients, and it is hoped that the work of Louck and Biedenharn[1] will be extended to more general cases. Since the algebraic construction for the most general $U(N)$ Racah coefficients is complicated, it may be useful to search for new relations or sum rules for the $U(N)$ Racah coefficients which have no analog for the simpler $SU(2)$ Racah coefficients.

It is the purpose of this note to exhibit such a summation relation. It has the *form* of the well-known orthogonality relations for the $U(N)$ Racah coefficients, or of a composition of recoupling transformations, except that the summation over column indices (for fixed row index) is over multiplicity labels *only*, for fixed $U(N)$ irreducible representation label within the column index. This relation is particularly simple if the representations $[f^2]$ and $[f^3]$ in the recoupling matrix for $[f^1]\times[f^2]\times[f^3] \to [f]$ are either "elementary," $[11...10...0] \equiv [1^k]$, or totally symmetric, $[k0...0] \equiv [k]$, or of the form $[kk...k0] \equiv [k^{N-1}]$. In this case the sum is completely independent of the multiplicity structure and can hence be calculated by permutation group techniques. This sum arises naturally[11] in applications to physical problems, since it is needed in the calculation of squares of matrix elements of operators averaged over the states of irreducible representations of $U(N)$, where these averages are needed in the study of spectroscopic problems using spectral distribution methods.[11,12] In Sec. 2 the summation relation for $U(N)$ Racah coefficients is related to the matrix element of a projection operator for the symmetric group. Section 3 takes up the calculational tools needed to evaluate this matrix element, including a transformation to nonstandard representations of S_n. The details of the calculation are exhibited through some illustrative examples in Sec. 4. Finally, results are tabulated for all cases of the recoupling matrix for $[f]\times[f^2]\times[f^3] \to [f]$, where the $U(N)$ irreducible representations $[f^2]$ are of the form $[1^{N-1}]$ or $[1^{N-2}]$ or $[2^{N-1}]$, and $[f^3]$ is of the form $[1]$ or $[1^2]$ or $[2]$, which are the $U(N)$ irreducible representations needed to construct all one- and two-body operators through the coupling $[f^2]\times[f^3]$.

2. THE SUMMATION RELATION

For present purposes it will be convenient to use a notation for the $U(N)$ Racah coefficient which is a straightforward generalization of that for the angular momentum calculus for $SU(2)$ and give the Racah coefficient in unitary form, the U coefficient, which is given by the recoupling matrix

$$U([f^1][f^2][f][f^3];[f^{12}]\rho^{12}\rho^{12,3};[f^{23}]\rho^{23}\rho^{1,23}) \\ = \langle((([f^1]\times[f^2])[f^{12}]\rho^{12}\times[f^3])[f]\rho^{12,3} \\ \times | ([f^1]\times([f^2]\times[f^3])[f^{23}]\rho^{23})[f]\rho^{1,23}\rangle. \tag{1}$$

Here, the irreducible representation labels $[f^s] \equiv [f^s_{iN}]$ are given by the partition numbers f^s_{iN}, $i=1,\ldots,N$, which specify the number of squares in the ith row of the Young tableau describing the representation $[f^s]$ of $U(N)$. The multiplicity labels ρ^{st} are needed whenever the Wigner coupling of $[f^s]$ with $[f^t]$ can yield a specific representation $[f^{st}]$ with d-fold multiplicity, $d>1$. Since the result of this investigation will involve a summation over multiplicity labels only, it is convenient to use a separate symbol for the multiplicity label and avoid the more elegant notation of references,[1-6] even though the Biedenharn–Louck canonical structure has been adopted for the $U(N)$ Wigner–Racah algebra. Note that the column index for the unitary transformation matrix is specified by both the irreducible representation label $[f^{23}]$ and the multiplicity labels ρ^{23} and $\rho^{1,23}$; similarly for the row index. In the notation of Louck and Biedenharn,[1] the above U coefficient is the matrix element of the $U(N)$ Racah invariant operator

$$\left\{\begin{pmatrix}[f^{23}]\\(\Gamma^{1,23})\end{pmatrix}\begin{pmatrix}(\Gamma^{23})\\ [f^3]\\(\Gamma^{12,3})\end{pmatrix}\begin{pmatrix}[f^2]\\(\Gamma^{12})\end{pmatrix}\right\} \tag{2}$$

connecting states of irreducible representation $[f^1]$ (on the right) to states $[f]$ (on the left). Here, the labels Γ^{st} include both the multiplicity labels ρ^{st} and the shift indices, $\Delta_i = f_{iN}^{st}$, which indicate how many of the squares of the Young tableau for $[f^t]$ have been added to the ith row of the tableau for $[f^s]$ to make the tableau for $[f^{st}]$.

For recoupling transformations in which the representations $[f^2]$ and $[f^3]$ are restricted to be "elementary" $[1^k]$, totally symmetric $[k]$, or of the form $[k^{N-1}]$, only the multiplicity label $\rho^{1,23}$ is needed. (The other Wigner couplings are free of multiplicity; whenever a multiplicity label ρ is unnecessary it will be omitted.) In this case the sum

$$\begin{aligned}&\sum_{\rho^{1,23}} U([f^1][f^2][f][f^3];[f^{12}]\_\_;[f^{23}]\_\rho^{1,23})\\ &\quad\times U([f^1][f^{2\prime}][f][f^{3\prime}];[f^{12\prime}]\_\_;[f^{23}]\_\rho^{1,23})\end{aligned} \tag{3}$$

can be evaluated by permutation group techniques. Note that with $[f^{2\prime}]=[f^2]$, $[f^{3\prime}]=[f^3]$, and a summation over *both* $\rho^{1,23}$ *and* $[f^{23}]$, the above would have become merely one of the orthonormality relations for the U coefficients. The above sum over ρ only, however, is a simple function of the irreducible representation labels $[f^1],[f^{23}],[f]$; $[f^2],[f^3],[f^{2\prime}],[f^{3\prime}]$; $[f^{12}]$ and $[f^{12\prime}]$. It is this function which is to be evaluated in this investigation.

To evaluate the sum of Eq. (3), it is convenient to introduce n-particle state vectors $|[f]\alpha;r_n r_{n-1} r_{n-2}\cdots r_1\rangle$ which are simultaneous base vectors for an irreducible representation of $U(N)$ and of the standard Young–Yamanouchi representation[13] of S_n (with $n=\sum_i f_{iN}$), where α stands for a complete set of subgroup labels for $U(N)$ (the Gel'fand labels f_{ij} with $i\le j=1,\ldots,N-1$, could be used, for example[1-6]), and where $r_n r_{n-1}\cdots r_1$ is a standard Yamanouchi symbol.[13] It will further be useful to transform to a nonstandard representation of S_n,[14,15] in which the group of k particles labeled n, $n-1,\ldots,n-k+1$ have a definite permutation symmetry, e.g., $[1^k]$ or $[k]$. Such a state vector can then be expanded in terms of $U(N)$ Wigner coefficients

$$\begin{aligned}&|[f]\alpha;\{r_n r_{n-1}\cdots r_{n-k+1}\}_{[f_1]} r_{n-k} r_{n-k-1}\cdots r_1\rangle\\ &\quad=\sum_{\alpha_1''\alpha_1}|[f_1'']\alpha_1''\rangle|[f_1]\alpha_1\rangle\langle[f_1'']\alpha_1''[f_1]\alpha_1|[f]\alpha\rangle,\end{aligned} \tag{4}$$

where the $U(N)$ Wigner coefficient, $\langle[f_1'']\alpha_1''[f_1]\alpha_1|[f]\alpha\rangle_\rho$, is the matrix element of the Wigner operator $\langle\ \rangle$, namely

$$\left\langle\begin{matrix}[f]\\(\alpha)\end{matrix}\right|\left\langle\begin{matrix}(\Gamma)\\ [f_1]\\(\alpha_1)\end{matrix}\right\rangle\left|\begin{matrix}[f_1'']\\(\alpha_1'')\end{matrix}\right\rangle \tag{5}$$

in the notation of references.[1-6] Note that $[f_1]$ is assumed to be of the form $[1^k]$, $[k]$, or $[k^{N-1}]$, so that no multiplicity label ρ is needed in the Wigner coefficient of Eq. (4). Note also that the representation $[f_1'']$ is determined uniquely by the Yamanouchi symbols $r_n,\ldots,r_{n-k+1}$. (The tableau for $[f_1'']$ is obtained from the tableau for $[f]$ by removing squares from rows $r_n, r_{n-1},\ldots$, and r_{n-k+1}.) By repeating this process for a second group of k' particles, and expressing the product of k-particle and k'-particle representations $[f_1]$ and $[f_1']$ in terms of coupled $U(N)$ representations $[f_0]$ for the $(k+k')$-particle state, the n-particle state vector can be expanded as

$$\begin{aligned}&|[f]\alpha;r(k,k')_{[f_1][f_1']}\cdots\rangle\\ &\equiv|[f]\alpha;\{r_n\cdots r_{n-k+1}\}_{[f_1]}\{r_{n-k}\cdots r_{n-k-k'+1}\}_{[f_1']} r_{n-k-k'}\cdots r_1\rangle\\ &=\sum_{\alpha_1''\alpha_1}\sum_{\alpha'\alpha_1'}\sum_{[f_0]\alpha_0}|[f']\alpha'\rangle|([f_1']\times[f_1])[f_0]\alpha_0\rangle\\ &\quad\times\langle[f']\alpha'[f_1']\alpha_1'|[f_1'']\alpha_1''\rangle\langle[f_1'']\alpha_1''[f_1]\alpha_1|[f]\alpha\rangle\\ &\quad\times\langle]f_1']\alpha_1'[f_1]\alpha_1|[f_0]\alpha_0\rangle.\end{aligned} \tag{6}$$

The sums over subgroup labels α_1, α_1', α_1'' for the product of three $U(N)$ Wigner coefficients can be expressed more simply in terms of $U(N)$ Racah coefficients by

$$\begin{aligned}&\sum_{\alpha_1''\alpha_1'\alpha_1}\langle[f']\langle[f']\ \alpha'[f_1']\ \alpha_1'|[f_1'']\ \alpha_1''\rangle\ \alpha_1'|[f_1'']\alpha_1''\rangle\\ &\quad\times\langle[f_1'']\alpha_1''[f_1]\alpha_1|[f]\alpha\rangle\langle[f_1']\alpha_1'[f_1]\alpha_1|[f_0]\alpha_0\rangle\\ &\quad=\sum_\rho\langle[f']\alpha'[f_0]\alpha_0|[f]\alpha\rangle_\rho\\ &\quad\times U([f'][f_1'][f][f_1];[f_1'']\_\_;[f_0]\_\rho).\end{aligned} \tag{7}$$

Since the Wigner coupling $[f']\times[f_0]$ is in general not free of multiplicity, both $U(N)$ Wigner and Racah coefficients are functions of the multiplicity label ρ, and the result involves a sum over this multiplicity label.

To obtain the relation for the sum of Eq. (3), consider the matrix element of a projection operator, $Y^{[f_0]}$,

$$\langle[f]\alpha;r'(k_2,k_2')_{[f_2][f_2']}\cdots|Y^{[f_0]}|[f]\alpha;r(k_1,k_1')_{[f_1][f_1']}\cdots\rangle$$

with $k_2+k_2'=k_1+k_1'$, where $Y^{[f_0]}$ is an operator, built from permutation operators for particles labeled $n,n-1,\ldots,n-k_1-k_1'+1$, which projects the representation with Young tableau $[f_0]$ out of an arbitrary (k_1+k_1')-particle state. By using Eqs. (6) and (7) for both state vectors, the matrix element of $Y^{[f_0]}$ can be expressed as

$$\langle [f]\alpha; r'(k_2,k_2')_{[f_2][f_2']}\dots | Y^{[f_0]} | [f]\alpha; r(k_1,k_1')_{[f_1][f_1']}\dots\rangle$$
$$= \sum_{\alpha'\alpha_0}\sum_{\rho\rho'} \langle [f']\alpha'[f_0]\alpha_0 | [f]\alpha\rangle_\rho$$
$$\times \langle [f']\alpha'[f_0]\alpha_0 | [f]\alpha\rangle_{\rho'}$$
$$\times U([f'][f_1'][f][f_1];[f_1'']\_\_;[f_0]\_\rho)$$
$$\times U([f'][f_2'][f][f_2];[f_2'']\_\_;[f_0]\_\rho'), \tag{8}$$

where we have used the property of the projection operator

$$\langle [f_0'']\alpha_0'' | Y^{[f_0]} | [f_0']\alpha_0'\rangle = \delta_{[f_0][f_0']}\delta_{[f_0][f_0'']}\delta_{\alpha_0'\alpha_0''} \tag{9}$$

to eliminate sums over $[f_0]$. From the orthonormality of the $U(N)$ Wigner coefficients

$$\sum_{\alpha'\alpha_0} \langle [f']\alpha'[f_0]\alpha_0 | [f]\alpha\rangle_\rho \langle [f']\alpha'[f_0]\alpha_0 | [f]\alpha\rangle_{\rho'} = \delta_{\rho\rho'} \tag{10}$$

we then obtain the desired summation relation

$$\sum_\rho U([f'][f_1'][f][f_1];[f_1'']\_\_;[f_0]\_\rho)$$
$$\times U([f'][f_2'][f][f_2];[f_2'']\_\_[f_0]\_\rho)$$
$$= \langle [f]\alpha; r(k_2,k_2')_{[f_2][f_2']}\dots | Y^{[f_0]} | [f]\alpha; r(k_1,k_1')_{[f_1][f_1']}\dots\rangle. \tag{11}$$

Since the projection operator, $Y^{[f_0]}$, serves only to project the representation $[f_0]$ out of an arbitrary (k_1+k_1')-particle state, its S_n subgroup character is completely immaterial. It could be constructed according to the Young–Yamanouchi–Rutherford[16] prescription; but it is usually much simpler to give it in symmetric or antisymmetric form[13] for some conveniently labeled tableau (not necessarily a standard labeling), since any normalized linear combination of $Y_i^{[f_0]}$'s with different S_n subgroup labels i will serve the purpose.

3. CALCULATIONAL TOOLS

The matrix element of $Y^{[f_0]}$ can be related to the basic matrix element of the transpositions $P_{m-1,m}$ in the standard Young–Yamanouchi representation[13]

$$\langle [f];\dots r_m = p, r_{m-1} = q\dots | P_{m-1,m} | [f];\dots r_m = p, r_{m-1} = q\dots\rangle$$
$$= 1/\tau_{pq},$$
$$\langle [f];\dots r_m = q, r_{m-1} = p\dots | P_{m-1,m} | [f];\dots r_m = p, r_{m-1} = q\dots\rangle$$
$$= [1 - 1/\tau_{pq}^2]^{1/2}, \tag{12}$$

where τ_{pq} is the "axial distance" between the squares labeled m and $m-1$ in the Young tableau,

$$\tau_{pq} = f_p^{(m)} - f_q^{(m-1)} - p + q \tag{13a}$$

and $f_i^{(m)}$ is the number of squares in row i of the m-particle tableau left, after particles labeled $n, n-1, \dots, m+1$ have been removed from the original n-particle tableau of shape $[f]$. If the symbol r_m is preceded by $\sigma(p)$ symbols with the label p and the symbol r_{m-1} is preceded by $\sigma(q)$ symbols with the label q,

$$\tau_{pq} = f_p - f_q - p + q - \sigma(p) + \sigma(q) \tag{13b}$$

where f_i now designates the number of squares in the ith row of the n-particle tableau, $[f]$ $(f_i \equiv f_{iN})$. Note that $\tau_{qp} = -\tau_{pq}$, and that $\tau_{pp} = +1$, since $\sigma(q) = \sigma(p) + 1$ in this case. Note also that τ_{pq} can be expressed in terms of differences of "partial hooks,"[1–6] $p_{iN} = f_{iN} - i + N$.

The transformation to nonstandard representations of S_n is particularly simple if the k particles to be singled out belong to the totally symmetric or antisymmetric representations of S_n. In this case the transformation coefficients have been given by Horie.[15] For the totally symmetric case in which the k particles labeled $n-m-1, \dots, n-m-k$ have been singled out

$$|r_n\dots r_{n-m}\{r_{n-m-1} = a_1, \dots, r_{n-m-k} = a_k\}_{[k]} r_{n-m-k-1}\dots r_1\rangle$$
$$= \sum_P P\left[\frac{1}{k!}\prod_{i<j=1}^{k}\left(1 + \frac{1}{\tau_{a_i a_j}}\right)\right]^{1/2}$$
$$\times |r_n\dots r_{n-m} a_1 a_2\dots a_k r_{n-m-k-1}\dots\rangle, \tag{14}$$

where the sum is over the $k!$ permutations P which permute the symbols $a_1, \dots, a_k$ in both the state vector and the coefficient. Similarly, for the totally antisymmetric case

$$|r_n\dots r_{n-m}\{r_{n-m-1} = a_1, \dots, r_{n-m-k} = a_k\}_{[1^k]} r_{n-m-k-1}\dots r_1\rangle$$
$$= \sum_P P(-1)^{p+1}\left[\frac{1}{k!}\prod_{i<j=1}^{k}\left(1 - \frac{1}{\tau_{a_i a_j}}\right)\right]^{1/2}$$
$$\times |r_n\dots r_{n-m} a_1 a_2\dots a_k r_{n-m-k-1}\dots\rangle \tag{15}$$

with p = even (odd) for even (odd) permutations, P.

It will also be useful to build state vectors antisymmetric (or symmetric) in one group of k particles, labeled $n-m-1, \dots, n-m-k$, from vectors antisymmetric (or symmetric) in the $k-1$ particles $n-m-2, \dots, n-m-k$ through antisymmetrizers A (or symmetrizers S), with

$$A(n-m-1, \dots, n-m-k) = \frac{1}{k!}\sum_P P(-1)^p$$

where the sum runs over the $k!$ permutations P of particles labeled $n-m-1, \dots, n-m-k$, and p = even (odd) for P = even (odd):

$$A(n-m-1, \dots, n-m-k)$$
$$\times |r_n\dots r_{n-m} a_1\{a_2 a_3\dots a_k\}_{[1^{k-1}]} r_{n-m-k-1}\dots r_1\rangle$$
$$= \left[\frac{1}{k}\prod_{i=2}^{k}\left(1 - \frac{1}{\tau_{a_1 a_i}}\right)\right]^{1/2}$$
$$\times |r_n\dots r_{n-m}\{a_1 a_2\dots a_k\}_{[1^k]} r_{n-m-k-1}\dots r_1\rangle. \tag{16}$$

The inverse transformation gives

$$|\dots\{a_1 a_2\dots a_k\}_{[1^k]}\dots\rangle$$
$$= \sum_{i=1}^{k}(-1)^{i+1}\left[\frac{1}{k}\prod_{\substack{j=1\\ j\neq i}}^{k}\left(1 - \frac{1}{\tau_{a_i a_j}}\right)\right]^{1/2}$$
$$\times |\dots a_i\{a_1\dots a_{i-1}a_{i+1}\dots a_k\}_{[1^{k-1}]}\dots\rangle \tag{17}$$

with analogous expressions for totally symmetric groups. Another useful relation involves the transposition operator which interchanges particles labeled $n-m-1$ and $n-m-k$ in a state vector antisymmetric in particles labeled $n-m-2, \dots n-m-k$:

$$P_{n-m-1,n-m-k}|\dots a_1\{a_2 a_3\dots a_k\}_{[1^{k-1}]}\dots\rangle$$
$$= \left[\prod_{i=2}^{k}\left(1 - \frac{1}{\tau_{a_1 a_i}^2}\right)\right]^{1/2}|\dots\{a_2 a_3\dots a_k\}_{[1^{k-1}]} a_1\dots\rangle$$

$$+\sum_{i=2}^{k}(-1)^i \frac{1}{\tau_{a_1a_i}}\left[\prod_{\substack{j=2\\ j\neq i}}^{k}\left(1-\frac{1}{\tau_{a_1a_j}}\right)\left(1-\frac{1}{\tau_{a_ja_i}}\right)\right]^{1/2}$$
$$\times|\ldots\{a_1\ldots a_{i-1}a_{i+1}\ldots a_k\}_{[1^{k-1}]}\,a_i\ldots\rangle. \tag{18a}$$

In the special case $a_1=a_2$ the above collapses to

$$P_{n-m-1,n-m-k}|\ldots a_1\{a_1a_3\ldots a_k\}_{[1^{k-1}]}\ldots\rangle$$
$$=|\ldots\{a_1a_3\ldots a_k\}_{[1^{k-1}]}a_1\ldots\rangle. \tag{18b}$$

The transformation to nonstandard representations of S_n is complicated in the case where the k-particle representations of S_n are other than one-dimensional. For a group of $(N-1)k$ particles of symmetry $[kk\ldots k0]\equiv[k^{N-1}]$, however, it is sufficient to construct a single $(N-1)k$-particle state of symmetry $[k^{N-1}]$, since the properties of $SU(N)$ insure that the Wigner coupling $[f]\times[k^{N-1}]$ is free of multiplicity. It will be convenient to choose this single state to transform according to the irreducible representation $[k^{N-1}]$ of S_n and to be in normal antisymmetric form in k distinct groups of $(N-1)$ particles each. (We shall use the notation $[k^{N-1}]_a$. Note that this state is a complicated linear combination of the Young–Yamanouchi base vectors for $[k^{N-1}]$.)

An $(N-1)$-particle state of symmetry $[1^{N-1}]$ in which the $N-1$ Yamanouchi symbols include the numbers 1 through N, with the *exception* of the specific number i, will be denoted by $\{\bar{i}\}_{[1^{N-1}]}$:

$$|\ldots\{\bar{i}\}_{[1^{N-1}]}\ldots\rangle\equiv|\ldots\{12\ldots i-1,i+1,\ldots N\}_{[1^{N-1}]}\ldots\rangle$$

In this notation the state $[k^{N-1}]_a$ can be expanded as

$$|[f]\ldots\{\bar{a}_1\bar{a}_2\ldots\bar{a}_k\}_{[k^{N-1}]_a}\ldots\rangle$$
$$=\sum_{\mathcal{P}}\mathcal{P}\left[\frac{1}{k!}\prod_{i<j=1}^{k}\left(1+\frac{1}{\tau_{\bar{a}_i\bar{a}_j}}\right)\right]^{1/2}$$
$$\times|[f]\ldots\{\bar{a}_1\}_{[1^{N-1}]}\{\bar{a}_2\}_{[1^{N-1}]}\ldots\{\bar{a}_k\}_{[1^{N-1}]}\ldots\rangle, \tag{19}$$

where the sum is over the $k!$ permutations $\mathcal{P}$ which permute the symbols $\bar{a}_i$. The coefficients follow from Eq. (14) and from conjugation properties under $SU(N)$. The state conjugate to $[f]$ transforms according to $SU(N)$ irreducible representation $[f^*]$ with $f_i^*=f_1-f_{N+1-i}$. If the state $[f(\bar{p})]$ is obtained by removing $N-1$ squares from the tableau for $[f]$, one from each *but* the pth row, then the irreducible representation conjugate to $[f(\bar{p})]$ is specified by the tableau $[f^*(N+1-p)]$ which is obtained from $[f^*]$ by removing one square from row $N+1-p$ of $[f^*]$. Thus the function $\tau_{\bar{a}_i\bar{a}_j}$ in Eq. (19) is related to axial distances in $[f]$ in the following way:

$$\tau_{\bar{a}_i\bar{a}_j}\equiv\tau^{[f]}_{\bar{a}_i\bar{a}_j}=\tau^{[f^*]}_{N+1-a_i,N+1-a_j}$$
$$=f^*_{N+1-a_i}-f^*_{N+1-a_j}+(N+1-a_j)-(N+1-a_i)$$
$$=(f_1-f_{a_i})-(f_1-f_{a_j})+a_i-a_j=f_{a_j}-f_{a_i}+a_i-a_j$$
$$=\tau_{a_ja_i}. \tag{20}$$

In particular, therefore,

$$|[f]r_n\ldots r_{n-m}\{\bar{p}\bar{q}\}_{[2^{N-1}]_a}r_{n-m-2N+1}\ldots r_1\rangle$$
$$=\left[\frac{1}{2}\left(1+\frac{1}{\tau_{qp}}\right)\right]^{1/2}|[f]r_n\ldots r_{n-m}\{12\ldots p-1,p+1,$$
$$\ldots N\}_{[1^{N-1}]}\{12\ldots q-1,q+1,\ldots N\}_{[1^{N-1}]}\ldots\rangle$$
$$+\left[\frac{1}{2}\left(1-\frac{1}{\tau_{qp}}\right)\right]^{1/2}|[f]r_n\ldots r_{n-m}\{12\ldots q-1,q+1,$$
$$\ldots N\}_{[1^{N-1}]}\{12\ldots p-1,p+1,\ldots N\}_{[1^{N-1}]}\ldots\rangle \tag{21}$$

with

$$\tau_{qp}=f_q-f_p-q+p-\sigma(q)+\sigma(p)$$

for $p\neq q$, where $\sigma(q)$ and $\sigma(p)$ are the number of occurrences of the row numbers q and p in the Yamanouchi symbols $r_n,\ldots r_{n-m}$ while $\tau_{qq}\equiv+1$.

In addition to the above results for nonstandard representations of S_n, the only other calculational tool needed is the value of the simple sum

$$\Sigma_m\equiv\sum_{i=1}^{m}\prod_{\substack{j=1\\ j\neq i}}^{m}\left(1-\frac{1}{\tau_{a_ia_j}}\right)=m. \tag{22}$$

The proof follows (see Ref. 1): Let $\tau_{a_ia_j}=(x_i-x_j)$. Then x_i are real numbers such that $(x_i-x_j)\neq 0$, for $i\neq j$ and Σ_m can be expressed as a contour integral in the complex z plane

$$\Sigma_m=-\frac{1}{2\pi i}\oint\prod_{j=1}^{m}\left(1-\frac{1}{(z-x_j)}\right)dz,$$

where the contour encloses the m simple poles $x_1,\ldots,x_m$. After expanding the product in the integrand, only the m terms of the form $1/(z-x_i)$ give a nonzero contribution to the contour integral.

4. ILLUSTRATIVE EXAMPLES

The methods used to evaluate the matrix elements of the projection operator $Y^{[f_0]}$ of Eq. (11) will be illustrated with two examples. Consider first the simplest (and trivial) case: $[f_1]=[f_2]=[1]$; $[f_1']=[f_2']=[1^{N-1}]$; $[f_0]=[21^{N-2}]$. This is a trivial example since the value of the sum of Eq. (3) follows in this case from the orthonormality relation for the U coefficients and the values for the U coefficients with $[f_0]=[1^N]\equiv[0\ldots0]$, which are given by simple dimension factors, see Ref. 1 and Eq. (43) below. Nevertheless, the example illustrates the techniques to be used in more complicated cases. In this example it is most convenient to choose the projection operator $Y^{[f_0]}$ in normal antisymmetric form

$$Y^{[21^{N-2}]}=\mathcal{N}A(n-1,\ldots,n-N+1)$$
$$\times\tfrac{1}{2}(1+P_{n,n-1})A(n-1,\ldots,n-N+1)$$

with a normalization factor $\mathcal{N}$ to be determined. In this case

$$\langle a_1\{a_2a_3\ldots a_N\}_{[1^{N-1}]}\ldots|Y^{[21^{N-2}]}|a_1\{a_2a_3\ldots a_N\}_{[1^{N-1}]}\ldots\rangle$$
$$=\tfrac{1}{2}\mathcal{N}\{1+\langle a_1\{a_2a_3\ldots a_N\}_{[1^{N-1}]}\ldots|$$
$$\times P_{n,n-1}|a_1\{a_2a_3\ldots a_N\}_{[1^{N-1}]}\ldots\rangle\}$$
$$=\tfrac{1}{2}\mathcal{N}\left\{1+\frac{1}{(N-1)}\sum_{i=2}^{N}\left[\frac{1}{\tau_{a_1a_i}}\prod_{\substack{j=2\\ j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)\right]\right\}, \tag{23}$$

where Eq. (17) has been used to uncouple the particle numbered $n-1$ from the antisymmetric group, together with the basic matrix element (12). By rewriting $1/\tau_{a_1a_i}=(1-1/\tau_{a_ia_1})-1$, the sum in Eq. (23) becomes

$$\left[\sum_{i=2}^{N}\prod_{\substack{j=1\\ j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)-\sum_{i=2}^{N}\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)\right]. \tag{24}$$

By adding and subtracting $\Pi_{j=2}^{N}(1-1/\tau_{a_1a_j})$ to the first term, both sums in (24) can be evaluated with the identity (22) to give

$$\left[N - \prod_{j=2}^{N}\left(1 - \frac{1}{\tau_{a_1 a_j}}\right) - (N-1)\right],$$

so that

$$\langle a_1\{a_2\ldots a_N\}_{[1^{N-1}]}|Y^{[21^{N-2}]}|a_1\{a_2\ldots a_N\}_{[1^{N-1}]}\ldots\rangle$$
$$=\sum_\rho U^2([f'][1^{N-1}][f][1];[f'']=[f(a_1)]\_\_;[21^{N-2}]\_\rho)$$
$$=\frac{\mathcal{N}}{2}\frac{N}{(N-1)}\left\{1-\frac{1}{N}\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_1a_j}}\right)\right\}. \tag{25}$$

The normalization factor $\mathcal{N}$ is most readily determined by choosing the labels a_i such that the Racah recoupling transformation collapses to a trivial one-dimensional unitary transformation. In the above example, with $a_2 = a_1 = 1$, $a_3a_4\ldots a_N = 23\ldots N-1$, and $[f] = [21^{N-2}]$, the representation $[f']$ is the scalar representation $[f'] = [0\ldots0]$, so that the square of the single surviving U coefficient in the sum of Eq. (25) has the value unity. In this case, with $a_2 = a_1$, $\tau_{a_1a_2} = +1$, and Eq. (25) is reduced to $1 = \mathcal{N}N/2(N-1)$. With this value for $\mathcal{N}$, arbitrary $[f]$, and $a_1a_2\ldots a_N$ = any permutation of $12\ldots N$, Eq. (25) yields

$$\sum_\rho U^2([f][1^{N-1}][f][1];[f(a_1)]\_\_;[21^{N-2}]\_\rho)$$
$$=\left\{1-\frac{1}{N}\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_1a_j}}\right)\right\}, \tag{26}$$

where $[f(a_1)]$ is the representation with a tableau obtained by removing one square from row a_1 of the tableau for $[f]$. Note that $[f'] = [f(12\ldots N)]$ (removal of one square from each row of $[f]$), is equivalent to $[f]$ in $SU(N)$.

As a second example consider the case $[f_1] = [f_2] = [2]$, $[f_1'] = [f_2'] = [2^{N-1}]$, $[f_0] = [21^{N-2}]$. With particles numbered $n-2$ through $n-2N+1$ already prepared with a permutation symmetry $[2^{N-1}]_a$ in the state vectors, it will now be sufficient to choose a projection operator $Y^{[f_0]}$ of the form

$$\mathcal{N}A(n-N-1,\ldots,n-2N+1)A(n-1,\ldots,n-N)S(n,n-2,$$
$$n-N-1)A(n-1,\ldots,n-N)A(n-N-1,\ldots,n-2N+1) \tag{27}$$

corresponding to the labeled tableau of Fig. 1. The symmetrizer

$$S=\tfrac{1}{2}(1+P_{n,n-2})\tfrac{1}{3}(1+P_{n,n-N-1}+P_{n-2,n-N-1})\tfrac{1}{2}(1+P_{n,n-2}) \tag{28}$$

when sandwiched in between the antisymmetrizers of (27) can be written in the form

$$S=\tfrac{1}{12}(1+P_{n,n-1})(1+2P_{n-1,n-N-1})(1+P_{n,n-1}). \tag{29}$$

The antisymmetrizers, when acting on a state vector of symmetry $|\{a_1a_2\}_{[2]}\{\bar a_1\bar a_2\}_{[2^{N-1}]}\ldots\rangle$ [see Eq. (21)], give

$$A(n-1,\ldots,n-N)A(n-N-1,\ldots,n-2N+1)$$
$$|\{a_1a_2\}_{[2]}\{\bar a_1\bar a_2\}_{[2^{N-1}]}\ldots\rangle$$
$$=\frac{1}{2}\left(1-\frac{1}{\tau_{a_1a_2}}\right)\left[\frac{1}{N}\left(1-\frac{1}{(\tau_{a_1a_2}+1)}\right)\prod_{j=3}^{N}\left(1-\frac{1}{\tau_{a_1a_j}}\right)\right]^{1/2}$$
$$\times|a_2\{a_1a_2\ldots a_N\}_{[1^N]}\{a_1a_3\ldots a_N\}_{[1^{N-1}]}\ldots\rangle$$
$$-\frac{1}{2}\left(1+\frac{1}{\tau_{a_1a_2}}\right)\left[\frac{1}{N}\left(1-\frac{1}{\tau_{a_2a_1}+1}\right)\prod_{j=3}^{N}\left(1-\frac{1}{\tau_{a_2a_j}}\right)\right]^{1/2}$$
$$\times|a_1\{a_1a_2\ldots a_N\}_{[1^N]}\{a_2a_3\ldots a_N\}_{[1^{N-1}]}\ldots\rangle, \tag{30}$$

where Eq. (16) has been used, after $|\{a_1a_2\}_{[2]}\ldots\rangle$ is expanded through Eq. (14). It is thus sufficient to evaluate the symmetrizer S, of Eq. (29), between states of the type $|a_1\{a_1a_2\ldots a_N\}_{[1^N]}\{b_1b_2\ldots b_{N-1}\}_{[1^{N-1}]}\ldots\rangle$, where

$$\langle a_1'\{a_1'a_2'\ldots a_N'\}_{[1^N]}\{b_1'b_2'\ldots b_{N-1}'\}_{[1^{N-1}]}|$$
$$\times S|a_1\{a_1a_2\ldots a_N\}_{[1^N]}\{b_1b_2\ldots b_{N-1}\}_{[1^{N-1}]}\rangle$$
$$=\tfrac{1}{6}\langle\ldots|(1+P_{n,n-1})|\ldots\rangle$$
$$+\tfrac{1}{6}\langle\ldots|(1+P_{n,n-1})P_{n-1,n-N}P_{n-N,n-N-1}P_{n-1,n-N}(1$$
$$+P_{n,n-1})|\ldots\rangle. \tag{31}$$

Now, using Eqs. (17) and (12), we have

$$(1+P_{n,n-1})|a_1\{a_1a_2\ldots a_N\}_{[1^N]}\ldots\rangle$$
$$=2\left[\frac{1}{N}\prod_{j=2}^{N}\left(1-\frac{1}{(\tau_{a_1a_j}-1)}\right)\right]^{1/2}|a_1a_1\{\bar a_1\}_{[1^{N-1}]}\ldots\rangle$$
$$+\sum_{i=2}^{N}(-1)^{i+1}\left[\frac{2}{N}\frac{(\tau_{a_1a_i}+1)}{(\tau_{a_1a_i}-1)}\prod_{\substack{j=2\\j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)\right]^{1/2}$$
$$\times|\{a_1a_i\}_{[2]}\{\bar a_i\}_{[1^{N-1}]}\ldots\rangle \tag{32}$$

while, using Eq. (18),

$$P_{n-1,n-N}(1+P_{n,n-1})|a_1\{a_1a_2\ldots a_N\}_{[1^N]}\ldots\rangle$$
$$=\left[\frac{1}{N}\prod_{j=2}^{N}\left(1+\frac{1}{(\tau_{a_1a_j}-1)}\right)\right]^{1/2}\left\{1+\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_1a_j}}\right)\right\}$$
$$\times|a_1\{\bar a_1\}_{[1^{N-1}]}a_1\ldots\rangle$$
$$+\sum_{i=2}^{N}(-1)^{i+1}\left[\frac{1}{N}\left(1-\frac{1}{(\tau_{a_1a_i}-1)}\right)\prod_{\substack{j=2\\j\neq i}}^{N}\left(1+\frac{1}{\tau_{a_ia_j}}\right)\right]^{1/2}$$
$$\times|a_1\{\bar a_i\}_{[1^{N-1}]}a_i\ldots\rangle$$
$$+\sum_{i=2}^{N}(-1)^{i+1}\left[\frac{1}{N}\prod_{\substack{j=1\\j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)\right]^{1/2}$$
$$\times|a_i\{\bar a_i\}_{[1^{N-1}]}a_1\ldots\rangle \tag{33}$$

where repeated use has been made of the identity (22) to simplify the coefficients of the three types of terms.

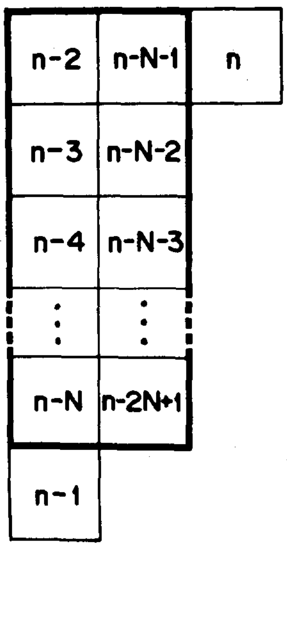

FIG. 1.

Next, with the shorthand notation $\tau_{a_1'a_i}=(\tau_{a_1a_i}-1)=-\tau_{a_ia_1'}=-(\tau_{a_ia_1}+1)$,

$$\langle a_1\{a_1\ldots a_N\}_{[1^N]}\ldots|(1+P_{n,n-1})|a_1\{a_1\ldots a_N\}_{[1^N]}\ldots\rangle$$
$$=\frac{1}{N}\left\{2\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_1'a_j}}\right)+2\sum_{i=2}^{N}\left(1-\frac{1}{\tau_{a_ia_1'}}\right)\prod_{\substack{j=2\\j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)-\sum_{i=2}^{N}\prod_{\substack{j=2\\j\neq i}}^{N}\left(1-\frac{1}{\tau_{a_ia_j}}\right)\right\}$$
$$=\frac{1}{N}\{2N-(N-1)\} \tag{34}$$

where the sums have again been evaluated with the identity (22). On the other hand,

$$\langle a_2\{a_1\ldots a_N\}_{[1^N]}\ldots|(1+P_{n,n-1})|a_1\{a_1\ldots a_N\}_{[1^N]}\ldots\rangle=0. \tag{35}$$

Finally, from (33), using (17) to uncouple the particle labeled $n-N-1$, we have

$$\langle a_1\{a_1\ldots a_N\}_{[1^N]}\{b_1\ldots b_{N-1}\}_{[1^{N-1}]}|(1+P_{n,n-1})P_{n-1,n-N}P_{n-N,n-N-1}$$
$$\times P_{n-1,n-N}(1+P_{n,n-1})|a_1\{a_1\ldots a_N\}_{[1^N]}\{b_1\ldots b_{N-1}\}_{[1^{N-1}]}\rangle$$
$$=\frac{1}{N(N-1)}\sum_{l=1}^{N-1}\prod_{\substack{j=1\\j\neq l}}^{N-1}\left(1-\frac{1}{\tau_{b_lb_j}}\right)F(b_l) \tag{36}$$

with

$$F(b_l)=\frac{1}{(\tau_{a_1b_l}-1)}\left\{N+2+\prod_{j=2}^{N}\left(1+\frac{1}{(\tau_{a_1a_j}-1)}\right)\right\}+\sum_{i=2}^{N}\frac{1}{\tau_{a_ib_l}}\left(1+\frac{1}{(\tau_{a_ia_1}-1)}\right)\prod_{\substack{j=2\\j\neq i}}^{N}\left(1+\frac{1}{\tau_{a_ia_j}}\right)$$
$$=\frac{(N+2)}{(\tau_{a_1b_l}-1)}+1-\left(1-\frac{1}{(\tau_{a_1b_l}-1)}\right)\prod_{j=2}^{N}\left(1-\frac{1}{\tau_{a_jb_l}}\right), \tag{37}$$

where $F(b_l)$ has again been simplified by the use of the identity (22). Finally, the very last product in (37) must always be zero, since $\tau_{a_1b_l}=2$ if $b_l=a_1$ (in this case b_l is preceded by two a_1's in the state vector); and $\tau_{a_jb_l}=+1$ if $b_l=a_j$ for any $j\geq 2$ (b_l is now preceded by one a_j). Hence

$$F(b_l)=\left\{1+\frac{(N+2)}{(\tau_{a_1b_l}-1)}\right\}$$

and in this form the sum over l in (36) can be performed to give

$$\sum_{l=1}^{N-1}\prod_{\substack{j=1\\j\neq l}}^{N-1}\left(1-\frac{1}{\tau_{b_lb_j}}\right)F(b_l)$$
$$=\left\{(2N+1)-(N+2)\prod_{l=1}^{N-1}\left(1-\frac{1}{(\tau_{a_1b_l}-1)}\right)\right\}. \tag{38}$$

Combining (31), (34), and (38), we have

$$\langle a_1\{a_1\ldots a_N\}_{[1^N]}\{b_1\ldots b_{N-1}\}_{[1^{N-1}]}|$$
$$\times S|a_1\{a_1\ldots a_N\}_{[1^N]}\{b_1\ldots b_{N-1}\}_{[1^{N-1}]}\rangle$$
$$=\frac{1}{6}\frac{(N+2)}{(N-1)}\left\{1-\frac{1}{N}\prod_{l=1}^{N-1}\left(1-\frac{1}{(\tau_{a_1b_l}-1)}\right)\right\}. \tag{39}$$

Two cases must be considered: First, if some b_j is equal to a_1, then this b_j is preceded by two a_1's, and $\tau_{a_1b_j}=+2$ for this j, so that the product in Eq. (39) is zero. On the other hand, if $b_i\neq a_1$, $i=1,\ldots,N-1$, then every b_i is preceded by a single a_i, and $\tau_{a_1b_i}=(\tau_{a_1a_i}+1)$. In the two cases, Eq. (39) gives

$$\langle a\{12\ldots N\}_{[1^N]}\{\bar{a}\}_{[1^{N-1}]}\ldots|$$
$$\times S|a\{12\ldots N\}_{[1^N]}\{\bar{a}\}_{[1^{N-1}]}\ldots\rangle$$
$$=\frac{1}{6}\frac{(N+2)}{(N-1)}\left\{1-\frac{1}{N}\prod_{\substack{i=1\\i\neq a}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\right\} \tag{40a}$$

where a is any number from 1 to N, while, with $a\neq b$,

$$\langle a\{12\ldots N\}_{[1^N]}\{\bar{b}\}_{[1^{N-1}]}\ldots|$$
$$\times S|a\{12\ldots N\}_{[1^N]}\{\bar{b}\}_{[1^{N-1}]}\ldots\rangle$$
$$=\frac{1}{6}\frac{(N+2)}{(N-1)}. \tag{40b}$$

Using similar techniques, we have

$$\langle b\{12\ldots N\}_{[1^N]}\{\bar{b}\}_{[1^{N-1}]}|S|a\{12\ldots N\}_{[1^N]}\{\bar{a}\}_{[1^{N-1}]}\ldots\rangle$$
$$=\frac{(N+2)}{6N(N-1)}\left[\left(1-\frac{1}{\tau_{ab}^2}\right)\prod_{\substack{i=1\\i\neq a\neq b}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\left(1-\frac{1}{\tau_{bi}}\right)\right]^{1/2}. \tag{40c}$$

Before the final result for the Racah summation relation can be written down, we need to evaluate the normalization factor $\mathcal{N}$ of Eq. (27). By choosing $[f]=[32\ldots21]$, so that $[f']=[0\ldots0]$, the matrix element of $Y^{[21^{N-2}]}$ again has the value unity. Thus

$$\langle[32^{N-2}1];\{1N\}_{[2]}\{\bar{N}\bar{N}\}_{[2^{N-1}]}|$$
$$\times Y^{[21^{N-2}]}|[32^{N-2}1];\{1N\}_{[2]}\{\bar{N}\bar{N}\}_{[2^{N-1}]}\rangle$$
$$=1=\mathcal{N}\frac{1}{2}\left(1+\frac{1}{\tau_{1N}}\right)\frac{1}{N}\left(1-\frac{1}{(\tau_{N1}-1)}\right)\prod_{i=2}^{N-1}\left(1-\frac{1}{\tau_{Ni}}\right)$$
$$\times\langle 1\{12\ldots N\}_{[1^N]}\{\bar{N}\}_{[1^{N-1}]}|$$
$$\times S|1\{12\ldots N\}_{[1^N]}\{\bar{N}\}_{[1^{N-1}]}\rangle, \tag{41}$$

with $\tau_{1N}=N+1=-\tau_{N1}$, $\tau_{Ni}=-(N-i+1)$, for $i\geq 2$, and the matrix element of S given by $(N+2)/6(N-1)$. Thus

$$1=\frac{\mathcal{N}}{2}\frac{(N+2)}{2N}\frac{(N+2)}{6(N-1)}.$$

Finally, with this value of $\mathcal{N}$, combining Eqs. (30), (40a), and (40c), the diagonal matrix element of $Y^{[21^{N-2}]}$ leads to the summation relation

$$\sum_\rho U^2([f][2^{N-1}][f][2];[f(ab)]\_\_;[21^{N-2}]\_\rho)$$
$$=\frac{1}{(N+2)}\left\{\frac{(\tau_{ab}+1)^2}{\tau_{ab}(\tau_{ab}-1)}\prod_{\substack{i\neq a\\i\neq b}}^{N}\left(1-\frac{1}{\tau_{bi}}\right)+\frac{(\tau_{ab}-1)^2}{\tau_{ab}(\tau_{ab}+1)}\prod_{\substack{i\neq a\\i\neq b}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)-\frac{4}{N}\prod_{\substack{i\neq a\\i\neq b}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\left(1-\frac{1}{\tau_{bi}}\right)\right\}, \tag{42}$$

where $[f'']=[f(ab)]$ is the representation with a tableau obtained by removing one square each from row a

2154

and b of the tableau for $[f]$. Note that in this particular example $[f']$ is again equivalent to $[f]$ in $SU(N)$, since $[f']$ is obtained from $[f]$ by removing two squares from *every* row of the tableau for $[f]$.

5. RESULTS

The summation relation (11) arises naturally in applications to physical problems[11] through the squares of matrix elements of operators, averaged over the states of an irreducible representation of $SU(N)$. [The reader is directed to Ref. 11 which shows in detail how the $SU(N)$ Racah sum, Eq. (11), arises in averaging the square of an operator or the product of two operators over the states of an irreducible representation of $SU(N)$. Reference 11 also gives a number of applications to nuclear spectroscopic problems using the methods of spectral distributions.] The operators of greatest interest are one- and two-body operators for which the representations $[f_1']$ and $[f_1]$ in the Racah recoupling matrix for $[f']\times[f_1']\times[f_1]\to[f]$ are of the type $[f_1'] = [1^{N-1}]$, $[1^{N-2}]$, or $[2^{N-1}]$; $[f_1]=[1]$, $[1^2]$, or $[2]$. Since $[f_1']$ is then a representation conjugate to a one- or two-particle representation, it will be natural to denote it by $[f_1'^*]$, with $[1^{N-2}]=[1^{2*}]$, for example. To eliminate trivial dimensional factors, it will also be useful to tabulate results in the form

$$\frac{\sum_\rho U([f'][f_1'^*][f][f_1];\,[f_1'']\_\_;[f_0]\_\rho)}{U([f'][f_1'^*][f'][f_1'];\,[f_1'']\_\_;[0]\_\_)} \times \frac{U([f'][f_2'^*][f][f_2];\,[f_2'']\_\_;[f_0]\_\rho)}{U([f'][f_2'^*][f'][f_2'];[f_2'']\_\_;[0]\_\_)}, \tag{11$'$}$$

where the U coefficients in the denominator, with $[f_0] = [0\ldots0]\equiv[0]$, equivalent to $[1^N]$ or $[2^N]$ in $SU(N)$ are given by trivial dimensional factors. By evaluating these coefficients by the techniques outlined in Secs. 3 and 4, the summation relation in the form of Eq. (11′) also becomes as much as possible independent of specific phase conventions for the $U(N)$ Wigner-Racah algebra. It will be useful to express the U coefficients with $[f_0]=[0]$ as functions of the axial distances in $[f]$. The needed coefficients are

$$U([f][1^{N-1}][f][1];\,[f(a)];\,[0]) = (-1)^{a+1}\left[\frac{1}{N}\prod_{i\neq a}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\right]^{1/2},$$

$$U([f][1^{N-2}][f][1^2];\,[f(ab)];[0]) = (-1)^{a+b}\left[\frac{2}{N(N-1)}\prod_{\substack{i\neq a\\ i\neq b}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\left(1-\frac{1}{\tau_{bi}}\right)\right]^{1/2},$$

$$U([f][2^{N-1}][f][2];\,[f(aa)];\,[0]) = \left[\frac{2}{N(N+1)}\prod_{i\neq a}^{N}\left(1-\frac{2}{\tau_{ai}}\right)\right]^{1/2},$$

$$U([f][2^{N-1}][f][2];\,[f(ab)];\,[0]) = (-1)^{a+b}\left[\frac{2}{N(N+1)}\prod_{\substack{i\neq a\\ i\neq b}}^{N}\left(1-\frac{1}{\tau_{ai}}\right)\left(1-\frac{1}{\tau_{bi}}\right)\right]^{1/2}, \tag{43}$$

where $[f(ab)]$ is again specified by a tableau obtained by removing one square each from row a and b of the tableau for $[f]$, with a and b any of the numbers from 1 to N. The possible irreducible representations $[f_0]$ for one- and two-body operators, can be read off from the direct products $[f_1'^*]\times[f_1]=\Sigma[f_0]$:

$$\begin{aligned}
&1.\ [1^{N-1}]\times[1]=[0]+[21^{N-2}],\\
&2.\ [1^{N-2}]\times[1^2]=[0]+[21^{N-2}]+[2^21^{N-4}],\\
&3.\ [2^{N-1}]\times[2]=[0]+[21^{N-2}]+[42^{N-2}],\\
&4.\ [2^{N-1}]\times[1^2]=[21^{N-2}]+[3^22^{N-3}],\\
&5.\ [1^{N-2}]\times[2]=[21^{N-2}]+[31^{N-3}].
\end{aligned} \tag{44}$$

Representations such as $[f']=[f(a_1^3a_2a_3\ldots a_{N-2})]$ or $[f']=[f(a_1^4a_2^2a_3^2\ldots a_{N-1}^2)]$ can be reached from $[f]$ by only a single operator $Y^{[f_0]}$. E.g., $[f_0]=[31^{N-1}]$ if $[f_1'^*][f_1] =[1^{N-2}][2]$, or $[f_0]=[42^{N-2}]$ if $[f_1'^*][f_1]=[2^{N-1}][2]$, for $[f']=[f(a_1^3a_2a_3\ldots a_{N-2})]$. {$[f']$ is the $SU(N)$ representation described by the tableau obtained by removing three squares from row a_1 and one each from rows a_2 through a_{N-2} from the tableau for $[f]$, where the a_i are any of the numbers from 1 through N; note that $[f(a_1^3a_2a_3\ldots a_{N-2})]$ is equivalent to $[f(a_1^4a_2^2a_3^2a_4^2\ldots a_{N-2}^2a_{N-1}a_N)]$ in $SU(N)$.} For such representations the summation relation (11) is trivial, since the Racah recoupling matrix becomes one-dimensional. The only nontrivial cases therefore involve irreducible representations:

$$\text{A. } [f']=[f(a_1^2a_2\ldots a_{N-1})] \equiv\{[f(a_1^3a_2^2\ldots a_{N-1}^2a_N)] \text{ in } SU(N)\}$$

and

$$\text{B. } [f']=[f] \equiv\{[f(a_1a_2\ldots a_N)]\equiv[f(a_1^2a_2^2\ldots a_N^2)] \text{ in } SU(N)\}.$$

Results for case A have been tabulated in Ref. 11. Results for case B (diagonal matrix elements) are collected in Table I. The summation relation, in the form (11′), with $[f_1'^*][f_1]=[f_2'^*][f_2]$ for the five types of one- and two-body operators enumerated in Eq. (44) are tabulated as cases 1–9. With $[f_0]=[21^{N-2}]$, it is possible to have $[f_1'^*][f_1]\neq[f_2'^*][f_2]$. There are ten such possibilities for one- and two-body operators which are tabulated as cases 10–19 in Table I.

TABLE I. The sums

$$\sum_\rho \frac{U([f][f_1'^*][f][f_1];[f_1''][f_0]\rho)}{U([f][f_1'^*][f][f_1'];[f_1''][0])}\times\frac{U([f][f_2'^*][f][f_2'];[f_2''][f_0]\rho)}{U([f][f_2'^*][f][f_2'];[f_2''][0])}$$

where

$$\Pi_a=\prod_{\substack{i\neq a\\ i\neq b}}(1-1/\tau_{ai}),\quad \Pi_b=\prod_{\substack{i\neq a\\ i\neq b}}(1-1/\tau_{bi}).$$

$[f_1''][f_2'']$	1. $[f_1'^*][f_1]=[f_2'^*][f_2]=[1^{N-1}][1];[f_0]=[21^{N-2}]$
$[f(a)][f(a)]$	$\left\{N\dfrac{\tau_{ab}}{(\tau_{ab}-1)\Pi_a}-1\right\}$
$[f(a)][f(b)]$	-1
$[f_1''][f_2'']$	2. $[f_1'^*][f_1]=[f_2'^*][f_2]=[1^{N-2}][1^2];\ [f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\dfrac{N(N-1)}{2(N-2)}\left\{\left(1+\dfrac{1}{\tau_{ab}}\right)\dfrac{1}{\Pi_b}+\left(1-\dfrac{1}{\tau_{ab}}\right)\dfrac{1}{\Pi_a}-\dfrac{4}{N}\right\}$
$[f(ac)][f(ab)]$	$\dfrac{N(N-1)}{2(N-2)}\left\{\left(1-\dfrac{1}{\tau_{ac}}\right)\dfrac{1}{\Pi_a}-\dfrac{4}{N}\right\}$
$[f(cd)][f(ab)]$	$-\dfrac{2(N-1)}{(N-2)}$
$[f_1''][f_2'']$	3. $[f_1'^*][f_1]=[f_2'^*][f_2]=[1^{N-2}][1^2];\ [f_0]=[2^21^{N-4}]$

Table I (*continued*)

$[f(ab)][f(ab)]$	$\frac{N}{(N-2)}\left\{1+\frac{(N-1)(N-2)}{2\Pi_a\Pi_b}-\frac{(N-1)}{2}\times\left[\left(1+\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_b}+\left(1-\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_a}\right]\right\}$
$[f(ac)][f(ab)]$	$\frac{N(N-1)}{2(N-2)}\left\{\frac{2}{(N-1)}-\frac{(\tau_{ac}-1)}{\tau_{ac}\Pi_a}\right\}$
$[f(cd)][f(ab)]$	$\frac{N}{(N-2)}$
$[f_1'][f_2']$	4. $[f_1^*][f_1]=[f_2^*][f_2]=[2^{N-1}][2];\ [f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{N(N+1)}{2(N+2)}\left\{\frac{(\tau_{ab}-1)^2}{\tau_{ab}(\tau_{ab}+1)}\frac{1}{\Pi_b}+\frac{(\tau_{ab}+1)^2}{\tau_{ab}(\tau_{ab}-1)}\frac{1}{\Pi_a}-\frac{4}{N}\right\}$
$[f(aa)][f(aa)]$	$\frac{2N(N+1)}{(N+2)}\left\{\frac{\tau_{ab}}{(\tau_{ab}-1)}\frac{1}{\Pi_a}-\frac{1}{N}\right\}$
$[f(aa)]$ or $\begin{matrix}[f(bb)]\\ [f(bc)]\end{matrix}$	$-\frac{2(N+1)}{(N+2)}$
$[f(aa)][f(ab)]$	$\frac{N(N+1)}{(N+2)}\left\{\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\frac{1}{\Pi_a}-\frac{2}{N}\right\}$
$[f(ab)][f(ac)]$	$\frac{N(N+1)}{2(N+2)}\left\{\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1+\frac{1}{\tau_{ac}}\right)\frac{1}{\Pi_a}-\frac{4}{N}\right\}$
$[f(ab)][f(cd)]$	$-\frac{2(N+1)}{(N+2)}$
$[f_1'][f_2']$	5. $[f_1][f_1]=[f_2][f_2]=[2^{N-1}][2];\ [f_0]=[42^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{N(N+1)}{2}\left\{\frac{1}{\Pi_a\Pi_b}+\frac{2}{(N+1)(N+2)}-\frac{1}{(N+2)}\left[\frac{(\tau_{ab}-1)^2}{\tau_{ab}(\tau_{ab}+1)}\frac{1}{\Pi_b}+\frac{(\tau_{ab}+1)^2}{\tau_{ab}(\tau_{ab}-1)}\frac{1}{\Pi_b}\right]\right\}$
$[f(aa)][f(aa)]$	$\frac{N}{(N+2)}\left\{1-2(N+1)\frac{\tau_{ab}}{(\tau_{ab}-1)}\frac{1}{\Pi_a}+\frac{(N+1)(N+2)}{2}\frac{\tau_{ab}}{(\tau_{ab}-2)}\prod_{\substack{i\neq a\\ i\neq b}}\frac{\tau_{ai}}{(\tau_{ai}-2)}\right\}$
$[f(aa)]$ or $\begin{matrix}[f(bb)]\\ [f(bc)]\end{matrix}$	$\frac{N}{(N+2)}$
$[f(aa)][f(ab)]$	$\frac{N}{(N+2)}\left\{1-\frac{(N+1)(\tau_{ab}+1)}{\Pi_a(\tau_{ab}-1)}\right\}$
$[f(ab)][f(ac)]$	$\frac{N}{(N+2)}\left\{1-\frac{(N+1)}{2}\frac{(\tau_{ab}+1)(\tau_{ac}+1)}{(\tau_{ab}-1)\tau_{ac}\Pi_a}\right\}$
$[f(ab)][f(cd)]$	$\frac{N}{(N+2)}$
$[f_1'][f_2']$	6. $[f_1^*][f_1]=[f_2^*][f_2]=[2^{N-1}][1^2];[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{(N+1)}{2}\left\{\left(1+\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_a}+\left(1-\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_b}\right\}$
$[f(ac)][f(ab)]$	$\frac{(N+1)}{2}\left[\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(cd)][f(ab)]$	0
$[f_1''][f_2'']$	7. $[f_1'^*][f_1]=[f_2'^*][f_2]=[2^{N-1}][1^2];\ [f_0]=[3^22^{N-3}]$
$[f(ab)][f(ab)]$	$\frac{N(N+1)}{2\Pi_a\Pi_b}\left\{1-\frac{1}{N}\left[\left(1-\frac{1}{\tau_{ab}}\right)\Pi_a+\left(1+\frac{1}{\tau_{ab}}\right)\Pi_b\right]\right\}$
$[f(ac)][f(ab)]$	$-\frac{(N+1)}{2}\left[\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(cd)][f(ab)]$	0

Table I (*continued*)

$[f_1'][f_2']$	8. $[f_1^*][f_1]=[f_2^*][f_2]=[1^{N-2}][2];\ [f_0]=[21^{N-2}]$.
$[f(ab)][f(ab)]$	$\frac{(N-1)}{2}\left\{\left(1+\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_a}+\left(1-\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_b}\right\}$
$[f(ac)][f(ab)]$	$\frac{(N-1)}{2}\left[\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(cd)][f(ab)]$	0
$[f_1'][f_2']$	9. $[f_1^*][f_1]=[f_2^*][f_2]=[1^{N-2}][2];\ [f_0]=[31^{N-3}]$
$[f(ab)][f(ab)]$	$\frac{N(N-1)}{2\Pi_a\Pi_a}\left\{1-\frac{1}{N}\left[\left(1-\frac{1}{\tau_{ab}}\right)\Pi_a+\left(1+\frac{1}{\tau_{ab}}\right)\Pi_b\right]\right\}$
$[f(ac)][f(ab)]$	$-\frac{(N-1)}{2}\left[\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(cd)][f(ab)]$	0
$[f_1'][f_2']$	10. $[f_1^*][f_1]=[1^{N-1}][1];\ [f_2^*][f_2]=[1^{N-2}][1^2];\ [f_0]=[21^{N-2}]$
$[f(a)][f(ab)]$	$N\left[\frac{N-1}{2(N-2)}\right]^{1/2}\left\{\frac{1}{\Pi_a}-\frac{2}{N}\right\}$
$[f(a)][f(bc)]$	$-\left[\frac{2(N-1)}{(N-2)}\right]^{1/2}$
$[f_1'][f_2']$	11. $[f_1^*][f_1]=[1^{N-1}][1];\ [f_2^*][f_2]=[2^{N-1}][2];\ [f_0]=[21^{N-2}]$
$[f(a)][f(aa)]$	$-\left[\frac{2(N+1)}{(N+2)}\right]^{1/2}\left\{\frac{N\tau_{ab}}{(\tau_{ab}-1)\Pi_a}-1\right\}$
$[f(a)][f(ab)]$	$-\left[\frac{N+1}{2(N+2)}\right]^{1/2}\left\{\frac{N(\tau_{ab}+1)}{(\tau_{ab}-1)\Pi_a}-2\right\}$
$[f(a)]$ or $\begin{matrix}[f(bb)]\\ [f(bc)]\end{matrix}$	$\left[\frac{2(N+1)}{(N+2)}\right]^{1/2}$
$[f_1'][f_2']$	12. $[f_1^*][f_1]=[1^{N-1}][1];\ [f_2^*][f_2]=[2^{N-1}][1^2];\ [f_0]=[21^{N-2}]$
$[f(a)][f(ab)]$	$-\left[\frac{N(N+1)}{2}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\frac{1}{\Pi_a}$
$[f(a)][f(bc)]$	0
$[f_1'][f_2']$	13. $[f_1^*][f_1]=[1^{N-1}][1];\ [f_2^*][f_2]=[1^{N-2}][2];\ [f_0]=[21^{N-2}]$
$[f(a)][f(ab)]$	$\left[\frac{N(N-1)}{2}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\frac{1}{\Pi_a}$
$[f(a)][f(bc)]$	0
$[f_1'][f_2']$	14. $[f_1^*][f_1]=[1^{N-2}][1^2];\ [f_2^*][f_2]=[2^{N-1}][2];\ [f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$-\left[\frac{N}{2}\frac{(N+1)(N-1)}{(N+2)(N-2)}\right]^{1/2}\left\{\left(1+\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_a}+\left(1-\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_b}-\frac{4}{N}\right\}$
$[f(ab)][f(ac)]$	$-\frac{N}{2}\left[\frac{(N+1)(N-1)}{(N+2)(N-2)}\right]^{1/2}\left\{\left(1+\frac{1}{\tau_{ac}}\right)\frac{1}{\Pi_a}-\frac{4}{N}\right\}$
$[f(ab)[f(aa)]$	$-N\left[\frac{(N+1)(N-1)}{(N+2)(N-2)}\right]^{1/2}\left\{\frac{1}{\Pi_a}-\frac{2}{N}\right\}$
$[f(ab)]$ or $\begin{matrix}[f(cc)]\\ [f(cd)]\end{matrix}$	$2\left[\frac{(N+1)(N-1)}{(N+2)(N-2)}\right]^{1/2}$

Table I (*continued*)

$[f_1'{}'][f_2'{}']$	15. $[f_1'{}^*][f_1]=[1^{N-2}][1^2]$; $[f_2'{}^*][f_2]=[2^{N-1}][1^2]$; $[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{1}{2}\left[\frac{N(N+1)(N-1)}{(N-2)}\left(1-\frac{1}{\tau_{ab}^2}\right)\right]^{1/2}\left\{\frac{1}{\Pi_b}-\frac{1}{\Pi_a}\right\}$
$[f(ab)][f(ac)]$	$-\frac{1}{2}\left[\frac{N(N+1)(N-1)}{(N-2)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(ab)][f(cd)]$	0
$[f_1'{}'][f_2'{}']$	16. $[f_1'{}^*][f_1]=[1^{N-2}][1^2]$; $[f_2'{}^*][f_2]=[1^{N-2}][2]$; $[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$-\frac{(N-1)}{2}\left[\frac{N}{(N-2)}\left(1-\frac{1}{\tau_{ab}^2}\right)\right]^{1/2}\left\{\frac{1}{\Pi_b}-\frac{1}{\Pi_a}\right\}$
$[f(ab)][f(ac)]$	$\frac{(N-1)}{2}\left[\frac{N}{(N-2)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(ab)][f(cd)]$	0
$[f_1'{}'][f_2'{}']$	17. $[f_1'{}^*][f_1]=[2^{N-1}][2]$; $[f_2'{}^*][f_2]=[2^{N-1}][1^2]$; $[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{(N+1)}{2}\left[\frac{N}{(N+2)}\left(1-\frac{1}{\tau_{ab}^2}\right)\right]^{1/2}\left\{\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\frac{1}{\Pi_a}-\frac{(\tau_{ab}-1)}{(\tau_{ab}+1)}\frac{1}{\Pi_b}\right\}$
$[f(aa)][f(ab)]$	$(N+1)\left[\frac{N}{(N+2)}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\frac{1}{\Pi_a}$
$[f(ac)][f(ab)]$	$\frac{(N+1)}{2}\left[\frac{N}{(N+2)}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\left(1+\frac{1}{\tau_{ac}}\right)\frac{1}{\Pi_a}$
$[f(cc)]$ or $[f(cd)]$ $[f(ab)]$	0
$[f_1'{}'][f_2'{}']$	18. $[f_1'{}^*][f_1]=[2^{N-1}][2]$; $[f_2'{}^*][f_2]=[1^{N-2}][2]$; $[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$\frac{1}{2}\left[\frac{N(N+1)(N-1)}{(N+2)}\left(1-\frac{1}{\tau_{ab}^2}\right)\right]^{1/2}\left\{\frac{(\tau_{ab}-1)}{(\tau_{ab}+1)}\frac{1}{\Pi_b}-\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\frac{1}{\Pi_a}\right\}$
$[f(aa)][f(ab)]$	$-\left[\frac{N(N+1)(N-1)}{(N+2)}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\frac{1}{\Pi_a}$
$[f(ac)][f(ab)]$	$-\frac{1}{2}\left[\frac{N(N+1)(N-1)}{(N+2)}\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\right]^{1/2}\times\left(1+\frac{1}{\tau_{ac}}\right)\frac{1}{\Pi_a}$
$[f(cc)]$ or $[f(cd)]$ $[f(ab)]$	0
$[f_1'{}'][f_2'{}']$	19. $[f_1'{}^*][f_1]=[2^{N-1}][1^2]$; $[f_2'{}^*][f_2]=[1^{N-2}][2]$; $[f_0]=[21^{N-2}]$
$[f(ab)][f(ab)]$	$-\frac{1}{2}[(N+1)(N-1)]^{1/2}\left\{\left(1+\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_a}+\left(1-\frac{1}{\tau_{ab}}\right)\frac{1}{\Pi_b}\right\}$
$[f(ac)][f(ab)]$	$-\frac{1}{2}\left[(N+1)(N-1)\frac{(\tau_{ab}+1)}{(\tau_{ab}-1)}\left(1-\frac{1}{\tau_{ac}^2}\right)\right]^{1/2}\frac{1}{\Pi_a}$
$[f(cd)][f(ab)]$	0

*Supported by U. S. National Science Foundation.

[1]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).

[2]L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13**, 1957 (1972).

[3]L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13**, 1985 (1972).

[4]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).

[5]E. Chacón, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13**, 577 (1972).

[6]J. D. Louck, Am. J. Phys. **38**, 3 (1970), presents a bibliography of earlier references.

[7]J. P. Draayer and Y. Akiyama, J. Math. Phys. **14**, 1904 (1973).

[8]Y. Akiyama and J. P. Draayer, Comput. Phys. Commun. **5**, 405 (1973).

[9]M. Moshinsky and E. Chacón, in *Spectroscopic and Group Theoretical Methods in Phyics*, edited by F. Bloch, *et al.* (North-Holland, Amsterdam, 1968).

[10]S. J. Ališauskas, A.-A. A. Jucys, and A. P. Jucys, J. Math. Phys. **13**, 1329 (1972).

[11]K. T. Hecht and J. P. Draayer, Nucl. Phys. A **223**, 285 (1974).

[12]F. S. Chang, J. B. French, and T. H. Thio, Ann. Phys. (N. Y.) **66**, 137 (1971).

[13]M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962), Chap. 7.

[14]I. G. Kaplan, Zh. Eksp. Teor. Fiz. **41**, 560, 790 (1961) [Sov. Phys. JETP **14**, 401, 568 (1962)].

[15]Hisashi Horie, J. Phys. Soc. Japan **19**, 1783 (1964).

[16]D. E. Rutherford, *Substitutional Analysis* (Edinburgh, U. P., Edinburgh, 1948).

Bäcklund transformations for certain nonlinear evolution equations

G. L. Lamb Jr.*

United Aircraft Research Laboratories, East Hartford, Connecticut 06108
(Received 10 April 1974)

Bäcklund transformations associated with the Korteweg–deVries (KdV), modified KdV, and nonlinear Schrödinger equations are derived by a method due to Clairin. Also, a Bäcklund transformation relating the KdV and modified KdV equations is obtained by the same technique.

1. INTRODUCTION

A number of nonlinear evolution equations such as the Korteweg—deVries (hereafter abbreviated KdV) equation, modified KdV equation, nonlinear Schrödinger equation, and sine—Gordon equation are known to share many remarkable properties.[1] Some understanding of their similarities has recently emerged with the discovery that these equations are members of a class of equations that can be solved by an inverse procedure.[2] Furthermore, it has been noted[3] that the pair of linear equations that are introduced in the course of effecting the solution by the inverse method are transformable to the Bäcklund transformations that are now known to be associated with certain of the evolution equations. Conversely, the Bäcklund transformations for the above-mentioned evolution equations each contain an equation with Riccati type nonlinearity. If one refrains from transforming these Riccati equations to second-order linear equations and follows the alternate procedure of replacing them by a pair of linear first-order equations, one finds that the resultant equations are of the type first introduced by Zakharov and Shabat[4] in their application of the inverse method to the nonlinear Schrödinger equation. It is the relevance of such linear equations to a number of other nonlinear evolution equations that has since been pointed out by Ablowitz *et al.*[2]

In the present paper it is shown that the relatively archaic notion of Bäcklund transformations[5] is a useful intermediate step in analyzing the currently popular nonlinear evolution equations and should be useful in developing an understanding of equations to arise in the future.

The derivation of the Bäcklund transformation also shows quite incidentally why modified KdV equations of the form $u_y + u^n u_x + u_{xxx} = 0$ with $n > 2$ can be expected to lie beyond the class of equations that admit of Bäcklund transformations of Riccati type with their associated linear eigenvalue problem and well-known spectral theory. The lack of soliton behavior exhibited by numerical solutions of these equations is undoubtedly another manifestation of this fact. Nonlinear partial differential equations with Bäcklund transformations that are beyond Riccati type in their nonlinear character can presumably be expected to yield to a corresponding analysis whenever a spectral theory can be associated with the relevant transformation equations.

Bäcklund transformations arose long ago in the study of surfaces in classical differential geometry.[6] Since this topic has been out of fashion for some time, a brief summary of the significance of the term Bäcklund transformation, and in particular the distinction between a Bäcklund transformation and a contact transformation, is perhaps in order.

Consider the two sets of variables related by

$$p = f(x', y', z', p', q'), \tag{1.1a}$$

$$q = \varphi(x', y', z', p', q') \tag{1.1b}$$

with $x = x'$, $y = y'$. (The notation $p = \partial z/\partial x$ and $q = \partial z/\partial y$ will be employed throughout the following development with a similar notation for the primed variables. The second derivatives introduced in the sequel will be denoted by $r = \partial^2 z/\partial x^2$, $s = \partial^2 z/\partial x \partial y$, and $t = \partial^2 z/\partial y^2$.) An explicit dependence upon z could be included in the functions f and φ (and indeed will be required in applications of the transformation theory presented in the body of this paper) but is unnecessary for present considerations.

The integrability condition for z requires $dp/dy = dq/dx$. This may be written as

$$\Omega = f_{y'} - \varphi_{x'} + f_{z'} q' - \varphi_{z'} p' + (f_{p'} - \varphi_{q'}) s' + f_{q'} t' + \varphi_{p'} r' = 0, \tag{1.2}$$

where subscripts indicate partial differention with respect to the indicated variable. The integrability requirement may be satisfied in either of two ways: either identically, in which case $f_{p'} - \varphi_{q'} = f_{q'} = \varphi_{p'} = 0$ and

$$f_{y'} - \varphi_{x'} + f_{z'} q' - \varphi_{z'} p' = 0 \tag{1.3}$$

or, since $\Omega = 0$ is actually a second-order equation of Monge—Ampere form, it may be satisfied by virtue of the fact that z' is a solution of this second-order equation. In the former case the transformation is a contact transformation while in the latter case it is a Bäcklund transformation. The reduction of the former case to the canonical form for a contact transformation is presented in detail by Forsyth.[7] The fact that the Monge—Ampere equation that arises in the latter case is not of the most general form, so that not all such equations can be expected to have Bäcklund transformations associated with them, may also be shown.[7] Needless to say, pairs of transformation equations that contain derivatives higher than the first may satisfy the integrability condition provided z' satisfies an appropriate higher order equation. Much of the following analysis will be addressed to the transformation of third-order equations. A straightforward extension of Eqs. (1) to include appropriate higher derivatives is therefore employed.

Liouville's equation $s = e^z$ provides a simple example of an equation for which the general solution may be obtained quite easily by means of a Bäcklund transformation. This equation has associated with it the Bäcklund transformation

$$p = p' - a\exp[\tfrac{1}{2}(z + z')], \tag{1.4a}$$

$$q = -q' - 2a^{-1}\exp[\tfrac{1}{2}(z - z')], \quad a = \text{const.} \tag{1.4b}$$

The integrability condition for z is satisfied if $s = e^z$ while the integrability condition for z' is satisfied if $s' = 0$. Also, elimination of z' between Eqs. (1.4) shows that z satisfies Liouville's equation while elimination of z shows that z' satisfies $s' = 0$. When the general solution of this latter equation is inserted into Eqs. (1.4), the solutions of the resulting first-order equations readily yield the general solution of Liouville's equation.

The sine–Gordon equation, which in characteristic coordinates may be written

$$s = \sin z, \tag{1.5}$$

has associated with the Bäcklund transformation,

$$\tfrac{1}{2}(p - p') = a\sin[\tfrac{1}{2}(z + z')], \tag{1.6a}$$

$$\tfrac{1}{2}(q + q') = a^{-1}\sin[\tfrac{1}{2}(z - z')], \tag{1.6b}$$

where a is an arbitrary constant. In this instance, however, z and z' both satisfy an equation of the form of Eq. (1.5). Although Eqs. (1.6) do not therefore lead to a simpler equation for which the general solution is known, they have been found useful. They lead to a so-called "theorem of permutability" by means of which four *particular* solutions of Eq. (1.5) are interrelated. This result enables one to obtain an infinite sequence of particular solutions to the sine–Gordon equation without additional use of quadrature. These particular solutions have been found useful in the many physical applications of Eq. (1.5). Only recently,[8] the pair of linear equations that can be related to either of Eqs. (1.6) have been solved by an inverse method to obtain the *general* solution to Eq. (1.5). The particular solutions obtained from the theorem of permutability are equivalent to the pure multisoliton solutions produced by the inverse method.

Of the nonlinear evolution equations under study at present, only the sine–Gordon equation, which is of Monge–Ampere form, was originally[5] known to possess a Bäcklund transformation which transforms the equation into itself. More recently such transformations have been found for the KdV [9] and modified KdV [10–12] equations. As noted above, they have also been obtained from the linear equations of the inverse method.[3]

Since the KdV and modified KdV equations are of third order, an extension of the usual Bäcklund transformation theory is necessary. The extension is minor, although the details of the calculation are somewhat tedious. One finds that the Bäcklund transformation for the third order equation consists of an equation of first order plus an equation of second order. The first-order equation is of Riccati form. In the present paper these transformations as well as one for the nonlinear Schrödinger equation are derived by a method due to Clairin.[13] Detailed derivation of the Bäcklund transformation for the sine–Gordon equation by this method has already appeared[14,15] and will not be repeated here.

The sine–Gordon equation has served as a useful guide in constructing Bäcklund transformations for some of the equations to be considered below; the following aspect of this equation should be noted: Pulse profiles are obtained as *derivatives* of solutions to the sine–Gordon equation.[16] In this instance then, the equation that is transformed into itself by a Bäcklund transformation is the equation for the *integral* over a pulse profile. In the search for such transformations for other nonlinear evolution equations, it has been found useful to exploit this result. Hence, in the following, Bäcklund transformations are actually found for equations satisfied by integrals over pulse profiles. The pulse profiles themselves, of course, satisfy the KdV and modified KdV equations. This procedure has not been found useful in dealing with the nonlinear Schrödinger equation, however, and in that case a Bäcklund transformation for the nonlinear Schrödinger equation itself is obtained.

The construction of the various Bäcklund transformations presented in the following sections has been found to depend upon the solution of overdetermined sets to equations (e.g., ten equations to be satisfied by six unknown functions in the case of the KdV equation). This is generally the case.[7] Hence, the equations that possess such transformations are of a somewhat special character. Why so many of the equations that describe nonlinear evolution processes happen to belong to this class is an apparently unanswered question at the present time. Furthermore, the results obtained are characterized by extensive specialization and hence leave open the question of the most general transformations that could be associated with these equations. Answers to such questions should be forthcoming when a more general understanding of Bäcklund transformations in nonlinear pulse propagation has been developed. Recent algebraic results by Gerber[17] and the interpretation of Bäcklund transformations within the framework of differential forms by Wahlquist and Estabrook[18] may provide initial steps toward such understanding.

To exhibit the relation between a Bäcklund transformation and the inverse method, the equivalence of the well-known Bäcklund transformation for the sine–Gordon equation and the linear equations of the inverse method is summarized here.[3] A similar analysis will be applied to each of the Bäcklund transformations to be obtained below.

Defining $\Gamma = \tan[(z + z')/4]$, one finds that Eq. (1.6a) becomes

$$\Gamma_x + a\Gamma - \tfrac{1}{2}p(1 + \Gamma^2) = 0. \tag{1.7}$$

Now, the Riccati equation

$$\Gamma_x + 2P\Gamma + Q\Gamma^2 + R = 0 \tag{1.8}$$

is equivalent (although not uniquely) to the pair of linear equations

$$w_{1x} + Pw_1 = -Rw_2, \tag{1.9a}$$

$$w_{2x} - Pw_2 = Qw_1, \tag{1.9b}$$

where $\Gamma = w_1/w_2$. Hence Eq. (1.7) may be replaced by

$$w_{1x} + \tfrac{1}{2}aw_1 = \tfrac{1}{2}pw_2, \tag{1.10a}$$

$$w_{2x}-\tfrac{1}{2}aw_2=-\tfrac{1}{2}pw_1, \tag{1.10b}$$

which have the structure of the equations solved previously by the two-component inverse method.

As a final example of Bäcklund transformation theory in nonlinear pulse propagation, a Bäcklund transformation relating the KdV to the modified KdV equation is developed. The expectation that such a transformation should exist is based upon a previous result of Miura.[19] Clairin's method has been used to construct the equation that must be adjoined to the Miura transformation to obtain a Bäcklund transformation.

2. KORTEWEG-DEVRIES EQUATION

The KdV equation will be written in the form

$$u_y+6uu_x+u_{xxx}=0. \tag{2.1}$$

To obtain an equation analogous to the sine–Gordon equation, one introduces a function representing an integral over the pulse profiles that satisfy Eq. (2.1). Setting

$$z(x,y)=\int_{-\infty}^{x}dx'u(x',y), \tag{2.2}$$

one finds that z satisfies

$$q+3p^2+\alpha=0. \tag{2.3}$$

Since interest centers around localized solutions, the arbitrary function arising from integration may be set equal to zero.

A functional form must now be chosen for the two lower order equations that will play the role of Bäcklund transformations for Eq. (2.3). Guided by the symmetry between z and z' in the Bäcklund transformations for the sine–Gordon equation, one may begin by incorporating such symmetry at the beginning of the present calculation. A bit of reflection as well as familiarity with the result of Clairin's work shows that a possible choice is

$$p=f(z,z',p), \tag{2.4a}$$

$$q=\tilde{\varphi}(z,z',q',r,r',p,p'). \tag{2.4b}$$

However, since

$$r=f_zf+f_{z'}p'+f_{p'}r', \tag{2.5}$$

one sees that Eq. (2.4b) is actually equivalent to

$$q=\varphi(z,z',q',p',r'). \tag{2.6}$$

The mixed second derivative of z may now be written in either of the forms

$$\frac{dp}{dy}=f_zq+f_{z'}q'+f_{p'}s' \tag{2.7}$$

or

$$\frac{dq}{dx}=\varphi_zp+\varphi_{z'}p'+\varphi_{q'}s'+\varphi_{p'}r'+\varphi_{r'}\alpha'. \tag{2.8}$$

One now requires the equality of these mixed derivatives as well as that z' satisfy Eq. (2.3). This is conveniently expressed by defining a function $\Omega(z,z',p,p',q,q',r',s')$ such that

$$\Omega=(f_{p'}-\varphi_{q'})s'+f_zq+f_{z'}q'-\varphi_zp-\varphi_{z'}p'-\varphi_{p'}r'+\varphi_{r'}(q'+3p'^2)=0. \tag{2.9}$$

Now

$$\Omega_{s'}=f_{p'}-\varphi_{q'}=0. \tag{2.10}$$

Since f is independent of q' and r', Eq. (2.10) also implies that

$$\varphi_{q'q'}=\varphi_{q'r'}=0 \tag{2.11}$$

Also,

$$\Omega_{q'}=f_{p'}f_z+f_{z'}-ff_{p'z}-p'f_{p'z'}-r'f_{p'p'}+\varphi_{r'}=0 \tag{2.12}$$

as well as

$$\Omega_{q'r'}=-f_{p'p'}+\varphi_{r'r'}=0. \tag{2.13}$$

From a consideration of the functional dependence allowed in f and φ by Eqs. (2.4), Eq. (2.13) implies

$$f_{p'p'}=\varphi_{r'r'}=a(z,z',p'), \tag{2.14}$$

where $a(z,z',p')$ is unknown. However, if this function is nonzero, a nonlinear dependence of f upon p' would result. While this might well lead to a valid transformation, it would destroy the expected symmetry of Eq. (2.4a) (as well as the Riccati–type nonlinearity anticipated for the result). Hence the subsequent analysis is restricted to the case in which $a(z,z',p')$ is assumed to vanish. (It is necessarily independent of p' since $\Omega_{r'r'r'}=-3f_{p'p'p'}=0$.)

Equations (2.14) then yield

$$f(z,z',p')=b(z,z')p'+c(z,z'), \tag{2.15a}$$

$$\varphi(z,z',q',p',r')=b(z,z')q'+\lambda(z,z',p')r'+\nu(z,z',p'), \tag{2.15b}$$

in which Eq. (2.11) has again been employed as well as Eq. (2.10). The functions c, λ, and ν arise in the integration process and are to be determined.

Returning to Eq. (2.9), one also finds that $\Omega_{r'r'}=-2\varphi_{r'p'}=0$ so that λ must be independent of p'. Also, one finds that $\Omega_{p'p'p'}=0$, which yields the following form for $\nu(z,z',p')$ in Eq. (2.15b):

$$\nu(z,z',p')=\nu_2(z,z')p'^2+\nu_1(z,z')p'+\nu_0(z,z'). \tag{2.16}$$

The subsequent analysis becomes extremely cumbersome if the general form for $b(z,z')$ is retained. Fortunately, this is unnecessary. Useful results are obtained by assuming $b(z,z')$ to be a constant.

The structure of the Bäcklund transformation has thus been reduced to

$$p=bp'+c, \tag{2.17a}$$

$$q=bq'+\lambda r'+\nu_2p'^2+\nu_1p'+\nu_0, \tag{2.17b}$$

where b is a constant while c, λ, and the three ν_i are as yet undetermined functions of z and z'.

Substituting Eqs. (2.17) into Ω and equating to zero the coefficient of each of the various dependences upon r', p', and q' in Ω (since these are independent), one obtains the following set of seven equations that must be satisfied by the five unknown functions and the constant b:

$$2\nu_2 = -(b\lambda_z + \lambda_{z'}), \tag{2.18a}$$

$$\lambda = -(bc_z + c_{z'}), \tag{2.18b}$$

$$\nu_1 = \lambda c_z - c\lambda_z, \tag{2.18c}$$

$$\nu_2 c_z - c\nu_{2z} + 3\lambda - b\nu_{1z} - \nu_{1z'} = 0, \tag{2.18d}$$

$$\nu_1 c_z - c\nu_{1z} - \nu_{0z'} - b\nu_{0z} = 0, \tag{2.18e}$$

$$b\nu_{2z} + \nu_{2z'} = 0, \tag{2.18f}$$

$$\nu_0 c_z - c\nu_{0z} = 0. \tag{2.18g}$$

In writing Eq. (2.9) for Ω the requirement that z' satisfy Eq. (2.3) has already been introduced. It must also be required that z satisfy this same equation. Calculation of the third derivative from Eq. (2.17a) shows that

$$\alpha = b\alpha' - \lambda r' + 2\nu_2 p'^2 + p'[2bcc_{zz} + 2cc_{zz'} + c_z(bc_z + c_{z'})] + c^2 c_{zz} + cc_z^2, \tag{2.19}$$

where the definition of ν_2 from Eqs. (2.18a, b) has been employed. The requirement that Eq. (2.3) also be satisfied by z now lead to

$$\nu_2 - b + b^2 = 0, \tag{2.20a}$$

$$bc_{zz} + 2b + c_{zz'} = 0, \tag{2.20b}$$

$$c^2 c_{zz} + cc_z^2 + \nu_0 + 3c^2 = 0. \tag{2.20c}$$

Equations (2.18) and (2.20) are the complete set of equations that must be satisfied. Such overdetermined systems are a characteristic of calculations dealing with Bäcklund transformations.

Equation (2.18g) implies that ν_0 may be written

$$\nu_0 = \psi(z')\, c(z, z'), \tag{2.21}$$

where $\psi(z')$ is to be determined. A first integral for Eq. (2.20c) may be obtained at this point. One finds

$$c_z^2 + 2c + \psi + Kc^{-2} = 0. \tag{2.22}$$

The consideration of elliptic functions may be avoided by setting $K = 0$. Equation (2.22) now implies

$$c_{zz} = -1. \tag{2.23}$$

From Eq. (2.20b),

$$c_{zz'} = -b, \tag{2.24}$$

and, finally, Eqs. (2.20a) plus (2.18a, b) yield

$$c_{z'z'} = 2b + b^2. \tag{2.25}$$

Integration of these last three equations yields

$$c(z, z') = m - \tfrac{1}{2}[z^2 + 2bzz' - b(2+b)z'^2] + kz + lz', \tag{2.26}$$

where k, l, and m are constants of integration. The choice of these constants is quite critical for the subsequent analysis. From Eq. (2.17a) one sees that solutions which are to vanish at $x = -\infty$, as is required by Eq. (2.2), will require $m = 0$. However, it is the derivative of z that yields u the solution of the KdV equation. Solutions with $m \neq 0$ may be used for this. It is such a solution that leads to previously reported results.[9] This point will be taken up again in connection with the theorem of permutability. The present calculation will be carried out for $m \neq 0$, $k = l = 0$.

From Eqs. (2.18) one now obtains

$$\lambda = 2b(z - z'), \tag{2.27a}$$

$$\nu_1 = -2bm - b(z^2 - 2zz' + b^2 z'^2), \tag{2.27b}$$

$$\nu_2 = b - b^2, \tag{2.27c}$$

$$\psi = -2m - 2b(1+b)z'^2. \tag{2.27d}$$

Equations (2.1a, b, c, f, g) and Eqs. (2.20) are now found to be satisfied identically while Eqs. (2.18d, e) require $b = -1$. The form of the Bäcklund transformation has now been completely determined and agrees with that of Wahlquist and Estabrook.[9] The result also takes the completely symmetric form[20]

$$p + p' = m - \tfrac{1}{2}(z - z')^2, \tag{2.28a}$$

$$q + q' = (z - z')(r - r') - 2(p^2 + pp' + p'^2). \tag{2.28b}$$

A. Theorem of permutability

As noted in the introduction, the Bäcklund transformation for the sine–Gordon equation was used long ago[6] to derive a theorem of permutability. This is a relation among the solutions that permits the construction of an infinite sequence of additional solutions without additional quadrature. For coherent optical pulse profiles, this result has been used[16] to construct solutions for $2n\pi$ pulses ($n = 0, 1, 2, \ldots$).

A similar situation exists in the case of other equations for which a Bäcklund transformation has been obtained. The result for the KdV equation was given by Wahlquist and Estabrook.[9] Their results are briefly summarized here along with an application to a particular example.

First, a single soliton solution may be obtained by noting that $z' = 0$ is a solution to the KdV equation. Then, using this result as z' in Eqs. (2.28), one finds that z satisfies

$$p = m - \tfrac{1}{2}z^2, \tag{2.29a}$$

$$q = zr - 2p^2 = -2mp. \tag{2.29b}$$

A solution of these equations is

$$z = (2m)^{1/2}\tanh[(m/2)^{1/2}(x - 2my)]. \tag{2.30}$$

As noted above, the nonzero value of m has led to a solution for z that does not vanish for $x = -\infty$. The corresponding solution of the KdV equation is

$$u = p = m\,\mathrm{sech}^2[(m/2)^{1/2}(x - 2my)]. \tag{2.31}$$

Unlike the sine–Gordon equation, Eq. (2.3) has solutions that are divergent. Equations (2.29) also have the solution

$$z^* = (2m)^{1/2}\coth[(m/2)^{1/2}(x - 2my)]. \tag{2.32}$$

In using the theorem of permutability to construct multisoliton solutions of the KdV equation, care must be taken to exclude such divergent results. Perhaps paradoxically, exclusion of divergent multisoliton expressions is carried out by a judicious usage of the divergent solution z^* given above.[9]

To obtain the theorem of permutability, one notes that Eq. (2.28a) may be interpreted as a transformation from a known solution z' to a new solution z_m which contains the constant m. Four such transformations may be written

$$p_{m_1}+p'=m_1-\tfrac{1}{2}(z_{m_1}-z')^2 \qquad (2.33a)$$

$$p_{m_2}+p'=m_2-\tfrac{1}{2}(z_{m_2}-z')^2, \qquad (2.33b)$$

$$p_{m_1m_2}+p_{m_1}=m_2-\tfrac{1}{2}(z_{m_1m_2}-z_{m_1})^2, \qquad (2.33c)$$

$$p_{m_1m_2}+p_{m_2}=m_1-\tfrac{1}{2}(z_{m_1m_2}-z_{m_2})^2. \qquad (2.33d)$$

In the latter pair of equations the two known solutions are those obtained in the previous two equations. One requires that the two latter transformation equations lead to the same final solution $z_{m_1m_2}$. Note that the same two constants have been used again but they have been interchanged so that m_2 is associated with z_{m_1} and vice versa. Subtracting the second from the first and the fourth from the third and finally eliminating $p_{m_1}-p_{m_2}$ among the two resulting equations yields

$$z_{m_1m_2}-z'=2(m_1-m_2)/(z_{m_1}-z_{m_2}). \qquad (2.34)$$

The possibility of divergence in this result may be removed by noting from Eqs. (2.30) and (2.32) that $z_1<(2m)^{1/2}<z_1^*$.

As an example, note that the choice $z'=0$ and $m_1=8$, $m_2=2$ with the divergent solution used for z_1 leads to

$$z_{m_1m_2}=6(2\coth A-\tanh B)^{-1}, \qquad (2.35)$$

where

$$A=2x-3y, \qquad (2.36a)$$

$$B=x-4y. \qquad (2.36b)$$

Differentiation of Eq. (2.35) then yields

$$u_3=12\,\frac{4\cosh 2B+\cosh 2A+3}{[\cosh(A+B)+3\cosh(A-B)]^2}, \qquad (2.37)$$

which agrees with a previously quoted[21] two-soliton solution for the KdV equation

The choice $m=0$, $l=k$ in Eq. (2.26) would have led to the single soliton result

$$z=k[1+\tanh\tfrac{1}{2}k(x-k^2y)], \qquad (2.38)$$

which has behavior at $x=-\infty$ that is consistent with Eq. (2.2). However, there appears to be no advantage in going to this solution since the theorem of permutability is not as concise as Eq. (2.34) and the divergent solutions are still required.

B. Relation to the inverse method

Since the first of the two Bäcklund transformation equations listed in Eqs. (2.28) is of Riccati type, the linearization procedure outlined in the introduction may be applied. By setting $\Gamma=z-z'$, Eq. (2.28a) becomes

$$\Gamma_x-\tfrac{1}{2}\Gamma^2+(m-2p)=0. \qquad (2.39)$$

Comparison with Eq. (1.8) shows that the appropriate linear system is

$$v_{1x}=(2p-m)v_2, \qquad (2.40a)$$

$$v_{2x}=-\tfrac{1}{2}v_1. \qquad (2.40b)$$

The linear transformation $w_1=-v_1+i\zeta v_2$, $w_2=-v_2$ yields the linear equations treated in Ref. 2.

In the present case the second-order equation is actually more convenient to analyze. One finds

$$v_{2xx}+(p-\tfrac{1}{2}m)v_2=0, \qquad (2.41)$$

where $p=\partial z/\partial x=u$, the solution of the KdV equation. Use of the inverse method to obtain p from this equation is now well known.[1]

3. MODIFIED KORTEWEG-DEVRIES EQUATION

A procedure quite similar to that of the previous section yields Bäcklund transformations associated with the modified KdV equation. This latter equation will be written in the form

$$u_y+6u^2u_x+u_{xxx}=0. \qquad (3.1)$$

The integral of the pulse profile now satisfies

$$q+2p^3+\alpha=0, \qquad (3.2)$$

where z is again related to u as in Eq. (2.2). The general functional forms chosen for the Bäcklund transformation are the same as those of Eqs. (2.4a) and (2.6). One finds that relations of the form of Eqs. (2.17) are again obtained. One may again choose b to be a constant. However, since z' now satisfies Eq. (3.2) instead of Eq. (2.3), Eqs. (2.18d) and (2.18f), which are the coefficients of p'^2 and p'^3 in the calculation of the previous section, must be modified. It is found that they must be replaced by

$$\nu_2c_z-c\nu_{2z}-b\nu_{1z}-\nu_{1z'}=0, \qquad (3.3a)$$

$$b\nu_{2z}+\nu_{2z'}-2\lambda=0. \qquad (3.3b)$$

The requirement that z also satisfy Eq. (3.2) leads to

$$b(b^2-1)=0, \qquad (3.4a)$$

$$\nu_2+2b^2c=0, \qquad (3.4b)$$

$$2bc+c_{zz'}+bc_{zz}=0, \qquad (3.4c)$$

$$cc_{zz}+c_z^2+2c^2+\psi(z')=0. \qquad (3.4d)$$

Equation (3.4a), which results from the cubic term in Eq. (3.2), imposes an additional restriction on the constant b. (It should be noted that the class of Bäcklund transformations being considered here does not have provision for powers of p greater than 3. Hence, modified KdV equations of the form $u_y+u^nu_x+u_{xxx}=0$ with $n>2$ will not have Bäcklund transformations of the "Riccati type" being considered here. The lack of soliton behavior observed in numerical solutions of these equations is undoubtedly related to this fact.) Combination of Eqs. (3.4b) and (3.4c), with ν_2 again obtained from Eqs. (2.18a, b), yields

$$b^2c_{zz}-c_{z'z'}=0. \qquad (3.5)$$

A first integral of the ordinary differential equation in Eq. (3.4d) is

$$c_z^2+c^2+\psi(z')=0, \qquad (3.6)$$

which implies

$$c_{zz}+c=0. \qquad (3.7)$$

Integration of Eqs. (3.7) and (3.5) yields

$$c(z,z')=c_1\exp(iv)+c_2\exp(-iv)+c_3\exp(iw)+c_4\exp(-iw), \qquad (3.8)$$

where

$$v = z + bz', \tag{3.9a}$$

$$w = z - bz', \tag{3.9b}$$

and the c_i are constants of integration. A result of sufficient generality for present purposes[22] is obtained by setting $c_3 = c_4 = 0$, $c_2 = -c_1$. Then

$$c(z, z') = a \sin v, \tag{3.10}$$

where $a = 2ic_1$. The remaining functions in the transformation equations are readily found. One obtains

$$\lambda = -2ab^3 \cos v, \tag{3.11a}$$

$$\nu_0 = -a^3 \sin v, \tag{3.11b}$$

$$\nu_1 = -2a^2 b^3, \tag{3.11c}$$

$$\nu_2 = -2ab^2 \sin v, \tag{3.11d}$$

$$\psi = -a^2. \tag{3.11e}$$

A Bäcklund transformation for the modified KdV equation is therefore

$$p = bp' + a \sin v, \tag{3.12a}$$

$$q = bq' - 2a[br' \cos v + p'^2 \sin v + \tfrac{1}{2}a(p + bp')], \tag{3.12b}$$

where, according to Eq. (3.4a), $b = \pm 1$. The solution $b = 0$ must obviously be discarded. A completely symmetric form of Eq. (3.12b) is

$$q = bq' - a[(r + br') \cos v + (p^2 + p'^2) \sin v], \quad b = \pm 1. \tag{3.13}$$

The choice of signs is expected since if z' is a solution of Eq. (3.2) so is $-z'$.

A. Theorem of permutability

The result given in Eq. (3.12a) is quite similar to the Bäcklund transformation equations for the sine–Gordon equation [cf. Eqs. (1.6)]. Since only one of the transformation equations is needed for the theorem of permutability, there is a corresponding similarity in the permutability relation as well. A calculation identical to that of Sec. 2A yields

$$\tan\left(\frac{z_3 - z_0}{2}\right) = b \left(\frac{a_1 + a_2}{a_1 - a_2}\right) \tan\left(\frac{z_1 - z_2}{2}\right), \quad b = \pm 1. \tag{3.14}$$

Starting from $z_0 = 0$, one finds

$$z_i = 2 \tan^{-1} \exp(\mu_i), \quad i = 1, 2, \tag{3.15}$$

where

$$\mu_i = a_i x - a_i^3 y + \delta_i, \tag{3.16}$$

in which δ_i is an integration constant. Equation (3.14) then yields

$$z_3 = \pm 2 \tan^{-1} \left[\left(\frac{a_1 + a_2}{a_1 - a_2}\right) \frac{\sinh\frac{1}{2}(\mu_1 - \mu_2)}{\cosh\frac{1}{2}(\mu_1 + \mu_2)}\right]. \tag{3.17}$$

A solution of this type has been obtained by Wadati[23] by employing the inverse method.

B. Relation to the inverse method

The change of variable $\Gamma = \tan[\frac{1}{2}(z + bz')]$ converts Eq. (3.12a) to the Riccati equation

$$\Gamma_x + a\Gamma - p(1 + \Gamma^2) = 0. \tag{3.18}$$

As noted in the introduction, use of Eqs. (1.8)–(1.10) leads to the linear equations considered by Ablowitz, *et al.*, viz.

$$w_{1x} + \tfrac{1}{2} a w_1 = p w_2, \tag{3.19a}$$

$$w_{2x} - \tfrac{1}{2} a w_2 = p w_1. \tag{3.19b}$$

Bäcklund transformation results equivalent to those presented here have recently been derived by Wadati[11] and Hirota[12] and have been obtained directly from the linear equations of the inverse method by Newell.[3]

4. NONLINEAR SCHRÖDINGER EQUATION

In treating the nonlinear Schrödinger equation, the introduction of an area function has not been found useful. (Indeed, the single soliton pulse profile for this equation is not the derivative of any particularly simple analytic function.) Proceeding, then, to a consideration of the nonlinear Schrödinger equation itself, the equation under study and its complex conjugate are taken in the form

$$iq + r + z^2 \bar{z} = 0, \tag{4.1a}$$

$$-i\bar{q} + \bar{r} + \bar{z}^2 z = 0, \tag{4.1b}$$

where the bar indicates complex conjugate. The general form adopted for the Bäcklund transformation is

$$p = f(z, \bar{z}, z', \bar{z}', p', \bar{p}'), \tag{4.2a}$$

$$q = \varphi(z, \bar{z}, z', \bar{z}', q', \bar{q}', p', \bar{p}'). \tag{4.2b}$$

In addition, one requires the complex conjugate transformation equations $\bar{p} = \bar{f}$, $\bar{q} = \bar{\varphi}$. Introducing the function Ω as in the previous examples and eliminating r and $\bar{r}$ by Eqs. (4.1), one finds that

$$\Omega_{s'} = f_{p'} - \varphi_{q'} = 0, \tag{4.3a}$$

$$\Omega_{\bar{s}'} = f_{\bar{p}'} - \varphi_{\bar{q}'} = 0, \tag{4.3b}$$

which reduces Ω to the form

$$\Omega = f_z \varphi + f_{z'} q' + f_{\bar{z}} \bar{\varphi} + f_{\bar{z}'} \bar{q}' - \varphi_z f - \varphi_{z'} p' - \varphi_{\bar{z}} \bar{f} - \varphi_{\bar{z}'} \bar{p}' + \varphi_{p'}(iq' + z'^2 \bar{z}') + \varphi_{\bar{p}'}(-i\bar{q}' + \bar{z}'^2 z') = 0. \tag{4.4}$$

From Eqs. (4.2) and (4.3)

$$\varphi_{q'q'} = \varphi_{q'\bar{q}'} = \varphi_{\bar{q}'\bar{q}'} = 0. \tag{4.5}$$

In addition, employing Eqs. (4.3),

$$\Omega_{q'q'} = 2i f_{p'p'} = 0, \tag{4.6a}$$

$$\Omega_{\bar{q}'\bar{q}'} = -2i f_{\bar{p}'\bar{p}'} = 0. \tag{4.6b}$$

Equations (4.3) then imply

$$\varphi_{q'p'} = \varphi_{\bar{q}'\bar{p}'} = 0. \tag{4.7}$$

Also, one finds

$$\Omega_{q'p'} = i\varphi_{p'p'} = 0, \tag{4.8a}$$

$$\Omega_{\bar{q}'\bar{p}'} = -i\varphi_{\bar{p}'\bar{p}'} = 0. \tag{4.8b}$$

Integration of Eqs. (4.7) and (4.3) yield

$$f_{p'} = g(Z, \bar{p}') = \varphi_{q'}, \tag{4.9}$$

where Z stands for the set of four dependent variables $z, z', \bar{z}, \bar{z}'$, and g is a function to be determined. Subsequent integration of Eq. (4.9) and use of Eqs. (4.6) leads to

$$f=kp'\bar{p}'+lp'+m\bar{p}'+n, \tag{4.10}$$

where k, l, m and n are arbitrary functions of Z. Furthermore, Eqs. (4.3) and (4.10) yield

$$\varphi=k(\bar{p}'q'+p'\bar{q}')+lq'+m\bar{q}'+\sigma p'\bar{p}'+\tau p'+\theta\bar{p}'+\chi, \tag{4.11}$$

where σ, τ, θ, and χ are also functions of Z.

The requirement that the unprimed variables satisfy Eq. (4.1) is now imposed. Calculating r from Eq. (4.10), one finds that

$$\begin{aligned}iq+r=-z^2\bar{z}=&-k\bar{p}'z'^2\bar{z}'+kp'(i\bar{q}'+\bar{r}')-lz'^2\bar{z}'+m(i\bar{q}'+\bar{r}')\\&+Ap'\bar{p}'+Bp'+C\bar{p}'+Dp'^2\bar{p}'+E\bar{p}'^2p'+Fp'^2\\&+G\bar{p}'^2+H,\end{aligned} \tag{4.12}$$

where Eqs. (4.1) in terms of the primed coordinates have been employed to write the coefficients of $k\bar{p}'$ and l in terms of Z. The variables p', $\bar{p}'$, $\bar{q}'$, $\bar{r}'$, and the Z are independent variables, and therefore Eq. (4.12) must be satisfied identically. Hence one sets $A, \ldots, G=0$, and $H=-z^2\bar{z}$. Since the coefficients of kp' and m can *not* be expressed in terms of Z by employing Eq. (4.1b), one must set $k=m=0$. For $k=m=0$, the requirement that Eq. (4.12) be satisfied indentically leads to

$$A=i\sigma+\bar{l}\,l_{\bar{z}}+l_{\bar{z}'}=0, \tag{4.13a}$$

$$B=i\tau+l_z n+l_{\bar{z}}\bar{n}+ln_z+n_{z'}=0, \tag{4.13b}$$

$$C=i\theta+n_{\bar{z}}\bar{l}+n_{\bar{z}'}=0, \tag{4.13c}$$

$$F=l_{z'}+ll_z=0, \tag{4.13d}$$

$$H=nn_z+\bar{n}n_{\bar{z}}-lz'^2\bar{z}'+i\chi=-z^2\bar{z}, \tag{4.13e}$$

as well as $D=E=G=0$. Equation (4.13d) may be satisfied by choosing l to be a constant. Then $\sigma=0$ by Eq. (4.13a). Such a solution has been found to be adequate for present purposes.[24]

The Bäcklund transformation has now been developed to the form

$$p=ap'+n, \tag{4.14a}$$

$$q=aq'+\tau p'+\theta\bar{p}'+\chi. \tag{4.14b}$$

In addition there are the complex conjugate expressions

$$\bar{p}=\bar{a}\bar{p}'+\bar{n}, \tag{4.15a}$$

$$\bar{q}=\bar{a}\bar{q}'+\bar{\tau}\bar{p}'+\bar{\theta}p'+\bar{\chi}. \tag{4.15b}$$

Substitution of these four expressions into the form of Ω given in Eq. (44.) leads to

$$\Omega=J\,q'+Kp'+L\bar{p}'+M+N\bar{q}'+Pp'^2+Qp'\bar{p}'+R\bar{p}'^2=0, \tag{4.16}$$

where the expressions $J, \ldots, R$ are functions of Z that are given below in the simplified form in which they are used in the subsequent analysis. Each such coefficient is again required to vanish identically. In particular, one finds

$$N=\bar{a}n_{\bar{z}}+n_{\bar{z}'}-i\theta=0. \tag{4.17}$$

In conjunction with Eq. (4.13c) this leads to $\theta=0$. With this simplification the vanishing of the coefficients in Eq. (4.16) yields the equations

$$J=an_z+n_{z'}+i\tau=0, \tag{4.18a}$$

$$K=\tau n_z-a\chi_z-n\tau_z-\chi_{z'}-\bar{n}\tau_{\bar{z}}=0, \tag{4.18b}$$

$$L=n_{\bar{z}}\bar{\tau}-\bar{a}\chi_{\bar{z}}-\chi_{\bar{z}'}=0, \tag{4.18c}$$

$$M=\chi n_z+\bar{\chi}n_{\bar{z}}-n\chi_z-\bar{n}\chi_{\bar{z}}+\tau z'^2\bar{z}'=0, \tag{4.18d}$$

$$N=\bar{a}n_{\bar{z}}+n_{\bar{z}'}=0, \tag{4.18e}$$

$$P=-(a\tau_z+\tau_{z'})=0, \tag{4.18f}$$

$$Q=-(\bar{a}\tau_{\bar{z}}+\tau_{\bar{z}'})=0, \tag{4.18g}$$

as well as $R=0$ identically.

Equation (4.13e) plus Eqs. (4.18 constitute a set of eight equations that must be satisfied by n, τ, χ and their complex conjugates. The solution is here outlined for the case $a=1$. (The result for $a=-1$ may be obtained by replacing z' by $-z'$ in the result to be obtained here.)

By introducing

$$w=z+z', \quad v=z-z', \tag{4.19}$$

Eqs. (4.18f, g) show that $\tau=\tau(v,\bar{v})$ while Eq. (4.18e) yields $n=n(w,v,\bar{v})$. Equation (4.18a) then leads to

$$n(w,v,\bar{v})=-\tfrac{1}{2}iw\tau(v,\bar{v})+\gamma(v,\bar{v}), \tag{4.20}$$

where γ is to be determined. From Eq. (4.18c)

$$2\chi=\bar{\tau}\bar{w}(-\tfrac{1}{2}iw\tau_{\bar{v}}+\gamma_{\bar{v}})+\zeta(w,v,\bar{v}), \tag{4.21}$$

where ζ is to be determined. From Eq. (4.18b),

$$\zeta=w(-\tfrac{1}{2}i\tau^2+\tau\gamma_v-\gamma\tau_v-\bar{\gamma}\tau_{\bar{v}})+\eta(v,\bar{v}), \tag{4.22}$$

where η is to be determined. Substitution into Eq. (4.13e) and separation of the resulting equation according to powers of w leads to

$$\tau\tau_v=\bar{v}, \tag{4.23a}$$

$$\gamma\tau_v+\bar{\gamma}\tau_{\bar{v}}=0. \tag{4.23b}$$

$$\bar{\tau}\tau_{\bar{v}}=-v, \tag{4.23c}$$

$$i\bar{\tau}\gamma_{\bar{v}}=0, \tag{4.23d}$$

$$-\tfrac{1}{2}i\tau\gamma+\gamma\gamma_v+\bar{\gamma}\gamma_{\bar{v}}+\tfrac{1}{2}i\eta=-\tfrac{1}{4}\bar{v}v^2. \tag{4.23e}$$

Equations (4.23a, c) are satisfied by

$$\tau=i(b-2v\bar{v})^{1/2}, \tag{4.24}$$

where b is real constant. Equation (4.23b) is satisfied by

$$\gamma=ikv, \tag{4.25}$$

where k is a real constant.

Equation (4.23e) then yields

$$\eta=iv(\tfrac{1}{2}|v|^2+k\tau-2k^2), \tag{4.26}$$

and Eqs. (4.26), (4.22), and (4.21) lead to

$$\chi=-kn+\tfrac{1}{2}\tau n+\tfrac{1}{4}iv(|w|^2+|v|^2). \tag{4.27}$$

Equation (4.18d), which has not been used in the analysis thus far, is now found to be satisfied indentically.

The Bäcklund transformation is therefore

$$p = p' - \tfrac{1}{2}iw\tau + ikv, \tag{4.28a}$$

$$q = q' + \tfrac{1}{2}\tau(p + p') - kn + \tfrac{1}{4}iv(|w|^2 + |v|^2), \tag{4.28b}$$

where

$$\tau = \pm i(b - 2|v|^2)^{1/2} \tag{4.29}$$

and b and k are arbitrary real constants.

While a theorem of permutability may be obtained from Eq. (4.28a) by following the procedure used in subsections 2A and 3A, the result appears to be too complex to be useful for computational purposes.

Connection with the inverse method

Contact with the inverse method may be made by defining

$$\Gamma = (b - 2|v|^2)^{1/2}/2^{1/2}v. \tag{4.30}$$

Equation (4.28a) and its complex conjugate yield

$$z[\Gamma_x + ik\Gamma + 2^{-1/2}(z\Gamma^2 + \bar{z})] = z'[\Gamma_x + ik\Gamma + 2^{-1/2}(z'\Gamma^2 + \bar{z}')]. \tag{4.31}$$

If one now sets either z or $z' = 0$ the other variable satisfies a Riccati equation. By following the procedure outlined in the introduction, this Riccati equation is equivalent to the pair of linear equations

$$w_{1x} + \tfrac{1}{2}ikw_1 = -2^{-1/2}\bar{z}w_2, \tag{4.32a}$$

$$w_{2x} - \tfrac{1}{2}ikw_2 = 2^{-1/2}zw_1, \tag{4.32b}$$

which are the linear equations for the inverse problem that has been associated with the nonlinear Schrödinger equation.[4,2]

5. RELATION BETWEEN KdV AND MODIFIED KdV EQUATIONS

The Bäcklund transformations constructed in the three previous sections are of a very special type in that they transform a given equation into itself. As noted in the Introduction, a more general usage of the transformation theory involves transformations between equations of different form.

The relation between the KdV and modified KdV equations discovered by Miura[19] leads one to suspect that these two equations may actually be related by a Bäcklund transformation. In the present section this is shown to be the case. Clairin's method is employed to obtain another equation that may be paired with the Miura transformation to complete the Bäcklund transformation.

To avoid the introduction of imaginary quantities, the KdV and modified KdV equations are written $q' + 6z'p' + \alpha' = 0$ and $q - 6z^2p + \alpha = 0$, respectively. The general form of the Bäcklund transformation will be taken to be

$$p = f(z, z'), \tag{5.1a}$$

$$q = \varphi(z, z', p', r'). \tag{5.1b}$$

The choice of the first equation is motivated by the known form of the Miura transformation (which in the present notation is $\pm p = z' + z^2$) while the choice of the second equation is dictated by the type of derivatives that appear when equality of mixed second partial derivatives is imposed and account is taken of the relation among these derivatives that is imposed by the KdV equation.

Following the procedure employed in the three previous sections, one defines

$$\Omega = f_z\varphi + f_{z'}q' - \varphi_z f - \varphi_{z'}p' - \varphi_{p'}r' + \varphi_{r'}(q' + 6z'p') = 0. \tag{5.2}$$

Then

$$\Omega_{q'} = f_{z'} + \varphi_{r'} = 0, \tag{5.3a}$$

$$\Omega_{q'p'} = \varphi_{r'p'} = 0, \tag{5.3b}$$

$$\Omega_{p'r'} = f_{z'z'} - \varphi_{p'p'} = 0, \tag{5.3c}$$

$$\Omega_{p'p'} = f_z f_{z'z'} - f_{zz'z'}f - 2\varphi_{z'p'} - p'f_{z'z'z'}, \tag{5.3d}$$

$$\Omega_{p'r'p'} = -\varphi_{p'p'p'} = 0, \tag{5.3e}$$

$$\Omega_{p'p'p'} = -3f_{z'z'z'} = 0. \tag{5.3f}$$

From Eqs. (5.3f) and (5.3c)

$$f_{z'z'} = \varphi_{p'p'} = a(z), \tag{5.4}$$

where the unknown function $a(z)$ arises from integration. It turns out that $a(z)$ must be set equal to zero to have $f(z, z')$ reduce to the Miura transformation. Then, integration of Eq. (5.4) in conjunction with Eq. (5.3a) yields

$$f_{z'} = g(z) = -\varphi_{r'} \tag{5.5}$$

as well as $\varphi_{p'} = \lambda(z, z', r')$ where both $g(z)$ and $\lambda(z, z'r')$ are to be determined. Equations (5.3b) and (5.3d) reduce the functional dependence of λ so that

$$\varphi_{p'} = \lambda(z). \tag{5.6}$$

Integration of this result and use of Eqs. (5.3a) and (5.5) leads to

$$\varphi = \lambda p' - gr' + \nu(z, z'). \tag{5.7}$$

The z' dependence is now determined by observing that

$$\Omega_{z'z'p'} = -\varphi_{z'z'z'} = 0 \tag{5.8}$$

and

$$\Omega_{z'z'z'} = 3(f_{zz'}\varphi_{z'z'} - \varphi_{zz'z'}f_{z'}) = 0. \tag{5.9}$$

Derivation of these results employs Eq. (5.3d) in the form $\varphi_{z'p'} = 0$, which follows from the vanishing of $a(z)$. Since $f_{z'} = g$, integration of Eq. (5.9) yields

$$\varphi_{z'z'} = \psi(z) \tag{5.10}$$

with

$$\psi(z) = Kg(z), \tag{5.11}$$

where K is a constant of integration. Integration of Eq. (5.10) yields a form for φ in which all but the z dependence has been determined. The Bäcklund transformation has now been developed to the form

$$p = gz' + h, \tag{5.12a}$$

$$q = \lambda p' - gr' + \tfrac{1}{2}Kgz'^2 + \chi z' + \theta, \tag{5.12b}$$

where g, h, λ, χ, and θ are as yet undetermined functions of z. Using Eqs. (5.12) in Eq. (5.2) and then setting equal to zero the coefficient of each of the different dependences upon the primed variables in Ω yields

$$g\lambda' - \lambda g' + g(6+K) = 0, \tag{5.13a}$$

$$g'\chi - \chi' g + \tfrac{1}{2}K(gh' - hg') = 0, \tag{5.13b}$$

$$g'\theta - \theta' g + \chi h' - h\chi' = 0, \tag{5.13c}$$

$$\chi = \lambda h' - h\lambda', \tag{5.13d}$$

$$\theta h' - h\theta' = 0, \tag{5.13e}$$

$$\lambda = hg' - gh'. \tag{5.13f}$$

Equations (5.13b) and (5.13f) lead to

$$K\lambda = 2(\chi g' - g\chi'). \tag{5.14}$$

Specialization of Eq. (5.12a) to the Miura transformation requires $g = \pm 1$. The construction of Ω has already incorporated the KdV equation. The requirement that Eqs. (5.12) yield $q - 6z^2 p + \alpha = 0$ leads to

$$h' + \lambda = 0, \tag{5.15a}$$

$$h'' + \tfrac{1}{2}K = 0, \tag{5.15b}$$

$$2hh'' + h'^2 + \chi - 6z^2 = 0, \tag{5.15c}$$

$$h'' h^2 + h'^2 h + \theta - 6z^2 h = 0. \tag{5.15d}$$

These equations are readily solved. One finds $h = \pm z^2$, $\lambda = -2z$, $\chi = \mp 2z^2$, $\theta = 0$, $K = -4$. Equations (5.13a, d, e) are also satisfied by these solutions. Finally, the Bäcklund transformation is

$$p = \pm (z' + z^2), \tag{5.16a}$$

$$q = \mp r' - 2(zp' + z'p). \tag{5.16b}$$

Equation (5.16a) is the Miura transformation.

ACKNOWLEDGMENTS

The author is indebted to F.B. Estabrook, R.M. Miura, A.N. Newell, and H.D. Wahlquist for their comments and corrections relating to earlier versions of this manuscript.

*Present address: Department of Mathematics and Optical Sciences Center, The University of Arizona, Tucson, Arizona 85721.

[1]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE 61, 1443 (1973).
[2]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. 31, 125 (1973).
[3]A.C. Newell (private communication).
[4]V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. 61, 118 (1971) [Sov. Phys. JETP 34, 62 (1972)].
[5]E. Goursat, *Le problème de Bäcklund*, Memorial Sci. Math., Fasc. 6 (Gauthier-Villars, Paris, 1925). (This volume contains an extensive bibliography of early papers on Bäcklund transformations.)
[6]L. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Reprinted, Dover, New York, 1960), pp. 284—90.
[7]A.R. Forsyth, *Theory of Differential Equations*, *Vol.* 6 (Reprinted, Dover, New York, 1959), Chap. 21.
[8]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. 30, 1262 (1973).
[9]H.D. Wahlquist and F.B. Estabrook, Phys. Rev. Lett. 31, 1386 (1973).
[10]G.L. Lamb, Jr., Phys. Lett. A 48, 73 (1974).
[11]M. Wadati (to be published).
[12]R. Hirota (to be published).
[13]M.J. Clairin, Ann. Toulouse 2e Sér. 5, 437 (1903).
[14]G.L. Lamb, Jr., in *In Honor of Philip M. Morse*, edited by H. Feshbach and K.U. Ingard (M.I.T. Press, Cambridge, Mass., 1969), p. 88 (Appendix A).
[15]W.F. Ames, *Nonlinear Partial Differential Equations in Engineering, Vol. II* (Academic, New York, 1972), pp. 44—48.
[16]G.L. Lamb, Jr., Rev. Mod. Phys. 43, 99 (1971).
[17]P.D. Gerber, IBM J. Res. Develop. 17, 314 (1973).
[18]H.D. Wahlquist and F.B. Estabrook (to be published).
[19]R.M. Miura, J. Math. Phys. 9, 1202 (1968).
[20]H.D. Wahlquist (private communication).
[21]N.J. Zabusky, Phys. Rev. 168, 124 (1968), Eq. (14).
[22]Note that if one also were to set $c_2 = 0$, a valid Bäcklund transformation would result, but the simplest solution of the resulting equations, which should correspond to the integral of the single soliton solution would diverge.
[23]M. Wadati, J. Phys. Soc. Japan 34, 1289 (1973).
[24]Investigation of other particular solutions obtained from the general solution $l = \phi(z - lz')$ may prove advantageous.

Orthogonal polynomials from the viewpoint of scattering theory*

K. M. Case

The Rockefeller University, New York, New York 10021
(Received 17 June 1974)

It is demonstrated that there is a close parallel between the theory of a class of orthogonal polynomials and scattering theory. In both cases a fundamental role is played by a particular solution of the basic difference (differential) equation which we call the Jost function. Under rather general conditions this function has simple analytic properties. It determines and is largely determined by either the asymptotic phases or the continuous part of the weight (spectral) function. Indeed this is more than an analogy. By appropriate limiting procedures one can pass from a result about orthogonal polynomials to one in scattering theory. Conversely, scattering theory throws considerable light on theorems about orthogonal polynomials.

I. INTRODUCTION

Two highly developed theories of mathematical physics are those of orthogonal polynomials and potential scattering. While much of the work on orthogonal polynomials predates that on scattering theory, the latter has been much more intensively investigated in the last 25 years. Recently[1] we have seen that the theory of orthogonal polynomials sheds considerable light on the inverse problem of scattering theory. Here we will turn the relationship around. We want to show that methods of scattering theory form a unified basis for obtaining the various properties of orthogonal polynomials. Very little in the way of new results are obtained. Our main attempt is to demonstrate that many diverse conclusions all readily follow from this particular point of view.

In order to make the relationship particularly clear we restrict ourself to polynomials defined over a region of compact support on the real axis. In particular, the weight function with respect to which our polynomials are to be orthogonal is continuous over a finite interval of the real axis. In addition, it may have a finite number of jump points outside of that interval.

Our program is the following: In Sec. II we briefly sketch the methods and results of the theory of potential scattering. The next section gives an appropriate formulation of the problem of orthogonal polynomials and indicates the relationship to the moment problem. The heart of this paper is Sec. IV in which it is demonstrated that the properties of these orthogonal polynomials can be determined completely paralleling the methods of scattering theory. Indeed by an appropriate limiting process we can at any point pass from a property of orthogonal polynomials to one of scattering theory. In Sec. V some applications of the general formalism are given. The moment problem is briefly discussed in Sec. VI.

II. SCATTERING THEORY[2]

We consider the Schrödinger equation with spherically symmetrical potential $q(x)$. Restricting attention to spherically symmetrical solutions (S waves), the equation (in appropriate units) is

$$-\left(\frac{1}{2}\frac{d^2}{dx^2}+q(x)\right)\psi(E,x)=E\psi,\quad 0\le x<\infty. \tag{II. 1}$$

Let us assume that $q(x)$ satisfies the conditions

$$\int_{-\infty}^{\infty} x\,|q(x)|\,dx<\infty,\quad\text{and}\quad \int_0^{\infty} x^2|q(x)|\,dx<\infty. \tag{II. 2}$$

Of particular interest are the "regular" solutions of (II. 1). These are the bounded solutions subject to

$$\psi(E,0)=0,\quad \frac{d\psi}{dx}(E,0)=C\ne 0. \tag{II. 3}$$

The square integrable such solutions (if they exist) belonging to eigenvalues E_i are the bound states. In addition, there are "scattering solutions" for all E, $0<E<\infty$. For large x these behave as

$$\psi(E,x)\sim A(E)\sin[\sqrt{2E}x+\delta(E)]. \tag{II. 4}$$

The direct problem is to determine the E_i and $\delta(E)$ $(0\le E<\infty)$, given $q(x)$. In the inverse problem the question is to find $q(x)$ given the E_i, $\delta(E)$, and constants ρ_i defined below.

To investigate the general properties it is convenient to introduce two other solutions $\psi_\pm(k,x)$ of Eq. (II. 1). With $k^2=2E$ these are defined by the boundary conditions

$$\lim_{x\to\infty}|\psi(k,x)-\exp(\pm ikx)|=0,\quad \mathrm{Im}k\begin{cases}\ge 0\\ \le 0\end{cases}. \tag{II. 4}$$

Convenient integral equations incorporating these conditions are readily obtained. In particular, for ψ_+ we have

$$\psi_+(k,x)=\exp(ikx)-\frac{2}{k}\int_x^{\infty}\sin k(x-y)q(y)\psi_+(y)\,dy. \tag{II. 5}$$

From the definitions we readily see that for $\mathrm{Im}k=0$,

$$\psi_+(k,x)=\psi_-(-k,x)$$

and

$$\psi_+^*(k,x)=\psi_-(k,x). \tag{II. 6}$$

Further, using the constancy of the Wronskian and the boundary conditions we can express the regular solution (for $\mathrm{Im}k=0$) in terms of $\psi_\pm$ as

$$\psi(k,x)=\frac{C}{2ik}[\psi_-(k,0)\psi_+(k,x)-\psi_+(k,0)\psi_-(k,x)]. \tag{II. 7}$$

[Note this implies $\psi(-k,x)=\psi(k,x)$ for $\mathrm{Im}k=0$.] Clearly, the Jost[3] function

$$f_+(k)=\psi_+(k,0) \tag{II. 8}$$

plays a very fundamental rule. Thus:

(1) For large x the solution (II. 7) behaves as

$$\psi(k,x) \sim (C\,|f_+|/k)\sin(kx+\delta) \tag{II. 9}$$

where

$$\delta = -\arg f_+(k), \tag{II. 10}$$

(i. e., the phase shift is just the negative of the argument of f_+).

(2) The bound states are at the zeros of $f_+(k)$ for Im$k \geqslant 0$. Indeed, if $f_+(k_i)=0$ we have

$$\psi(k_i,x) = \frac{\psi_+(k_i,x)C}{d/dx\,\psi_+(k_i,0)}\,. \tag{II. 11}$$

Using essentially only Green's identity, we can then establish that

(a) The discrete eigenvalues are such that Re$k_i = 0$, Im$k_i \geqslant 0$.

(b) They are simple.

(Alternatively, the E_i are real, negative, and simple.)

Further,

$$\int_0^\infty \psi(E,x)\psi(E',x)\,dx = \frac{\delta(EE')}{\rho'(E)}, \quad 0<E,E'<\infty, \tag{II. 12}$$

with

$$\rho'(E) = \frac{2}{\pi}\frac{\sqrt{2E}}{C^2|f_+(E)|^2}, \tag{II. 13}$$

$$\int_0^\infty \psi(E,x)\psi(E_i,x)\,dx = 0, \quad 0<E<\infty, \quad E_i \text{ a discrete eigenvalue}, \tag{II. 14}$$

and

$$\int_0^\infty \psi(E_i,x)\psi(E_j,x)\,dx = \delta(i,j)/\rho_i, \tag{II. 15}$$

with

$$\rho_i = \frac{-2(d/dx)\psi_+(E_i,0)}{C^2[df_+(E_i)/dE_i]}\,. \tag{II. 16}$$

The integral Eq. (II. 5) plus the assumed properties of $q(x)$ imply that $f_+(k)$ is analytic in the upper half k plane and continuous on the real axis. This implies the discrete eigenvalues are finite in number. The analyticity and Cauchy's theorem then enable us to prove the completeness theorem

$$\int_{-\infty}^\infty \psi(E,x)\psi(E,x')\,d\rho(E) = \delta(x-x'), \tag{II. 17}$$

where

$$d\rho(E) = \sum_i \rho_i\delta(E-E_i)\,dE, \quad -\infty<E\leqslant 0, \tag{II. 18}$$

and

$$d\rho(E) = \rho'(E)\,dE, \quad 0\leqslant E<\infty. \tag{II. 19}$$

In the inverse problem we are given the ρ_i, E_i ($i=1,2,\ldots,N$) and $\delta(E)$, $0<E<\infty$. [It may be noted that in virtue of Levinson's theorem[4] these are not completely independent. We have the relation $\delta(\infty)-\delta(0) = -2\pi N$.] From the above we see that we are thus given the phase of $f_+(k)$ on the real axis and the position of its zeros in the upper half plane. Using the analyticity of $f_+(k)$ and the fact [from Eq. (II. 5)] that $f_+(\infty)=1$, we can by simple function theoretic arguments (similar to those given in Sec. IV) determine $f_+(k)$ for Im$k>0$. From (II. 13) and (II. 16) we see we then know $d\rho(E)$ for $-\infty<E<\infty$. The solution of the Gelfand—Levitan[5] equation then leads to $q(x)$.

We will see that there is an analogous problem in the theory of orthogonal polynomials. There again we will want to construct a spectral function ρ from a knowledge of a phase and the position of certain zeros. The procedure is as above. First the Jost function is found and then ρ'.

However, an equally natural problem arising for orthogonal polynomials has an analog in our present discussion: How can we construct the Jost function given the E_i and $\rho'(E)$? In order to illustrate the close relation between ordinary scattering theory and the theory of orthogonal polynomials we here give a somewhat detailed discussion of this construction.

First consider the following slight modification of the Poisson-Jensen formula.[6] Suppose $h(k)$ is analytic in the upper half plane $h(\infty)=0$ and on the real axis $h(k')$ is continuous and $h^*(k') = h(-k')$. Then with k in the upper half plane we have by Cauchy's theorem

$$h(k) = \frac{1}{2\pi i}\int_{-\infty}^\infty \frac{h(k')\,dk'}{k-k'} = \frac{1}{2\pi i}\int_{-\infty}^\infty \frac{[h_r(k')+ih_m(k')]\,dk'}{k'-k} \tag{II. 20}$$

and

$$0 = \frac{1}{2\pi i}\int_{-\infty}^\infty \frac{[h_r(k')+ih_m(k')]\,dk'}{k'+k} = \frac{1}{2\pi i}\int_{-\infty}^\infty \frac{h_r(-k')+ih_m(-k')\,dk'}{-k'+k},$$

or

$$0 = \frac{1}{2\pi i}\int_{-\infty}^\infty \frac{[h_r(k')-ih_m(k')]\,dk'}{k'-k}\,. \tag{II. 21}$$

Adding Eqs. (II. 20, 21) yields

$$h(k) = \frac{1}{\pi i}\int_{-\infty}^\infty \frac{h_r(k')\,dk'}{k'-z}\,. \tag{II. 22}$$

Now if $i\kappa_i$ ($i=1,2,\ldots,N$) are the zeros of $f_+(k)$ in the upper half plane we know from the properties enumerated earlier that $\ln\{f_+(k)\prod_{i=1}^N[(k+i\kappa_i)/(k-i\kappa_i)]\}$ has all the properties required for the representation of Eq. (II. 22).

$$\therefore f_+(k) = \prod_{i=1}^N \frac{(k-i\kappa_i)}{(k+i\kappa_i)}\exp\left(\frac{1}{\pi i}\int_{-\infty}^\infty \frac{\ln|f_+(k')|\,dk'}{k'-k}\right), \tag{II. 23}$$

since $|(k'+i\kappa_i)/(k'-i\kappa_i)|=1$, for k' on the real axis. Finally, to express f_+ in terms of ρ' it is convenient to use the evenness of $\ln|f_+|$ to write the integral as one over $E'=k'^2/2$. Then from the relation of $|f_+|$ to ρ' we obtain

$$f_+(k) = \prod_{i=1}^N \frac{(k-i\kappa_i)}{(k+i\kappa_i)}\exp\left(-\frac{k}{\pi i}\int_0^\infty \frac{\ln\{(\pi/2)[C^2\rho'(E')/\sqrt{2E'}]\}\,dE'}{\sqrt{2E'}(E'-k^2/2)}\right). \tag{II. 24}$$

Among other possibilities this equation may be regarded as a source of sum rules. Thus imagine we have some independent means of calculating the behavior of f_+ for large k [say from the integral Eq. (II. 5)]. Suppose this has the form

$$f_+(k) \sim 1 + \frac{C_1}{k} + \frac{C_2}{k^2} + \cdots\,. \tag{II. 25}$$

Then expanding Eq. (II. 24) in powers of $1/k$ we have sum rules relating moments of $\ln\rho'$ to the coefficients C_i.

As a simple example consider the case where there are no bound states. Then as an approximation to Eq. (II. 5) (valid for large k) we have

$$\psi_+(k,x) \approx \exp(ikx) - \frac{2}{k}\int_x^\infty \sin k(x-y)q(y)\exp(iky)\,dy. \tag{II. 26}$$

Then

$$f_+(k) = \psi_+(k,0) \approx 1 + \frac{i}{k}\int_0^\infty q(y)\,dy. \tag{II. 27}$$

Thus

$C_1 = i\int_0^\infty q(y)\,dy.$

On the other hand, expanding Eq. (II. 24) (to order $1/k$) we find

$$f_+(k) \approx 1 - \frac{i}{k}\frac{2}{\pi}\int_0^\infty \frac{\ln\{[(\pi/2)C^2\rho'(E')]/\sqrt{2E'}\}}{\sqrt{2E'}}\,dE'. \tag{II. 28}$$

Comparing Eqs. (II. 27) and (II. 28) yields the relation

$$\int_0^\infty q(y)\,dy = -\frac{2}{\pi}\int_0^\infty \ln\left(\frac{(\pi/2)C^2(E')}{\sqrt{2E'}}\right)\frac{dE'}{\sqrt{2E'}}. \tag{II. 29}$$

Finally, we would like to point out the relationship of the functions $\psi_\pm$ to certain other solutions of the Schrödinger equation. (The analog of these other solutions play a significant role in the classical discussions of orthogonal polynomials.)

Thus to the functions $\psi(E_i,x)$, $i=1,\ldots,N$, $\psi(E,x)$, $0\le E<\infty$, we associate another set of functions defined by

$$\phi(E,x) = \frac{1}{2}\int_{-\infty}^\infty \frac{\psi(E',x)\,d\rho(E')}{E'-E}. \tag{II. 30}$$

These are at first defined for all nonreal E but then also for real E (two of them) by considering the limiting boundary values.

It may be noted that these associated functions satisfy the Schrodinger equation (II. 1) for all $x \ne 0$.

Indeed,

$$\begin{aligned}
&\left(-\frac{1}{2}\frac{d^2}{dx^2}+q(x)\right)\phi(E,x)\\
&=\frac{1}{2}\int_{-\infty}^\infty \left(-\frac{1}{2}\frac{d^2}{dx^2}+q(x)\right)\frac{\psi(E',x)\,d\rho(E')}{E'-E}\\
&=\frac{1}{2}\int_{-\infty}^\infty \frac{E'\psi(E',x)\,d\rho(E')}{E'-E}\\
&=\frac{1}{2}\int_{-\infty}^\infty \frac{(E'-E+E)\psi(E',x)\,d\rho(E')}{E'-E}\\
&=E\frac{1}{2}\int_{-\infty}^\infty \frac{\psi(E',x)\,d\rho(E')}{E'-E}+\frac{1}{2}\int_{-\infty}^\infty \psi(E',x)\,d\rho(E').
\end{aligned}$$

But putting $x'=0$, $\psi(E',0)=C$ in Eq. (II. 17), we see that

$\int_{-\infty}^\infty \psi(E',x)\,d\rho(E') = \delta(x)/C.$

Therefore,

$$\left(-\frac{1}{2}\frac{d^2}{dx^2}+q(x)\right)\phi(E,x) = E\phi(E,x)+\frac{\delta(x)}{2C}. \tag{II. 31}$$

Inserting the expressions of Eq. (II. 13) and (II. 16) and the form of Eq. (II. 7) for $\psi(E,x)$, $0\le E<\infty$, we find explicitly that

$$\psi_+(k,x) = Cf_+(k)\phi(k,x). \tag{II. 32}$$

III. ORTHOGONAL POLYNOMIALS AND THE MOMENT PROBLEM

Let us collect some well known properties of orthogonal polynomials.[7]

Suppose we are given some nondecreasing function $\rho(\lambda)$ defined on the real axis. We are to find polynomials $\psi(\lambda,n)$ such that

(i) $\psi(\lambda,n)$ is a polynomial of exact degree n and its leading coefficient is positive.

(ii) The orthonormality relations hold:

$$\int_{-\infty}^\infty \psi(\lambda,n)\psi(\lambda,m)\,d\rho(\lambda) = \delta(m,n). \tag{III. 1}$$

The construction is a straightforward application of the Hilbert–Schmidt procedure. Indeed, if we denote the moments of ρ by s_m, i.e.,

$$s_m = \int_{-\infty}^\infty \lambda^m\,d\rho(\lambda), \tag{III. 2}$$

the result is

$$\psi(\lambda,n) = \frac{1}{\sqrt{D_{n-1}D_n}}\begin{vmatrix} s_0 & s_1 & \cdots & s_n\\ s_1 & s_2 & \cdots & s_{n+1}\\ \cdots & \cdots & \cdots & \cdots\\ s_{n-1} & s_n & \cdots & s_{2n-1}\\ 1 & \lambda & \cdots & \lambda^n\end{vmatrix}, \quad n=1,2,3,\cdots, \tag{III. 3}$$

where

$$D_n = \begin{vmatrix} s_0 & s_1 & \cdots & s_n\\ s_1 & s_2 & \cdots & s_{n+1}\\ \cdots & \cdots & \cdots & \cdots\\ s_n & s_{n+1} & \cdots & s_{2n}\end{vmatrix}. \tag{III. 4}$$

[The Eq. (III. 3) also holds for $n\ge 0$ provided we define D_{-1} as $(s_0)^2$.] From this explicit result it is readily shown that these polynomials satisfy the three term recursion relation

$$\begin{aligned}&a(n+1)\psi(\lambda,n+1)+b(n)\psi(\lambda,n)+a(n)\psi(\lambda,n-1)\\&=\lambda\psi(\lambda,n),\quad n=1,2,3\cdots,\end{aligned} \tag{III. 5}$$

where

$b(n) = \int_{-\infty}^\infty \lambda\psi^2(\lambda,n)\,d\rho(\lambda)$

and

$$a(n+1) = \int_{-\infty}^\infty \lambda\psi(\lambda,n)\psi(\lambda,n+1)\,d\rho(\lambda), \tag{III. 6}$$

or more explicitly

$$a(n+1) = \sqrt{D_{n+1}D_{n-1}/D_n^2}. \tag{III. 7}$$

We note that Eq. (III. 5) also holds for $n=0$ provided we define $a(0)\psi(\lambda,-1)$ to be zero.

Thus, in principle the procedure is so: Given $\rho(\lambda)$ we compute the moments s_m. From Eq. (III. 3) we then have the $\psi(\lambda,n)$. In the inverse (moment) problem we are to find ρ given the moments. The usual discussion[7] of this problem makes considerable use of the polynomials $\psi(\lambda,n)$ and another set of polynomials $Q(\lambda,n)$ defined by

$$Q(\lambda,n)=\int_{-\infty}^{\infty}\frac{\psi(\lambda,n)-\psi(\lambda',n)}{\lambda-\lambda'}\,d\rho(\lambda'). \qquad \text{(III. 8)}$$

It may be noted that the $Q(\lambda,n)$ also satisfy Eq. (III. 5) for $n\geq 1$.

IV. SCATTERING THEORY AND ORTHOGONAL POLYNOMIALS

We take Eq. (III. 5) for $n\geq 0$ with the initial condition $a(0)\psi(\lambda,-1)=0$ and $\psi(\lambda,0)=C=1/\sqrt{s_0}$ as fundamental for our discussion of orthogonal polynomials. Further, we restrict attention to the case when $a(\infty)$ and $b(\infty)$ exist and the limits are approached at least as fast as $1/n^2$. (As will be seen later this is the situation when the support of $d\rho(\lambda)$ is compact.)

Denote as "regular" those solutions of Eq. (III. 5) with the given initial conditions which for a fixed λ are bounded as $n\to\infty$. With the assumed conditions it is readily seen that such solutions exist for all λ such that

$$b(\infty)-2a(\infty)\leq\lambda\leq b(\infty)+2a(\infty). \qquad \text{(IV. 1)}$$

These solutions are conveniently described by z such that

$$\lambda=b(\infty)+a(\infty)[z+z^{-1}]. \qquad \text{(IV. 2)}$$

The statement then is that the Jacobi matrix formed from the $a(n)$, $b(n)$ has a continuous spectrum for λ in the interval described by Eq. (IV. 1) or, alternatively, for z lying on the unit circle ($z=e^{i\theta}$). In addition, there may be some discrete eigenvalues λ_i corresponding to square summable solution of Eq. (III. 5). We will show that these eigenvalues are:

(i) real

(ii) simple,

(iii) finite in number, and

(iv) lie outside or at the edge of the continuum.

(In z they are real and within the unit circle or at $z=\pm 1$.) These results imply that the $\rho(\lambda)$ used to form our orthogonal polynomials has only a finite number of jumps outside the interval of Eq. (IV. 1) plus a continuous part in the interval.

Some preliminaries: Let $\psi^{(1)}(\lambda,n)$, $\psi^{(2)}(\lambda,n)$ be two solutions of Eq. (III. 5). Then we have the analog of the Wronskian theorem, namely

$$W[\psi^{(1)},\psi^{(2)}]=a(n)[\psi^{(1)}(\lambda,n-1)\psi^{(2)}(\lambda,n)-\psi^{(2)}(\lambda,n-1)\psi^{(1)}(\lambda,n)] \qquad \text{(IV. 3)}$$

is independent of n. Introduce two auxiliary solutions $\psi_\pm$ of Eq. (III. 5) defined for $|z|\leq 1$ by the boundary conditions

$$\lim_{n\to\infty}|\psi_\pm-z^{\pm n}|\to 0. \qquad \text{(IV. 4)}$$

Further, use Eq. (III. 5) to define $f_\pm(z)$ as

$$f_\pm(z)=a(0)\psi_\pm(z,-1). \qquad \text{(IV. 5)}$$

[We choose to call $f_+(z)$ the Jost function since it will be seen to play the same role for the orthogonal polynomials as the function $f_+(k)$ does for scattering theory.]

Some properties are obvious. On the unit circle

$$\psi_+(z,n)=\psi_-^*(z,n)=\psi_-(z^{-1},n). \qquad \text{(IV. 6)}$$

Since the $\psi_\pm$ are linearly independent we can express the continuum regular solutions as linear combinations of them. Using the constancy of the "Wronskian" and the boundary conditions on ψ and $\psi_\pm$, we obtain

$$\psi(\lambda,n)=\frac{C}{a(\infty)2i\sin\theta}[f_-(z)\psi_+(z,n)-f_+(z)\psi_-(z,n)],\quad z=e^{i\theta}. \qquad \text{(IV. 7)}$$

In particular, it may be noted that the linearly independent functions are obtained for z running over the upper half of the unit circle (i.e., $0\leq\theta\leq\pi$).

As in scattering theory the Jost function determines the asymptotic behavior of the continuum functions. Thus from Eq. (IV. 4) we see that as $n\to\infty$

$$\psi(\lambda,n)\sim\frac{C|f_+(z)|}{a(\infty)\sin\theta}\sin(n\theta+\delta), \qquad \text{(IV. 8)}$$

where

$$\delta(\theta)=-\arg f_+(z).$$

Also as in Sec. II the zeros of $f_+(z)$ (within or on the unit circle) determine the discrete eigenvalues. Thus if $f_+(z_i)=0$, then

$$\psi(\lambda_i,n)=C[\psi_+(z_i,n)/\psi_+(z_i,0)] \qquad \text{(IV. 9)}$$

To investigate $f_+(z)$ further we can parallel the path of Sec. II and obtain "integral equations" for the functions $\psi_\pm$ and ψ. For this purpose it is useful to transform Eq. (III. 5). Let us introduce a new function $\overline{\psi}(\lambda,n)$ by

$$\overline{\psi}(\lambda,n)=\psi(\lambda,n)/\sqrt{g(n)}$$

where

(i) $\lim_{n\to\infty}g(n)=1$,

(ii) $g(n)g(n-1)=a^2(\infty)/a^2(n)$. (IV. 10)

Then Eq. (III. 5) becomes

$$a(\infty)[\overline{\psi}(\lambda,n+1)+\overline{\psi}(\lambda,n-1)]+(b(\infty)-\lambda)\overline{\psi}(\lambda,n)=\gamma(n), \qquad \text{(IV. 11)}$$

with

$$\gamma(n)=[\lambda(g(n)-1)+b(\infty)-b(n)g(n)]\overline{\psi}(\lambda,n). \qquad \text{(IV. 12)}$$

It is readily verified that an appropriate $g(n)$ is

$$g(n)=\frac{\prod_{i=1}^{\infty}[a^2(n+2i)/a^2(\infty)]}{\prod_{i=1}^{n}[a^2(n+2i-1)/a^2(\infty)}. \qquad \text{(IV. 13)}$$

Then (using Green's functions) integral equations for the solutions of Eq. (IV. 11) for various boundary conditions can be immediately written down. For example, for $\overline{\psi}_+$ we have

$$\overline{\psi}_+(z,n)=z^n+\sum_{m=n+1}\frac{[z^{n-m}-z^{(n-m)}]}{a(\infty)(z-z^{-1})}\gamma(m) \qquad \text{(IV. 14)}$$

where $\gamma(m)$ is as in Eq. (IV. 12) with $\overline{\psi}$ replaced by $\overline{\psi}_+$. Then by standard techniques we can use this equation to demonstrate that $\overline{\psi}_+$ (and thus ψ_+) is analytic within the unit z circle and continuous on it. Thus, from the definition

$$f_+(z) = [\lambda - b(0)]\psi_+(z,0) - a(1)\psi(z,1) \quad \text{(IV. 15)}$$

we conclude that $f_+(z)$ is analytic within the unit circle except perhaps for a simple pole at $z=0$. The pole does indeed exist since from Eq. (IV. 14) or Eq. (III. 5) we can show that for small z

$$\psi_+(z,n) \approx z^n \prod_{i=1}^{\infty} \frac{a(\infty)}{a(n+i)} . \quad \text{(IV. 16)}$$

Then from (IV. 15) we conclude that the residue of $f_+(z)$ at $z=0$ is

$$a(\infty) \prod_{i=1}^{\infty} \frac{a(\infty)}{a(i)} \neq 0. \quad \text{(IV. 17)}$$

The analyticity of f_+ implies only a finite number of zeros there. Hence follows our statement that there are only a finite number of discrete λ_i. On the unit circle a zero of f_+ is by Eq. (IV. 6), also a zero of f_-. Then $\psi(\lambda, n)$ given by Eq. (IV. 7), is identically zero—unless the denominator also vanishes. Thus, in addition to the finite number of discrete z_i in the unit circle, there are at most zeros at $z=\pm 1$.

The analytic properties of the functions $\psi(\lambda,n)$ which satisfy Eq. (III. 5) and the initial conditions $a(0)\psi(\lambda,-1)=0$, $\psi(\lambda,0)=C$ are even simpler to obtain. Since $\psi(\lambda,n)$ is a polynomial of order n in λ it is analytic in z within the unit circle except for a possible pole at $z=0$. Direct calculation for small z yields

$$\psi(\lambda,n) \approx C z^{-n} \prod_{i=1}^{n} \frac{a(\infty)}{a(i)} . \quad \text{(IV. 17')}$$

Further, information is obtained from the discrete form of Green's identity. Thus let $\psi^{(1)}(\lambda,n)$, $\psi^{(2)}(\lambda',n)$ be any two solutions of Eq. (III. 5). Familiar manipulation yields

$$\begin{aligned}(\lambda-\lambda')\sum_{n=0}^{N}\psi^{(1)}(\lambda,n)\psi^{(2)}(\lambda',n) &= a(N+1)[\psi^{(2)}(\lambda',N)\psi^{(1)}(\lambda,N+1)-\psi^{(1)}(\lambda,N)\psi^{(2)}(\lambda',N+1)] \\ &\quad + a(0)[\psi^{(1)}(\lambda,-1)\psi^{(2)}(\lambda',0)-\psi^{(2)}(\lambda',-1)\psi^{(1)}(\lambda',0)].\end{aligned} \quad \text{(IV. 18)}$$

Suppose first $\lambda=\lambda_i$ (one of the discrete eigenvalues) and $\lambda'=\lambda_i^*$. The right-hand side of Eq. (IV. 17) vanishes as $N\to\infty$ in virtue of the assumed summability of $\psi(\lambda_i,n)$ and the initial conditions

$$\therefore \operatorname{Im}\lambda_i \sum_{n=0}^{\infty} |\psi(\lambda_i,n)|^2 = 0.$$

The λ_i are real. [We therefore can choose the $\psi(\lambda_i,n)$ to be real.]

Second, let $\lambda=\lambda_i$, $\psi^{(1)}=\psi(\lambda_i,n)$ and $\psi^{(2)}(\lambda',n)=\psi_+(\lambda',n)$. Letting $N\to\infty$ and using the boundary conditions yields

$$\sum_{n=0}^{\infty}\psi(\lambda_i,n)\psi_+(\lambda',n) = \frac{f_+(\lambda')C}{\lambda'-\lambda_i} . \quad \text{(IV. 19)}$$

Passing to the limit $\lambda'\to\lambda_i$ and using Eq. (IV. 9), we see that

$$\sum_{n=0}^{\infty} |\psi(\lambda_i,n)|^2 = 1/\rho_i, \quad \text{(IV. 20)}$$

where

$$\rho_i = \frac{\psi_+(z_i,0)}{C^2(df_+/d\lambda)_{\lambda_i}} . \quad \text{(IV. 21)}$$

In particular, this shows $(df_+/d\lambda)_{\lambda_i}\neq 0$—and hence the zeros of f_+ (and hence the λ_i) are simple.

Obviously, also if $\lambda_i\neq\lambda_j$ the sum of the products $\psi^*(\lambda_i,n)\psi(\lambda_j,n)$ are zero, i.e.,

$$\sum_{n=0}^{\infty}\psi^*(\lambda_i,n)\psi(\lambda_j,n) = \frac{\delta(\lambda_i,\lambda_j)}{\rho_i} . \quad \text{(IV. 22)}$$

Finally, in Eq. (IV. 17) let $\psi^{(1)}$ and $\psi^{(2)}$ be two of the regular solutions of the continuum. The terms on the right for $n=0$ vanish in virtue of the initial conditions. For N large we can use Eq. (IV. 8) to evaluate the contribution from the upper limit. The result is

$$\sum_{n=0}^{\infty}\psi(\lambda,n)\psi(\lambda',n) = \frac{\delta(\lambda-\lambda')}{\rho'(\lambda)}, \quad \lambda,\lambda'\in \text{continuum} \quad \text{(IV. 23)}$$

where

$$\rho'(\lambda) = a(\infty)\sin\theta/\pi C^2|f_+|^2 \quad \text{(IV. 24)}$$

Two things may be noted:

(1) Equations (IV. 23) and (IV. 22) together with the obvious relation

$$\sum_{n=0}^{\infty}\psi(\lambda_i,n)\psi(\lambda',n)=0 \quad \begin{cases}\lambda_i \text{ a discrete eigenvalue,} \\ \lambda' \text{ in the continuum,}\end{cases} \quad \text{(IV. 25)}$$

though derived as orthogonality relations are actually the completeness theorem for the polynomials.

(2) The notation is intentional. It will now be shown that the ρ_i and ρ' are indeed related to the weight function $\rho(\lambda)$ used to define the polynomials.

Consider

$$I = \frac{1}{2\pi i}\oint G(\lambda,n;m)\,d\lambda, \quad \text{(IV. 26)}$$

where

$$\begin{aligned} G(\lambda,n;m) &= -\psi(\lambda,n)\psi_+(\lambda,m), \quad n\leq m \\ &= -\psi_+(\lambda,n)\psi(\lambda,m), \quad n\geq m. \end{aligned} \quad \text{(IV. 27)}$$

Here the path in λ is such that z goes around the unit circle. This integral may be expressed in two ways. One is to write it in terms of an integral over λ between $b(\infty)-2a(\infty)$ and $b(\infty)+2a(\infty)$. Alternately, it can be evaluated by residues. The resulting identity is

$$\int_{-\infty}^{\infty}\psi(\lambda,n)\psi(\lambda,m)\,d\rho(\lambda) = \delta(n,m), \quad \text{(IV. 28)}$$

where

$$\begin{aligned} d\rho(\lambda) &= \rho'(\lambda)\,d\lambda, \quad b(\infty)-2a(\infty)\leq\lambda\leq b(\infty)+2(\infty) \\ &= \sum_i \rho_i\delta(\lambda-\lambda_i)\,d\lambda, \quad \lambda \text{ not as above.} \end{aligned}$$

Here the ρ' and ρ_i are indeed as given by Eqs. (IV. 24) and (IV. 22).

Thus we have seen that the weight function for our orthogonal polynomials is closely related to function f_+ which determined many of the properties of the polynomials. Thus:

(i) The values of f_+ on the unit circle determine the continuous part of the weight function $[\rho'(\lambda)]$ by means of Eq. (IV. 24).

(ii) These values of f_+ also determine the number of jump points λ_i. Indeed, by the principle of the argument, the change in phase of f_+ as one goes around the unit circle is $2\pi(N-1)$ where N is the number of zeros of f_+ within the unit circle (the discrete eigenvalues). The 1 is due to the simple pole at the origin,[8] i. e.,

$$\Delta\delta = -2\pi(N-1). \tag{IV. 29}$$

(We recognize a discrete version of Levinson's theorem.[4])

(iii) The actual position of the zeros of f_+ within the unit circle determine the λ_i.

However, in the theory of orthogonal polynomials the problem must be turned around. There $\rho(\lambda)$ is given and f_+ should be found. This can be done using a version of the Poisson-Jensen formula. Thus suppose we have a function $h(z)$ with the properties:

(i) $h(z)$ is analytic within and continuous on the unit circle;

(ii) $h(z)$ is real, i. e., $h^*(z) = h(z^*)$;

(iii) $h(0)$ is real.

Then there is a representation of the form

$$h(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_r(1,\theta)\left(\frac{\exp(i\theta')+z}{\exp(i\theta')-z}\right)d\theta', \quad |z|<1. \tag{IV. 30}$$

It may be noted from what has been stated previously one can choose

$$h(z) = \ln\, zf_+(z)\left(\frac{\Pi_{i=1}^{N}(1-z_i z)}{\Pi_+(z_i-z)\,\Pi_-(z-z_i)}\right). \tag{IV. 31}$$

The requirements are satisfied. (Here $\Pi_\pm$ means products over all z_i in the unit circle subject to $z_i \gtrless 0$.) Therefore,

$$f_+(z) = \frac{\Pi_+(z-z)\,\Pi_-(z-z_i)}{z\,\Pi_i(1-z_i z)}\exp\frac{1}{2\pi} \times \int_{-\pi}^{\pi}\ln|f_+(\theta')|\,\frac{(\exp(i\theta')+z)}{(\exp(i\theta')-z)}\,d\theta'. \tag{IV. 32}$$

Here we have simplified using

$$|z'| = 1 = |1-z_i z'|/|z_i - z'|.$$

It is illuminating to express the integral here in terms of the variables

$$\lambda' = b(\infty) + 2a(\infty)\cos\theta'$$

and

$$\lambda = b(\infty) + a(\infty)(z+z^{-1}).$$

We then find

$$f_+(z) = \frac{\Pi_+(z_i-z)\,\Pi_-(z-z_i)}{z\,\Pi_i(1-z_i z)}\exp I, \tag{IV. 33}$$

with

$$I = \frac{z-z^{-1}}{2\pi}\int_{b(\infty)-2a(\infty)}^{b(\infty)+2a(\infty)} \frac{\ln|f(\lambda')|\,d\lambda'}{\sin\theta'(\lambda'-\lambda)}.$$

Now from Eq. (IV. 24) we see that for the range of integration indicated here

$$\ln|f_+| = -\tfrac{1}{2}\ln\left(\frac{\pi C^2\rho'(\lambda')}{a(\infty)\sin\theta'}\right) \tag{IV. 34}$$

and therefore

$$I = -\frac{(z-z^{-1})}{4\pi}\int_{b(\infty)-2a(\infty)}^{b(\infty)+2a(\infty)}\frac{d\lambda'}{\sin\theta'(\lambda'-\lambda)} \times \ln[\pi C^2\rho'(\lambda')/a(\infty)\sin\theta']. \tag{IV. 35}$$

Thus Eqs. (IV. 33) and (IV. 35) give the Jost function explicitly in terms of the continuous part of the weight function and the position of its jumps.

It may be noted that the construction here is very similar to that encountered in the discrete inverse scattering problem.[1] Here we have constructed f_+ given its absolute value in the continuous range and the position of the discrete eigenvalues. There the problem is to construct f_+ from $\arg f_+$ in the continuous range and the position of the discrete eigenvalues.

Finally, to compare with the usual treatment of orthogonal polynomials let us determine the relation of the functions ψ_+ to "associated functions" $\phi(\lambda, n)$ where

$$\phi(\lambda,n) = \int_{-\infty}^{\infty}\frac{\psi(\lambda',n)\,d\rho(\lambda')}{\lambda-\lambda'}, \quad n \geq -1. \tag{IV. 36}$$

[Note that this implies $\phi(\lambda,-1)\equiv 0$.] It is readily verified that in virtue of the recursion formula for $\psi(\lambda,n)$ and the initial conditions that

$$a(n+1)\phi(\lambda,n+1) + b(n)\phi(\lambda,n) + a(n)\phi(\lambda,n-1) = \lambda\phi(\lambda,n) + \delta(n,0)/C, \quad n \geq 0. \tag{IV. 37}$$

Thus for $n \geq 1$ the $\phi(\lambda,n)$ satisfy exactly the same three term recursion relation as do the $\psi(\lambda,n)$ and $\psi_\pm(\lambda,n)$.

Consider

$$J(z) = \frac{1}{2\pi i}\oint_c \frac{\psi_+(z')}{f_+(z')}\frac{dz'}{z'}\frac{[z'-z'^{-1}]}{[z'+z'^{-1}-(z+z^{-1})]}. \tag{IV. 38}$$

(Here c is the unit circle.) Introducing variables λ' and λ as before, we can (using the results obtained previously) rewrite this as

$$J(z) = C\int_{b(\infty)-2a(\infty)}^{b(\infty)+2a(\infty)}\frac{\psi(\lambda',n)\rho'(\lambda')\,d\lambda'}{\lambda-\lambda'}.$$

On the other hand, $J(z)$ can be evaluated using Cauchy's theorem. The integrand is analytic except for simple poles at the zeros of f_+ and at $z'=z$. The resulting identity then gives the relation

$$\psi^+(z) = Cf_+(z)\phi(\lambda,n). \tag{IV. 39}$$

V. APPLICATIONS

A. Asymptotic formulas

We are now in a position to answer the following question. Given $\rho(\lambda)$ what is the behavior for large n of the appropriately constructed polynomials? Using Eq. (IV. 33) we construct f_+. The asymptotic behavior is then obtained from Eq. (IV. 8).

As a very simple example let us consider the Legendre polynomials. Then

$$d\rho(\lambda) = d\lambda, \quad -1 \leq \lambda \leq 1$$
$$= 0, \quad |\lambda| > 1. \tag{V. 1}$$

Thus the ρ_i are zero—no discrete jump points. Also,

$$b(\infty) - 2a(\infty) = -1$$

and

$$b(\infty)+2a(\infty)=+1. \quad \text{(V.2)}$$

Therefore,

$$b(\infty)=0, \quad a(\infty)=\tfrac{1}{2}. \quad \text{(V.3)}$$

Further,

$$C=1/\sqrt{\textstyle\int_{-\infty}^{\infty}d\rho(\lambda)}=1/\sqrt{2}. \quad \text{(V.4)}$$

The Eq. (IV.33) becomes

$$f_+(z)=z^{-1}\exp\left(\frac{(z-z^{-1})}{4\pi}\int_{-1}^{1}\frac{\ln[\sqrt{1-\lambda'^2}/\pi]d\lambda'}{\sqrt{1-\lambda'^2}\,(\lambda'-\lambda)}\right)$$

$$\lambda=(z+z^{-1})/2, \quad \text{(V.5)}$$

or

$$f_+(z)=[(z^{-2}-1)/2\pi]^{1/2}. \quad \text{(V.6)}$$

Inserting in Eq. (IV.8) then gives

$$\psi(\lambda,n)\sim\left(\frac{2}{\pi\sin\theta}\right)^{1/2}\cos[(n+1/2)\theta-\pi/4], \quad \text{(V.7)}$$

which is the conventional result when we recognize that our polynomials are normalized, i.e.,

$$\psi(\lambda,n)=\sqrt{(2n+1)/2}P_n(\lambda), \quad \text{(V.8)}$$

where $P_n(\lambda)$ are the ordinary Legendre polynomials.

Asymptotic formulas for the associated functions are also readily found. This is done by inserting the asymptotic formula of Eq. (IV.8) into the definition of Eq. (IV.36). In the present example

$$\phi(\lambda,n)=\sqrt{2(2n+1)}Q_n(\lambda) \quad \text{(V.9)}$$

where $Q_n(\lambda)$ are the usual associated Legendre functions. Our procedure then gives the usual asymptotic formulae for the Q_n.

B. Sum rules

From the explicit expression for f_+ in terms of ρ' we can obtain a number of identities which the coefficients $a(n)$, $b(n)$ satisfy. Thus suppose we expand $f_+(z)$ in a Laurent series around the origin. We know this has the form

$$f_+(z)=C_{-1}/z+C_0+C_1z+\cdots. \quad \text{(V.10)}$$

From Eq. (IV.33) we obtain explicit expressions for the C_i in terms of the z_i and various moments of $\ln\rho'$. On the other hand, starting with either Eq. (IV.14) or the recursion relation plus boundary conditions, we can calculate the C_i in terms of the $a(n)$ and $b(n)$. Thus Eq. (IV.17) tells us that

$$C_{-1}=a(\infty)\prod_{i=1}^{\infty}\frac{a(\infty)}{a(i)}. \quad \text{(V.11)}$$

But expanding Eq. (IV.33) for small z gives

$$C_{-1}=\prod_{+}z_i\prod_{-}(-z_i)\exp-\left(\frac{1}{4\pi a(\infty)}\right.$$
$$\left.\times\int_{b(\infty)-2a(\infty)}^{b(\infty)+2a(\infty)}\frac{d\lambda'\ln\pi C^2\rho'(\lambda')/a(\infty)\sin\theta'}{\sin\theta'}\right). \quad \text{(V.12)}$$

Equating the expressions in Eqs. (V.11) and (V.12) then gives $\prod_{i=1}^{\infty}a(i)$ in terms of the z_i and an integral of $\ln\rho'(\lambda')$.

As an example again consider the Legendre polynomials. Then from Eq. (V.6)

$$C_{-1}=(2\pi)^{-1/2}. \quad \text{(V.13)}$$

From (V.11) we then have [since $a(\infty)=1/2$].

$$(2\pi)^{-1/2}=\tfrac{1}{2}\prod_{i=1}^{\infty}1/[2a(i)]. \quad \text{(V.14)}$$

Of course, since we *know* the $a(i)$ for the Legendre case we can check this directly. Thus

$$2a(n)=n/\sqrt{(n^2-1/4)} \quad \text{(V.15)}$$

and Eq. (V.14) is the identity

$$(2\pi)^{-1/2}=\frac{1}{2}\prod_{n=1}^{\infty}\frac{\sqrt{n^2-1/4}}{n}. \quad \text{(V.16)}$$

The coefficient C_0 can also be calculated directly with the result

$$\frac{C_0}{C_{-1}}=-\sum_{i=0}^{\infty}\frac{b'(i)}{b(\infty)}. \quad \text{(V.17)}$$

From Eq. (IV.33) we find the constant term in the Laurent expansion. This combined with Eq. (V.17) then gives $\sum_{i=0}^{\infty}b'(i)$ in terms of integrals of $\ln\rho'$ and the z_i. We note that in many cases, for example the Legendre polynomials, this particular identity is trivial. Indeed, if $d\rho(\lambda)=d\rho(-\lambda)$ then the $b(n)\equiv 0$ and $f_+(z)$ has a Laurent expansion in odd powers of z only. Our identity here is merely that zero equals zero.

In the case

$$b(n)\equiv 0 \quad \text{(V.18)}$$

let us proceed to calculate C_1. We obtain

$$\frac{C_1}{C_{-1}}=-\frac{1}{\pi}\int_0^{\pi}\ln\left(\frac{C^2\rho'(\lambda')}{a(\infty)\sin\theta'}\right)\cos2\theta'\,d\theta'. \quad \text{(V.19)}$$

[Here $\lambda'=a(\infty)\cos\theta'$.]

Again as an example we consider the Legendre case. Then from (V.19) or even more simply by expanding (V.6), we obtain $C_1/C_{-1}=-1/2$.

$$\therefore -\frac{1}{2}=\sum_{n=1}^{\infty}\left(1-\frac{a^2(i)}{a^2(\infty)}\right). \quad \text{(V.20)}$$

In terms of the known values of the coefficients this is the identity

$$-\frac{1}{2}=\sum_{n=1}^{\infty}\left(1-\frac{n^2}{n^2-1/4}\right), \quad \text{(V.21)}$$

which, of course, can be verified directly.

It may be remarked that the sum rules involving C_0 and C_1 are the direct analog of the sum rule given by Eq. (II.29). Indeed, by an appropriate limiting procedure one can pass from either of the present sum rules to Eq. (II.29).

C. Szego-Kac formulas

Szego[9] has given formulas for the determinants of Toeplitz matrices. Kac has given some generalizations.[10] Here we show that the present formalism very readily yields analogous results for the determinants (D_n) of the Hankel forms described in Sec. III. Indeed, the result is merely a rewording of the first sum rule discussed in that section.

We want to discuss the behavior of the D_n of Sec. III in the limit $n\to\infty$.

Rewriting our previous results, we have

$$R=\lim_{z\to 0} zf_+(z)=\frac{(\Pi_+ z_i)\Pi_-(-z_i)}{\pi^{1/2}}\exp I, \tag{V.22}$$

with

$$I=-\frac{1}{2\pi}\int_0^\pi \ln\left(\frac{C^2\rho'}{a(\infty)}\right)d\theta'. \tag{V.23}$$

Alternately, Eq. (IV.17) tells us that

$$R=a(\infty)\prod_{i=1}^{\infty}\frac{a(\infty)}{a(i)}.$$

Let

$$R(n)=a(\infty)\prod_{i=1}^{n}\frac{a(\infty)}{a(i)}. \tag{V.24}$$

[Then $R\equiv R(\infty)$.]

As has been noted [Eq. (III.7)]

$$a(n)=\sqrt{D_nD_{n-2}/D_{n-1}D_{n-1}}.$$

Thus

$$R(n)=a^{n+1}(\infty)\sqrt{D_{n-1}/D_0D_n}. \tag{V.25}$$

From this we readily conclude that

$$\lim_{n\to\infty}(D_n)^{1/n}/a^n(\infty)=a^3(\infty)/D_0R^2 \tag{V.26}$$

This is just the analog for our present problem of the Szego formula for Toeplitz forms.

VI. REMARKS ON THE MOMENT PROBLEM

Since as indicated earlier there is a close connection between orthogonal polynomials and the moment problem it may be useful to see what can be said in the present context.

The problem we consider is the following: Suppose we are given the moments of a distribution function $\rho(\lambda)$. Further, $\rho'(\lambda)$ is continuous over a given finite segment of the real axis. It may in addition have jumps at a finite number of fixed points outside that interval. What is $\rho(\lambda)$? We assume the necessary and sufficient conditions on the moments are satisfied so that a solution exists.[7] (In the present case it is then also unique.[7])

It is then easiest to write down the solution and verify it by inspection. Thus suppose we have a weight function $\sigma(\lambda)$ with precisely the desired points of support. Let $\phi^0(\lambda,n)$ be the corresponding orthogonal polynomials. (Later we indicate how to construct such.) The solution of the problem is clearly

$$d\rho(\lambda)=d\sigma(\lambda)\sum_{n=0}^{\infty}\phi_0(\lambda,n)\int_{-\infty}^{\infty}\phi_0(\lambda',n)\,d\rho(\lambda'). \tag{VI.1}$$

Further, if the sum over n is restricted to terms up to N, the same formula gives a solution to the restricted moment problem—i.e., to determine a $\rho(\lambda)$ which will have a prescribed first N moments.

As a trivial example let us consider the following: There are no jump points and the interval is between -1 and $+1$. The moments are

$$\int_{-1}^{1}\lambda^n\,d\rho(\lambda)=\frac{1}{n+1},\quad n \text{ even},$$
$$=0 \qquad n \text{ odd}. \tag{VI.2}$$

Choose

$$\sigma(\lambda)=0,\quad \lambda<-1,$$
$$=\frac{2}{\pi}\int_{-1}^{\lambda}\sqrt{1-u^2}\,du,\quad -1\le\lambda\le 1, \tag{VI.3}$$
$$=1,\quad \lambda>1.$$

Since this is the weight function for the Tchebycheff polynomials of the second kind we have

$$\phi^0(\lambda,n)=\frac{\sin(n+1)\theta}{\sin\theta},\quad \lambda=\cos\theta. \tag{VI.4}$$

Then

$$\int_{-1}^{1}\phi^0(\lambda,n)=\frac{2}{n+1},\quad n \text{ even}$$
$$=0,\qquad n \text{ odd}.$$

Therefore,

$$d\rho(\lambda)=0,\qquad \lambda<-1$$
$$=d\lambda\sum_{n\text{ even}}\frac{4}{\pi}\frac{\sin(n+1)\theta}{(n+1)},\quad -1\le\lambda\le 1 \tag{VI.5}$$
$$=0,\qquad \lambda>1.$$

But

$$\sum_{n\text{ even}}\frac{4}{\pi}\frac{\sin(n+1)\theta}{n+1}$$

is just the Fourier sine series for the function which is 1 in the interval $0<\theta<\pi$.

$$\therefore d\rho(\lambda)=d\lambda,\quad -1<\lambda<1, \tag{VI.6}$$
$$=0,\quad \text{otherwise}$$

(as of course we knew beforehand).

How do we construct an appropriate $\sigma(\lambda)$ and the associated $\phi^0(\lambda,n)$?

First suppose there are no jump points and the interval is $-1\le\lambda\le 1$. The Legendre weight function $\sigma'(\lambda)=1$ and the associated normalized Legendre polynomials is one possibility. A second is the weight function $\sigma'(\lambda)=(2/\pi)\sqrt{1-\lambda^2}$ and the associated Tchebycheff polynomials $\sin(n+1)\theta/\sin\theta$.

If there are no jumps but the interval is other than -1 to $+1$, we can, with appropriate translation and stretching of coordinates, still use the Legendre or Tchebycheff polynomials.

Finally, to include discrete jump points we use the discrete form of the Gelfand–Levitan equation.[1]

Let $\bar\sigma(\lambda)$, $\bar\phi^0(\lambda,n)$ be appropriate quantities just for the continuous interval. We extend $\bar\sigma(\lambda)$ to an appropriate function $\sigma(\lambda)$ which includes the jumps by writing

$$\begin{aligned} d\sigma(\lambda) &= d\bar{\sigma}(\lambda), \quad \lambda \text{ in the continuum} \\ &= \sum_i \sigma_i \delta(\lambda - \lambda_i)\, d\lambda_i, \quad |\lambda_i| > 1. \end{aligned} \tag{VI.7}$$

Here the σ_i are arbitrary positive constants. What are the corresponding $\phi_0(\lambda, n)$? Note that these are to be polynomials orthogonal with weight $\sigma(\lambda)$. As such they can be constructed by the Hilbert–Schmidt procedure from the linearly independent polynomials $\bar{\phi}_0(\lambda, n)$. Thus there are constants $K(n,m)$ such that

$$\phi_0(\lambda, n) = \sum_{m=0}^{n} K(n,m)\bar{\phi}_0(\lambda, m). \tag{VI.8}$$

The requirement that the $\phi_0(\lambda, n)$ are orthogonal polynomials means they are orthogonal to all polynomials of lower order. In particular,

$$\int_{-\infty}^{\infty} \phi^0(\lambda, n)\bar{\phi}_0(\lambda, m)\, d\sigma(\lambda) = 0, \quad n > m. \tag{VI.9}$$

The normalization condition on the $\phi^0(\lambda, n)$ is

$$\int_{-\infty}^{\infty} \phi^0(\lambda, n)^2\, d\sigma(\lambda) = 1. \tag{VI.10}$$

Inserting the expansion of Eq. (VI.8) into Eqs. (VI.9) and (VI.10) yields the "discrete Gelfand–Levitan equations"[1]:

$$\kappa(n,m) + g(n,m) + \sum_{l=0}^{n-1} \kappa(n,l) g(l,m) = 0 \quad (n > m) \tag{VI.11}$$

and

$$\frac{1}{K(n,n)^2} = 1 + g(n,n) + \sum_{l=0}^{n-1} \kappa(n,l) g(l,n) \tag{VI.12}$$

where

$$\kappa(n,m) = K(n,m)/K(n,n)$$

and

$$g(n,m) = \int_{-\infty}^{\infty} \bar{\phi}_0(\lambda, n)\bar{\phi}_0(\lambda, m)\, d[\sigma(\lambda) - \bar{\sigma}(\lambda)]$$

Having chosen $d\sigma(\lambda)$ as in Eq. (VI.7) makes $g(n,m)$ rather simple. Indeed,

$$g(n,m) = \sum_i \sigma_i \bar{\phi}^0(\lambda_i, n)\bar{\phi}_0(\lambda_i, m). \tag{VI.13}$$

The procedure then is so: Having specified $\bar{\sigma}(\lambda)$ we have the $\bar{\phi}^0$. Choosing σ_i we have $g(n,m)$. Then Eq. (VI.11) is solved for $\kappa(n,m)$. From the Eq. (VI.12) determine $K(n,n)(>0)$ and hence $K(n,m)$. The ϕ^0 then follow from (VI.8). This procedure is not as complicated as one might think. We illustrate with an example.

Suppose the continuum is $-1 \le \lambda \le 1$. In addition there is to be a single jump point at $\lambda_i (|\lambda_i| > 1)$. We choose

$$\begin{aligned} d\bar{\sigma}(\lambda) &= \frac{2}{\pi}\sqrt{1-\lambda^2}\, d\lambda, \quad |\lambda| \le 1 \\ &= 0, \qquad\qquad\qquad |\lambda| > 1. \end{aligned} \tag{VI.14}$$

The $\bar{\phi}^0(\lambda; n)$ are then $\sin(n+1)\theta/\sin\theta$, $\lambda = \cos\theta$. Choose a positive constant σ_1. From Eq. (VI.13)

$$g(n,m) = \sigma_1 \bar{\phi}^0(\lambda_1, n)\bar{\phi}^0(\lambda_1, n).$$

The solutions of the resulting Gelfand–Levitan equations are

$$\kappa(n,m) = \frac{-\sigma_1 \bar{\phi}^0(\lambda_1, n)\bar{\phi}^0(\lambda_1, m)}{1 + \frac{1}{2}\sigma_1[\bar{\phi}^0(\lambda_1, n)\mathring{\bar{\phi}}^0(\lambda_1, n-1) - \mathring{\bar{\phi}}^0(\lambda_1, n-1)\bar{\phi}^0(\lambda_1, n)]} \tag{VI.15}$$

and

$$K(n,n) = \frac{\sqrt{1 + \frac{1}{2}\sigma_1[\mathring{\bar{\phi}}^0(\lambda_1, n)\bar{\phi}^0(\lambda_1, n-1) - \mathring{\bar{\phi}}^0(\lambda_1, n-1)\phi^0(\lambda_1, n)]}}{1 + \frac{1}{2}\sigma_1[\mathring{\bar{\phi}}^0(\lambda_1, n+1)\bar{\phi}^0(\lambda_1, n) - \mathring{\bar{\phi}}^0(\lambda_1, n)\bar{\phi}^0(\lambda_1 \circ n+1)]} \tag{VI.16}$$

(Here $\mathring{\bar{\phi}}^0$ denotes the derivative with respect to λ.)

VII. CONCLUSIONS

It is hoped that it has been demonstrated that there is a very close parallel between the theory of orthogonal polynomials and scattering theory. In both a fundamental role is played by the value of a particular solution of the basic (differential or difference) equation. We have called this the Jost function in both cases. Under fairly general conditions this function has simple analytic properties. It determines the asymptotic behavior of the solutions of our equations and the continuous part of the spectral (weight) function. Conversely, the asymptotic phases plus the position of the bound states (jump points) determines the Jost function. Alternatively, an explicit expression for the Jost function can be written in terms of the spectral (weight) function.

We would like to stress that this is more than an analogy. Every result we have obtained for orthogonal polynomials of the type considered here leads by an appropriate limiting procedure to a result in scattering theory.

*This work was supported in part by the Air Force Office of Scientific Research, Grant 722187.

[1]K. M. Case and M. Kac, J. Math. Phys. 14, 594 (1973).
[2]Here we merely give a brief sketch with no proofs. For details we refer the reader to the literature, e.g., see R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[3]R. Jost and W. Kohn, Phys. Rev. 87, 977 (1952).
[4]N. Levinson, Phys. Rev. 75, 1445 (1949).
[5]I. M. Gelfand and B. M. Levitan, Izv. Akad. Nauk, USSR 15, 309 (1951) [Am. Math. Soc. Transl. 1, 253 (1956)].
[6]Compare E.C. Titchmarsh, *The Theory of Functions* (Oxford U. P., London, 1939).
[7]See, for example, N.I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, Edinburgh, 1965).
[8]For simplicity we here assume that no λ_i is at the edge of the continuum. If not the usual modifications of Levinson's theorem (Ref. 4) apply.
[9]G. Szegö, Communications du Seminaire Mathematique de l'Université de Lünd, tome supplémentaire dédié A M. Riesz, 228 (1952).
[10]M. Kac, Duke Math. J. 21, 501 (1954).

A remark on the Green's function for the face-centered cubic lattices

Koichi Mano

Air Force Cambridge Research Laboratories, Bedford, Massachusetts 01730

(Received 15 July 1974)

The lattice Green's function $G(2p,0,0)$ for the face-centered cubic lattices, which was obtained by Inoue as a linear combination of the F_4 function of Appell, is shown to be expressible as the product of the pth derivatives of two complete elliptic integrals of the first kind.

In a recent paper Inoue[1] has demonstrated that the knowledge on $G(2p, 0, 0)$ ($p = 0$ or positive integer) and $G(2, 2, 0)$ suffices to determine the entire family of the Green's function, $G(l, m, n)$ [defined by Eq. (1) below] for the face-centered cubic lattices. Based on this observation she concentrated on $G(2p, 0, 0)$ and evaluated it as a linear combination of finite number of F_4 functions of Appell.

In this paper we would like to show that her final result for $G(2p, 0, 0)$ can be expressed as a simple product of the pth derivatives of two complete elliptic integrals of the first kind.

The case for $l = 2p$, $m = n = 0$ of the Green's function $G(l, m, n)$ for the face-centered cubic lattices, defined by

$$G(l,m,n) = \frac{1}{\pi^3}\iiint_0^1 \frac{\cos lx \cos my \cos nz}{E - i\delta - \cos x\cos y - \cos x \cos z - \cos y \cos z}, \tag{1}$$

has been evaluated (for $E > 3$) by Inoue as

$$G(2p, 0, 0) = \frac{1}{2^{2p}(1+E)^{1+p/2}E^{p/2}p!} \sum_{j=0}^{p}\sum_{\mu=0}^{p-j}\sum_{\mu'=0}^{j} C^{p-j,j}_{\mu,\mu'} \times F_4(\mu+\tfrac{1}{2};\mu'+\tfrac{1}{2};p-j+1,j+1;X_+,X_-). \tag{2}$$

In Eq. (2), the coefficient C and the variables $X_\pm$ are given, respectively, by

$$C^{p-j,j}_{\mu,\mu'} = (-1)^j \binom{p}{j}^2\binom{p-j}{\mu}\left(\frac{1}{2}\right)_\mu\left(\frac{1}{2}\right)_{p-j-\mu}\binom{j}{\mu'}\left(\frac{1}{2}\right)_{\mu'}\left(\frac{1}{2}\right)_{j-\mu'} \tag{3}$$

and

$$X_\pm = (\sqrt{1+E} \mp \sqrt{E})^2/(1+E)^2, \tag{4}$$

with the usual Pochhammer notation $(a)_n = \Gamma(a+n)/\Gamma(a)$.

In order to achieve our goal we first note the following integral representation for F_4 [2]:

$$\begin{aligned} F_4(\alpha;\beta;\gamma,\gamma';\xi(1-\eta),\eta(1-\xi)) &= \frac{\Gamma(\gamma)\Gamma(\gamma')}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma-\alpha)\Gamma(\gamma'-\beta)} \\ &\times \iint_0^1 du\,dv\,u^{\alpha-1}v^{\beta-1}(1-u)^{\gamma-\alpha-1}(1-v)^{\gamma'-\beta-1}(1-u\xi)^{\alpha-\gamma-\gamma'-1} \\ &\times(1-v\eta)^{\beta-\gamma-\gamma'-1} \\ &\times(1-u\xi-v\eta)^{\gamma+\gamma'-\alpha-\beta-1} \\ &(\mathrm{Re}\,\alpha>0,\ \mathrm{Re}\,\beta>0,\ \mathrm{Re}(\gamma-\alpha)>0,\ \mathrm{Re}(\gamma'-\beta)>0). \end{aligned} \tag{5}$$

Then, by reinstating the $\Gamma(a+n)/\Gamma(a)$ form for the symbol $(a)_n$ and by using Eq. (5) with $X_+ = \xi(1-\eta)$ and $X_- = \eta(1-\xi)$, we can rewrite the sums, which we will call S hereafter, of Eq. (2) as

$$\begin{aligned} S &= \frac{(p!)^2}{\pi^2}\iint_0^1 du\,du^{-1/2}v^{-1/2}(1-u)^{-1/2}(1-v)^{-1/2} \\ &\times(1-u\xi)^{-p-1/2}(1-v\eta)^{-p-1/2}(1-u\xi-v\eta)^p \\ &\times\sum_{j=0}^{p}(-1)^j(1-u)^{p-j}(1-v)^j\sum_\mu \frac{1}{\mu!\,\Gamma(p-j-\mu+1)} \\ &\times\left(\frac{u(1-u\xi)}{(1-u)(1-u\xi-v\eta)}\right)^\mu \sum_{\mu'}\frac{1}{\mu'!\,\Gamma(j-\mu'+1)} \\ &\times\left(\frac{v(1-v\eta)}{(1-v)(1-u\xi-v\eta)}\right)^{\mu'}. \end{aligned} \tag{6}$$

If we use the relation

$$\frac{\Gamma(z+1)}{\Gamma(z-n+1)} = (-1)^n\frac{\Gamma(-z+n)}{\Gamma(-z)} \quad (n = 0 \text{ or positive integer}) \tag{7}$$

to rewrite $1/\Gamma(j-\mu'+1)$, the μ' sum yields

$$\begin{aligned} &\frac{1}{j!}\sum_{\mu'}\frac{(-j)_{\mu'}}{\mu'!}\left(-\frac{v(1-v\eta)}{(1-v)(1-u\xi-v\eta)}\right)^{\mu'} \\ &= \frac{1}{j!}\left(1+\frac{v(1-v\eta)}{(1-v)(1-u\xi-v\eta)}\right)^j = \frac{1}{j!}\,\frac{(1-u\xi-v\eta+uv\xi)^j}{(1-v)^j(1-u\xi-v\eta)^j}. \end{aligned} \tag{8}$$

Analogously, the μ sum in Eq. (6) gives

$$\frac{1}{(p-j)!}\,\frac{(1-u\xi-v\eta+uv\eta)^{p-j}}{(1-u)^{p-j}(1-u\xi-v\eta)^{p-j}}. \tag{9}$$

When Eqs. (8) and (9) are substituted into Eq. (6) it will be seen that the j sum is again of the same type (i.e., binomial) as for the μ' and μ sums, with the result

$$\frac{[uv(\eta-\xi)]^p}{p!\,(1-u\xi-v\eta)^p}. \tag{10}$$

Collecting the results found in the above, we obtain

$$\begin{aligned} S &= \frac{(\eta-\xi)^p p!}{\pi^2}\iint_0^1 du\,dv\,u^{p-1/2}v^{p-1/2}(1-u)^{-1/2}(1-v)^{-1/2} \\ &\times(1-u\xi)^{-p-1/2}(1-v\eta)^{-p-1/2}. \end{aligned} \tag{11}$$

Since the u and v integrals of Eq. (11) are separable and each gives a ${}_2F_1$ function, we have

$$S = \frac{(\eta-\xi)^p[(\frac{1}{2})_p]^2}{p!}\,{}_2F_1\left(\begin{matrix}\frac{1}{2}+p,\frac{1}{2}+p;\xi\\ 1+p\end{matrix}\right){}_2F_1\left(\begin{matrix}\frac{1}{2}+p,\frac{1}{2}+p;\eta\\ 1+p\end{matrix}\right). \tag{12}$$

If we note the relation

$$\left(\frac{d}{dz}\right)^p {}_mF_n\begin{pmatrix}a_1,\ldots,a_m;z\\ b_1,\ldots,b_n\end{pmatrix} = \frac{(a_1)_p\cdots(a_m)_p}{(b_1)_p\cdots(b_n)_p} \times {}_mF_n\begin{pmatrix}a_1+p,\ldots,a_m+p;z\\ b_1+p,\ldots,b_n+p\end{pmatrix},$$

we may further simplify the product of the two ${}_2F_1$'s of Eq. (12) as

$$\left(\frac{p!}{[(\frac{1}{2})_p]^2}\right)^2\left(\frac{d}{d\xi}\right)^p\left(\frac{d}{d\eta}\right)^p {}_2F_1\begin{pmatrix}\frac{1}{2},\frac{1}{2};\xi\\ 1\end{pmatrix} {}_2F_1\begin{pmatrix}\frac{1}{2},\frac{1}{2};\eta\\ 1\end{pmatrix}$$
$$= \left(\frac{p!}{[(\frac{1}{2})_p]^2}\cdot\frac{2}{\pi}\right)^2 \frac{d^pK(\sqrt{\xi})}{d\xi^p}\,\frac{d^pK(\sqrt{\eta})}{d\eta^p}, \qquad (13)$$

where $K(\cdot)$ stands for the complete elliptic integral of the first kind.

Now we recall that ξ and η in the above expressions are to be determined from $\xi(1-\eta)=X_+$ and $\eta(1-\xi)=X_-$. This means that they are to be identified with k_+^2 and k_-^2, respectively, of Ref. 1, and hence $\eta-\xi$ is to be replaced by $4E^{1/2}(1+E)^{1/2}/(1+E)^2$.

With this value for $\eta-\xi$ we summarize our results, Eqs. (12) and (13), as

$$G(2p,0,0)=\frac{[(\frac{1}{2})_p]^2}{(1+E)^{2p+1}(p!)^2}\,{}_2F_1\begin{pmatrix}\frac{1}{2}+p,\frac{1}{2}+p;k_+^2\\ 1+p\end{pmatrix} \times {}_2F_1\begin{pmatrix}\frac{1}{2}+p,\frac{1}{2}+p;k_-^2\\ 1+p\end{pmatrix} \qquad (14)$$

$$=\frac{4}{(1+E)^{2p+1}[\pi(\frac{1}{2})_p]^2}\,\frac{d^pK(k_+)}{d(k_+^2)^p}\,\frac{d^pK(k_-)}{d(k_-^2)^p}. \qquad (15)$$

That Eqs. (14) and (15) give identical result to that of Ref. 1 can readily be checked for $p=0$ and 1. Thus, for $p=0$, they give $[4/\pi^2(1+E)]K(k_+)K(k_-)$ which is Eq. (3.17a) of Ref. 1. For this case we may of course go back to Eq. (2) which gives $G(0,0,0)\propto F_4(\frac{1}{2};\frac{1}{2};1,1;X_+,X_-)$. If we use Eq. (5) for this F_4, we find also that $G(0,0,0) \propto {}_2F_1(\frac{1}{2},\frac{1}{2};1;\xi){}_2F_1(\frac{1}{2},\frac{1}{2};1;\eta)\propto K(\sqrt{\xi})K(\sqrt{\eta})$. For $p=1$, Eq. (15) with the relation

$$\frac{dK(k)}{dk}=\frac{E(k)}{kk'^2}-\frac{K(k)}{k},$$

where $k'^2=1-k^2$ and $E(k)$ stands for the complete elliptic integral of the second kind, yields

$$\frac{4}{(1+E)^3\pi^2k_+k_-}\left(\frac{E(k_+)}{k_+(1-k_+^2)}-\frac{K(k_+)}{k_+}\right)\left(\frac{E(k_-)}{k_-(1-k_-^2)}-\frac{K(k_-)}{k_-}\right),$$

which coincides with Eq. (3.17b) of Ref. 1.

[1]M. Inoue, J. Math. Phys. 15, 704 (1974).
[2]A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, p. 230.

Statistical theory of effective electrical, thermal, and magnetic properties of random heterogeneous materials. III. Perturbation treatment of the effective permittivity in completely random heterogeneous materials

Motoo Hori

Department of Applied Physics, Tokyo Institute of Technology, Meguroku, Tokyo 152, Japan

Fumiko Yonezawa

Belfer Graduate School of Science, Yeshiva University, 2495 Amsterdam Avenue, New York, New York 10033

Department of Applied Physics, Tokyo Institute of Technology, Meguroku, Tokyo 152, Japan
(Received 28 December 1973)

Perturbation expansion series are derived for the effective permittivity of completely random heterogeneous materials. The formulation is performed by regarding a completely random material as a limiting case of an isotropic cell material. It is emphasized that, in order to obtain a physically reasonable and mathematically correct result, the "exclusion effect" must be taken into account in the averaging procedure. Prescription for evaluating the perturbation coefficient of an arbitrary order is given and explicit forms of leading terms are presented. The results bear a wide variety of applications in calculating effective physical constants such as dielectric constant, magnetic permeability, electrical and thermal conductivity, and diffusion constant. It is mentioned that the idea and formulation in this article are important for more general approximations (to be studied in the succeeding Paper IV) especially in connection with the problem of electron localization in some disordered systems.

1. INTRODUCTION

This paper is a sequel to two previous papers,[1,2] which will hereafter be referred to as I and II. In I, a general perturbation formulation was developed for the effective permittivity of random heterogeneous materials that are statistically homogeneous but not necessarily statistically isotropic.[3] Furthermore, the second-order and third-order perturbation terms were calculated explicitly on the basis of the modified cell model. Unfortunately, however, it is difficult to evaluate higher-order perturbation terms for cell materials. In II, therefore, we derived upper and lower bounds of the effective permittivity, taking account of the third-order perturbation effects.[4] The present and succeeding papers[5] (henceforth called III and IV, respectively) deal with the higher-order perturbation calculation for the effective permittivity of a completely random heterogeneous material, which we shall regard as a limiting case of a spherical-cell material.

Now, in order to explain effectively the purposes and meanings of our papers III and IV, let us briefly describe the present status of research in this field. Historically speaking, the problem of evaluating effective or overall properties of inhomogeneous systems has repeatedly attracted many authors not only in physics but also in various other fields[6] and it is truly amazing to know that essentially the same idea or method has been reinvented and used quite independently by a number of people in different fields. One of the typical examples is the "effective-medium (EM) theory," which we shall treat in IV. In solid state physics, this is termed the coherent-potential approximation (CPA) which itself has been attained through various independent and rather different approaches.[7] In the study of classical mixtures, the EM theory is introduced to obtain the effective physical constants on the basis of the self-consistent local field concept.[8] In contrast to the theory of dilute suspensions,[9] the validity of the self-consistent method is not restricted to systems with low concentrations of inclusions, but it may serve as a good approximation for a whole range of the concentration.

Recently classical treatments of random media have been brought into the spotlight in connection with the discussion of electron localization in some inhomogeneous materials.[10] The EM theories especially are regarded as very useful approximations for the study of overall physical properties in some parameter region of disordered systems.[11,12] It is also pointed out that the EM theory for scalar conductivity, wonderfully as it works for some parameter regions, does not hold good near the critical percolation concentrations for a random mixture of conducting and insulating materials.

In view of these facts, one of our purposes is to observe a more direct relationship between the EM theory and the CPA. Although it has been mentioned that the fundamental philosophy characterizing the EM theory is akin to that for the CPA,[6,11] there has been no detailed proof to show that the mathematical structures of both theories are identical. Since the validity of the CPA has been so well established, the equivalence between the EM theory and the CPA guarantees the usefulness of the EM theory. Another purpose of ours is to see whether we can improve the EM theory so that the results are also available near the critical concentration region. These purposes will be completed in IV, where a proof of the above-described equivalence is given and an approximation much better than the EM theory is proposed. Meanwhile, we give some preparatory formulations in this article, bearing these final aims in mind.

Readers are not necessarily required to be acquainted with the preceding papers I and II. In Sec. 2, some brief review of I and II is given together with the explanation of notations and the formal expansion series of the effective permittivity. In Sec. 3, we are mainly concerned with the perturbation methods for the effective permittivity in completely random media. A particular attention is paid to the concept and importance of the "exclusion effect" in the process of evaluating the average of the perturbation terms.[13] We give a general prescription to construct the averaged perturbation terms of arbitrary orders when the exclusion effects are adequately taken into account. The results are compared with Kröner's perturbation solution[14–16] in which exclusion effect is neglected. Explicit forms of the first few terms in the perturbation series are obtained in Sec. 4. A summary and some preparatory remarks on the next paper IV are given in Sec. 5. It is suggested that the exclusion effect becomes significant especially when a partial summation of the perturbation series is employed as an approximate solution to the effective permittivity. Detailed discussion of criteria for summing up some important terms in the perturbation series will be postponed until we have a thorough thought of it in IV. To readers who are not specifically interested in the detailed mathematical analyses, it is suggested to skip Secs. 2 to 4 and study the summary in Sec. 5 before they proceed to IV.

It must be noted that, although all formulations in what follows are carried out in the language of dielectric constant, the whole argument and result hold right for any other kind of physical constant defined as a proportionality factor between a solenoidal vector and an irrotational vector. The problem of scalar conductivity especially is of great physical interest.

2. GENERAL FORMULATION AND PRELIMINARY RESULTS

We consider a random heterogeneous material with spatially fluctuating permittivity $\epsilon(\mathbf{r})$ subjected to a constant-average electric field. Assume that the medium is statistically homogeneous and its volume V is infinite. Statistical homogeneity means that all of the many-point moments of $\epsilon(\mathbf{r})$ do not depend upon the absolute position of the points but only upon their relative configuration. Denote ensemble averaging by the brackets $\langle\ \rangle$ and let $\epsilon'(\mathbf{r})=\epsilon(\mathbf{r})-\langle\epsilon\rangle$. Then we can write

$$\begin{aligned}\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\cdots\epsilon'(\mathbf{r}_n)\rangle&=\langle\epsilon'(0)\epsilon'(\mathbf{r}_{12})\cdots\epsilon'(\mathbf{r}_{1n})\rangle\\&=\langle\epsilon'^n\rangle f(\mathbf{r}_{12},\mathbf{r}_{13},\ldots,\mathbf{r}_{1n})=\langle\epsilon'^n\rangle g(\mathbf{r}_{12},\mathbf{r}_{23},\ldots,\mathbf{r}_{n-1,n}),\end{aligned}\tag{2.1}$$

where $\mathbf{r}_{ij}$ designates the relative position $\mathbf{r}_j-\mathbf{r}_i$ and $g(\mathbf{r}_{12},\mathbf{r}_{23},\ldots,\mathbf{r}_{n-1,n})$ represents the normalized n-point correlation function of $\epsilon(\mathbf{r})$.

As shown in I, the effective permittivity tensor ϵ^*_{ij} for such a heterogeneous material with homogeneous statistics may be expanded in a series of the form

$$\epsilon^*_{ij}=\langle\epsilon\rangle\left(\delta_{ij}-\sum_{n=2}^{\infty}(-1)^n A^{(n)}_{ij}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n}\right),\tag{2.2}$$

δ_{ij} indicating the Kronecker delta. The nth-order perturbation coefficient $A^{(n)}_{ij}$ is related to the n-point correlation function $g(\mathbf{r}_{12},\mathbf{r}_{23},\ldots,\mathbf{r}_{n-1,n})$ by

$$A^{(n)}_{ij}=\left(-\frac{1}{4\pi}\right)^{n-1}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n-1,n}\times\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^{n-1}g(\mathbf{r}_{12},\mathbf{r}_{23},\ldots,\mathbf{r}_{n-1,n})}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n-1,n;j}}.\tag{2.3}$$

Here, for example, $d\omega_{12}$ is a volume element in the space of $\mathbf{r}_{12}$, and the summation convention has been employed. When the coordinate axes are chosen to coincide with the principal axes of ϵ^*_{ij},

$$\epsilon^*_i=\langle\epsilon\rangle\left(1-\sum_{n=2}^{\infty}(-1)^n A^{(n)}_i\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n}\right).\tag{2.4}$$

Needless to say, ϵ^*_i and $A^{(n)}_i$ are the eigenvalues of the tensors ϵ^*_{ij} and $A^{(n)}_{ij}$, respectively.

For a statistically isotropic medium the effective permittivity tensor ϵ^*_{ij} reduces to a scalar ϵ^* such that

$$\epsilon^*=\langle\epsilon\rangle\left(1-\sum_{n=2}^{\infty}(-1)^n A^{(n)}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n}\right),\tag{2.5}$$

and

$$A^{(n)}=\frac{1}{3}\left(-\frac{1}{4\pi}\right)^{n-1}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n-1,n}\times\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^{n-1}g(\mathbf{r}_{12},\mathbf{r}_{23},\ldots,\mathbf{r}_{n-1,n})}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n-1,n;i}}.\tag{2.6}$$

In particular, $A^{(2)}$ is given by

$$A^{(2)}=-\frac{1}{3\cdot 4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial g(r_{12})}{\partial x_{12,i}}=\frac{1}{3}\tag{2.7}$$

which is an immediate consequence of

$$A^{(2)}_{ii}=-\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial g(\mathbf{r}_{12})}{\partial x_{12,i}}=1.\tag{2.8}$$

Boundary conditions under which Eq. (2.8) holds are expressed in spherical coordinates as

$$\lim_{r_{12}\to\infty} r_{12}\frac{\partial g(r_{12},\theta_{12},\phi_{12})}{\partial r_{12}}=0,\tag{2.9a}$$

$$\lim_{r_{12}\to\infty}\frac{\partial g(r_{12},\theta_{12},\phi_{12})}{\partial\theta_{12}}=0,\tag{2.9b}$$

$$\lim_{r_{12}\to\infty}\frac{\partial g(r_{12},\theta_{12},\phi_{12})}{\partial\phi_{12}}=0.\tag{2.9c}$$

According to Miller,[17,18] a symmetric cell material is defined as a random multiphase material that satisfies the following requirements:

(i) The material space is completely covered by nonoverlapping cell;

(ii) cells are distributed in a manner such that the material is statistically homogeneous;

(iii) the material property ϵ of a cell is statistically independent of the material property of any other cell;

(iv) the conditional probabilities of n points being and n' points not being in the same cell of a particular material, given that one point is in a cell of that material, are the same for each material.

Henceforth we shall restrict ourselves to a symmetric cell material composed of cells of uniform shape, size, and orientation. The independence hypothesis (iii) asserts that the two-point moment $\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle$ vanishes whenever both of the points $\mathbf{r}_1$ and $\mathbf{r}_2$ are not contained in the same cell; that is,

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle=\langle\epsilon'^2\rangle g(\mathbf{r}_{12})=\langle\epsilon'^2\rangle P(\mathbf{r}_1,\mathbf{r}_2), \tag{2.10}$$

where $P(\mathbf{r}_1,\mathbf{r}_2)$ stands for the probability that the two points $\mathbf{r}_1$ and $\mathbf{r}_2$ fall into the same cell. Likewise, if we denote by $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ the probability that all the points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ lie in the same cell,

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\rangle=\langle\epsilon'^3\rangle g(\mathbf{r}_{12},\mathbf{r}_{23})=\langle\epsilon'^3\rangle P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3). \tag{2.11}$$

Substitution of Eqs. (2.10) and (2.11) into Eq. (2.3) yields

$$A_{ij}^{(2)}=-\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial P(\mathbf{r}_1,\mathbf{r}_2)}{\partial x_{12,j}}, \tag{2.12}$$

$$A_{ij}^{(3)}=\frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{\partial^2 P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)}{\partial x_{12,k}\partial x_{23,j}}. \tag{2.13}$$

It was pointed out in I that the second-order and third-order perturbation coefficients may be represented as functions of the cell shape. Recall that $A_{ij}^{(2)}$ is equal to the magnetometric demagnetization tensor[19,20] of the cell. In terms of the point-function demagnetization tensor[20] $L_{ij}(\mathbf{r})$, we obtain

$$A_{ij}^{(2)}=\frac{1}{v}\int_v d\omega L_{ij}(\mathbf{r}), \tag{2.14}$$

$$A_{ij}^{(3)}=\frac{1}{v}\int_v d\omega L_{ik}(\mathbf{r})L_{kj}(\mathbf{r}), \tag{2.15}$$

v being the cell volume. When the medium consists of ellipsoidal cells, the point-function demagnetization tensor $L_{ij}(\mathbf{r})$ becomes constant throughout the cell volume, so that

$$A_{ij}^{(2)}=L_{ij},\quad A_{ij}^{(3)}=L_{ik}L_{kj}. \tag{2.16}$$

Especially, for an isotropic cell material comprising cells of spherical shape,

$$A^{(2)}=1/3,\quad A^{(3)}=1/9. \tag{2.17}$$

Higher-order perturbation coefficients are by far more difficult to compute than the second- or third-order perturbation coefficient. Take $A_{ij}^{(4)}$ as an example. The four-point moment $\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle$ assumes $\langle\epsilon'^4\rangle$ when four points are in the same cell, $\langle\epsilon'^2\rangle^2$ when two pairs of points are in two different cells, and zero otherwise. Let $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ be the probability that the points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$ are in the same cell, $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ the probability that the two pairs of points, $(\mathbf{r}_1,\mathbf{r}_2)$ and $(\mathbf{r}_3,\mathbf{r}_4)$, are in two different cells, and so on. Then we have

$$\begin{aligned}\langle\epsilon'(\mathbf{r}_1)&\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle\\&=\langle\epsilon'^4\rangle g(\mathbf{r}_{12},\mathbf{r}_{23},\mathbf{r}_{34})=\langle\epsilon'^4\rangle P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)\\&\quad+\langle\epsilon'^2\rangle^2[P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)+P(\mathbf{r}_1,\mathbf{r}_4;\mathbf{r}_2,\mathbf{r}_3)\\&\quad+P(\mathbf{r}_1,\mathbf{r}_3;\mathbf{r}_2,\mathbf{r}_4)].\end{aligned} \tag{2.18}$$

Consequently, $A_{ij}^{(4)}$ can be separated into four parts as

$$A_{ij}^{(4)}=A_{1,ij}^{(4)}+\frac{\langle\epsilon'^2\rangle^2}{\langle\epsilon'^4\rangle}(A_{2,ij}^{(4)}+A_{3,ij}^{(4)}+A_{4,ij}^{(4)}). \tag{2.19}$$

Here

$$A_{1,ij}^{(4)}=-\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}\times\frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \tag{2.20}$$

$$A_{2,ij}^{(4)}=-\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}\times\frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \tag{2.21}$$

$$A_{3,ij}^{(4)}=-\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}\times\frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_4;\mathbf{r}_2,\mathbf{r}_3)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \tag{2.22}$$

$$A_{4,ij}^{(4)}=-\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}\times\frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_3;\mathbf{r}_2,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}. \tag{2.23}$$

We notice that the probability $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ concerns the geometry of a single cell, while the quantities like $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ refer to the mutual relation between different cells. For this reason, $A_{1,ij}^{(4)}$ is calculated analogously to $A_{ij}^{(2)}$ or $A_{ij}^{(3)}$, provided that identical ellipsoidal cells are uniformly oriented. As expressions corresponding to Eqs. (2.16) and (2.17), we get

$$A_{1,ij}^{(4)}=L_{ik}L_{kh}L_{hj}, \tag{2.24}$$

$$A_1^{(4)}=1/27, \tag{2.25}$$

the proof of which will be presented in Appendix A. On the contrary, it is practically impossible to determine $A_{2,ij}^{(4)}$, $A_{3,ij}^{(4)}$, and $A_{4,ij}^{(4)}$ exactly. In addition, Eqs. (2.24) and (2.25) are easily generalized to

$$\begin{aligned}A_{1,ij}^{(n)}&=\left(-\frac{1}{4\pi}\right)^{n-1}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n-1,n}\\&\quad\times\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^{n-1}P(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n)}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n-1,n;j}}\\&=L_{ik}L_{kh}\cdots L_{mj}\quad[(n-1)\text{-fold product}],\end{aligned} \tag{2.26}$$

$$A_1^{(n)}=1/3^{n-1}. \tag{2.27}$$

In II we derived effective permittivity bounds involving the three-point correlation functions. For cell materials composed of ellipsoidal cells these bounds are

$$\epsilon_i^*\leq\langle\epsilon\rangle\left(1-\frac{(L_{(i)}\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2)^2}{L_i\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2+L_{(i)}^2\langle\epsilon'^3\rangle/\langle\epsilon\rangle^3}\right), \tag{2.28}$$

$$\epsilon_i^*\geq\frac{1}{\langle1/\epsilon\rangle}\left\{1-(1-L_{(i)})^2\left(1-\frac{1}{\langle1/\epsilon\rangle\langle\epsilon\rangle}\right)^2\times\left[(1-2L_i+L_{(i)}^2)\left(1-\frac{1}{\langle1/\epsilon\rangle\langle\epsilon\rangle}\right)+(L_i-L_{(i)}^2)\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2}\right]^{-1}\right\}^{-1}, \tag{2.29}$$

where L_i signifies the depolarizing factor of the ellipsoid and the index (i) indicates no summation on i. The above inequalities are simplified for spherical-cell materials to

$$\epsilon^* \le \langle\epsilon\rangle \left(1 - \frac{\langle\epsilon'^2\rangle^2/\langle\epsilon\rangle^4}{3\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2 + \langle\epsilon'^3\rangle/\langle\epsilon\rangle^3}\right), \tag{2.30}$$

$$\epsilon^* \ge \frac{1}{\langle 1/\epsilon\rangle}\left\{1 - 2\left(1 - \frac{1}{\langle 1/\epsilon\rangle\langle\epsilon\rangle}\right)^2 \times \left[2\left(1 - \frac{1}{\langle 1/\epsilon\rangle\langle\epsilon\rangle}\right) + \frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2}\right]^{-1}\right\}^{-1}. \tag{2.31}$$

3. COMPLETELY RANDOM MEDIA AND EXCLUSION EFFECT

Now, as a prototype of classical mixtures, let us consider a completely random heterogeneous material or a perfectly disordered composite material.[14–16] By a completely random material we mean a random inhomogeneous material in which physical constants at different points are statistically independent. For instance, the two-point moment $\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle$ becomes

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle = \langle\epsilon'^2\rangle g(\mathbf{r}_{12}) = \langle\epsilon'^2\rangle\,\delta_{\mathbf{r}_{12}}, \tag{3.1}$$

where $\delta_{\mathbf{r}_{12}}$ is a null function such that

$$\delta_{\mathbf{r}_{12}} = \begin{cases} 1 & \text{for } \mathbf{r}_1 = \mathbf{r}_2, \\ 0 & \text{for } \mathbf{r}_1 \ne \mathbf{r}_2. \end{cases} \tag{3.2}$$

Similarly,

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\rangle = \langle\epsilon'^3\rangle g(\mathbf{r}_{12}, \mathbf{r}_{23}) = \langle\epsilon'^3\rangle\,\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}. \tag{3.3}$$

It should be noted that the completely random material is not only statistically homogeneous but also statistically isotropic.

This model of completely random mixtures has been proposed by Kröner[14] to analyze the elastic behavior of composite materials. For later convenience, we rephrase here Kröner's treatment in the language of the dielectric constant. Because of the statistical isotropy of the medium, Eq. (2.7) is valid for $g(\mathbf{r}_{12}) = \delta_{\mathbf{r}_{12}}$; namely,

$$A^{(2)} = -\frac{1}{4\pi}\int_V d\omega_{12}\,\frac{x_{12,i}}{r_{12}^3}\,\frac{\partial\delta_{\mathbf{r}_{12}}}{\partial x_{12,(i)}} = \frac{1}{3}. \tag{3.4}$$

From Eq. (3.3) we find

$$A^{(3)} = \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\,\frac{x_{12,i}}{r_{12}^3}\,\frac{x_{23,k}}{r_{23}^3}\,\frac{\partial^2\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}}{\partial x_{12,k}\partial x_{23,(i)}} = \tfrac{1}{3}\delta_{ik}\cdot\tfrac{1}{3}\delta_{k(i)} = \tfrac{1}{9}. \tag{3.5}$$

Let us express the four-point moment as

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle = \langle\epsilon'^4\rangle g(\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{34}) = \langle\epsilon'^4\rangle\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}} + \langle\epsilon'^2\rangle^2(\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}} + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}} + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}), \tag{3.6}$$

and put

$$\frac{\partial^3\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}} = 0, \tag{3.7a}$$

$$\frac{\partial^3\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}} = 0, \tag{3.7b}$$

$$\frac{\partial^3\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}} = 0. \tag{3.7c}$$

The assumptions (3.6) and (3.7) are exactly what Kröner implicitly used in obtaining his solution although he himself noticed this only later.[21] Insertion of Eq. (3.6) into Eq. (2.6) gives

$$A^{(4)} = 1/27. \tag{3.8}$$

More generally,

$$A^{(n)} = 1/3^{n-1}, \tag{3.9}$$

which leads to

$$\frac{\epsilon^*}{\langle\epsilon\rangle} = 1 - \frac{\langle\epsilon'^2\rangle}{3\langle\epsilon\rangle^2} + \frac{\langle\epsilon'^3\rangle}{9\langle\epsilon\rangle^3} - \frac{\langle\epsilon'^4\rangle}{27\langle\epsilon\rangle^4} + \frac{\langle\epsilon'^5\rangle}{81\langle\epsilon\rangle^5} - \cdots = 1 - \frac{1}{\langle\epsilon\rangle}\left\langle\frac{\epsilon'^2}{3\langle\epsilon\rangle + \epsilon'}\right\rangle = \left\langle\frac{1 + 4\epsilon'/3\langle\epsilon\rangle}{1 + \epsilon'/3\langle\epsilon\rangle}\right\rangle. \tag{3.10}$$

A careful study of the above formulation of Kröner suggests that the following three steps must be reconsidered:

(i) Equation (3.6) is not perfect because it neglects the "exclusion effect." For example, let us take the first term $\langle\epsilon'^2\rangle^2\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}$ in the parentheses on the right-hand side of Eq. (3.6). Since the case where the four points coincide is counted in the first term $\langle\epsilon'^4\rangle\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}$, this must be strictly excluded from any of the three terms in the parentheses in order to avoid overcounting. Accordingly, the above-mentioned term corresponds to the situation in which neither $\mathbf{r}_1$ nor $\mathbf{r}_2$ is identical with $\mathbf{r}_3$ or $\mathbf{r}_4$. This condition is expressed by $(1 - \delta_{\mathbf{r}_{23}})$ and the corrected form of Eq. (3.6) is given by

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle = \langle\epsilon'^4\rangle\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}} + \langle\epsilon'^2\rangle^2[\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}(1 - \delta_{\mathbf{r}_{23}}) + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}(1 - \delta_{\mathbf{r}_{12}}) + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}(1 - \delta_{\mathbf{r}_{12}})]. \tag{3.11}$$

By means of cumulants or semiinvariants, Eq. (3.11) is rewritten as

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle = \langle\epsilon^4\rangle_c\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}} + \langle\epsilon^2\rangle_c^2(\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}} + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}} + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}). \tag{3.12}$$

Here $\langle\epsilon^2\rangle_c = \langle\epsilon'^2\rangle$ and $\langle\epsilon^4\rangle_c = \langle\epsilon'^4\rangle - 3\langle\epsilon'^2\rangle^2$ imply the second-order and fourth-order cumulants of $\epsilon(\mathbf{r})$, respectively. A similar care must be taken of the exclusion effect for any higher-order term and it is shown that the whole procedure is reduced to replacing moments by the corresponding cumulants. This exclusion effect is exactly the crucial point which has led to the CPA through a diagram technique in the quantum mechanics of disordered binary alloys.[13]

(ii) Although Eq. (3.7a) is easily proved to be true, it is not straightforward to see whether or not Eqs. (3.7b) and (3.7c) hold. As we shall check later, there is some trouble concerning Eqs. (3.7b) and (3.7c).

(iii) In the calculations of terms such as Eqs. (3.7a) to (3.7c) where the derivatives of the null functions $\delta_{\mathbf{r}}$ are required, a careful treatment of $\delta_{\mathbf{r}}$ is necessary so that no ambiguity sneaks in to the formulation. Let $\delta_{\mathbf{r}}$ be represented by

$$\delta_{\mathbf{r}} = \lim_{\rho\to+0} I_\rho(r), \tag{3.13}$$

where

$$I_\rho(r) = \begin{cases} 1 & \text{for } r<\rho, \\ 0 & \text{for } r>\rho. \end{cases} \tag{3.14}$$

By way of explanation consider the third-order perturbation coefficient and set

$$A^{(3)} = \lim_{\rho_{12},\rho_{23}\to+0} \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3} \times \frac{\partial^2 I_{\rho_{12}}(r_{12})\, I_{\rho_{23}}(r_{23})}{\partial x_{12,k}\partial x_{23,(i)}}, \tag{3.15a}$$

$$A^{(3)\prime} = \lim_{\rho_{12},\rho_{13}\to+0} \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3} \times \frac{\partial^2 I_{\rho_{12}}(r_{12})\, I_{\rho_{13}}(r_{13})}{\partial x_{12,k}\partial x_{23,(i)}}, \tag{3.15b}$$

$$A^{(3)\prime\prime} = \lim_{\rho_{12},\rho_{13},\rho_{23}\to+0} \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3} \times \frac{\partial^2 I_{\rho_{12}}(r_{12})\, I_{\rho_{13}}(r_{13})\, I_{\rho_{23}}(r_{23})}{\partial x_{12,k}\partial x_{23,(i)}}. \tag{3.15c}$$

Formally we can write

$$\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}} = \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{13}} = \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{23}}, \tag{3.16}$$

but $A^{(3)}$, $A^{(3)\prime}$, $A^{(3)\prime\prime}$ do not always coincide with one another (see Sec. 4).

For the purpose of avoiding these complications and dealing with the problem in a safer mathematical framework, we employ a slightly different point of view. Let us regard the completely random heterogeneous material as a special case of the symmetric cell material where constituting cells are of spherical shape and infinitesimal size. In the limit as the cell radius tends to zero, we have

$$P(\mathbf{r}_1, \mathbf{r}_2) = \delta_{\mathbf{r}_{12}}, \tag{3.17}$$

$$P(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}, \tag{3.18}$$

$$P(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}, \tag{3.19}$$

$$P(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4) = \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}(1-\delta_{\mathbf{r}_{23}}), \tag{3.20}$$

$$P(\mathbf{r}_1, \mathbf{r}_4; \mathbf{r}_2, \mathbf{r}_3) = \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}(1-\delta_{\mathbf{r}_{12}}), \tag{3.21}$$

$$P(\mathbf{r}_1, \mathbf{r}_3; \mathbf{r}_2, \mathbf{r}_4) = \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}(1-\delta_{\mathbf{r}_{12}}), \tag{3.22}$$

and so on, which enable us to evaluate $A^{(n)}$ for arbitrary n. Detailed calculation based on this idea will be carried out in the subsequent section.

4. DETERMINATION OF LOW-ORDER PERTURBATION TERMS

With those instructions given in Sec. 3, we shall first investigate the second-order perturbation coefficient $A^{(2)}$ for a completely random heterogeneous material. For a symmetric cell material consisting of spherical cells, it has already been demonstrated that $A^{(2)}$ is equal to $\frac{1}{3}$, irrespective of the cell size. Since the completely random heterogeneous material is a limiting case of the isotropic cell material, $A^{(2)}=\frac{1}{3}$ applies also to the completely random case. The same result is attained by using Eqs (3.13) and (3.17) formally; that is,

$$\begin{aligned} A^{(2)} &= -\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial \delta_{\mathbf{r}_{12}}}{\partial x_{12,(i)}} \\ &= -\lim_{\rho_{12}\to+0}\frac{1}{3\cdot 4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial I_{\rho_{12}}(r_{12})}{\partial x_{12,i}} \\ &= -\lim_{\rho_{12}\to+0}\frac{1}{3\cdot 4\pi}\int_0^{2\pi} d\phi_{12}\int_0^{\pi} d\theta_{12}\sin\theta_{12} \times \int_0^{\infty} dr_{12}\frac{dI_{\rho_{12}}(r_{12})}{dr_{12}} \\ &= 1/3. \end{aligned} \tag{4.1}$$

Next let us deal with the third-order perturbation coefficient $A^{(3)}$. For the sake of illustration, we calculate $A^{(3)}$ by three different methods described in Eqs. (3.15a)–(3.15c). From Eq. (3.15a) it follows that

$$\begin{aligned} A^{(3)} &= \lim_{\rho_{12}\to+0}\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial I_{\rho_{12}}(r_{12})}{\partial x_{12,k}} \times \lim_{\rho_{23}\to+0}\frac{1}{4\pi}\int_V d\omega_{23}\frac{x_{23,k}}{r_{23}^3}\frac{\partial I_{\rho_{23}}(r_{23})}{\partial x_{23,(i)}} \\ &= A^{(2)}\delta_{ik}\cdot A^{(2)}\delta_{k(i)} = 1/9. \end{aligned} \tag{4.2}$$

This agrees with the correct value obtained by virtue of the cell model. On the other hand, the limits in Eqs. (3.15b) and (3.15c) depend upon the manner in which $\rho_{12}, \rho_{13}, \rho_{23}$ approach zero, so that $A^{(3)\prime}$ or $A^{(3)\prime\prime}$ cannot be uniquely determined.

In Appendix B we prove that

$$\begin{aligned} J_{\rho_{13},ij}(\mathbf{r}_{12}) &= -\frac{1}{4\pi}\int_V d\omega_{23}\frac{x_{23,i}}{r_{23}^3}\frac{\partial I_{\rho_{13}}(r_{13})}{\partial x_{23,j}} \\ &= C_{\rho_{13}}(r_{12})\frac{x_{12,i}}{r_{12}}\frac{x_{12,j}}{r_{12}} + D_{\rho_{13}}(r_{12})\delta_{ij}, \end{aligned} \tag{4.3}$$

where

$$C_{\rho_{13}}(r_{12}) = \begin{cases} 0 & \text{for } r_{12}<\rho_{13} \\ -\rho_{13}^3/r_{12}^3 & \text{for } r_{12}>\rho_{13}, \end{cases} \tag{4.4}$$

$$D_{\rho_{13}}(r_{12}) = \begin{cases} 1/3 & \text{for } r_{12}<\rho_{13} \\ \rho_{13}^3/3r_{12}^3 & \text{for } r_{12}>\rho_{13}. \end{cases} \tag{4.5}$$

Substituting Eqs. (4.3)–(4.5) into Eq. (3.15b), we get

$$\begin{aligned} A^{(3)\prime} &= -\lim_{\rho_{12},\rho_{13}\to+0}\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial I_{\rho_{12}}(r_{12})\, J_{\rho_{13},k(i)}(r_{12})}{\partial x_{12,k}} \\ &= \frac{1}{9} - \frac{2}{3}\lim_{\rho_{12},\rho_{13}\to+0}\int_0^\infty \frac{I_{\rho_{12}}(r_{12})C_{\rho_{13}}(r_{12})}{r_{12}}\, dr_{12}. \end{aligned} \tag{4.6}$$

Especially, when $\rho_{12}=\rho_{13}=\rho$, Eq. (4.6) reduces to

$$A^{(3)\prime} = 1/9. \tag{4.7}$$

Nevertheless, Eq. (3.15c) does not produce $A^{(3)\prime\prime}=1/9$ even though we put $\rho_{12}=\rho_{13}=\rho_{23}=\rho$.

In view of the results mentioned above, we shall evaluate $A^{(4)}$ in accordance with the following conventions: (i) Introduce Eqs. (3.19)–(3.22) into Eqs. (2.20)–(2.23), and express $A^{(4)}$ in terms of $\delta_{\mathbf{r}_{12}}$, $\delta_{\mathbf{r}_{13}}$, $\delta_{\mathbf{r}_{14}}$, $\delta_{\mathbf{r}_{23}}$,

$\delta_{\mathbf{r}_{24}}, \delta_{\mathbf{r}_{34}}$; (ii) contract each product of δ's to the simplest form with respect to $\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{34}$ and arrange the subscripts in order whenever possible; for example, use $\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}$ instead of $\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}$, $\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}$, $\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}$, etc.; (iii) replace $\delta_{\mathbf{r}_{12}}, \delta_{\mathbf{r}_{13}}, \cdots$ by the characteristic functions $I_\rho(r_{12}), I_\rho(r_{13}), \cdots$ having a common value of ρ; (iv) take the limits as $\rho \to +0$.

For convenience we rewrite Eq. (2.19) as

$$A^{(4)}\langle\epsilon'^4\rangle = A_1^{(4)}\langle\epsilon'^4\rangle + (A_2^{(4)} + A_3^{(4)} + A_4^{(4)})\langle\epsilon'^2\rangle^2$$
$$= B_1^{(4)}\langle\epsilon^4\rangle_c + (B_2^{(4)} + B_3^{(4)} + B_4^{(4)})\langle\epsilon^2\rangle_c^2, \tag{4.8}$$

where

$$A_1^{(4)} = B_1^{(4)}, \tag{4.9}$$

$$A_2^{(4)} = B_2^{(4)} - B_1^{(4)}, \tag{4.10}$$

$$A_3^{(4)} = B_3^{(4)} - B_1^{(4)}, \tag{4.11}$$

$$A_4^{(4)} = B_4^{(4)} - B_1^{(4)}. \tag{4.12}$$

It is readily seen that

$$A_1^{(4)} = B_1^{(4)}$$
$$= -\lim_{\rho\to+0}\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}$$
$$\times\frac{\partial^3 I_\rho(r_{12})I_\rho(r_{23})I_\rho(r_{34})}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,(i)}} = \frac{1}{27}, \tag{4.13}$$

which is equivalent to Eq. (2.25) valid for a spherical-cell material. As to $B_2^{(4)}$, straightforward calculation shows

$$B_2^{(4)} = -\lim_{\rho\to+0}\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}$$
$$\times\frac{\partial^3 I_\rho(r_{12})I_\rho(r_{34})}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,(i)}} = 0, \tag{4.14}$$

whence

$$A_2^{(4)} = -1/27. \tag{4.15}$$

To calculate $A_3^{(4)}$ we contract $\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}$ as

$$\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}} = \delta_{\mathbf{r}_{12}+\mathbf{r}_{23}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{23}} = \delta_{\mathbf{r}_{12}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{23}}; \tag{4.16}$$

then

$$B_3^{(4)} = -\lim_{\rho\to+0}\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}$$
$$\times\frac{\partial^3 I_\rho(|\mathbf{r}_{12}+\mathbf{r}_{34}|)I_\rho(r_{23})}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,(i)}}. \tag{4.17}$$

Proceeding in the same way as we derived Eqs. (2.36) and (2.38) of I, we have

$$\frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{34,h}}{r_{34}^3}\frac{\partial^2 I_\rho(|\mathbf{r}_{12}+\mathbf{r}_{34}|)}{\partial x_{12,k}\partial x_{34,i}}$$
$$= \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{34,i}}{r_{34}^3}\frac{\partial^2 I_\rho(|\mathbf{r}_{12}+\mathbf{r}_{34}|)}{\partial x_{12,k}\partial x_{34,h}}$$
$$= -\frac{1}{4\pi}\int_V d\omega\frac{x_k}{r^3}\frac{\partial I_\rho(r)}{\partial x_h}. \tag{4.18}$$

Accordingly,

$$B_3^{(4)} = \tfrac{1}{3}A^{(2)}\delta_{kh}\cdot A^{(2)}\delta_{kh} = 1/9, \tag{4.19}$$

$$A_3^{(4)} = 2/27. \tag{4.20}$$

In Appendix C it will be verified that $B_3^{(4)} = 1/9$ can also be deduced without utilizing Eq. (4.16).

Finally, we are concerned with the numerical estimation of $B_4^{(4)}$. By definition,

$$B_4^{(4)} = -\lim_{\rho\to+0}\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3}$$
$$\times\frac{\partial^3 I_\rho(r_{13})I_\rho(r_{24})}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,(i)}}$$
$$= -\lim_{\rho\to+0}\frac{1}{4\pi}\int_V d\omega_{23}\frac{x_{23,k}}{r_{23}^3}\frac{\partial J_{\rho,ik}(\mathbf{r}_{23})J_{\rho,h(i)}(\mathbf{r}_{23})}{\partial x_{23,h}}. \tag{4.21}$$

Analogously to the derivation of Eq. (B11), we obtain

$$\frac{1}{4\pi}\int_V d\omega_{23}\frac{x_{23,k}}{r_{23}^3}\frac{\partial J_{\rho,ik}(\mathbf{r}_{23})J_{\rho,hi}(\mathbf{r}_{23})}{\partial x_{23,h}}$$
$$= \frac{1}{4\pi}\int_V\frac{d\omega_{23}}{r_{23}^2}\left(\frac{d}{dr_{23}}[C_\rho(r_{23}) + D_\rho(r_{23})]^2\right.$$
$$\left.+\frac{2C_\rho(r_{23})[C_\rho(r_{23}) + 2D_\rho(r_{23})]}{r_{23}}\right)$$
$$= -\frac{1}{9} + 2\int_0^\infty\frac{C_\rho(r_{23})[C_\rho(r_{23}) + 2D_\rho(r_{23})]}{r_{23}}dr_{23}, \tag{4.22}$$

so that

$$B_4^{(4)} = \frac{1}{27} - \frac{2}{9}\lim_{\rho\to+0}\rho^6\int_\rho^\infty\frac{dr_{23}}{r_{23}^7} = 0, \tag{4.23}$$

$$A_4^{(4)} = -1/27. \tag{4.24}$$

As a result, the fourth-order term in the perturbation series becomes

$$-A^{(4)}\frac{\langle\epsilon'^4\rangle}{\langle\epsilon\rangle^4} = -\frac{\langle\epsilon^4\rangle_c + 3\langle\epsilon^2\rangle_c}{27\langle\epsilon\rangle^4} = -\frac{\langle\epsilon'^4\rangle}{27\langle\epsilon\rangle^4}. \tag{4.25}$$

The procedure to calculate higher-order perturbation coefficients proceeds in like manner. For the five-point moment we have

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\epsilon'(\mathbf{r}_5)\rangle$$
$$= \langle\epsilon'^5\rangle\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}} + \langle\epsilon'^3\rangle\langle\epsilon'^2\rangle[\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{45}}(1-\delta_{\mathbf{r}_{34}})$$
$$+ \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}}(1-\delta_{\mathbf{r}_{23}}) + \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{25}}\delta_{\mathbf{r}_{34}}(1-\delta_{\mathbf{r}_{23}})$$
$$+ \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{45}}\delta_{\mathbf{r}_{23}}(1-\delta_{\mathbf{r}_{12}}) + \delta_{\mathbf{r}_{15}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}(1-\delta_{\mathbf{r}_{12}})$$
$$+ \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{35}}(1-\delta_{\mathbf{r}_{23}}) + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{45}}(1-\delta_{\mathbf{r}_{12}})$$
$$+ \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{25}}(1-\delta_{\mathbf{r}_{12}}) + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{35}}(1-\delta_{\mathbf{r}_{12}})$$
$$+ \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{35}}\delta_{\mathbf{r}_{24}}(1-\delta_{\mathbf{r}_{12}})]. \tag{4.26}$$

By the use of cumulants, Eq. (4.26) is transformed into

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\epsilon'(\mathbf{r}_5)\rangle$$
$$= \langle\epsilon^5\rangle_c\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}} + \langle\epsilon^3\rangle_c\langle\epsilon^2\rangle_c[\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{45}} + \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}}$$
$$+ \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{25}}\delta_{\mathbf{r}_{34}} + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{45}}\delta_{\mathbf{r}_{23}} + \delta_{\mathbf{r}_{15}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}} + \delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{35}}$$
$$+ \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{45}} + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{25}} + \delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{35}} + \delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{35}}\delta_{\mathbf{r}_{24}}], \tag{4.27}$$

where $\langle\epsilon^3\rangle_c = \langle\epsilon'^3\rangle$ and $\langle\epsilon^5\rangle_c = \langle\epsilon'^5\rangle - 10\langle\epsilon'^3\rangle\langle\epsilon'^2\rangle$. The fifth-order perturbation coefficient is therefore given by

$$A^{(5)}\langle\epsilon'^5\rangle = A_1^{(5)}\langle\epsilon'^5\rangle + \sum_{m=2}^{11}A_m^{(5)}\langle\epsilon'^3\rangle\langle\epsilon'^2\rangle$$
$$= B_1^{(5)}\langle\epsilon^5\rangle_c + \sum_{m=2}^{11}B_m^{(5)}\langle\epsilon^3\rangle_c\langle\epsilon^2\rangle_c, \tag{4.28}$$

where

$$A_m^{(5)} = \begin{cases} B_m^{(5)} & \text{for } m=1, \\ B_m^{(5)} - B_1^{(5)} & \text{for } 2 \le m \le 11. \end{cases} \quad (4.29)$$

In Appendix D we prove that

$$B_1^{(5)} = 1/81, \quad (4.30a)$$

$$B_2^{(5)} = B_3^{(5)} = 0, \quad (4.30b)$$

$$B_4^{(5)} = B_5^{(5)} = B_6^{(5)} = 1/27, \quad (4.30c)$$

$$B_7^{(5)} = B_8^{(5)} = B_9^{(5)} = B_{10}^{(5)} = 0. \quad (4.30d)$$

As for the last coefficient $B_{11}^{(5)}$ corresponding to $\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{35}}\delta_{\mathbf{r}_{24}}$, we are not successful in determining it. In analogy to $B_4^{(4)}$, however, we assume that

$$B_{11}^{(5)} = 0. \quad (4.30e)$$

Then,

$$A_1^{(5)} = 1/81, \quad (4.31a)$$

$$A_2^{(5)} = A_3^{(5)} = -1/81, \quad (4.31b)$$

$$A_4^{(5)} = A_5^{(5)} = A_6^{(5)} = 2/81, \quad (4.31c)$$

$$A_7^{(5)} = A_8^{(5)} = \cdots = A_{11}^{(5)} = -1/81. \quad (4.31d)$$

Accordingly, the fifth-order perturbation term becomes

$$A^{(5)} \frac{\langle \epsilon'^5 \rangle}{\langle \epsilon \rangle^5} = \frac{\langle \epsilon^5 \rangle_c + 9\langle \epsilon^3 \rangle_c \langle \epsilon^2 \rangle_c}{\langle \epsilon \rangle^5} = \frac{\langle \epsilon'^5 \rangle - \langle \epsilon'^3 \rangle \langle \epsilon'^2 \rangle}{\langle \epsilon \rangle^5}. \quad (4.32)$$

The effective permittivity ϵ^* of a completely random material is thus written in the form

$$\frac{\epsilon^*}{\langle \epsilon \rangle} = 1 - \frac{\langle \epsilon^2 \rangle_c}{3\langle \epsilon \rangle^2} + \frac{\langle \epsilon^3 \rangle_c}{9\langle \epsilon \rangle^3} - \frac{\langle \epsilon^4 \rangle_c + 3\langle \epsilon^2 \rangle_c^2}{27\langle \epsilon \rangle^4} + \frac{\langle \epsilon^5 \rangle_c + 9\langle \epsilon^3 \rangle_c \langle \epsilon^2 \rangle_c}{81\langle \epsilon \rangle^5} - \cdots, \quad (4.33)$$

or

$$\frac{\epsilon^*}{\langle \epsilon \rangle} = 1 - \frac{\langle \epsilon'^2 \rangle}{3\langle \epsilon \rangle^2} + \frac{\langle \epsilon'^3 \rangle}{9\langle \epsilon \rangle^3} - \frac{\langle \epsilon'^4 \rangle}{27\langle \epsilon \rangle^4} + \frac{\langle \epsilon'^5 \rangle - \langle \epsilon'^3 \rangle \langle \epsilon'^2 \rangle}{81\langle \epsilon \rangle^5} - \cdots. \quad (4.34)$$

Up to the fourth order, the expression agrees with Kröner's formula (3.10). This agreement is due to the fact that $A_2^{(4)} + A_3^{(4)} + A_4^{(4)} = 0$ on the right-hand side of Eq. (4.8) and is only accidental. The two formulations [Eqs. (3.10) and (4.34)] start to differ from each other at the fifth order; the higher the order of the term, the larger the discrepancy becomes. This discrepancy is absolutely serious when an infinite sum of the expansion series is required.

In fact, it is easily checked that Eq. (3.10) cannot be used in the whole range of parameters. By way of illustration we consider the effective permittivity of a completely random two-phase medium. Let the two constituents have permittivities ϵ_1 and ϵ_2 and occupy fractions of the total volume v_1 and $v_2 = 1 - v_1$. Then Kröner's formula reduces to

$$\epsilon^* = [\epsilon_2 + (\epsilon_1 - \epsilon_2)v_1] \left(1 - 3(\epsilon_1 - \epsilon_2)^2 \times \frac{v_1(1-v_1)}{[\epsilon_1 + 2\epsilon_2 + 2(\epsilon_1 - \epsilon_2)v_1][3\epsilon_2 + 2(\epsilon_1 - \epsilon_2)v_1]}\right). \quad (4.35)$$

If $\epsilon_1 \to \infty$ or $\epsilon_2 = 0$, we have

$$\epsilon^*/\epsilon_1 = -v_1(1 - 7v_1)/2(1 + 2v_1), \quad (4.36)$$

which fails to give a nonnegative solution for the concentration region $v_1 < 1/7$. It should also be remarked that the power series in Eq. (3.10) is not convergent for $v_1 < 1/4$.

5. SUMMARY AND CONCLUDING REMARKS

In this article, we have shown how the effective permittivity ϵ^* of a completely random inhomogeneous material is calculated by the perturbation method. As usual, the procedure to be followed consists of the three steps:

(i) Expand a local electric field $E_i(\mathbf{r})$ in a perturbation series;

(ii) average in the ensemble sense each term of the expansion series of $\epsilon(\mathbf{r})E_i(\mathbf{r})$;

(iii) resum the averaged perturbation terms and determine the effective constant ϵ^* defined by $\epsilon^*\langle E_i \rangle = \langle \epsilon(\mathbf{r})E_i(\mathbf{r}) \rangle$.

Out of these three processes, the first two have been discussed in I and III.

It is emphasized that the exclusion effect plays a very important role in the averaging process. This exclusion effect originates from the fact that physical quantities associated with different points are statistically independent in a completely random material and accordingly the average can be taken independently. When quantities associated with two or more different points are considered for ensemble averaging, the possibility that some of these points coincide must be strictly excluded in the mathematical formulation. This is not a trivial problem because the neglect of the exclusion effect not only gives mathematically incorrect formulations but also results in physically unreasonable solutions.

With the exclusion effect being taken into account, the prescription for evaluating the n-point moment $\langle \epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2) \cdots \langle \epsilon'(\mathbf{r}_n) \rangle$ is described as follows:

(i) Divide n variables $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ into m subsets ($m < n$) such that at least two variables are assigned to each subset.

(ii) When ν_1 variables are allotted to the first subset, ν_2 to the second, and so on, calculate

$$\epsilon(n, \{\nu_m\}) = \langle \epsilon^{\nu_1} \rangle_c \langle \epsilon^{\nu_2} \rangle_c \cdots \langle \epsilon^{\nu_m} \rangle_c, \quad (5.1)$$

where $\{\nu_m\}$ denotes the way of partitioning n variables into m subsets as indicated in the above.

(iii) By means of a product of such null functions as $\delta_{\mathbf{r}_{ij}}$, express the condition that variables belonging to the same subset all coincide. The product must be constructed so that the variable subscripts are arranged in order whenever possible [see the instruction (ii) in Sec. 4]. Write the product of δ's thus calculated as $F_n(\{\nu_m\}; \mathbf{r}_{12}, \mathbf{r}_{23}, \ldots, \mathbf{r}_{n-1,n})$.

(iv) Determine the n-point moment through the relation

$$\langle \epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2) \cdots \epsilon'(\mathbf{r}_n) \rangle = \sum_{\{\nu_m\}} F_n(\{\nu_m\}; \mathbf{r}_{12}, \mathbf{r}_{23}, \ldots, \mathbf{r}_{n-1,n}) \epsilon(n, \{\nu_m\}), \quad (5.2)$$

where the summation $\sum_{\{\nu_m\}}$ is taken over all possible partitions of n variables into subsets of at least two components.

The effective permittivity ϵ^* is obtained by substituting Eq. (5.2) into Eq. (2.6). The result is represented as the sum of infinite perturbation terms in the form

$$\epsilon^* = \langle\epsilon\rangle + \sum_{n=2}^{\infty} \frac{\sum_{\{\nu_m\}} B^{(n)}_{\{\nu_m\}} \epsilon(n, \{\nu_m\})}{(-\langle\epsilon\rangle)^{n-1}}, \tag{5.3}$$

with

$$B^{(n)}_{\{\nu_m\}} = \frac{1}{3}\left(-\frac{1}{4\pi}\right)^{n-1} \int_V d\omega_{12} \int_V d\omega_{23} \cdots \int_V d\omega_{n-1,n} \times \frac{x_{12,i}}{r_{12}^3} \frac{x_{23,k}}{r_{23}^3} \cdots \frac{\partial^{n-1} F_n(\{\nu_m\}; \mathbf{r}_{12}, \mathbf{r}_{23}, \ldots \mathbf{r}_{n-1,n})}{\partial x_{12,k} \partial x_{23,h} \cdots \partial x_{n-1,n;i}}. \tag{5.4}$$

As for the third step of resummation as mentioned at the beginning of this section, it is practically impossible to sum up all the infinite terms in Eq. (5.3) because each perturbation term is written only formally and is not subject to summation. What is usually tried is to content oneself with a partial summation of the whole series and invoke that the partial sum serves as a good approximation when most important terms are picked up for the summation. Criteria for selecting important terms out of the right-hand side of Eq. (5.3) will be fully discussed in IV.

ACKNOWLEDGMENT

The authors are greatly indebted to Professor E. Kröner for valuable comments concerning the manuscript and for bringing Refs. 3, 4, and 21 to their attention.

APPENDIX A

Let us seek $A^{(4)}_{1,ij}$ and prove Eq. (2.24). In the same way as we got Eq. (I.3.23),

$$P(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \frac{1}{v} \int_v d\omega_{20} I(\mathbf{r}_{10}) I(\mathbf{r}_{30}) I(\mathbf{r}_{40}). \tag{A1}$$

The integrand on the right is the product of the characteristic functions defined in Sec. 3B of I. Substitution of Eq. (A1) into Eq. (2.20) yields

$$A^{(4)}_{1,ij} = -\frac{1}{(4\pi)^3 v} \int_v d\omega_{20} \int_V d\omega_{12} \int_V d\omega_{23} \int_V d\omega_{34} \frac{x_{12,i}}{r_{12}^3} \frac{x_{23,k}}{r_{23}^3} \times \frac{x_{34,h}}{r_{34}^3} \frac{\partial I(\mathbf{r}_{10})}{\partial x_{12,k}} \frac{\partial^2 I(\mathbf{r}_{30}) I(\mathbf{r}_{40})}{\partial x_{23,h} \partial x_{34,j}}. \tag{A2}$$

From Eq. (I.3.13) it follows that

$$L_{ik}(\mathbf{r}_{20}) = \frac{1}{4\pi} \int_V d\omega_{10} \frac{\partial}{\partial x_{10,i}} \left(\frac{1}{r_{12}}\right) \frac{\partial I(\mathbf{r}_{10})}{\partial x_{10,k}} = -\frac{1}{4\pi} \int_V d\omega_{12} \frac{x_{12,i}}{r_{12}^3} \frac{\partial I(\mathbf{r}_{10})}{\partial x_{12,k}}. \tag{A3}$$

The point-function demagnetization tensor $L_{ik}(\mathbf{r}_{20})$ is constant within an ellipsoidal cell, so that

$$A^{(4)}_{1,ij} = \frac{L_{ik}}{(4\pi)^2 v} \int_v d\omega_{20} \int_V d\omega_{23} \int_V d\omega_{34} \frac{x_{23,k}}{r_{23}^3} \frac{x_{34,h}}{r_{34}^3} \times \frac{\partial^2 I(\mathbf{r}_{30}) I(\mathbf{r}_{40})}{\partial x_{23,h} \partial x_{34,j}}. \tag{A4}$$

Since

$$P(\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \frac{1}{v} \int_v d\omega_{20} I(\mathbf{r}_{30}) I(\mathbf{r}_{40}), \tag{A5}$$

Eqs. (2.13), (2.16), and (A4) lead to

$$A^{(4)}_{1,ij} = L_{ik} A^{(3)}_{kj} = L_{ik} L_{kh} L_{hj}, \tag{A6}$$

which completes the proof of Eq. (2.24).

APPENDIX B

The integral appearing in Eq. (4.3),

$$J_{\rho_{13},ij}(\mathbf{r}_{12}) = -\frac{1}{4\pi} \int_V d\omega_{23} \frac{x_{23,i}}{r_{23}^3} \frac{\partial I_{\rho_{13}}(r_{13})}{\partial x_{23,j}}, \tag{B1}$$

is an isotropic tensor and has the form

$$J_{\rho_{13},ij}(\mathbf{r}_{12}) = C_{\rho_{13}}(r_{12}) \frac{x_{12,i}}{r_{12}} \frac{x_{12,j}}{r_{12}} + D_{\rho_{13}}(r_{12}) \delta_{ij}. \tag{B2}$$

Accordingly,

$$\begin{aligned} J_{\rho_{13},ii}(r_{12}) &= -\frac{1}{4\pi} \int_V d\omega_{23} \frac{x_{23,i}}{r_{23}^3} \frac{\partial I_{\rho_{13}}(r_{13})}{\partial x_{23,i}} \\ &= -\frac{1}{4\pi} \int_0^{2\pi} d\phi_{23} \int_0^{\pi} d\theta_{23} \sin\theta_{23} \times \int_0^{\infty} dr_{23} \frac{\partial I_{\rho_{13}}(r_{13})}{\partial r_{23}} \\ &= I_{\rho_{13}}(r_{12}) = C_{\rho_{13}}(r_{12}) + 3D_{\rho_{13}}(r_{12}). \end{aligned} \tag{B3}$$

For the evaluation of $C_{\rho_{13}}(r_{12})$ and $D_{\rho_{13}}(r_{12})$ choose $\mathbf{r}_{12}$ to lie along the x_3 axis; then

$$\frac{\partial r_{13}}{\partial x_{13,3}} = \frac{x_{13,3}}{r_{13}} = \cos\theta_{13} = u_{13}. \tag{B4}$$

From Eqs. (B1) and (B4) we obtain

$$\begin{aligned} J_{\rho_{13},33}(r_{12}) &= -\frac{1}{4\pi} \int_V d\omega_{13} \frac{x_{23,3}}{r_{23}^3} \frac{\partial I_{\rho_{13}}(r_{13})}{\partial x_{13,3}} \\ &= -\frac{1}{4\pi} \int_0^{2\pi} d\phi_{13} \int_{-1}^{1} du_{13} \int_0^{\infty} dr_{13} \times \frac{r_{13}^2 u_{13}(r_{13}u_{13} - r_{12})}{(r_{13}^2 + r_{12}^2 - 2r_{12}r_{13}u_{13})^{3/2}} \frac{dI_{\rho_{13}}(r_{13})}{dr_{13}}. \end{aligned} \tag{B5}$$

Considering that

$$\int_{-1}^{1} \frac{u_{13}(r_{13}u_{13} - r_{12})}{(r_{13}^2 + r_{12}^2 - 2r_{12}r_{13}u_{13})^{3/2}} du_{13} = \begin{cases} -4r_{13}/3r_{12}^3 & \text{for } r_{13} < r_{12}, \\ 2/3r_{13}^2 & \text{for } r_{13} > r_{12}, \end{cases} \tag{B6}$$

Eq. (B5) becomes

$$J_{\rho_{13},33}(r_{12}) = \frac{2}{3r_{12}^3} \int_0^{r_{12}} r_{13}^3 dI_{\rho_{13}}(r_{13}) + \tfrac{1}{3} I_{\rho_{13}}(r_{12}) = C_{\rho_{13}}(r_{12}) + D_{\rho_{13}}(r_{12}). \tag{B7}$$

The solutions of Eqs. (B3) and (B7) are

$$C_{\rho_{13}}(r_{12}) = \frac{1}{r_{12}^3} \int_0^{r_{12}} r_{13}^3 dI_{\rho_{13}}(r_{13}) = \begin{cases} 0 & \text{for } r_{12} < \rho_{13}, \\ -\rho_{13}^3/r_{12}^3 & \text{for } r_{12} > \rho_{13}, \end{cases} \tag{B8}$$

$$D_{\rho_{13}}(r_{12})=\frac{I_{\rho_{13}}(r_{12})}{3}-\frac{1}{3r_{12}^3}\int_0^{r_{12}} r_{13}^3\,dI_{\rho_{13}}(r_{13})$$
$$=\begin{cases}1/3 & \text{for } r_{12}<\rho_{13},\\ \rho_{13}^3/3r_{12}^3 & \text{for } r_{12}>\rho_{13}.\end{cases} \tag{B9}$$

Now we can confirm the validity of Eq. (4.6). Because

$$\frac{\partial J_{\rho_{13},ki}(\mathbf{r}_{12})}{\partial x_{12,k}}=\left(\frac{dC_{\rho_{13}}(r_{12})}{dr_{12}}+\frac{dD_{\rho_{13}}(r_{12})}{dr_{12}}+\frac{2C_{\rho_{13}}(r_{12})}{r_{12}}\right)\frac{x_{12,i}}{r_{12}}, \tag{B10}$$

we have

$$\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}{}^3}\frac{\partial I_{\rho_{12}}(r_{12})J_{\rho_{13},ki}(\mathbf{r}_{12})}{\partial x_{12,k}}$$
$$=\frac{1}{4\pi}\int_V \frac{d\omega_{12}}{r_{12}{}^2}\left(\frac{d}{dr_{12}}I_{\rho_{12}}(r_{12})[C_{\rho_{13}}(r_{12})+D_{\rho_{13}}(r_{12})]+\frac{2I_{\rho_{12}}(r_{12})C_{\rho_{13}}(r_{12})}{r_{12}}\right)$$
$$=-\frac{1}{3}+2\int_0^\infty \frac{I_{\rho_{12}}(r_{12})C_{\rho_{13}}(r_{12})}{r_{12}}\,dr_{12}. \tag{B11}$$

Combination of Eqs. (3.15b) and (B11) gives

$$A^{(3)\prime}=\frac{1}{9}-\frac{2}{3}\lim_{\rho_{12},\rho_{13}\to+0}\int_0^\infty\frac{I_{\rho_{12}}(r_{12})C_{\rho_{13}}(r_{12})}{r_{12}}\,dr_{12}. \tag{B12}$$

APPENDIX C

Defining $K_{\rho,kh}(\mathbf{r}_{23})$ by

$$K_{\rho,kh}(\mathbf{r}_{23})=\frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}{}^3}\frac{x_{34,h}}{r_{34}{}^3}\times\frac{\partial^2 I_\rho(r_{14})}{\partial x_{12,k}\partial x_{34,i}}, \tag{C1}$$

we may express $B_3^{(4)}$ as

$$B_3^{(4)}=-\lim_{\rho\to+0}\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}{}^3}\frac{x_{23,k}}{r_{23}{}^3}\frac{x_{34,h}}{r_{34}{}^3}\times\frac{\partial^3 I_\rho(r_{14})I_\rho(r_{23})}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,(i)}}$$
$$=-\lim_{\rho\to+0}\frac{1}{3\cdot4\pi}\int_V d\omega_{23}\frac{x_{23,k}}{r_{23}{}^3}\frac{\partial I_\rho(r_{23})K_{\rho,kh}(\mathbf{r}_{23})}{\partial x_{23,h}}. \tag{C2}$$

In a similar manner to that developed in Sec. 2C of I [see Eqs. (I.2.36)–(I.2.40)], Eq. (C1) is transformed into

$$K_{\rho,kh}(\mathbf{r}_{23})=\frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}{}^3}\frac{x_{34,i}}{r_{34}{}^3}\times\frac{\partial^2 I_\rho(|\mathbf{r}_{12}+\mathbf{r}_{34}+\mathbf{r}_{23}|)}{\partial x_{12,k}\partial x_{34,h}}$$
$$=-\frac{1}{4\pi}\int_V d\omega\frac{x_k}{r^3}\frac{\partial I_\rho(|\mathbf{r}+\mathbf{r}_{23}|)}{\partial x_h}=J_{\rho,kh}(\mathbf{r}_{23}), \tag{C3}$$

where $\mathbf{r}=\mathbf{r}_{12}+\mathbf{r}_{34}$. Note that Eq. (C3), together with Eqs. (B1) and (C1), implies

$$J_{\rho,kh}(\mathbf{r}_{23})=-\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}{}^3}\frac{\partial J_{\rho,hi}(\mathbf{r}_{13})}{\partial x_{12,k}}. \tag{C4}$$

Making use of Eqs. (B11), (C2), and (C3), we arrive at

$$B_3^{(4)}=-\lim_{\rho\to+0}\frac{1}{3\cdot4\pi}\int_V d\omega_{23}\frac{x_{23,k}}{r_{23}{}^3}\frac{\partial I_\rho(r_{23})J_{\rho,kh}(\mathbf{r}_{23})}{\partial x_{23,h}}$$
$$=\frac{1}{9}-\frac{2}{3}\lim_{\rho\to+0}\int_0^\infty\frac{I_\rho(r_{23})C_\rho(r_{23})}{r_{23}}\,dr_{23}$$
$$=1/9. \tag{C5}$$

APPENDIX D

Equations (4.30a) and (4.30b) are almost self-evident; for instance,

$$B_1^{(5)}=\frac{1}{(4\pi)^4}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\int_V d\omega_{45}\frac{x_{12,i}}{r_{12}{}^3}\frac{x_{23,k}}{r_{23}{}^3}\times\frac{x_{34,h}}{r_{34}{}^3}\frac{x_{45,l}}{r_{45}{}^3}\frac{\partial^4\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}}}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,l}\partial x_{45,(i)}}$$
$$=A^{(2)}\delta_{ik}\cdot A^{(2)}\delta_{kh}\cdot A^{(2)}\delta_{hl}\cdot A^{(2)}\delta_{l(i)}=1/81. \tag{D1}$$

To prove Eqs. (4.30c) we employ

$$\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{25}}\delta_{\mathbf{r}_{34}}=\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}+\mathbf{r}_{45}}\delta_{\mathbf{r}_{34}}, \tag{D2a}$$
$$\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{45}}\delta_{\mathbf{r}_{23}}=\delta_{\mathbf{r}_{12}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}}\delta_{\mathbf{r}_{23}}, \tag{D2b}$$
$$\delta_{\mathbf{r}_{15}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}=\delta_{\mathbf{r}_{12}+\mathbf{r}_{45}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}, \tag{D2c}$$

which lead to

$$B_4^{(5)}=B_5^{(5)}=B_6^{(5)}=A^{(2)}B_3^{(4)}=1/27. \tag{D3}$$

Furthermore, it is readily seen that

$$\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{35}}=\delta_{\mathbf{r}_{12}}\delta_{\mathbf{r}_{23}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{34}+\mathbf{r}_{45}}, \tag{D4a}$$
$$\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{24}}\delta_{\mathbf{r}_{45}}=\delta_{\mathbf{r}_{12}+\mathbf{r}_{23}}\delta_{\mathbf{r}_{23}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{45}}, \tag{D4b}$$
$$\delta_{\mathbf{r}_{13}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{25}}=\delta_{\mathbf{r}_{12}+\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}}\delta_{\mathbf{r}_{23}+\mathbf{r}_{45}}, \tag{D4c}$$
$$\delta_{\mathbf{r}_{14}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{35}}=\delta_{\mathbf{r}_{12}+\mathbf{r}_{34}}\delta_{\mathbf{r}_{23}}\delta_{\mathbf{r}_{34}+\mathbf{r}_{45}}, \tag{D4d}$$

whence

$$B_7^{(5)}=B_8^{(5)}=B_9^{(5)}=B_{10}^{(5)}=A^{(2)}B_4^{(4)}=0. \tag{D5}$$

[1]M. Hori, J. Math. Phys. **14**, 514 (1973).
[2]M. Hori, J. Math. Phys. **14**, 1942 (1973).
[3]An approach similar to I was made independently by R. Zeller and P.H. Dederichs, Phys. Status Solidi B **55**, 831 (1973).
[4]A similar approach is also found in P.H. Dederichs and R. Zeller, Z. Phys. **259**, 103 (1973).
[5]M. Hori and F. Yonezawa, J. Math. Phys. **16**, (1975) (to be published).
[6]J.A. Krumhansl, in *Amorphous Magnetism*, edited by H.O. Hooper and A.M. de Graaf (Plenum, New York, 1973), p. 15.
[7]F. Yonezawa and K. Morigaki, Suppl. Prog. Theor. Phys. **53**, 1 (1973).
[8]A comprehensive list of references on the EM theory will be found in Ref. 5.
[9]See, for example, Z. Hashin, in *Mechanics of Composite Materials*, edited by F.W. Wendt, H. Liebowitz, and N. Perrone (Pergamon, Oxford, 1970), p. 201.
[10]M.H. Cohen and J. Jortner, Phys. Rev. Lett. **30**, 699 (1973).
[11]S. Kirkpatrick, Phys. Rev. Lett. **27**, 1722 (1971).
[12]M.H. Cohen and J. Jortner, Phys. Rev. Lett. **30**, 696 (1973).
[13]F. Yonezawa, Prog. Theor. Phys. **40**, 734 (1968).
[14]E. Kröner, J. Mech. Phys. Solids **15**, 319 (1967).
[15]E. Kröner, Phys. Status Solidi **41**, 499 (1970).
[16]E. Kröner, Phys. Status Solidi B **49**, K23 (1972).
[17]M.N. Miller, J. Math. Phys. **10**, 1988 (1969).
[18]M.N. Miller, J. Math. Phys. **10**, 2005 (1969); **12**, 1057 (1971).
[19]W.F. Brown Jr., *Magnetostatic Principles in Ferromagnetism* (North-Holland, Amsterdam, 1962), p. 49.
[20]R. Moskowitz and E. Della Torre, IEEE Trans. Mag. **2**, 739 (1966).
[21]E. Kröner, Int. J. Engng. Sci. **11**, 171 (1973).

Power statistics for wave propagation in one-dimension and comparison with radiative transport theory. II

W. Kohler

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

G. C. Papanicolaou*

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 10 April 1974)

We consider the one-dimensional problem of a slab having a random index of refraction and illuminated from within by a point source. We compute the expected value and the fluctuations of both the total power and power flux. These quantities, which are functions of the slab width, source location, and observation point, are determined in the limit of weak refractive index fluctuations and large slab thickness. We compare the expected values of total intensity and flux with the predictions of radiative transport theory. We also compare the results of both theories with numerical simulations.

1. INTRODUCTION AND SUMMARY

This work is a continuation and extension of previous work by us,[1] which we shall refer to as I in the text. Except in the derivations of Sec. 4, where we rely on some of the analysis developed in I, the description of the problem and the results are self-contained here.

We consider a one-dimensional medium with a random index of refraction that fluctuates slightly from an expected value of unity. This medium occupies the interval $[0, l]$. The regions to the right of l and to the left of 0 are assumed to have a constant index of refraction equal to one. A time-harmonic point source is located within the interval $[0, l]$. We are interested in the statistical properties of the resulting wave field throughout the medium. More specifically, we are interested in the mean value or expectation of the total power (intensity) of the waves, the expectation of the power flux and the fluctuations of the intensity and flux about their mean values.

The formulation of the above as a transmission line problem was carried out in I, and it leads to the same mathematical considerations. This is also true for the propagation of the fundamental mode in a waveguide with random inhomogeneities. All computations presented in this work will deal with what was termed the matched case in I, i. e., the medium in the absence of random perturbations, has an index of refraction equal to unity everywhere on $(-\infty, \infty)$. The extension of the new results to the mismatched case, however, can be carried out without difficulty as we indicate in Sec. 4.

We study the above problem in the asumptotic limit of weak fluctuations of the refractive index and large slab thicknesses. The fluctuations are characterized by a small parameter ϵ while the thickness $l \sim 1/\epsilon^2$. The wavelength in the unperturbed medium and the correlation length of the random inhomogeneities are assumed to be of order one relative to ϵ. We shall refer to the asymptotic limit as the diffusion limit. A formal description of this limit is given in Ref. 2. More mathematical descriptions are presented in Refs. 3, 4, while additional references are cited in Refs. 3, 4 and I. References 5 and 6 can be consulted for related information.

Our results are the following. First, we compute in the diffusion limit the expectation of the total power or intensity of the wave field as a function of the scaled width of the slab of the random medium, the scaled source location and the scaled observation point. From this expression, in turn, we determine the expectation of the power flux. Thus, we generalize the results of I, wherein the source was located at the left end of the slab, i.e., radiation was incident from the left. This latter problem has also been treated by Gazaryan[7] and Lang.[8] Rubin[9,10] has considered the analogous problem of wave propagation through a one-dimensional randomly disordered crystal while Halperin[11] has also dealt with a similar problem in his calculation of the spectral density for a particle in a one-dimensional random potential.

Our second result is the computation in the diffusion limit of the fluctuations of the total power about its mean value as a function of the scaled slab width, scaled source location, and scaled observation point. We again use this expression to determine, as a special case, the fluctuations of the power flux. These quantities, which were not computed in I, provide important insights into the basic nature of wave propagation in random media. Marcuse[12] has also computed power fluctuations but within the forward scattering approximation, and so his results differ from ours.

We compare our results for the expected total power with the predictions of radiative transport theory. This is a phenomenological theory, due to Schuster,[13] that leads to simple equations for the total intensity and flux of radiation through an inhomogeneous medium. As in I, we find discrepancies between the stochastic and transport theories and conclude that radiative transport theory in one-dimension cannot be derived from a stochastic wave theory in the diffusion limit as one might expect from physical considerations (cf. references in I). Furthermore, intensity fluctuations in the interior can be so large as to render the mean intensity in the interior a relatively unimportant quantity.

We also compare our results with the results of numerical simulations. The predictions of the stochastic theory for the mean value and fluctuations of both the total power and power flux are found to be in good agreement with the simulation data.

Section 2 presents the formulation of the problem and a delineation of our results. Section 3 compares these diffusion limit results both with those of radiative transport theory and the results of numerical simulations. Graphs are presented which illustrate the behavior of the quantities of interest; these graphs are discussed in Sec. 3. We also briefly indicate how our results can be applied to the case where the physical configuration remains fixed while frequency or wavenumber is permitted to vary. Section 4 presents a derivation of the results. This derivation relies both upon theorems established in Refs. 3, 4 and also upon the formulation in I. Therefore, some details are omitted.

We take this opportunity to refer to the work of Besieris and Tappert[14] in connection with the pulse problem discussed in Sec. 9 of I. Our formula for the pulse-spreading factor [below (I. 9.47)] agrees, up to a factor 3/2, with their results. In our formula, as well as in (I. 9.44), the factor $\int_0^\infty R(s)\cos 2ks\,ds$ is set equal to one so that (I. 9.44) is in fact dimensionally correct. Besieris and Tappert treat the problem in the forward scattering approximation but, as (I. 9.8) and (I. 9.9) indicate, backscattering is negligible in the diffusion limit as well.

2. FORMULATION OF THE PROBLEM AND STATEMENT OF RESULTS

Let $u(x)$ denote the complex-valued scalar wave field at location $x \in (-\infty, \infty)$ with the time dependence $\exp(-i\omega t)$ omitted throughout. We assume that $u(x)$ satisfies the following equation and boundary conditions:

$$\frac{d^2u(x)}{dx^2} + k^2[1+\epsilon\mu(x)]u(x) = i2k\,\delta(x-y), \quad 0 \le x, y \le l, \tag{2.1}$$

$$u(x) = T_+\exp(ikx), \quad x \ge l, \quad u(x) = T_-\exp(-ikx), \quad x \le 0, \tag{2.2}$$

$$u(x) \text{ and } \frac{du(x)}{dx} \text{ continuous}, \tag{2.3}$$

$$E\{\mu(x)\} = 0, \quad R(z) = E\{\mu(x+z)\mu(x)\}. \tag{2.4}$$

Here k is the free space wavenumber, $\mu(x)$ is a wide-sense stationary random process satisfying (2.4)[15] (where $E\{\cdot\}$ denotes expected value) and ϵ is a small parameter characterizing the fluctuations of the refractive index. $T_+(y,l)$ and $T_-(y,l)$ are the complex-valued right and left transmission coefficients. These coefficients depend upon y, the source location, and l, the width of the random medium, as does the wave field $u = u(x,y,l)$. In general, we will not display this dependence upon y and l explicitly.

As in (I. 2.15)[16] we define the complex valued functions $A(x,y,l)$ and $B(x,y,l)$ by

$$u(x) = \exp(ikx)A(x) + \exp(-ikx)B(x),$$
$$\frac{du(x)}{dx} = ik[\exp(ikx)A(x) - \exp(-ikx)B(x)] \tag{2.5}$$

so that

$$A(x) = \tfrac{1}{2}\exp(-ikx)\left(u(x) + \frac{1}{ik}\frac{du(x)}{dx}\right),$$
$$B(x) = \tfrac{1}{2}\exp(ikx)\left(u(x) - \frac{1}{ik}\frac{du(x)}{dx}\right). \tag{2.6}$$

We interpret $A(x)$ and $B(x)$ as the "slowly varing" complex amplitudes of right and left propagating waves whose sum composes the wave field $u(x)$. From (2.1)–(2.3) and (2.5) we see that $A(x)$ and $B(x)$ satisfy the following stochastic boundary value problem:

$$\frac{dA(x)}{dx} = \frac{\epsilon ik\mu(x)}{2}[A(x) + B(x)\exp(-i2kx)],$$
$$\frac{dB(x)}{dx} = \frac{-\epsilon ik\mu(x)}{2}[\exp(i2kx)A(x) + B(x)], \quad 0 \le x \le l, \ x \ne y, \tag{2.7}$$

$$A(y+0,y,l) - A(y-0,y,l) = \exp(-iky),$$
$$B(y+0,y,l) - B(y-0,y,l) = -\exp(iky), \quad 0 \le y \le l, \tag{2.8}$$

$$A(0,y,l) = B(l,y,l) = 0. \tag{2.9}$$

The arguments $y+0$ and $y-0$ in jump conditions (2.8) refer to the limits as x tends to y from the right and left respectively. Note that $A(x,y,l)$, $B(x,y,l)$, and $u(x,y,l)$ are random functions which depend upon ϵ. We shall sometimes use a superscript, i.e., $A^{(\epsilon)}$, $B^{(\epsilon)}$, $u^{(\epsilon)}$, to denote this dependence.

From (2.6) and (2.7)–(2.9) it follows that

$$|T_+^{(\epsilon)}(y,l)|^2 = |A^{(\epsilon)}(x,y,l)|^2 - |B^{(\epsilon)}(x,y,l)|^2$$
$$= \frac{1}{i2k}\left(\bar{u}^{(\epsilon)}\frac{du^{(\epsilon)}}{dx} - u^{(\epsilon)}\frac{d\bar{u}^{(\epsilon)}}{dx}\right), \quad x > y, \tag{2.10}$$

$$|T_-^{(\epsilon)}(y,l)|^2 = -|A^{(\epsilon)}(x,y,l)|^2 + |B^{(\epsilon)}(x,y,l)|^2$$
$$= -\frac{1}{2ik}\left(\bar{u}^{(\epsilon)}\frac{du^{(\epsilon)}}{dx} - u^{(\epsilon)}\frac{d\bar{u}^{(\epsilon)}}{dx}\right), \quad x < y. \tag{2.11}$$

The functions $|T_+|^2$ and $|T_-|^2$ represent the power flux to the right and to the left of the source, respectively. Since the medium is lossless, the two fluxes are indepen dent of the location of the observation point and depend only upon the location of the source point and the width of the random medium. We define the total power or intensity by

$$\mathcal{J}^{(\epsilon)}(x,y,l) = |A^{(\epsilon)}(x,y,l)|^2 + |B^{(\epsilon)}(x,y,l)|^2$$
$$= \tfrac{1}{2}\left(|u^{(\epsilon)}|^2 + \frac{1}{k^2}\left|\frac{du^{(\epsilon)}}{dx}\right|^2\right). \tag{2.12}$$

This quantity depends on the observation point, source point, and the width of the random medium. Note that when $\epsilon = 0$, i.e., there are no random inhomogeneities, then $u(x) = \exp(ik|x-y|)$, $|A|^2 - |B|^2 = \operatorname{sgn}(x-y)$, and $|A|^2 + |B|^2 = 1$. It follows from (2.9) that

$$|T_+^{(\epsilon)}(y,l)|^2 = \mathcal{J}^{(\epsilon)}(l,y,l), \tag{2.13}$$

$$|T_-^{(\epsilon)}(y,l)|^2 = \mathcal{J}^{(\epsilon)}(0,y,l), \quad 0 \le y \le l. \tag{2.14}$$

Therefore, it is not necessary to compute the power fluxes separately since they can be obtained from $\mathcal{J}^{(\epsilon)}$.

By letting $y \downarrow 0$, $y \le x \le l$ in (2.7)–(2.9) we recover problem (I. 2.6), (I. 2.7) with $\Gamma_g = \Gamma_l = 0$. Therefore, $\mathcal{J}^{(\epsilon)}(x,0,l)$ is the function that was considered in I, i.e., total power or intensity as a function of the observation point and width of the random medium, with plane wave illumination of the medium from the left. Note that $\mathcal{J}^{(\epsilon)}(l,0,l)$ is the power transmission coefficient for this configuration. All relevant information about power transport is contained, therefore, in the random function $\mathcal{J}^{(\epsilon)}(x,y,l)$, $0 \le x, y \le l$.

We shall now state our results. Let τ, ξ, and η be defined as follows:

$$\tau=\epsilon^2 l, \quad \xi=\epsilon^2 y-\tau/2, \quad \eta=\epsilon^2 x-\tau/2. \tag{2.15}$$

These variables are the scaled width of the random medium, the scaled distance of the source from the midpoint, and the scaled distance of the observation point from the midpoint, respectively. The limit $\epsilon \downarrow 0$ with τ, ξ, and η fixed is called the diffusion limit. The mean power or intensity in the diffusion limit is defined by

$$MJ(\tau,\xi,\eta)=\lim_{\epsilon\downarrow 0} E\{J^{(\epsilon)}([\tau/2+\eta]/\epsilon^2,[\tau/2+\xi]/\epsilon^2,\tau/\epsilon^2)\}. \tag{2.16}$$

As our first result, we assert that this limit exists and that MJ is given by the following formula:

$$MJ(\tau,\xi,\eta)=\exp(3\alpha\tau/4-\alpha|\xi-\eta|)\int_{-\infty}^{\infty}\frac{\exp(-t^2\alpha\tau)\pi\sinh\pi t}{t\cosh^2\pi t}$$
$$\times[(t^2+\tfrac14)\cos 2t\alpha(\xi+\eta)+(t^2-\tfrac14)\cos 2t\alpha(\tau-|\xi-\eta|)$$
$$+t\sin 2t\alpha(\tau-|\xi-\eta|)]\,dt, \tag{2.17}$$

where $\tau\geq 0$, $-\tau/2\leq\xi,\eta\leq\tau/2$, and

$$\alpha=\tfrac12 k^2\int_0^\infty R(s)\cos 2ks\,ds. \tag{2.18}$$

From (2.17) and (2.18) it follows that MJ depends on k and the correlation function $R(s)$ [cf. (2.4)] through the parameter α which is the value of the power spectrum of $k\mu$ at wavenumber $2k$. For brevity, we refer to MJ as a function of τ, ξ, and η although it actually is a function of $\alpha\tau$, $\alpha\xi$, and $\alpha\eta$. Observe that MJ is a symmetric function of ξ and η (i.e., it obeys the principle of reciprocity) and is invariant under the transformation $\xi\to-\xi$, $\eta\to-\eta$.

When $\xi=-\tau/2$, i.e., the source is at the left end of the random medium, we recover formula (I. 6.32)[17] with $\theta_g=\theta_l=0$.

$$MJ(\tau,-\tau/2,\eta)=\exp(\alpha\tau/4-\alpha\eta)\int_{-\infty}^{\infty}\frac{\exp(-t^2\alpha\tau)\pi t\sinh\pi t}{\cosh^2\pi t}$$
$$\times\left(\cos t\alpha(\tau-2\eta)+\frac{\sin t\alpha(\tau-2\eta)}{2t}\right)dt,$$
$$-\tau/2\leq\eta\leq\tau/2. \tag{2.19}$$

Because of the symmetry $MJ(\tau,-\tau/2,\eta)=MJ(\tau,\eta,-\tau/2)$ it follows that $MJ(\tau,\xi,-\tau/2)$ is given by (2.19) with ξ replacing η. However, from (2.14) we conclude that

$$MJ(\tau,\xi,-\tau/2)=\lim_{\epsilon\downarrow 0}E\{|T_-^{(\epsilon)}([\tau/2+\xi]/\epsilon^2,\tau/\epsilon^2)|^2\}. \tag{2.20}$$

Thus, the mean power flux to the left as a function of source location and slab width coincides in the diffusion limit with the mean total power in the interior at the former source location when the slab is now excited by a source at the left end. Analogous consequences of reciprocity exist for T_+, but no additional computations are needed in view of obvious symmetry about the midpoint of the slab. When $\xi=-\tau/2$ and $\eta=\tau/2$, we obtain the mean power transmission coefficient (cf. references in I, Sec. 7).

$$MJ(\tau,-\tau/2,\tau/2)=\exp(-\alpha\tau/4)\int_{-\infty}^{\infty}\exp(-t^2\alpha\tau)\frac{\pi t\sinh\pi t}{\cosh^2\pi t}dt\,. \tag{2.21}$$

Our second result concerns the fluctuations of the total intensity $J^{(\epsilon)}$ in the diffusion limit. We define $KJ(\tau,\xi,\eta)$ as follows:

$$KJ(\tau,\xi,\eta)=\lim_{\epsilon\downarrow 0}E\{(J^{(\epsilon)}([\tau/2+\eta]/\epsilon^2,[\tau/2+\xi]/\epsilon^2,\tau/\epsilon^2))^2\}. \tag{2.22}$$

The fluctuation in the total power is then given by

$$FJ(\tau,\xi,\eta)=[KJ(\tau,\xi,\eta)-(MJ(\tau,\xi,\eta))^2]^{1/2}, \tag{2.23}$$
$$\tau\geq 0,\ -\tau/2\leq\xi,\ \eta\leq\tau/2.$$

We find that

$$KJ(\tau,\xi,\eta)=\frac{\exp(15\alpha\tau/4-4\alpha|\eta-\xi|)}{8}\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\exp(-t^2\alpha\tau)$$
$$\times\frac{(t^2+\frac14)}{(t^2+1)^2}\Big((t^2+\tfrac34)\exp(-2\alpha(\tau-2\sigma\eta))+(t^2+\tfrac54)$$
$$\times\cos 2t\alpha(\tau-2\sigma\eta)+(t^2+\tfrac34)\frac{\sin 2t\alpha(\tau-2\sigma\eta)}{t}\Big)$$
$$\times\Big[(t^2+\tfrac14)\exp(-2\alpha(\tau+2\sigma\xi))+3\Big((t^2+\tfrac54)$$
$$\times\cos 2t\alpha(\tau+2\sigma\xi)+(t^2+\tfrac34)\frac{\sin 2t\alpha(\tau+2\sigma\xi)}{t}\Big)\Big]\,dt,$$
$$\sigma\equiv\operatorname{sgn}(\eta-\xi). \tag{2.24}$$

Observe that $KJ(\tau,\xi,\eta)$ is invariant under the transformation $\xi\to-\xi$, $\eta\to-\eta$. However, as Eq. (2.24) indicates, the second moment of the total intensity is not invariant under an interchange of ξ and η, the source and observation points.

When $\xi=-\tau/2$, (2.24) reduces to

$$KJ(\tau,-\tau/2,\eta)=\frac{\exp(7\alpha\tau/4-4\alpha\eta)}{2}\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}$$
$$\times\exp(-t^2\alpha\tau)\frac{(t^2+\frac14)}{(t^2+1)}\ (t^2+\tfrac34)\exp(-2\alpha(\tau-2\eta))+(t^2+\tfrac54)$$
$$\times\cos 2t\alpha(\tau-2\eta)+(t^2+\tfrac34)\frac{\sin 2t\alpha(\tau-2\eta)}{t}\ dt\,. \tag{2.25}$$

From (2.25), (2.19), and (2.23), we obtain the power fluctuations at observation point η, $-\tau/2\leq\eta\leq\tau/2$, when the source is at the left end; this solves the power fluctuation problem associated with I. Finally, when $\xi=-\tau/2$ and $\eta=\tau/2$, we obtain

$$KJ(\tau,-\tau/2,\tau/2)=\exp(-\alpha\tau/4)\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}$$
$$\times\exp(-t^2\alpha\tau)\,(t^2+\tfrac14)\,dt. \tag{2.26}$$

Therefore, from (2.26), (2.21), and (2.23) we obtain the fluctuation in the mean power transmission coefficient.

In Figs. 1a–12a we plot the mean total power or intensity MJ and in Figs. 1b–12b the intensity fluctuations FJ as functions of the observation point $\alpha\eta$ for a selection of slab widths and source locations. In Figs. 13a and 13b we plot the mean power transmission coefficient and its fluctuations as a function of slab thickness with the source at the left end. Observe that we have only considered source locations over half the slab width (i.e., $\xi=-\tau/2,-\tau/4,0$). Since MJ, KJ, and therefore

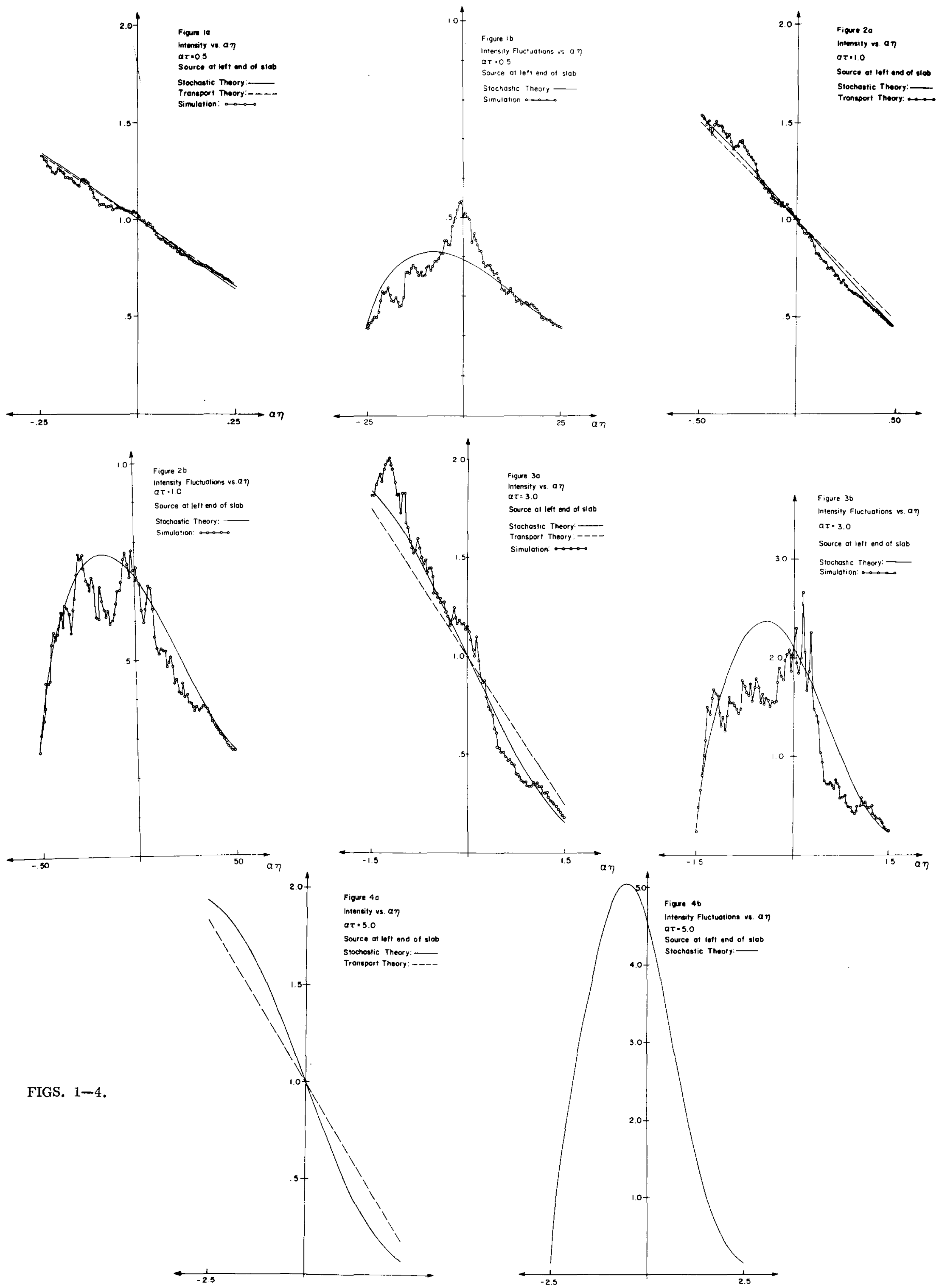
Figure 1a
Intensity vs. $\alpha\eta$
$\alpha\tau=0.5$
Source at left end of slab
Stochastic Theory: ——
Transport Theory: – – – –
Simulation: o–o–o–o
Figure 1b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau=0.5$
Source at left end of slab
Stochastic Theory ——
Simulation o–o–o–o
Figure 2a
Intensity vs. $\alpha\eta$
$\alpha\tau=1.0$
Source at left end of slab
Stochastic Theory: ——
Transport Theory: o–o–o
Figure 2b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau=1.0$
Source at left end of slab
Stochastic Theory: ——
Simulation: o–o–o–o
Figure 3a
Intensity vs. $\alpha\eta$
$\alpha\tau=3.0$
Source at left end of slab
Stochastic Theory: ——
Transport Theory: – – – –
Simulation: o–o–o–o
Figure 3b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau=3.0$
Source at left end of slab
Stochastic Theory: ——
Simulation: o–o–o–o
Figure 4a
Intensity vs. $\alpha\eta$
$\alpha\tau=5.0$
Source at left end of slab
Stochastic Theory: ——
Transport Theory: – – – –
Figure 4b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau=5.0$
Source at left end of slab
Stochastic Theory: ——
$\alpha\eta$

FIGS. 1—4.

FJ are invariant under the transformation $\eta \to -\xi$, $\eta \to -\eta$, the plots for the corresponding source locations in the other half of the slab (i.e., $\xi = \tau/4, \tau/2$) can be obtained by reflecting the presented curves about the vertical axis.

On these graphs, we have superimposed the results of random simulations conducted to verify the theory; the agreement is good. Observe that for the three source locations considered, the maximum value of the mean intensity occurs at the source point. As one moves away from the source point, the mean intensity decreases monotonically. For a fixed slab thickness, the value of the mean intensity at the source point increases as the source point is moved from the slab end toward the center. If, on the other hand, we increase the slab thickness while keeping the relative source position fixed (i.e., $\xi/\tau = \text{const.}$) the mean intensity at the source point is again seen to increase.

When the slab is excited at the left end, the intensity fluctuations at the left and right ends of the slab are equal (cf. Figs. 1b–4b). This is to be expected since the intensities at the left and right ends of the slab correspond to one plus the power reflection coefficient and the power transmission coefficient, respectively. Since the random medium is nondissipative, the two intensities must sum to two; thus the variance of the two slab end intensities must be equal. Observe that the peak fluctuations (with source at left end) occur in the slab interior between the left end and center. As the slab width $\alpha\tau$ is increased, the peak fluctuation also increases.

When the source is positioned in the slab interior, the peak fluctuations occur at the source point and for a given slab thickness, they are considerably greater than those occuring when the slab is excited at the end. Moreover, as the slab thickness is increased with the relative source position held fixed (i.e., $\xi/\tau = \text{const.}$), the fluctuations build up very rapidly.

On the basis of the results described above and displayed in the figures we may conclude the following:

For $\alpha\tau$ small, say less than one, the intensity fluctuations are relatively small and the mean intensity behaves in much the same way as the predictions of radiative transport theory (cf. Sec. 3). For $\alpha\tau > 2$, however, the intensity fluctuations in the interior can be very large, especially when the source is also located in the interior. In this case, neither the mean intensity of the stochastic theory nor that of the phenomenological transport theory give any insight into the extremely fluctuating character of the fields. On the other hand, away from the source point and near the slab extremities, the intensity fluctuations remain moderate even when $\alpha\tau$ is large. This is particularly true when the slab is illuminated at one end. Therefore, the mean of the power reflection and transmission coefficients (cf. Fig. 13) are stable quantities and the comparison with transport theory is meaningful. The enormous size of the fluctuations in the interior was unexpected and indeed surprising.

In Sec. 3, we shall discuss these results in more detail and compare them with transport theory.

3. COMPARISON WITH RADIATIVE TRANSPORT THEORY AND DESCRIPTION OF THE RESULTS

Radiative transport theory is a phenomenological theory that views the propagation and scattering of radiation as an incoherent process. This theory was first applied to the one-dimensional problem by Schuster.[13] We shall now outline the transport theory analog of problem (2.1)–(2.4) and compare the corresponding solution with the results of Sec. 2, as we did in Sec. 8 of I.

Let us assume that a scattering medium, occupying the interval $[-\tau/2, \tau/2]$, is excited from within by a point source whose distance from the center of the slab is denoted by ξ. Let η represent the distance of the observation point from the center of the slab. We shall assume a steady state condition and a conservative medium. Let $I^+(\tau, \xi, \eta)$ and $I^-(\tau, \xi, \eta)$ represent the intensities of radiation at location η, propagating in the positive and negative η directions, respectively. Assume that over an interval of length $d\eta$, there occurs a backscattering of radiation equal to $\alpha I^+ d\eta$ and a forward scattering equal to $\alpha I^- d\eta$. Then, a conservation of energy argument leads to the following equations:

$$\frac{d}{d\eta} I^+ = \frac{d}{d\eta} I^- = -\alpha (I^+ - I^-), \tag{3.1}$$

$$I^+(\tau, \xi, -\tau/2) = I^-(\tau, \xi, \tau/2) = 0, \tag{3.2}$$

$$I^\pm(\tau, \xi, \xi + 0) - I^\pm(\tau, \xi, \xi - 0) = \pm 1. \tag{3.3}$$

Boundary conditions (3.2) are a simple consequence of the fact that scattering only occurs in the interval $[-\tau/2, \tau/2]$. The notation $\xi + 0$ and $\xi - 0$ used in jump condition (3.3) again refers to limits as η approaches ξ from the right and left, respectively. The point source at location ξ is assumed to emit radiation of unit intensity. Note that the transport coefficient in (3.1) is the parameter α defined by (2.18). This choice has been justified, on the one hand, in an *a priori* manner by the heuristic arguments of Marcuse.[18] On the other hand, this choice will also be dictated in an *a posteriori* manner by the comparison of stochastic and transport theoretic predictions for small values of $\alpha\tau$.

We shall use the subscript s (for Schuster) to denote the transport theoretic quantities of interest. It follows readily from (3.1)–(3.3) that:

$$\begin{aligned} MJ_s(\tau, \xi, \eta) &\equiv I^+ + I^- \\ &= [1 + \alpha(\tau + 2\sigma\xi)][1 + \alpha(\tau - 2\sigma\eta)]/(1 + \alpha\tau), \\ \sigma &\equiv \operatorname{sgn}(\eta - \xi). \end{aligned} \tag{3.4}$$

Noting (2.10)–(2.14), we also obtain

$$|T_{+_s}(\xi, \tau)|^2 \equiv I^+ - I^- = \frac{1 + \alpha(\tau + 2\xi)}{1 + \alpha\tau} = MJ_s(\tau, \xi, \tau/2), \quad \eta \geq \xi, \tag{3.5}$$

$$|T_{-_s}(\xi, \tau)|^2 \equiv I^- - I^+ = \frac{1 + \alpha(\tau - 2\xi)}{1 + \alpha\tau} = MJ_s(\tau, \xi, -\tau/2), \quad \eta \leq \xi. \tag{3.6}$$

We again can obtain the right and left-directed power flux by an evaluation of the total intensity at the right and left slab ends, respectively.

Let us now compare the mean total intensities MJ and MJ_s when the slab thickness is small. Assume that

$\alpha\tau \ll 1$ with $-\tau/2 \leq \xi,\ \eta \leq \tau/2$. Define

$$\beta_n \equiv \int_{-\infty}^{\infty} \frac{\pi t^{2n+1} \sinh \pi t}{\cosh^2 \pi t} dt, \quad n = 0, 1, 2, \cdots. \tag{3.7}$$

Using the fact that $\beta_0 = 1$, $\beta_1 = 3/4$, and $\beta_2 = 25/16$, we find the expansions of MJ and MJ_s agree to terms of order $(\alpha\tau)^3$. Specifically

$$MJ(\tau, \xi, \eta) = 1 + \alpha\tau - 2\alpha|\eta - \xi| - 4\alpha^2 \xi\eta + O[(\alpha\tau)^3]$$
$$= MJ_s(\tau, \xi, \eta) + O[(\alpha\tau)^3]. \tag{3.8}$$

In Figs. 1–13, we present a graphical comparison of the predictions of the stochastic and radiative transport theories. The stochastic theoretic curves are drawn as solid lines while the transport theory curves are drawn as dashed lines. Also displayed on these graphs are the results of numerical simulations conducted to verify these theoretical results. Figures 1a–12a compare the intensities, i.e., MJ and MJ_s, as a function of the observation point for a variety of slab thicknesses and source locations. Figures 1b–12b present the intensity fluctuations FJ as a function of the observation point for the same selection of slab thicknesses and source locations. Note that we have not exhibited a transport-theoretic analog of FJ. Figure 13a compares the stochastic and transport theoretic transmission coefficients, i.e., $MJ(\tau, -\tau/2, \tau/2)$ and $MJ_s(\tau, -\tau/2, \tau/2)$, as functions of slab thickness while Fig. 13b presents the transmission coefficient fluctuations $FJ(\tau, -\tau/2, \tau/2)$.

Figures 1–4 correspond to a source location at the left end of the slab. For this configuration, both the stochastic and transport theoretic intensities attain their maximum values at the source point and are monotonically decreasing functions of the observation point. Both intensities are equal to unity at the slab center. For small values of $\alpha\tau$, the two intensity curves practically coincide; this is to be expected in view of (3.8). As $\alpha\tau$ increases, however, the effects of multiple scattering become more pronounced. Both intensity curves tend asymtotically toward the values 2 and 0 at the left and right slab ends, respectively. However, as $\alpha\tau$ increases, the transport theoretic intensity remains a linear function of the observation point while the stochastic intensity exhibits an increasingly nonlinear behavior. Note that, in all cases, for both theories, the intensities at the two slab ends sum to two. This is to be expected since the intensity at the left end equals one plus the power reflection coefficient while the intensity at the right end equals the power transmission coefficient. Since the random medium is conservative, the reflection and transmission coefficients must sum to unity.

In Figures 1b–4b, we present the intensity fluctuations FJ as a function of the observation point $\alpha\eta$ for the source point at the left end of the slab. As we noted in Sec. 2, the fluctuations at the two slab ends are equal. This is due to the fact that they represent the standard deviations of two random variables whose sum is a constant, i.e., 2. Observe that the largest intensity fluctuations occur in the left half of the slab; these fluctuations increase as the slab thickness increases. The occurrence of large fluctuations in the half of the slab nearest the incident excitation is, as the graphs indicate, supported by our numerical simulations. This phenomenon is also in qualitative agreement with observations made by Frisch, Froeschle, Scheidecker, and Sulem[19] based on numerical simulations that they conducted.

Figures 5–8 correspond to a source location midway between the left end and the center of the slab, i.e., $\xi = \tau/4$. Observe that the sum of the intensities at the two slab ends, i.e., $MJ(\tau, \xi, -\tau/2) + MJ(\tau, \xi, \tau/2)$ and $MJ_s(\tau, \xi, -\tau/2) + MJ_s(\tau, \xi, \tau/2)$, again equlas 2. This phenomenon occurs for an arbitrary interior source location. We shall now show that in the stochastic case there follows, as a simple consequence of jump condition (2.8), the fact that the right and left power fluxes are independent of the observation point [cf. (2.10)–(2.14)], and the diffusion limit. From (2.10), (2.11) it follows that

$$|T_+^{(\epsilon)}|^2 + |T_-^{(\epsilon)}|^2 = |A^{(\epsilon)}(y+0, y, l)|^2 - |A^{(\epsilon)}(y-0, y, l)|^2$$
$$+ |B^{(\epsilon)}(y-0, y, l)|^2 - |B^{(\epsilon)}(y+0, y, l)|^2. \tag{3.9}$$

Using (2.8) and (2.13)–(2.15), we obtain

$$J^{(\epsilon)}(\tau/\epsilon^2, [\tau/2+\xi]/\epsilon^2, \tau/\epsilon^2) + J^{(\epsilon)}(0, [\tau/2+\xi]/\epsilon^2, \tau/\epsilon^2)$$
$$= 2 + 2\,\mathrm{Re}\{[\overline{A^{(\epsilon)}}(\tau/2+\xi-0]/\epsilon^2, [\tau/2+\xi]/\epsilon^2, \tau/\epsilon^2)$$
$$+ B^{(\epsilon)}([\tau/2+\xi+0]/\epsilon^2, [\tau/2+\xi]/\epsilon^2, \tau/\epsilon^2)]$$
$$\times \exp[ik(\tau/2+\xi)/\epsilon^2]\}. \tag{3.10}$$

When the diffusion limit (2.16) is applied, the rapid phase variations annihilate the expected value of the second term on the right side of (3.10) and we obtain the aforementioned result. The argument for the transport theoretic case follows immediately from jump condition (3.3) and Eqs. (3.5)–(3.6).

For the source located midway between the left end and center of the slab, the peak intensity occurs at the source point. Observe that, as the slab thickness increases, the effects of multiple scattering again become increasingly important and the peak intensity predicted by the stochastic theory grows much faster than that predicted by transport theory (cf. Figs. 5a–8a). Figures 5b–8b display the intensity fluctuations corresponding to this source configuration. Observe that the intensity fluctuations at the slab ends are not equal in this case. Note also the discontinuity in the fluctuations at the source point that is very apparent for the smaller slab thicknesses (Figs. 5b, 6b) but which effectively disappears for the thicker slabs (Figs. 7b, 8b), i.e., when the source is located in the deep interior.

Figures 8–12 display the intensity and fluctuation variations corresponding to a source located at the center of the slab. The graphs display the obviously required symmetry with respect to the slab center. The peak intensity and peak fluctuations both occur at the source point. For a given slab thickness, this source location produces the largest peak values. Note, moreover, that these peak values become very large for moderate valyes of $\alpha\tau$; a peak intensity of 30 and a peak fluctuation of 5000 occur for $\alpha\tau$ equal to 5.

Figures 13a and 13b show the variation of the power transmission coefficient and power transmission coefficient fluctuations, respectively, as a function of $\alpha\tau$ for a source located at the left end of the slab. For the stochastic theory, the transmission coefficient is given by

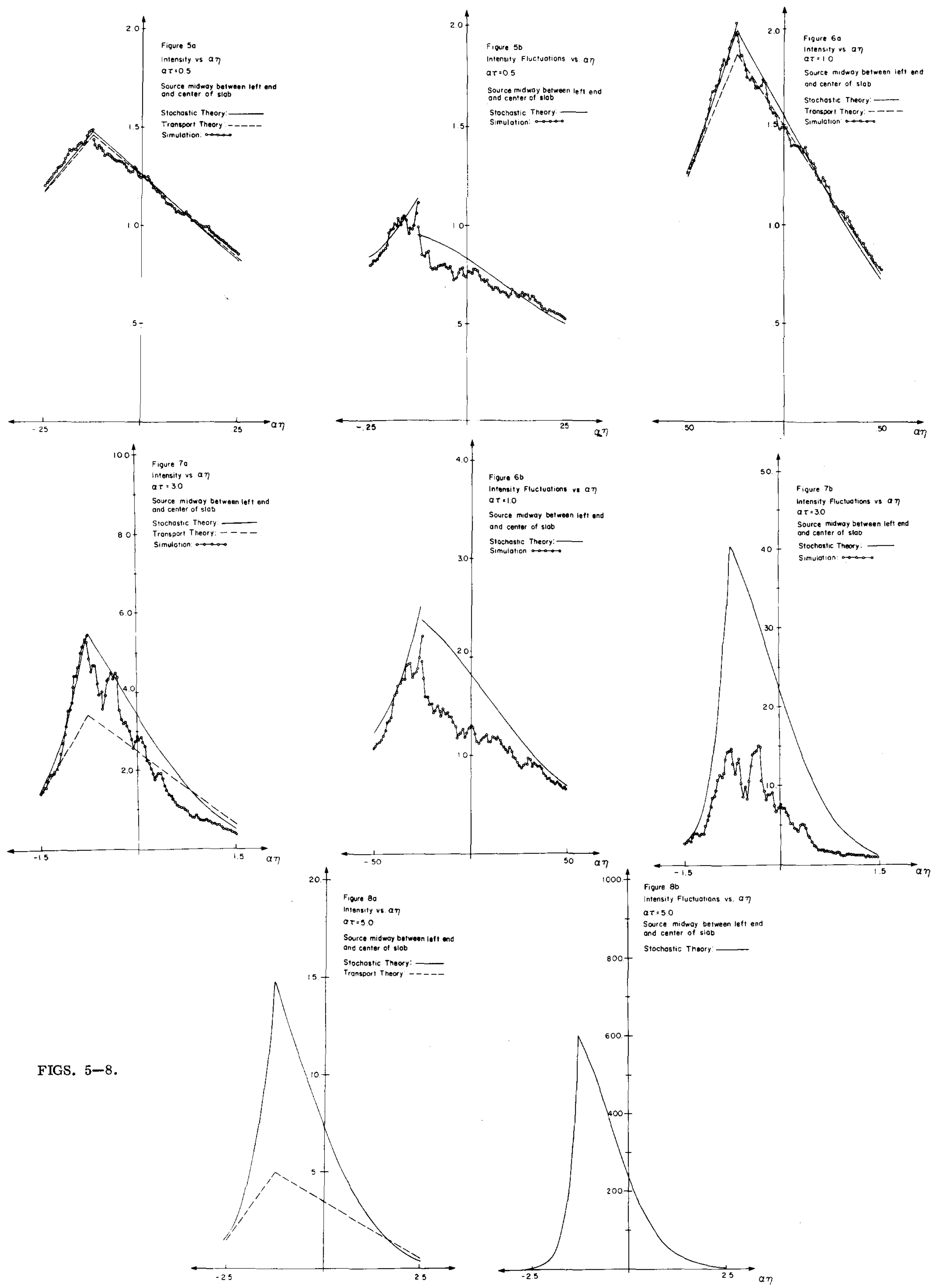
Figure 5a
Intensity vs αη
ατ=0.5
Source midway between left end and center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 5b
Intensity Fluctuations vs. αη
ατ=0.5
Source midway between left end and center of slab
Stochastic Theory:
Simulation:
Figure 6a
Intensity vs αη
ατ=1.0
Source midway between left end and center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 7a
Intensity vs αη
ατ=3.0
Source midway between left end and center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 6b
Intensity Fluctuations vs αη
ατ=1.0
Source midway between left end and center of slab
Stochastic Theory:
Simulation
Figure 7b
Intensity Fluctuations vs αη
ατ=3.0
Source midway between left end and center of slab
Stochastic Theory:
Simulation:
Figure 8a
Intensity vs. αη
ατ=5.0
Source midway between left end and center of slab
Stochastic Theory:
Transport Theory
Figure 8b
Intensity Fluctuations vs. αη
ατ=5.0
Source midway between left end and center of slab
Stochastic Theory:
αη

FIGS. 5—8.

(2.21); the transport theoretic transmission coefficient can be obtained by setting $\xi = -\tau/2$ in (3.5). We have

$$MJ_s(\tau, \tau/2, \tau/2) = 1/(1+\alpha\tau). \tag{3.11}$$

The stochastic transmission coefficient decreases exponentially while the transport theoretic transmission coefficient decreases algebraically. The fluctuations also decrease with increasing $\alpha\tau$. As $\alpha\tau$ increases, therefore, the transmission coefficient approaches zero in probability. This behavior has also been established by Sulem and Frisch[20] when the index of refraction is a random telegraph process and, in fact, convergence is with probability one. For values of $\alpha\tau$ greater than 8, the fluctuations are less than the corresponding spread between stochastic and transport theory intensity predictions. In the light of these observations, one would expect reasonably good agreement between stochastic theory predictions and numerical simulation at the right end of the slab. Moreover, the simulated results should discriminate between the two theories. Simulation results of this sort have been reported by Morrison.[21] The simulations plotted in Figures 1–4 and 13 also behave in this anticipated manner.

In performing the numerical simulations, the expected values were approximated by computing an average over 100 realizations. Each of these realizations in turn was a slab consisting of 2000 sections (i.e., of unscaled length 2000). Within each realization, the index of refraction was assumed to be a two-state random process, with states $\sqrt{1\pm\epsilon}$. The initial state (i.e., the value of the process at the left end of the slab) was chosen randomly; subsequent switching of states occured randomly at intervals which were (approximately) exponentially distributed. In the computations, the average number of sections between changes of the index of refraction was varied from 2.5 to 10, while a wavenumber of 0.5 was used throughout. The parameter ϵ, therefore, was not specified directly but rather was determined by the other variables. Typically, ϵ fell within the range $0.1 \le \epsilon \le 0.3$.

The simulations were beset by two difficulties, the strongly fluctuating nature of the process being simulated and the inherent limitations of the discrete approximating model. Note, in particular, the failure of the simulation model to generate the fluctuations predicted in Figs. 7b and 11b. In general, however, the agreement between the simulated results and stochastic theory is good, and we feel that these simulations amply demonstrate the applicability of the stochastic theory.

Throughout this discussion, we have assumed that frequency (or wavenumber k) is fixed while the spatial variables change. Note, however, from (2.21)–(2.24), (3.4) and the graphs, that MJ, FJ, and MJ_s are functions of $\alpha\tau$, $\alpha\xi$, and $\alpha\eta$, where α is defined by (2.18). Consequently, we could equally well adopt the point of view that the spatial variables are fixed and wavenumber is variable. For fixed τ, ξ, η and a particular correlation function, we could use our graphical data to determine the variation of MJ, FJ, and MJ_s as functions of wavenumber through $\alpha(k)$; observe from (2.1), however, that we would have to account for the fact that our source strength is frequency-dependent.

4. DERIVATION OF THE RESULTS

The derivation that we shall present will rely on Secs. 3–6 of I. In addition to the theory of Ref. 3 (Theorem 3 of Ref. 3) which we used in I, we shall now also apply an improved version of that theory.[4] For the problem being considered, we need the improved theory to conclude that the limit theorem for the propagator matrices (Sec. 4 of I) holds for certain unbounded functions of these matrices. In fact, in I the condition of Theorem 3 (Ref. 3) that $f(g)$ be bounded was violated. With the improved theory, however, such conditions are no longer needed. Hence, the results of I, up to Sec. 9, are rigorously correct. The pulse propagation results of Sec. 9, though, still require additional theoretical considerations because of complications that were overlooked; we shall not pursue this matter here. In the analysis that follows we shall point out where the improved theory is needed.

Let $m(x)$ denote the 2×2 matrix-valued stochastic process:

$$m(x) = \frac{ik\mu(x)}{2}\begin{pmatrix} 1 & \exp(-i2kx) \\ -\exp(i2kx) & -1 \end{pmatrix}. \tag{4.1}$$

Let η_1, η_2, η_3 be defined as the following 2×2 matrices:

$$\eta_1 = \tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \eta_2 = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \eta_3 = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}. \tag{4.2}$$

We can express $m(x)$ in terms of η_1, η_2, η_3 as follows:

$$m(x) = k\mu(x)\eta_1 + (k\mu(x)\sin 2kx)\eta_2 + (k\mu(x)\cos 2kx)\eta_3. \tag{4.3}$$

Note that η_1, η_2, and η_3 constitute a representation of the Lie algebra $su(1,1)$ with commutation relations

$$\eta_1\eta_2 - \eta_2\eta_1 = [\eta_1, \eta_2] = \eta_3, \quad [\eta_1, \eta_3] = -\eta_2, \quad [\eta_2, \eta_3] = -\eta_1. \tag{4.4}$$

Thus, $m(x)$ is a stochastic process with values in $su(1,1)$.

Let $Y(x,y)$ denote the 2×2 matrix solution of the initial value problem:

$$\frac{dY}{dx}(x,y) = \epsilon m(x) Y(x,y),$$
$$Y(y,y) = I \text{ (}2\times 2\text{ identity matrix)}, \quad x \ge y. \tag{4.5}$$

In view of (4.3), $Y(x,y)$ is a stochastic process with values in $SU(1,1)$, the group of 2×2 matrices of the form

$$Y = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix}, \quad |a|^2 - |b|^2 = 1. \tag{4.6}$$

We decompose $Y(l,0)$ into the product

$$\begin{aligned} Y(l,0) &= Y_3(l,y)Y_2(y,x)Y_1(x,0), \quad 0 \le x \le y \le l, \\ Y(l,0) &= Y_3(l,x)Y_2(x,y)Y_1(y,0), \quad 0 \le y \le x \le l. \end{aligned} \tag{4.7}$$

The matrices Y_1, Y_2, and Y_3, when viewed as functions of their first argument, are solutions of (4.5) which equal the identity matrix when their two arguments coincide. Equation (4.7) is simply an expression of the propagator property (cf. I, Sec. 3). To simplify the notation, we omit the arguments and write

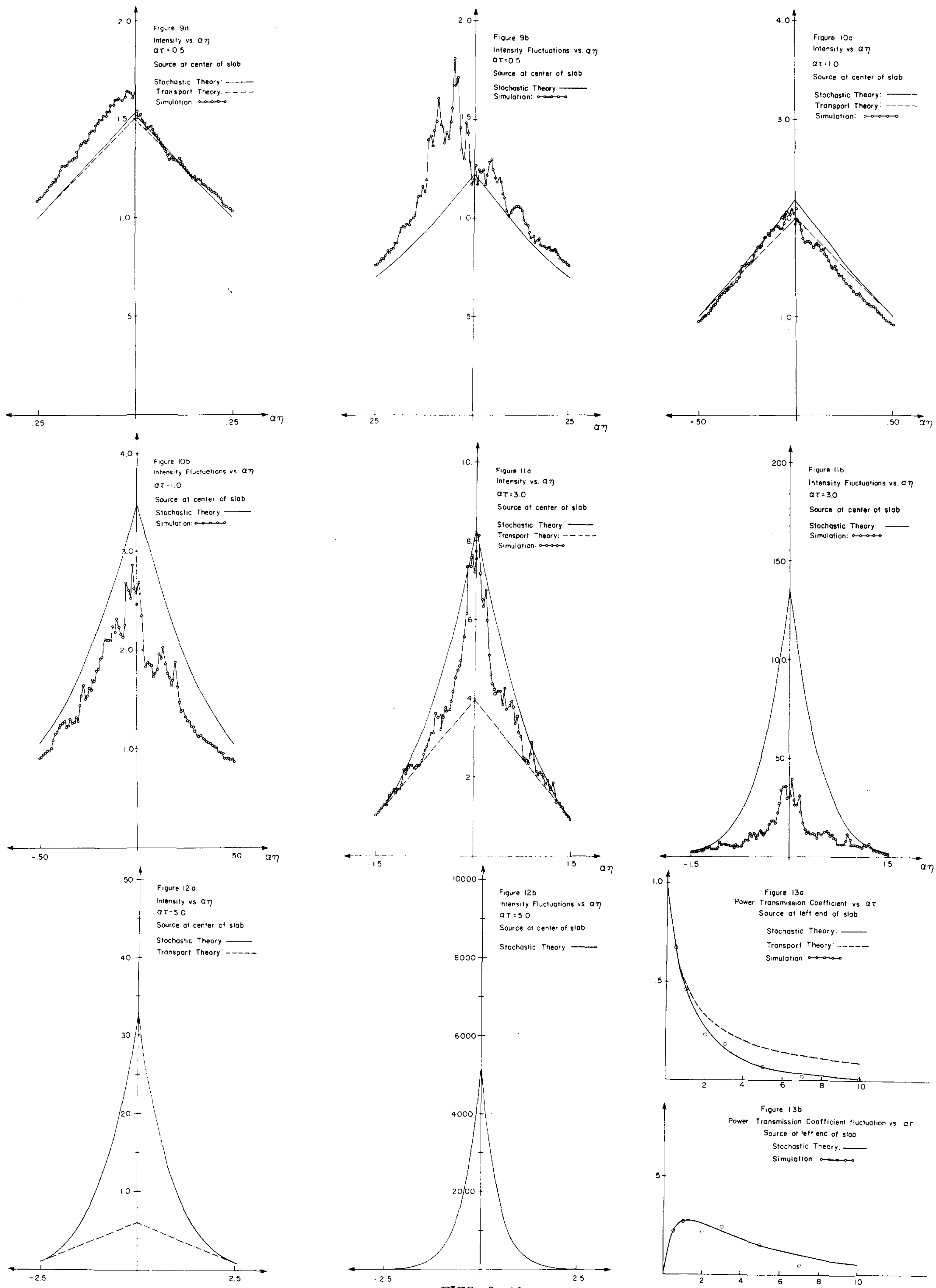
Figure 9a
Intensity vs. $\alpha\eta$
$\alpha\tau = 0.5$
Source at center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 9b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau = 0.5$
Source at center of slab
Stochastic Theory:
Simulation:
Figure 10a
Intensity vs. $\alpha\eta$
$\alpha\tau = 1.0$
Source at center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 10b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau = 1.0$
Source at center of slab
Stochastic Theory:
Simulation:
Figure 11a
Intensity vs. $\alpha\eta$
$\alpha\tau = 3.0$
Source at center of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 11b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau = 3.0$
Source at center of slab
Stochastic Theory:
Simulation:
Figure 12a
Intensity vs. $\alpha\eta$
$\alpha\tau = 5.0$
Source at center of slab
Stochastic Theory:
Transport Theory:
Figure 12b
Intensity Fluctuations vs. $\alpha\eta$
$\alpha\tau = 5.0$
Source at center of slab
Stochastic Theory:
Figure 13a
Power Transmission Coefficient vs. $\alpha\tau$
Source at left end of slab
Stochastic Theory:
Transport Theory:
Simulation:
Figure 13b
Power Transmission Coefficient fluctuation vs. $\alpha\tau$
Source at left end of slab
Stochastic Theory:
Simulation:
$\alpha\eta$

FIGS. 9—13.

$$Y_j=\begin{pmatrix} a_j & b_j \\ \bar b_j & \bar a_j \end{pmatrix}, \quad |a_j|^2-|b_j|^2=1, \quad j=1,2,3. \tag{4.8}$$

The solution of boundary value problem (2.7)–(2.9) can be expressed in terms of Y_1, Y_2, and Y_3 as follows:

$$\left.\begin{aligned} A&=\frac{b_1[\bar a_3\exp(iky)-\bar b_3\exp(-iky)]}{\bar b_3(a_2b_1+b_2\bar a_1)+\bar a_3(\bar b_2 b_1+\bar a_2\bar a_1)} \\ B&=\frac{-\bar a_1[\bar a_3\exp(iky)-\bar b_3\exp(-iky)]}{\bar b_3(a_2b_1+b_2\bar a_1)+\bar a_3(\bar b_2 b_1+\bar a_2\bar a_1)} \end{aligned}\right\}, \quad 0\le x\le y\le l, \tag{4.9}$$

$$\left.\begin{aligned} A&=\frac{\bar a_3[\bar a_1\exp(-iky)+b_1\exp(iky)]}{b_1(\bar b_2\bar a_3+a_2\bar b_3)+\bar a_1(b_2\bar b_3+\bar a_2\bar a_3)} \\ B&=\frac{-b_3[a_1\exp(-iky)+b_1\exp(iky)]}{b_1(\bar b_2\bar a_3+a_2\bar b_3)+\bar a_1(b_2\bar b_3+\bar a_2\bar a_3)} \end{aligned}\right\}, \quad 0\le y\le x\le l\cdot \tag{4.10}$$

These formulas generalize (I. 3.9) and (I. 3.10) for the matched case.

We shall now introduce the notation:

$$\begin{pmatrix} a_{21} & b_{21} \\ \bar b_{21} & \bar a_{21} \end{pmatrix}=\begin{pmatrix} a_2 & b_2 \\ \bar b_2 & \bar a_2 \end{pmatrix}\begin{pmatrix} a_1 & b_1 \\ \bar b_1 & \bar a_1 \end{pmatrix}=\begin{pmatrix} a_2a_1+b_2\bar b_1 & a_2 b_1+b_2\bar a_1 \\ \bar b_2 a_1+\bar a_2\bar b_1 & \bar b_2 b_1+\bar a_2\bar a_1 \end{pmatrix}, \tag{4.11}$$

$$\begin{pmatrix} a_{32} & b_{32} \\ \bar b_{32} & \bar a_{32} \end{pmatrix}=\begin{pmatrix} a_3 & b_3 \\ \bar b_3 & \bar a_3 \end{pmatrix}\begin{pmatrix} a_2 & b_2 \\ \bar b_2 & \bar a_2 \end{pmatrix}=\begin{pmatrix} a_3a_2+b_3\bar b_2 & a_3 b_2+b_3\bar a_2 \\ \bar b_3 a_2+\bar a_3\bar b_2 & \bar b_3 b_2+\bar a_3\bar a_2 \end{pmatrix}. \tag{4.12}$$

Observe that the denominator of (4.9) is $\bar b_3 b_{21}+\bar a_3\bar a_{21}$ and the denominator of (4.10) is $b_1\bar b_{32}+\bar a_1\bar a_{32}$. We shall also introduce polar coordinates, as in Sec. 5 of I, by defining

$$\begin{aligned} a_j&=\exp[i(\phi_j+\psi_j)/2]\cosh(\theta_j/2), \\ b_j&=\exp[i(\phi_j-\psi_j)/2]\sinh(\theta_j/2), \\ j&=1,2,3. \end{aligned} \tag{4.13}$$

Similarly, let

$$\begin{aligned} a_{21}&=\exp[i(\phi_{21}+\psi_{21})/2]\cosh(\theta_{21}/2), \\ b_{21}&=\exp[i(\phi_{21}-\psi_{21})/2]\sinh(\theta_{21}/2), \\ a_{32}&=\exp[i(\phi_{32}+\psi_{32})/2]\cosh(\theta_{32}/2), \\ b_{32}&=\exp[i(\phi_{32}-\psi_{32})/2]\sinh(\theta_{32}/2). \end{aligned} \tag{4.14}$$

The following useful relations, which constitute the law of cosines in the hyperbolic disc, are a consequence of (4.11) and (4.12):

$$\begin{aligned} \cosh\theta_{21}&=\cosh\theta_1\cosh\theta_2+\cos(\phi_1+\psi_2)\sinh\theta_1\sinh\theta_2, \\ \cosh\theta_{32}&=\cosh\theta_2\cosh\theta_3+\cos(\phi_2+\psi_3)\sinh\theta_2\sinh\theta_3. \end{aligned} \tag{4.15}$$

We are primarily interested in $\mathcal{J}^{(\epsilon)}(x,y,l)$, which is defined by (2.12). Using (4.9)–(4.16), we express it as the following function of the polar coordinates:

$$\mathcal{J}^{(\epsilon)}(x,y,l)=\frac{2\cosh\theta_1[\cosh\theta_3-\cos(2ky-\psi_3)\sinh\theta_3]}{1+\cosh\theta_3\cosh\theta_{21}+\cos(\psi_3+\phi_{21})\sinh\theta_3\sinh\theta_{21}}, \quad 0\le x\le y\le l, \tag{4.16}$$

$$\mathcal{J}^{(\epsilon)}(x,y,l)=\frac{2\cosh\theta_3[\cosh\theta_1+\cos(2ky+\phi_1)\sinh\theta_1]}{1+\cosh\theta_1\cosh\theta_{32}+\cos(\phi_1+\psi_{32})\sinh\theta_1\sinh\theta_{32}}, \quad 0\le y\le x\le l. \tag{4.17}$$

These formulas generalize (I. 5.12) and (I. 5.13) for the matched case.

Note that

$$|\mathcal{J}^{(\epsilon)}(x,y,l)|\le\cosh\theta_1(\cosh\theta_3+|\sinh\theta_3|), \quad 0\le x\le y\le l, \tag{4.18}$$

or

$$|\mathcal{J}^{(\epsilon)}(x,y,l)|\le\cosh\theta_3(\cosh\theta_1+|\sinh\theta_1|), \quad 0\le y\le x\le l. \tag{4.19}$$

From (4.13) it follows that:

$$|\mathcal{J}^{(\epsilon)}(x,y,l)|\le(|a_1|^2+|b_1|^2)(|a_3|+|b_3|)^2, \quad 0\le x\le y\le l, \tag{4.20}$$

$$|\mathcal{J}^{(\epsilon)}(x,y,l)|\le(|a_3|^2+|b_3|^2)(|a_1|+|b_1|)^2 \quad 0\le y\le x\le l. \tag{4.21}$$

$\mathcal{J}^{(\epsilon)}$ can therefore be bounded by absolute moments of the elements of the propagator matrices. These are precisely the kind of estimates required in the improved theory of Ref. 4. Theorem 3 of Ref. 3, on the other hand required uniform boundedness. We can thus proceed now as in Secs. 4 and 5 of I without difficulty.

To facilitate application of the limit theorem, expressions (4.18) and (4.19) for $\mathcal{J}^{(\epsilon)}$ will be further transformed. We record the following facts about the Legendre functions $P_\nu^m(u)$ that are needed[22]:

$$\frac{d}{du}\left((u^2-1)\frac{d}{du}P_\nu^m(u)\right)-\frac{m^2}{u^2-1}P_\nu^m(u)=\nu(\nu+1)P_\nu^m(u), \quad u>1, \tag{4.22}$$

$$\frac{2}{u+1}=\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}P_{-1/2+it}(u)\,dt, \tag{4.23}$$

$$P_\nu(\cosh\xi)=\sum_{m=-\infty}^{\infty}\frac{\Gamma(\nu-|m|+1)}{\Gamma(\nu+|m|+1)}P_\nu^{|m|}(\cosh\theta)\times P_\nu^{|m|}(\cosh\tilde\theta)\exp[im(\phi+\psi)], \tag{4.24}$$

$$\cosh\xi=\cosh\theta\cosh\tilde\theta+\cos(\phi+\psi)\sinh\theta\sinh\tilde\theta, \tag{4.25}$$

$$P_\nu^m(u)=P_{-\nu-1}^m(u), \tag{4.26}$$

$$uP_\nu^{|m|}(u)=\frac{1}{2\nu+1}[(\nu-|m|+1)P_{\nu+1}^{|m|}(u)+(\nu+|m|)P_{-\nu-1}^{|m|}(u)], \tag{4.27}$$

$$\frac{4}{(u+1)^2}=\int_{-\infty}^{\infty}(t^2+\tfrac14)\frac{\pi t\sinh\pi t}{\cosh^2\pi t}P_{-1/2+it}(u)\,dt, \quad u\ge1. \tag{4.28}$$

Only (4.23) and (4.28) require a brief comment. Note that if we set $u=1$ and use the fact that $P_\nu(1)=1$, we obtain the relations $\beta_0=1$ and $\beta_1+\frac14\beta_0=1$, respectively [cf. (3.7)]. It is well known that (4.23) follows from the Mehler transform.[22] We can derive (4.28) from (4.23) by observing that the solution $g(\tau,u)$ of the equation

$$\frac{\partial}{\partial\tau}g=\frac{\partial}{\partial u}\left((u^2-1)\frac{\partial}{\partial u}g\right), \quad u>1, \quad g(0,u)=\frac{1}{1+u} \tag{4.29}$$

has the integral representation

$$g(\tau,u)=\int_0^\infty \exp[-(t^2+\tfrac14)\tau]P_{-1/2+it}(u)\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\,dt. \quad (4.30)$$

Therefore, we have

$$g_\tau(0,u)=\left[(u^2-1)\left(\frac{1}{1+u}\right)'\right]'=\frac{-2}{(1+u)^2}. \quad (4.31)$$

By combining (4.31) and (4.30), we obtain (4.28).

We shall analyze (4.16) in detail; the analysis of (4.17) follows in the same way. Observe that we can write

$$\cosh\theta_{321}=\cosh\theta_3\cosh\theta_{21}+\cos(\psi_3+\phi_{21})\sinh\theta_3\sinh\theta_{21}, \quad (4.32)$$

where the triple subscript is an obvious extension of the notation introduced in (4.14) and (4.15). By using (4.23), Eq. (4.16) can be rewritten as follows:

$$J^{(\epsilon)}=\cosh\theta_1[\cosh\theta_3-\cos(2ky-\psi_3)\sinh\theta_3]\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\times P_{-1/2+it}(\cosh\theta_{321})\,dt,\quad 0\le x\le y\le l. \quad (4.33)$$

Addition theorem (4.24) when applied to (4 yields

$$J^{(\epsilon)}=\cosh\theta_1[\cosh\theta_3-\cos(2ky-\psi_3)\sinh\theta_3]\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\times\left(\sum_{m=-\infty}^{\infty}\frac{\Gamma(\nu-|m|+1)}{\Gamma(\nu+|m|+1)}P_\nu^{|m|}(\cosh\theta_3)P_\nu^{|m|}(\cosh\theta_{21})\times\exp(im(\phi_{21}+\psi_3))\right)dt,\quad \nu=-\tfrac12+it \quad (4.34)$$

To decompose $P_\nu^{|m|}(\cosh\theta_{21})$, we require a generalization of addition theorem (4.24). This, in turn, necessitates the introduction of generalized Legendre functions. All necessary information about such functions, including addition theorems and recurrence relations, is given by Vilenkin[23] (in Chap. VI). We record here the required addition theorem:

$$\exp(im\phi_{21})\,P_\nu^{|m|}(\cosh\theta_{21})=\sum_{n=-\infty}^{\infty}\frac{\Gamma(\nu+|m|+1)}{\Gamma(\nu+n+1)}\exp(-in(\phi_1+\psi_2)+im\phi_2)\times P^\nu_{|m|,n}(\cosh\theta_2)\,P_\nu^n(\cosh\theta_1). \quad (4.35)$$

The generalized Legendre function P^ν_{mn} satisfies the differential equation:

$$\frac{d}{du}\left((u^2-1)\frac{d}{du}P^\nu_{mn}(u)\right)-\left(\frac{m^2+n^2-2mnu}{u^2-1}\right)P^\nu_{mn}(u)=\nu(\nu+1)P^\nu_{mn}(u),\quad u>1, \quad (4.36)$$

and also the relation

$$P^\nu_{0n}(\cosh\theta)=\frac{\Gamma(\nu-n+1)}{\Gamma(\nu+1)}P_\nu^m(\cosh\theta). \quad (4.37)$$

Therefore, when $m=0$, addition theorem (4.35) reduces to (4.24).

We now use (4.35) in (4.34) to obtain the desired representation for $J^{(\epsilon)}$ (when $0\le x\le y\le l$):

$$J^{(\epsilon)}=\cosh\theta_1[\cosh\theta_3-\cos(2ky-\psi_3)\sinh\theta_3]\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\times\left(\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{\Gamma(\nu-|m|+1)}{\Gamma(\nu+n+1)}\exp(im(\phi_2+\psi_3))\exp(-in(\phi_1+\psi_2))\times P_\nu^n(\cosh\theta_1)\,P^\nu_{|m|,n}(\cosh\theta_2)\,P_\nu^{|m|}(\cosh\theta_3)\right)dt,\quad \nu=-\tfrac12+it. \quad (4.38)$$

A similar analysis, utilizing (4.28), can be performed to obtain the following representation for $(J^{(\epsilon)})^2$ (when $0\le x\le y\le l$)

$$(J^{(\epsilon)})^2=\cosh^2\theta_1[\cosh\theta_3-\cos(2ky-\psi_3)\sinh\theta_3]^2\int_{-\infty}^{\infty}(t^2+\tfrac14)\times\frac{\pi t\sinh\pi t}{\cosh^2\pi t}\left(\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{\Gamma(\nu-|m|+1)}{\Gamma(\nu+n+1)}\exp(im(\phi_2+\psi_3))\times\exp(-in(\phi_1+\psi_2))P_\nu^n(\cosh\theta_1)P^\nu_{|m|,n}(\cosh\theta_2)\times P_\nu^{|m|}(\cosh\theta_3)\right)dt,\quad \nu=-\tfrac12+it. \quad (4.39)$$

We now apply the diffusion limit, i.e., (2.16) and (2.22), to (4.38) and (4.39) to obtain MJ and KJ, respectively. Thus, we take expectation or statistical average of $J^{(\epsilon)}$ and $(J^{(\epsilon)})^2$ with respect to the limiting probability distribution of the propagators. As explained in Sec. 4 of I, the propagator matrices corresponding to nonoverlapping intervals become statistically independent in the diffusion limit. Up to this point, our analysis could be applied to the mismatched problem by adjusting the initial value of Y_1 and the final value of Y_3 (cf. Sec. 3 of I). In the sequel, we shall consider only the matched case; the formulas for the more general mismatched case are unwieldy but not particularly difficult to obtain.

Observe that the angle ψ_1 is absent in both (4.38) and (4.39). As explained in Sec. 5 of I, the limiting transition density for ϕ is uniformly distributed over $[0,2\pi]$ and is independent of θ_1. (Recall that we are dealing with the matched case.) Therefore, when we average (4.38) and (4.39) with respect to ϕ_1, only the $n=0$ term survives. Moreover, since ψ_2 appears only in the combination $(\phi_1+\psi_2)$, this average with respect to ϕ_1 also eliminates functional dependence upon ψ_2. We next average with respect to ϕ_2. The same argument applies to this case also; the transition density for ϕ_2 is uniformly distributed over $[0,2\pi]$ and independent of θ_2. When this average is performed, only the $m=0$ term survives. We next average with respect to the uniformly distributed angle ψ_3. If we use angular brackets $\langle\cdot\rangle$ to denote averaging with respect to the angles ϕ_1, ϕ_2, and ψ_3, then (4.38) and (4.39) yield

$$\langle J\rangle=\cosh\theta_1\cosh\theta_3\int_{-\infty}^{\infty}\frac{\pi t\sinh\pi t}{\cosh^2\pi t}P_\nu(\cosh\theta_1)\times P_\nu(\cosh\theta_2)\,P_\nu(\cosh\theta_3)\,dt, \quad (4.40)$$

$$\langle J^2\rangle=\cosh^2\theta_1\left(\frac{3\cosh^2\theta_3-1}{2}\right)\int_{-\infty}^{\infty}(t^2+\tfrac14)\frac{\pi t\sinh\pi t}{\cosh^2\pi t}P_\nu(\cosh\theta_1)\times P_\nu(\cosh\theta_2)P_\nu(\cosh\theta_3)\,dt,\quad \nu=-\tfrac12+it. \quad (4.41)$$

To obtain MJ and KJ, we must average (4.40) and (4.41) with respect to the limiting distributions of θ_1, θ_2, and θ_3. Since we are in the matched case, it follows from (I. 5.21) [or directly from (I. 5.16)] that in the diffusion limit

$$E\{P_\nu(\cosh\theta(\tau))\} = \exp(\nu(\nu+1)\alpha\tau). \tag{4.42}$$

In (4.40) and (4.41) we must take the expected value of functions of the form $u^j P_\nu(u)$, $j=0,1,2$, where $u=\cosh\theta$. Notice, however, that by using (4.26) and (4.27) we can rewrite $uP_\nu(u)$ and $u^2P_\nu(u)$ as linear combination of Legendre functions with different degrees. We use the scaled variables defined by (2.15) in specifying the arguments of the random functions θ_1, θ_2, and θ_3. For the case being considered, i.e., $0 \leqslant x \leqslant y \leqslant l$, these arguments are $\tau/2+\eta$, $\xi-\eta$ and $\tau/2-\xi$, respectively. Equation (4.42) is used repeatedly and after some straightforward but lengthy computations, (2.17) is obtained from (4.40) and (2.24) from (4.41). By starting with (4.17), the analysis for the other case, i.e., $0 \leqslant y \leqslant x \leqslant l$, follows in basically the same way as the case we have discussed.

Gazaryan[7] studied the configuration corresponding to the matched case with a source location at the left end of the slab ($\xi = -\tau/2$). He observed that the total intensity in the interior, i.e., $MJ(\tau, -\tau/2, \eta)$, satisfies the heat equation in the variables τ and η. Moreover, Eq. (2.19) with $\tau=0$, can be recast into the following very simple initial condition:

$$MJ(0,0,\eta) = 1 - \tanh\alpha\eta - \alpha\eta\,\mathrm{sech}^2\alpha\eta. \tag{4.43}$$

Using the fundamental solution to the heat equation, Gazaryan obtained an alternate representation for $MJ(\tau, -\tau/2, \eta)$. His observation was generalized and used extensively in Sec. 6 of I.

For the general configuration, where the source location is permitted to vary, one can show that $MJ(\tau, \xi, \eta)$ satisfies the following partial differential equation:

$$\partial_{\alpha\tau}MJ = \tfrac{1}{4}\{\tfrac{1}{2}(\partial^2_{\alpha\eta,\alpha\eta} + \partial^2_{\alpha\xi,\alpha\xi}) - \sigma(\partial_{\alpha\eta} - \partial_{\alpha\xi})\}MJ,$$
$$\sigma = \mathrm{sgn}(\eta - \xi). \tag{4.44}$$

However, KJ does not satisfy such a simple partial differential equation. Moreover, the simplicity of the approach as a whole is apparently lost in this more general problem. Consequently, the approach was not pursued.

ACKNOWLEDGMENT

We wish to thank Uriel Frisch for bringing the interior source problem to our attention and J.B. Keller for useful comments and suggestions.

*Research supported by the Air Force Office of Scientific Research under Grant No. AFOSR-71-2013.

[1]W. Kohler and G.C. Papanicolaou, J. Math. Phys. **14**, 1733 (1973).
[2]G.C. Papanicolaou and J.B. Keller, SIAM J. Appl. Math. (Soc. Ind. Appl. Math.) **21**, 287 (1971).
[3]G.C. Papanicolaou and S.R.S. Varadhan, Comm. Pure Appl. Math. **26**, 497 (1973).
[4]G.C. Papanicolaou and W. Kohler, Comm. Pure Appl. Math. **27**, (1974) (to be published).
[5]J.A. Morrison and J. McKenna, SIAM—AMS Proc. 6, 97 (1973).
[6]A. Brissaud and U. Frisch, J. Math. Phys. **15**, 524 (1974).
[7]Yu. L. Gazaryan, Zh. Eksp. Teor. Fiz. **56**, 1856 (1969) [Sov. Phys. JETP **29**, 996 (1969)].
[8]R.H. Lang, J. Math. Phys. **14**, 1921 (1973).
[9]R.J. Rubin, J. Math. Phys. **9**, 2252 (1968).
[10]R.J. Rubin, J. Math. Phys. **11**, 1857 (1970).
[11]B.I. Halperin, Phys. Rev. **139**, A104 (1965).
[12]D. Marcuse, Bell Sys. Tech. J. **51**, 1793 (1972).
[13]A. Schuster, Astrophys. J. **21**, 1 (1905).
[14]I.M. Besieris and F.D. Tappert, J. Appl. Phys. **44**, 2119 (1973).
[15]For the validity of the diffusion limit it is required that $\mu(x)$ be almost surely bounded and mixing in a sufficiently strong sense (cf. Refs. 1, 3, and 4). Model media with $\mu(x)$ a finite state ergodic Markov chain are included in our formulation.
[16]This denotes Eq. (2.15) in I.
[17]The leading minus sign on the third line of (I.6.32) should be a plus sign.
[18]D. Marcuse, IEEE Trans. MTT **20**, 541 (1972).
[19]U. Frisch, C. Froeschle, J.-P. Scheidecker, and P.L. Sulem, Phys. Rev. A **8**, 1416 (1973).
[20]P.L. Sulem and U. Frisch, J. Plasma Phys. **8**, 217 (1972).
[21]J.A. Morrison, IEEE Trans. MTT **22**, 126 (1974).
[22]W. Magnus, F. Oberhettinger and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[23]N.J. Vilenkin, *Special Functions and the Theory of Group Representations*, 22, Transl. of Math. Monographs (AMS, Providence, R.I., 1968).

Proof of the charge superselection rule in local relativistic quantum field theory*

F. Strocchi

INFN, Sezione di Pisa, Scuola Normale Superiore, Pisa, Italy

A. S. Wightman

Joseph Henry Laboratory of Physics, Princeton University, Princeton, New Jersey 08540

(Received 22 July 1974)

The paper interprets and proves the charge superselection rule within the framework of local relativistic field theory as the statement that the charge operator commutes with all quasilocal observables. Once the basic formalism expressing the property of locality of the observables has been accepted, the proof is an elementary application of Gauss law relating the electric charge in a region to the flux of electric field through the boundary of the region. Most of the paper is devoted to the evidence that the indefinite metric formalism and its accompanying definitions of gauge, gauge transformation, and gauge invariance are internally coherent and consistent with the evidence from free field theory and the renormalized perturbation theory of coupled fields. The paper closes with speculations on analogous explanations of the baryon and lepton superselection rules within the framework of gauge models of strong and weak interactions.

1. INTRODUCTION

In its most general form, a *superselection rule* for a quantum mechanical theory can be defined as any restriction on what is observable in the theory. In its traditional more restricted form, a superselection rule is specified by an Hermitian operator A commuting with all observables of the theory, and the requirement that no observed states Φ of the theory are nontrivial superpositions $\alpha_1\Phi_1 + \alpha_2\Phi_2$, $\alpha_1\alpha_2 \neq 0$, of eigenstates Φ_1, Φ_2 belonging to distinct eigenvalues of A.[1] There are standard invariance arguments for the existence of some superselection rules. For example, the *univalence superselection rule*, which says that $(-1)^{2J}$ commutes with all observables where J is the total angular momentum operator can be deduced in any rotationally invariant theory in which spinors are not observables.[2] Nevertheless, there did not exist any systematic theoretical framework in which the existence of superselection rules had a natural place until the work of Haag and Kastler.[3] Without going into the details of their proposal let us recall how superselection rules appear in it.

The basic construct of the Haag–Kastler theory is the *quasilocal algebra* $\mathfrak{A}$ and its associated subalgebras $\mathfrak{A}(\mathcal{O})$, the *local algebras* of bounded space–time regions $\mathcal{O}$. The *local observables* attached to the region $\mathcal{O}$ are self-adjoint elements of $\mathfrak{A}(\mathcal{O})$ and $\mathfrak{A}(\mathcal{O})$ is supposed to be generated by such observables. A state ω on $\mathfrak{A}$ is a positive linear functional on $\mathfrak{A}$ normalized to 1 at the identity element of $\mathfrak{A}$:

$$\omega(\mathbb{1}) = 1.$$

Each such state determines uniquely a cyclic representation of $\mathfrak{A}$ and conversely. (For the definition of this representation by the so-called GNS construction see, for example, Ref. 4.) Not all cyclic representations are physically relevant, and much effort has gone into locating appropriate restrictive criteria to exclude pathological or physically irrelevant representations.[5–8] We will not concern ourselves here with the precise statement of such criteria. Instead, we will simply use the phrase *physically admissible* to indicate a state selected according to some appropriate criteria. The important point is that, once the definition of physically admissible has been fixed, the superselection sectors of a Haag–Kastler theory are given by the unitarily inequivalent physically admissible representations of the quasilocal algebra. More precisely, one chooses one representation from each unitary equivalence class. The Hilbert space of the full physical theory is the direct sum of the Hilbert spaces in which these representations are realized. Thus, in a sense, *a Haag–Kastler theory predicts its own superselection rules.*

For example, in a theory in which electric charge defines the only superselection rule the quasilocal algebra ought to have one unitary equivalence class of physically admissible representations for each integer value of the charge, and no others. The Hilbert space $\mathcal{H}$ of such a theory is a direct sum $\oplus_{n=-\infty}^{\infty}\mathcal{H}_n$, where $\mathcal{H}_n$ is the subspace whose vectors have charge n.

In recent work Doplicher, Haag, and Roberts have gone much further in developing this idea, giving a detailed analysis of the construction of the superselection sectors in the context of a theory that has a group of local gauge transformations.[6–8] As the authors emphasize, this work applies only to theories in which long-range forces are absent (theories with a mass gap). It was natural to work out this case first, since it is technically easier and the general problems in this subject are formidable indeed. However, to treat the superselection rule of electric charge, as we do in the present paper, one must go further since massless photons must be dealt with.

In fact, as is well known (see, for example, Ref. 6, p. 3 or Refs. 9–11) but not so widely appreciated, to treat quantum electrodynamics without abandoning a description in terms of local fields, one has to generalize the basic framework of Haag and Kastler in an essential way: the notion of state described above does not suffice to provide the representations necessary for the construction of the electromagnetic vector potential as a local field nor of charged fields as local fields. A formalism with indefinite metric of the type introduced by Gupta and Bleuler is necessary. Since this result is essential for the justification of the assumptions we make in the following, let us recall some of the evidence that supports it.

Suppose one attempts to construct a theory describing a free electromagnetic field by a field operator $F_{\mu\nu}$ and a vector potential operator A_μ satisfying

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.1}$$

and that one accepts the usual assumptions:

(1) $F_{\mu\nu}$ and A_μ are operator valued distributions on Minkowski space with values which are unbounded operators in a Hilbert space $\mathcal{H}$.

(2) In $\mathcal{H}$ there is a continuous representation of the restricted Poincaré group $\{a, A\} \to U(a,\Lambda)$ such that

$$U(a,\Lambda)F_{\kappa\lambda}(x)U(a,\Lambda)^{-1} = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda F_{\mu\nu}(\Lambda x + a). \tag{1.2}$$

There is a sesquilinear form $\langle \cdot , \cdot \rangle$ such that $U(a,\Lambda)$ is unitary with respect to $\langle \cdot , \cdot \rangle$

$$\langle U(a,\Lambda)\Phi, U(a,\Lambda)\Psi\rangle = \langle \Phi, \Psi\rangle. \tag{1.3}$$

For A_μ one assumes initially only a transformation law under translations

$$U(a,1)A_\mu(x)U(a,1)^{-1} = A_\mu(x+a). \tag{1.4}$$

(3) There exists a unique vector Ψ_0 in $\mathcal{H}$ invariant under $U(a,\Lambda)$

$$U(a,\Lambda)\Psi_0 = \Psi_0. \tag{1.5}$$

Theorem 1.1[9]: In a local quantum field theory with fields $F_{\mu\nu}$, ϕ_α, $\alpha = 1, \ldots, n$, in which (1), (2), (3) hold, the assumptions

$$\partial^\mu F_{\mu\nu} = 0 \tag{1.6}$$

and

$$U(0,\Lambda)A_\mu(x)U(0,\Lambda)^{-1} = \Lambda^\nu{}_\mu A_\nu(\Lambda x) \tag{1.7}$$

imply

$$\langle \Psi_0, F_{\kappa\lambda}(x)F_{\mu\nu}(y)\Psi_0\rangle = 0.$$

This theorem shows that the validity of Maxwell's equations (with vanishing electric current) together with the Lorentz transformation law of the vector potential leads to a trivial theory.

A second result of this kind is

Theorem 1.2[11]: In a local quantum field theory with fields $F_{\mu\nu}$, ϕ_α, $\alpha = 1, \ldots, n$, in which (1), (2), (3) hold, the assumptions

$$\partial^\mu F_{\mu\nu} = 0$$

and

$$(\Psi_0, [A_\mu(x), A_\nu(y)]\Psi_0) = 0 \tag{1.8}$$

for spacelike $x-y$, i.e., $(x-y)^2 < 0$, imply

$$\langle \Psi_0, F_{\kappa\lambda}(x)F_{\mu\nu}(y)\Psi_0\rangle = 0.$$

The net effect of these results is to show that a theory of the free electromagnetic field that maintains the Maxwell equations as operator identities must use a vector potential that is both nonlocal and not a Lorentz convariant vector field. That is precisely what happens in Coulomb gauge quantum electrodynamics.

The Gupta–Bleuler formalism evades these difficulties by abandoning the Maxwell equations as operator identities. It insists on locality and Lorentz invariance but replaces (1.6) by

$$\partial^\mu F_{\mu\nu} = -\partial_\nu(\partial^\lambda A_\lambda) \tag{1.9}$$

valid in a linear space $\mathcal{H}$. $\mathcal{H}$ has a distinguished subspace $\mathcal{H}'$ on which the modified Lorentz condition

$$\partial^\mu A_\mu^{(-)}\Psi = 0 \tag{1.10}$$

holds. $\mathcal{H}$ is equipped with a sesquilinear Hermitian form $\langle \Phi, \Psi\rangle$ which is semidefinite on $\mathcal{H}'$.

$$\langle \Phi, \Phi\rangle \geq 0, \quad \Phi \in \mathcal{H}' \tag{1.11}$$

(sesquilinear means $\langle \Phi, \Psi\rangle$ is antilinear in Φ and linear in Ψ, and Hermitian means $\langle \Phi, \Psi\rangle = \overline{\langle \Psi, \Phi\rangle}$). $\mathcal{H}'$ in turn has a distinguished subspace $\mathcal{H}''$ consisting of the vectors Φ of zero length

$$\langle \Phi, \Phi\rangle = 0. \tag{1.12}$$

Vectors describing physical states are elements of the quotient space $\mathcal{H}'/\mathcal{H}'' = \mathcal{H}_{\text{phys}}$. Maxwell's equations (1.6) are valid in the sense that

$$\langle \Phi, \partial^\mu F_{\mu\nu}\Psi\rangle = 0 \tag{1.13}$$

for $\Phi, \Psi \in \mathcal{H}'$.

It appears that many physicists find the nonphysical aspects of the Gupta–Bleuler formalism repellent,[12] so it is natural to ask whether the difficulties involved in satisfying Maxwell's equations using a vector potential are special to the *free* electromagnetic field. The following makes it clear that analogous troubles arise in the presence of charges. [Recall that a field Φ carries a charge q if $\exp(i\alpha Q)\phi\exp(-i\alpha Q) = \exp(i\alpha q)\phi$ where Q is the electric charge operator and α is any real number.]

Theorem 1.3[13]: In any quantum field theory in which a charged field ϕ is defined as a local field in a Hilbert space equipped with a nondegenerate sesquilinear form $\langle \cdot , \cdot \rangle$ Maxwell's equation

$$\partial_\mu F^{\mu\nu} = j^\nu, \quad \partial_\mu {}^*F^{\mu\nu} = 0, \tag{1.14}$$

cannot hold as operator equations in $\mathcal{H}$. Moreover, if $\mathcal{H}'$ is a linear manifold $\subset \mathcal{H}$, stable under $\mathcal{A}^\nu = \partial_\mu F^{\mu\nu} - j^\nu$,

$$\mathcal{A}^\nu \mathcal{H}' \subset \mathcal{H}', \tag{1.15}$$

and such that

$$\langle \Phi, \Phi\rangle \geq 0 \quad \text{for } \Phi \in \mathcal{H}' \tag{1.16}$$

and

$$\langle \Phi, \mathcal{A}^\nu\Psi\rangle = 0 \quad \text{for all } \Phi, \Psi \in \mathcal{H}', \tag{1.17}$$

then $\langle \cdot , \cdot \rangle$ cannot be strictly positive on $\mathcal{H}'$ nor nonnegative on $\mathcal{H}$ unless

$$\langle \Phi, \phi(f)\Psi\rangle = 0 \quad \text{for all } \Phi, \Psi \in \mathcal{H}'. \tag{1.18}$$

The mathematical consistency of the Gupta–Bleuler construction of the free Maxwell theory is well known. (see, for example, Ref. 14, pp. 169–72.) One can object to it at most on aesthetic grounds. Its extension to a theory with nonvanishing electric current has at least the same status as any other nontrivial renormalizable local relativistic quantum field theory: Its renormalized perturbation series are known to exist term by term. (The idea of the proof modulo infrared divergences goes back to Dyson[15]; the proof that with the appropriate definitions there are no infrared divergences in the Green's functions and retarded functions is a recent result.[16]) Thus, if one seeks to maintain the ideas on which Haag and Kastler base their assumption about local algebras of observables and at the same time incorporate the laws of electrodynamics it is not unnatural to use an indefinite metric formalism.

The purpose of the present paper is to answer the question: Given quantum electrodynamics in an indefinite metric formalism, what is the status of the charge superselection rule? We will show that it is a *theorem* of the theory: *All quasilocal observables necessarily commute with the total charge.*

As the reader will see, once the definitions have been fixed, the proof is an elementary application of Gauss' theorem of electrostatics. If we simply ignore the difficulties indicated in Theorems 1.1, 1.2, and 1.3 and work in the Coulomb gauge, it goes as follows: If A is a local operator, i.e., an operator representing an observable of the bounded region $\mathcal{O}$ of space, what has to be proved is

$$[Q,A]=0 \tag{1.19}$$

where Q is the electric charge. Now

$$Q=\lim_{R\to\infty}\int_{|\mathbf{x}|\le R}\rho(\mathbf{x},t)\,d^3x \tag{1.20}$$

and

$$\operatorname{div}\mathcal{E}=\rho. \tag{1.21}$$

Thus

$$\begin{aligned}[Q,A]&=\lim_{R\to\infty}\int_{|\mathbf{x}|\le R}[\rho(\mathbf{x},t),A]\,d^3x\\&=\lim_{R\to\infty}\int_{|\mathbf{x}|=R}[\mathcal{E}(\mathbf{x},t),A]\cdot ds.\end{aligned} \tag{1.22}$$

But for R sufficiently large, the electric field is evaluated only outside the region $\mathcal{O}$ and therefore the commutator should vanish, and consequently Q should commute with A.

Why does this not settle the matter and eliminate the need for the following rather long argument? In our opinion the answer is threefold. First of all, the argument depends on the use of Maxwell's equation (1.21) which cannot hold in any local or covariant gauge. On the other hand, in the Coulomb gauge in which (1.21) *is* valid, the argument (1.20)—(1.22) is illegitimate or at least highly dubious because renormalization constants appear in the equation relating the charge density and the total charge:

$$[\operatorname{div}\mathcal{E}(t,\mathbf{x}),\psi_c(t,\mathbf{y})]=-e_{\text{ren}}Z_3^{-1}\delta(\mathbf{x}-\mathbf{y})\psi_c(t,\mathbf{x}). \tag{1.23}$$

[See Ref. 17, pp. 107—9. The relevant equation (6.42c) contains a misprint; there should be a factor e_{ren} on the right-hand side.) Thus, in fact, the limit on the right-hand side of (1.20) does not exist if $Z_3=0$ as is generally regarded as likely except possibly for isolated values of e_{ren}. If $Z_3\neq 0$, the limit would exist but would not yield the charge unless $Z_3=1$, which is generally regarded as unlikely except when $e_{\text{ren}}=0$. These statements hold even if the limit is understood as holding only for matrix elements between appropriately chosen states dense in the Hilbert space of states. Secondly, as will be discussed in somewhat more detail later, the Coulomb gauge may have appreciably worse ultraviolet properties than the local covariant gauges. The argument for the validity of these statements may be based either on perturbation theory or formal canonical field theory. All the needed formulas are to be found in Ref. 17, pp. 94—112. Thirdly, even if one accepts this argument in the Coulomb gauge, one has the problem of obtaining the same result in local covariant gauges, since only in those gauges is it possible to do field theory in an explicitly local and covariant form consistent with the style of the Haag—Kastler theory.

The reader will see below that the proof in any local covariant gauge requires only slight modifications of the above questionable proof. There are no new ideas involved, just a conscientious use of the theory in a local covariant gauge. The fact that the revised proof is still very simple does not make the result any less true, nor, in the opinion of the authors, less significant.

In the course of correspondence on the subject of this paper with Haag, we learned of Ref. 18, which we had previously overlooked. In it the above explanation (1.19)—(1.22) of the origin of the charge superselection rule in Maxwell's equations and the locality of observables was clearly outlined. (see especially p. 34 where the following statement occurs: "... Die physikalisch wesentliche Aussage der superselection rule der Ladung ist, grob gesprochen, dass bereits die Ladung, welche in einem endlichen Raumgebiet enthalten ist, mit allen Observablen innerhalb dieses Gebietes kommutiert.") We also learned in correspondence with Haag and Swieca that in the middle sixties the latter had formulated and proved the statement: "In quantum electrodynamics there are no local charge carrying fields," but it remained unpublished. All this work was done in Coulomb gauge but as the authors made clear to us, they felt that was only a matter of convenience. As we have just stated we believe the argument to be unsound in the Coulomb gauge but its modification in local covariant gauges to be sound. However, we do not wish thereby to claim originality for the ideas involved in the argument. (see, in particular, Ref. 19, p. 228.) The purpose of the present paper will be achieved if it makes very plain to the reader that the charge superselection rule is, in local relativistic field theory, a consequence of the dynamical laws governing the behavior of charges.

If one has recognized the dynamical origin of the charge superselection rule in Maxwell's equations it is natural to ask whether the baryon superselection rule and the (presumed) lepton superselection rule have analogous explanations in the special structure of strong and weak dynamics. The answer is yes in appropriately chosen local gauge theories if the solutions of the theories exist and have the properties currently attributed to them.

In outline, the paper is arranged as follows. In Sec. 2 we introduce notions of gauge, gauge transformation, and gauge invariance and test their effectiveness on the available evidence in free field theory and perturbation theory. In Sec. 3 the charge superselection rule is stated and proved. Section 4 contains an account, necessarily speculative, of possible dynamical explanations of the baryon and lepton superselection rules.

2. GAUGES, GAUGE TRANSFORMATIONS, AND VARIOUS NOTIONS OF GAUGE INVARIANCE

It is the purpose of this section to motivate and then define precisely *gauge*, *gauge transformation*, and four notions of gauge invariance. We go to considerable lengths to verify that the definitions in question work smoothly for the free electromagnetic field in most of the special gauges we are aware of. The definitions of *strict gauge invariance*, *gauge invariance*, and *weak gauge invariance* we adopt were, to the best of our knowledge, first introduced by Symanzik,[17] see also Ref. 9. As will be seen in the following section, it is a fourth notion, *gauge independence*, that is needed in the proof of the charge superselection rule.

The reader who is uninterested in motivation and circumstantial evidence for the reasonableness of definitions is advised to proceed directly to Sec. 3 after reading the Definitions, Lemmas, and Propositions.

Gauge invariance in classical and Schrödinger theory

In the classical magnetohydrodynamics of a charged fluid, the equations of motion are Maxwell's equations

$$\partial_\mu F^{\mu\nu} = j^\nu, \quad \partial_\mu {}^*F^{\mu\nu} = 0, \tag{2.1}$$

$$j^\mu = \frac{q}{m}\rho v^\mu \tag{2.2}$$

and the Maxwell–Lorentz equations

$$\partial_\mu(\rho v^\mu v^\nu) = F^{\nu\mu} j_\mu. \tag{2.3}$$

Here ρ is the mass density of the fluid and v^μ its 4-velocity. In this theory the equations of motion are expressed directly in terms of the electromagnetic field $F_{\mu\nu}$ and the observables ρ, v^μ. The introduction of a vector potential is purely a matter of mathematical convenience. Independence of the choice of gauge has no dynamical consequences that are not already explicit in the equations of motion. If all the observables of the theory are regarded as functions of the quantities ρ, v, F, they are automatically independent of the choice of gauge.

In a classical field theory in which some of the basic fields carry charge the situation is different. For example, in the theory of a coupled Maxwell and charge scalar field the Maxwell equations (2.1) are supplemented by

$$j^\mu = \frac{q\hbar}{2m} i(\bar\phi(\partial^\mu + iq(\hbar c)^{-1}A^\mu)\phi - (\partial^\mu - iq(\hbar c)^{-1}A^\mu)\bar\phi\phi) \tag{2.4}$$

and

$$(\partial^\mu + iq(\hbar c)^{-1}A^\mu)(\partial_\mu + iq(\hbar c)^{-1}A_\mu)\phi + \left(\frac{mc}{\hbar}\right)^2\phi = 0. \tag{2.5}$$

The invariance of the dynamics under a change of gauge is expressed by the fact that the equations retain their form under the replacement

$$\phi \to \tilde\phi = \exp[iq(\hbar c)^{-1}\chi]\phi, \tag{2.6}$$

$$A_\mu \to \tilde A_\mu = A_\mu - \partial_\mu\chi, \tag{2.7}$$

i.e., the dynamics is invariant under the gauge transformation (2.6) and (2.7). In such a theory all observables are expressible as functions of A_μ and the charged fields and it is natural to impose the physical requirement that they too be invariant under gauge transformation. In the example of the charged scalar field satisfying (2.4) and (2.5), the reasonableness of the requirement of gauge invariance for all observables is further buttressed by the fact that the theory can be rewritten (at least formally) in terms of the manifestly gauge invariant quantities, j, F,

$$s = \phi^*\phi, \tag{2.8}$$

and $\dot s$, the time derivative of s.[20] Thus, the electrodynamics of a charge scalar field when rewritten in this way becomes a manifestly gauge independent theory like magnetohydrodynamics (2.1), (2.2), (2.3).

Is every theory of charged fields that is invariant under gauge transformations such a manifestly gauge invariant theory in disguise? The answer is not known, but it seems plausible that the answer ought to be yes. To the extent that these examples are a guide, it is not unreasonable to accept the requirement of the invariance of dynamical equations under gauge transformations as a physical restriction on classical field theories.

Next let us turn to the analogous situation in n-body Schrödinger theory. There the state of the system is described by a wavefunction $\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_n, t)$ satisfying the (in general, time-dependent) Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(t) = H(t)\Psi(t) \tag{2.9}$$

with

$$H(t) = \sum_{j=1}^n \frac{1}{2m_j}\left(\mathbf{p}_j - \frac{q_j}{c}\mathbf{A}(\mathbf{x}_j, t)\right)^2 + \sum_{j=1}^n q_j\phi(\mathbf{x}_j, t) + V(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

The vector potential $\mathbf{A}$ and the scalar potential ϕ are regarded as given functions of $\mathbf{x}$ and t. The analogs of the transformations (2.6) and (2.7) are here

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, t) \to \exp\left(i\sum_{j=1}^n q_j(\hbar c)^{-1}\chi(\mathbf{x}_j, t)\right)\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_n, t),$$

$$\mathbf{A}(\mathbf{x}, t) \to \mathbf{A}(\mathbf{x}, t) + \nabla\chi(\mathbf{x}, t), \tag{2.10}$$

$$\phi(\mathbf{x}, t) \to \phi(\mathbf{x}, t) - \frac{\partial\chi}{\partial t}(\mathbf{x}, t),$$

and all presently existing evidence points to the idea that all observables are invariant under such a change of gauge, as the time-dependent Schrödinger equation is.

As is well known, the presence of suitable vector and scalar potentials in the Schrödinger equation gives rise to subtle interference phenomena (Ehrenburg–Siday–Aharanov–Bohm effects[21,22]), effects which apparently depend on $\mathbf{A}$ and ϕ rather than on the electromagnetic field strengths $\mathcal{E}$ and $\mathcal{B}$ at the location of the electron. These phenomena are relevant here because at first sight their existence appears to contradict the conjecture made above that the equations of motion for gauge invariant theories can be rewritten in equivalent manifestly gauge invariant form as equations of motion for local quantities.

To understand the solution of this apparent paradox, it suffices to consider the case of one particle ($n=1$). Then it is natural to introduce the local gauge-invariant quantities, the density

$$\rho(\mathbf{x},t)=q\,|\Psi(\mathbf{x},t)|^2, \tag{2.11}$$

the current

$$\mathbf{j}(\mathbf{x},t)=\frac{q\hbar}{2mi}\left(\overline{\Psi(\mathbf{x},t)}\left(\nabla-\frac{qi}{\hbar c}\mathbf{A}(\mathbf{x},t)\right)\Psi(\mathbf{x},t)\right.$$
$$\left.-\left(\nabla+\frac{qi}{\hbar c}\mathbf{A}(\mathbf{x},t)\right)\overline{\Psi(\mathbf{x},t)}\Psi(\mathbf{x},t)\right) \tag{2.12}$$

and the stress tensor

$$\textcircled{\scriptsize H}=\left(\frac{\hbar}{2m}\right)^2\left[\left(\nabla+\frac{qi}{\hbar c}\mathbf{A}\right)\overline{\Psi}\left(\nabla-\frac{qi}{\hbar c}\mathbf{A}\right)\Psi\right.$$
$$+\left(\nabla-\frac{qi}{\hbar c}\mathbf{A}\right)\Psi\left(\nabla+\frac{qi}{\hbar c}\mathbf{A}\right)\overline{\Psi}$$
$$-\overline{\Psi}\left(\nabla-\frac{qi}{\hbar c}\mathbf{A}\right)\left(\nabla-\frac{qi}{\hbar c}\mathbf{A}\right)\Psi$$
$$\left.-\left(\nabla+\frac{qi}{\hbar c}\mathbf{A}\right)\left(\nabla+\frac{qi}{\hbar c}\mathbf{A}\right)\overline{\Psi}\Psi\right]. \tag{2.13}$$

The equation of continuity for ρ and $\mathbf{j}$ follows from the Schrödinger equation. It can be regarded as an equation of motion for ρ,

$$\frac{\partial\rho}{\partial t}=-\nabla\cdot\mathbf{j}. \tag{2.14}$$

An equation of motion for $\mathbf{j}$ similarly follows from the Schrödinger equation. It is the differential form of Newton's law of motion

$$\frac{\partial\mathbf{j}}{\partial t}=\frac{1}{m}\left[\rho\mathcal{E}+\mathbf{j}\times\mathcal{B}\right]-q\nabla\cdot\textcircled{\scriptsize H}. \tag{2.15}$$

If $\textcircled{\scriptsize H}$ is written as a function of ρ and $\mathbf{j}$,

$$q\textcircled{\scriptsize H}=\frac{1}{\rho}\left[\mathbf{j}\mathbf{j}-\left(\frac{\hbar}{2m}\right)^2\nabla\rho\nabla\rho\right]+\left(\frac{\hbar}{2m}\right)^2\nabla\nabla\rho, \tag{2.16}$$

the pair of equations (2.14) and (2.15) provide the required manifestly gauge-invariant local substitute for the Schrödinger equation. How then can one understand the Aharanov—Bohm paradoxes? What Aharanov and Bohm described are situations in which an interference effect [one giving rise to a change in $\rho(\mathbf{x},t)$] is produced by the introduction of potentials which are essentially constant along the path where the charged particle moves (and therefore produce no electromagnetic field there). They are nonconstant and therefore do produce a field elsewhere, where the charged particle does not go. According to Aharanov and Bohm, this shows that in quantum theory the potentials can produce physical effects not expressible in terms of local action by the fields. Looked at from the point of view of (2.14) and (2.15) this behavior is very strange because they are perfectly local equations in which the potentials do not occur. The explanation is that the description of Aharanov and Bohm is over-idealized at a decisive point. (There are many "explanations,"[23] including some rather close to ours[24]; we give one sufficient for our present purposes.) The solution of the Schrödinger equation always has a tail which runs into the region of nonvanishing field and that field, by purely local manifestly gauge-invariant action, produces the effect. It will not do to argue from finite propagation speed that the effect will not be felt elsewhere soon enough; in Schrödinger theory effects can be propagated instantaneously. If one keeps the wavefunction out of the region of nonvanishing field by fiat (or by introducing an infinitely large repulsive potential), making ρ and the normal component of $\mathbf{j}$ vanish on the boundary, there is still a vestige of the old physical effect in the necessity of specifying the tangential component of $\mathbf{j}$. For the case of an electron scattered by a magnetized wisker, for example, one has

$$\int_L\frac{\mathbf{j}}{\rho}\cdot d\mathbf{r}=\int_L\mathbf{A}(\mathbf{x})\cdot d\mathbf{r}=\text{magnetic flux}$$

where L is a loop around the whisker. Thus, the arbitrariness in the tangential component of $\mathbf{j}$ provides the freedom to produce (or not produce, if the flux is zero) the ESAB Effects.

The resolution of the apparent paradox for the many particle case ($n>1$) goes along similar lines. Formally, it is efficient to follow Dashen and Sharp,[25] by introducing the second quantization of ρ and $\mathbf{j}$. We will not pursue the matter but rather turn to the formulation of gauge invariance in the quantum theory of fields, our real objective.

In quantum field theory proper, one treats both the electromagnetic field and the matter fields quantum mechanically, and one has to consider anew the formulation of gauge invariance. As we have already recounted in Sec. 1, to describe a vector potential for a free electromagnetic field in a local or relativistically invariant way one needs a Gupta—Bleuler formalism. Gauge invariance in this context has some new features.

The Gupta-Bleuler gauges for the free electromagnetic field

There is a completely consistent formulation of the quantum electrodynamics of the free electromagnetic field in terms of the electromagnetic field operators $F_{\mu\nu}$ which does not use a vector potential at all. (In fact, this was the first manifestly covariant construction of a quantized electromagnetic field by Jordan and Pauli.[26]) However, in order to be able to formulate gauge invariance in the conventional way one has to introduce a field operator $\mathbf{A}_\mu$ for the vector potential and to give meaning to the transformation (2.7). Here a question arises. Is the gauge field χ to be interpreted as a quantized field or as a real-valued function multiplying the identity operator? The former interpretation is necessary if the vector potential, both before and after gauge transformation, is required to transform under a representation of the Poincaré group as

$$U(a,\Lambda)A_\mu(x)U(a,\Lambda)^{-1}=\Lambda^\nu{}_\mu A_\nu(\Lambda x+a) \tag{2.17}$$

and only it will be considered for the moment. We will make some remarks on the alternative later, when we treat the case of the coupled fields.

In order to display the operator explicitly let us recall the construction of the Gupta—Bleuler formalism for a free electromagnetic field in a little more detail. The first step is the definition of the Hilbert space $\mathcal{H}^{(1)}$ whose vectors are all 4-component functions $\{\Phi_\mu; \mu = 0, 1, 2, 3\}$ defined on the mantle C_+ of the future light cone and square integrable with respect to the invariant measure $d\Omega_0(k) = d^3k/k^0$, $k^0 = [\mathbf{k}^2]^{1/2}$. The scalar product in $\mathcal{H}^{(1)}$ is

$$(\Phi, \Psi)^{(1)} = \int_{C_+} d\Omega_0(k) \sum_{\mu=0}^{3} \overline{\Phi_\mu(k)} \Psi_\mu(k). \tag{2.18}$$

An indefinite sesquilinear (i. e., antilinear in its first argument, linear in its second) form is defined by

$$\langle \Phi, \Psi \rangle^{(1)} = \int_{C_+} d\Omega_0(k) \overline{\Phi_\mu(k)} \{-g^{\mu\nu}\} \Psi_\nu(k). \tag{2.19}$$

This form is nonnegative for Φ's that satisfy the auxiliary condition

$$k^\mu \Phi_\mu(k) = 0 \tag{2.20}$$

for almost all k. Such Φ constitute a closed subspace $\mathcal{H}^{(1)\prime}$ of $\mathcal{H}^{(1)}$. $\mathcal{H}^{(1)\prime}$ in turn has a closed subspace $\mathcal{H}^{(1)\prime\prime}$ consisting of all vectors of $\mathcal{H}^{(1)\prime}$ that have zero length

$$\langle \Phi, \Phi \rangle^{(1)} = 0. \tag{2.21}$$

They are of the form

$$\Phi^\mu(k) = g(k) k^\mu. \tag{2.22}$$

The vectors of the quotient space $\mathcal{H}^{(1)}_{\text{phys}} = \mathcal{H}^{(1)\prime}/\mathcal{H}^{(1)\prime\prime}$ describe one-photon states.

The full Hilbert space of the theory is the direct sum

$$\mathcal{H} = \bigoplus_{n=0}^{\infty} \mathcal{H}^{(n)} \tag{2.23}$$

where $\mathcal{H}^{(0)}$ is a one-dimensional Hilbert space and $\mathcal{H}^{(n)}$ is the symmetric tensor product

$$\mathcal{H}^{(n)} = (\mathcal{H}^{(1)\otimes n})_s \tag{2.24}$$

of $\mathcal{H}^{(1)}$ with itself n times. The elements of $\mathcal{H}$ are therefore given by sequences

$$\Phi = \{\Phi^{(0)}, \Phi^{(1)}, \Phi^{(2)}, \cdots\} \tag{2.25}$$

where $\Phi^{(0)}$ is a complex number and

$$\Phi^{(n)}_{\mu_1\mu_2\cdots\mu_n}(k_1 \cdots k_n) \tag{2.26}$$

is defined for $k_1 \cdots k_n \in C_+$ and is symmetric under simultaneous permutations of $\mu_1 \cdots \mu_n$ and $k_1 \cdots k_n$.

$\mathcal{H}$ is a Hilbert space when equipped with the scalar product $(\cdot, \cdot)$ induced by the scalar product $(\cdot, \cdot)^{(1)}$ on $\mathcal{H}^{(1)}$. Explicitly,

$$(\Phi, \Psi) = \overline{\Phi^{(0)}} \Psi^{(0)} + \sum_{n=1}^{\infty} \int \cdots \int \left(\prod_{j=1}^{n} d\Omega_0(k_j) \right) \times \sum_{\mu_1\cdots\mu_n} \overline{\Phi_{\mu_1\cdots\mu_n}(k_1 \cdots k_n)} \Psi_{\mu_1\cdots\mu_n}(k_1 \cdots k_n). \tag{2.27}$$

Similarly, there is an indefinite sesquilinear form on $\mathcal{H}$ induced by $\langle \cdot, \cdot \rangle^{(1)}$ on $\mathcal{H}^{(1)}$. Explicitly,

$$\langle \Phi, \Psi \rangle = \overline{\Phi^{(0)}} \Psi^{(0)} + \sum_{n=1}^{\infty} \int \cdots \int \left(\prod_{j=1}^{n} d\Omega_0(k_j) \right) \times (-1)^n \overline{\Phi^{\mu_1\cdots\mu_n}(k_1 \cdots k_n)} \Psi_{\mu_1\cdots\mu_n}(k_1 \cdots k_n). \tag{2.28}$$

The form $\langle \cdot, \cdot \rangle$ is nonnegative on the subspace $\mathcal{H}'$, defined by

$$k_1^{\mu_1} \Phi^{(n)}_{\mu_1\cdots\mu_n}(k_1 \cdots k_n) = 0 \tag{2.29}$$

for all $n \geq 1$ and almost all $k_1 \cdots k_n$. The Φ of $\mathcal{H}'$ have amplitudes $\Phi^{(n)}_{\mu_1\cdots\mu_n}(k_1 \cdots k_n)$ whose $\mu_1 \cdots \mu_n$ dependence is that of a product of polarization vectors, i. e., the vectors are orthogonal to k (k itself is among them). $\mathcal{H}''$ is the subspace of $\mathcal{H}'$ consisting of vectors of zero length:

$$\langle \Phi, \Phi \rangle = 0. \tag{2.30}$$

When $\Phi \in \mathcal{H}''$, $\Phi^{(n)}$ is proportional to k_μ in its dependence at least one of its indices μ.

The vector potential operator $A(M,f)$, smeared with a 4-component test function f, is expressed symbolically

$$A(M;f) = \int d^4x f^\mu(x) A_\mu(M;x) \tag{2.31}$$

and defined by

$$A(M;f) = a(\Pi_+(M)f) + a^*(\Pi_-(M)f). \tag{2.32}$$

Here a and a^* are annihilation and creation operators defined for $g \in \mathcal{H}^{(1)}$ by

$$(a(g)\Phi)^{(n)}_{\mu_1\cdots\mu_n}(k_1 \cdots k_n) = -\sqrt{n+1} \int d\Omega_0(k) \times g^\mu(k) \Phi^{(n+1)}_{\mu\mu_1\cdots\mu_n}(kk_1 \cdots k_n), \tag{2.33}$$

$$(a^*(g)\Phi)^{(n)}_{\mu_1\cdots\mu_n}(k_1 \cdots k_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [g_{\mu_j}(k_j) \times \Phi^{(n-1)}_{\mu_1\cdots\hat{\mu}_j\cdots\mu_n}(k_1 \cdots k_n)], \tag{2.34}$$

while the mapping $\Pi_\pm(M)$ are defined for $k \in C_+$ by

$$(\Pi_\pm(M)f)_\mu(k) = \sqrt{\pi} [\hat{f}_\mu(\pm k) - Mk_\mu k^\nu \hat{f}_\nu(\pm k)]. \tag{2.35}$$

M is an arbitrary real number whose presence reflects the arbitrariness in the choice of gauge as will be seen shortly.

An elementary calculation shows that

$$[A(M;f), A(M;g)]_- = -\pi \int d\Omega_0(k) \hat{f}_\mu(k) [g^{\mu\nu} - 2Mk^\mu k^\nu] \hat{g}_\nu(-k), \tag{2.36}$$

i. e., expressed symbolically

$$[A_\mu(M;x), A_\nu(M;y)] = -\{g_{\mu\nu} + 2M \partial_\mu \partial_\nu\} i^{-1} D(x-y), \tag{2.37}$$

where

$$D(x) = D^{(+)}(x) + D^{(-)}(x), D^{(-)}(x) = -D^{(+)}(-x) \tag{2.38}$$

and

$$D^{(+)}(x) = \frac{i}{2(2\pi)^3} \int_{C_+} d\Omega_0(k) \exp(-ik \cdot x). \tag{2.39}$$

The $A_\mu(M,x)$ for different M are related to one another by the gauge transformation

$$A_\mu(M;x) = A_\mu(0,x) - M \partial_\mu \chi(x), \tag{2.40}$$

where $\chi(x)$ is defined by

$$(\chi(f)\Phi)^{(n)}_{\mu_1\cdots\mu_n}(k_1 \cdots k_n)$$

$$= i\sqrt{\pi}\left(\sqrt{n+1}\int d\Omega_0(k)\hat{f}(k)k^\mu \Phi^{(n+1)}_{\mu\mu_1\cdots\mu_n}(kk_1\cdots k_n) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{f}(-k_j)k_{j\mu_j}\Phi^{(n-1)}_{\mu_1\cdots\widehat{\mu_j}\cdots\mu_n}(k_1\cdots\widehat{k}_j\cdots k_n)\right). \tag{2.41}$$

The field χ satisfies

$$\Box\chi(x) = 0. \tag{2.42}$$

Consequently, $\partial^\mu A_\mu(M,x)$ is independent of M. The Maxwell equation (1.8) is valid for every M with a right-hand side independent of M, and auxiliary condition

$$\partial_\mu A^{\mu(-)}(f)\Phi = 0, \tag{2.43}$$

that defines $\mathcal{H}'$ is independent of M, although the negative frequency part of the vector potential, $A^{(-)}(M,f) = a(\Pi_+(M)f)$, itself is dependent on M.

The gauges given explicitly in (2.40) are the basic covariant gauges of the Gupta–Bleuler theory of a free electromagnetic field. They are parametrized by the single real parameter M. The original papers of Gupta and Bleuler[27,28] give this construction for $M = 0$. The general covariant gauge apparently first appeared in the quantum electrodynamics of an interacting electromagnetic field where M is a function of the charge e that goes to zero with e.[29] It is (and was then) evident that M may also be nonvanishing when $e = 0$, as it is in (2.40). We will denote all these gauges Gupta–Bleuler gauges. On the other hand, we will use the phrase *indefinite metric formalism* to describe the structure consisting of the three spaces $\mathcal{H}$, $\mathcal{H}'$, $\mathcal{H}''$, the sesquilinear form, $\langle\cdot,\cdot\rangle$, and the associated field operators. Gupta and Bleuler were the first to use an indefinite metric formalism to obtain a Gupta–Bleuler gauge, but the formalism is far more general.

Gauges and gauge transformations for the free electromagnetic field

The Gupta–Bleuler gauges are far from exhausting the gauges that have been found useful in field theory. For example, one has the Coulomb or radiation gauge, the Landau gauge, and many more, some of which are explicitly discussed below. It is not difficult to see that, in general, the vector potentials of these other gauges are not connected with $A_\mu(M,x)$ nor each other by transformations of the form (2.7), $\tilde{A}_\mu(x) = A_\mu(x) - \partial_\mu\chi(x)$. [for example, in the Coulomb gauge $\partial^\mu F_{\mu\nu} = 0$ is an operator identity on the Hilbert space of physical states whereas in the Gupta–Bleuler theory the modified Maxwell equation (1.8) holds on $\mathcal{H}$. If the Coulomb field vector potential were connected with the Gupta–Bleuler vector potential by (2.7), they would yield the same $F_{\mu\nu}$.]

When the circumstance that not every gauge transformation is of the form (2.7) is combined with the theorems of the Introduction, one is forced to recognize that the formulation of gauge invariance is really quite different in classical theory and quantum electrodynamics. Our next task is to give a formulation of the notion of gauge and of gauge transformation for the free electromagnetic field that is sufficiently general that all the standard gauges are related to one another by the proposed notion of gauge transformation.

We begin with the definition of a notion of gauge.

Definition 2.1: A quantization of the free Maxwell equations by means of a vector potential A_μ, or, briefly, a *gauge* is specified by

(a) An operator valued distribution A_μ in a Hilbert space $\mathcal{H}$.

(b) A representation U of the Poincaré group in $\mathcal{H}$.

(c) A sesquilinear form $\langle\cdot,\cdot\rangle$ on $\mathcal{H}$ with respect to which the representation U is unitary.

(d) A distinguished subspace $\mathcal{H}' \subset \mathcal{H}$ such that

(i) the restriction of the sesquilinear form $\langle\cdot,\cdot\rangle$ to $\mathcal{H}'$ is bounded and nonnegative,

$$\langle\Psi,\Psi\rangle \geq 0 \quad \text{for } \Psi\in\mathcal{H}'; \tag{2.44}$$

(ii) the operators $F_{\mu\nu}(f) = \int F_{\mu\nu}(x)f(x)\,d^4x$, if $f\in\mathcal{S}$, defined in terms of A_μ by (1.1), are local, leave $\mathcal{H}'$ invariant,

$$F_{\mu\nu}(f)\mathcal{H}' \subset \mathcal{H}', \tag{2.45}$$

and satisfy

$$\langle\Phi, \partial_\mu F^{\mu\nu}(f)\Psi\rangle = 0 \tag{2.46}$$

for all $\Phi,\Psi\in\mathcal{H}'$ with Ψ in the domain of $\partial^\lambda F^{\mu\nu}(f)$;

(iii) the representation U leaves $\mathcal{H}'$ invariant, and the subspace $\mathcal{H}''$ of $\mathcal{H}'$ also; $\mathcal{H}''$ consists of the vectors in $\mathcal{H}'$ of zero length $\langle\Phi,\Phi\rangle = 0$; as usual $\mathcal{H}_{\text{phys}} = \mathcal{H}'/\mathcal{H}''$. [It is also assumed that there exists a vector Ψ_0, the vacuum vector, which is invariant under the representation U, and lies in $\mathcal{H}'$. Ψ_0 is a cyclic vector for the vector potential operators and is the unique vector of $\mathcal{H}$ invariant under the translations $U(a,1)$.]

(iv) for all $\Phi\in\mathcal{H}'$, the Fourier transform of $\langle\Phi, F_{\mu\nu}(x)\Psi_0\rangle$ has support contained in the closure of the future light cone $\bar{V}_+$.

We anticipate a remark that will be justified later in Lemma 2.2. Call two vectors Φ_1 and $\Phi_2\in\mathcal{H}'$ equivalent when $\Phi_1 - \Phi_2 \in\mathcal{H}''$. Denote the equivalence class of Φ_j by $[\Phi_j]$. Then as a consequence of $\mathcal{P}(F_{\mu\nu}(f))\mathcal{H}'\subset\mathcal{H}'$, the matrix elements of polynomials $\mathcal{P}$ in the smeared electromagnetic fields

$$\langle\Phi, \mathcal{P}(F_{\mu\nu}(f))\Psi\rangle$$

are constant as Φ varies over $[\Phi]$ and Ψ over $[\Psi]$.

For brevity a gauge will be denote $\{A_\mu, \mathcal{H}, \langle\cdot,\cdot\rangle, \mathcal{H}'\}$ with the understanding that $\mathcal{H}$ has a representation of the Poincaré group associated with it.

In all gauges known to us, the spectral condition holds in the following form stronger than (iv):

(iv′) For all $\Phi\in\mathcal{H}'$, the Fourier transform of $\langle\Phi, A_\mu(x)A_\nu(y)\Psi_0\rangle$ has support contained in $\bar{V}_+$ in the momenta conjugate to the variables $(x+y)/2$ and $x-y$.

While this definition has been tailored to fit the covariant Gupta–Bleuler gauges, it is general enough to include the Coulomb gauge. In that case, $\mathcal{H} = \mathcal{H}' = \mathcal{H}_{\text{phys}}$, $\mathcal{H}'' = \{0\}$, $\langle\cdot,\cdot\rangle = (\cdot,\cdot)$, and the Maxwell equations hold as operator equations. On the other hand, it is not obvious that such a gauge as the Landau gauge actually conforms to the definition. That will be shown below.

With the definition of gauge in hand, one comes to a natural notion of gauge transformation.

Definition 2.2: A *generalized gauge tranformation* is an ordered pair consisting of two gauges

$$\{A_{1\mu},\mathcal{H}_1,\langle\cdot,\cdot\rangle_1,\mathcal{H}_1'\} \quad\text{and}\quad \{A_{2\mu},\mathcal{H}_2,\langle\cdot,\cdot\rangle_2,\mathcal{H}_2'\}$$

together with a bijection g of $\mathcal{H}_{1\text{phys}}$ onto $\mathcal{H}_{2\text{phys}}$ such that

$$\text{(i)}\ \langle\Phi_1,\mathcal{P}(F_{1\mu\nu}(f))\Psi_1\rangle=\langle\Phi_2,\mathcal{P}(F_{2\mu\nu}(f))\Psi_2\rangle \tag{2.47}$$

for all $\Phi_1,\Psi_1\in\mathcal{H}_1'$ and $\Phi_2,\Psi_2\in\mathcal{H}_2'$ with

$$[\Psi_2]=g[\Psi_1],\quad [\Phi_2]=g[\Phi_1], \tag{2.48}$$

and Ψ_1,Ψ_2 in the domains of $\mathcal{P}(F_{1\mu\nu}(f))$ and $\mathcal{P}(F_{2\mu\nu}(f))$, respectively.

$$\text{(ii)}\ [\Psi_{20}]=g[\Psi_{10}]. \tag{2.49}$$

Note that the left- and right-hand sides of (2.47) are independent of which representative vectors Φ and Ψ are chosen from the equivalence classes $[\Phi]$ and $[\Psi]$. Note further that even if $\mathcal{H}_1=\mathcal{H}_2$ and $\mathcal{H}_1'=\mathcal{H}_2'$, we need not have

$$F_{1\mu\nu}(f)=F_{2\mu\nu}(f) \tag{2.50}$$

in $\mathcal{H}$ for the two gauges connected by a gauge transformation. (see example 4 below.)

Among the generalized gauge transformations that lead from a fixed gauge to other gauges, there is an important subclass defined as follows.

Definition 2.3: A *special gauge transformation* is a generalized gauge transformation for which the Hilbert space $\mathcal{H}$, its subspace $\mathcal{H}'$, the sesquilinear form $\langle\cdot,\cdot\rangle$, and the representation U do not change and the bijection g of $\mathcal{H}_{1\text{phys}}$ into $\mathcal{H}_{\text{phys}}$ is the identity.

For special gauge transformations, it is not difficult to see that the validity of Eq. (2.50) implies that the mapping $A_{1\mu}\to A_{2\mu}$ may be written in the form

$$A_{2\mu}(x)=A_{1\mu}(x)-\partial_\mu\chi(x).$$

Not all the special gauge transformations have this property. For a subclass of special gauge transformations we can establish (2.50).

Proposition 2.1: Any special gauge transformation leading from a local and covariant gauge to a local and covariant gauge

$$\{A_{1\mu},\mathcal{H},\langle\ ,\ \rangle,\mathcal{H}'\}\to\{A_{2\mu},\mathcal{H},\langle\ ,\ \rangle,\mathcal{H}'\}$$

with the properties that

(i) the commutators

$$[A_{1\mu},F_{1\rho\sigma}],\quad [A_1,F_{2\rho\sigma}]$$

are multiples of the identity operator,

(ii) the spectral condition holds in the stronger form (iv′),

(iii) the vacuum Ψ_{01} is a cyclic vector for A_1 satisfies (2.50).

Proof: Consider the two-point functions

$$\langle\Psi_0,A_{1\mu}(x)F_{1\rho\sigma}(y)\Psi_0\rangle \tag{2.51}$$

and

$$\langle\Psi_0,A_{1\mu}(x)F_{2\rho\sigma}(y)\Psi_0\rangle. \tag{2.52}$$

The assumed covariance implies according to the Lemma 1 of Ref. 9, that these may be written as

$$(\partial_\rho g_{\mu\sigma}-\partial_\sigma g_{\mu\rho})F_1(x), \tag{2.53}$$

$$(\partial_\rho g_{\mu\sigma}-\partial_\sigma g_{\mu\rho})F_2(x), \tag{2.54}$$

respectively, where F_1 and F_2 are Lorentz invariant distributions.

Now the states Ψ_0 and $\Phi\equiv F_{1\mu\nu}(f)\Psi_0$ belong to $\mathcal{H}'$ and thus condition (2.47) gives

$$\langle\Phi,F_{1\rho\sigma}(y)\Psi_0\rangle=\langle\Phi,F_{2\rho\sigma}(y)\Psi_0\rangle. \tag{2.55}$$

From the equality of (2.51) and (2.53), and of (2.52) and (2.54), together with this equality (2.55), we get that all second derivatives of F_1 are equal to those of F_2 and therefore, up to a linear function, F_1 and F_2 are equal. The linear function is constant since it must be Lorentz invariant and the constant is insignificant in the evaluation of (2.53) and (2.54), so we may assume $F_1=F_2$, i.e.;

$$\langle\Psi_0,A_{1\mu}(F_{1\rho\sigma}-F_{2\rho\sigma})\Psi_0\rangle=0. \tag{2.56}$$

Under the assumption that $[A_{1\mu},F_{1\rho\sigma}]$ and $[A_{1\mu},F_{2\rho\sigma}]$ have commutators that are multiples of the identity, one can express

$$\langle\Psi_0,A_{1\mu}\cdots A_{1\nu}(F_{1\rho\sigma}-F_{2\rho\sigma})\Psi_0\rangle \tag{2.57}$$

in terms of a sum of products that always include a factor (2.56). Thus, (2.57) is zero. Since the vacuum is a cyclic vector for $A_{1\mu}$,

$$(F_{1\rho\sigma}-F_{2\rho\sigma})\Psi_0=0.$$

Thus, by locality and a Reeh—Schlieder theorem one concludes that $F_{1\rho\sigma}$ and $F_{2\rho\sigma}$, coincide. (For further details of this argument see p. 165 of Streater—Wightman.[14])

There is a practical sufficient criterion for a special gauge transformation that starts from the Gupta—Bleuler gauge.

Proposition 2.2: If $A_{1\mu}(x)=A_\mu(0;x)$ and $A_{2\mu}=A_\mu$ has the two-point function

$$\langle\Psi_0,A_\mu(x)A_\nu(y)\Psi_0\rangle$$
$$=-[8\pi^3]^{-1}\int d^4k\,\delta(k^2)\theta(k)\exp[-ik\cdot(x-y)]$$
$$\times\{g_{\mu\nu}+f_{\mu\nu}(k)+f_{\nu\mu}(k)+f_\mu{}^\rho(k)f_{\nu\rho}(k)\}, \tag{2.58}$$

where $f_{\mu\nu}$ satisfies

$$f_{\mu\nu}\mathcal{S}(\mathbb{R}^4)\subset L^2(d^4k\,\delta(k^2)\theta(k)) \tag{2.59}$$

and

$$f_{\nu\mu}(k)k^\nu=k_\mu f(k) \tag{2.60}$$

on C_+, then $A_{2\mu}$ is unitary equivalent to a field connected with $A_{1\mu}$ by special gauge transformation. In particular, if $f_{\mu\nu}(k)$ is of the form $f_{\mu\nu}(k)=F_\nu(k)k_\mu$ and

$$F_\mu\mathcal{S}(\mathbb{R}^4)\subset L^2(d^4k\,\delta(k^2)\theta(k)), \tag{2.61}$$

then the special gauge transformation may be written in the form (2.7).

Proof: Consider the field $B(f)$ defined by

$$(B(f)\Psi)^{(n)}_{\mu_1\cdots\mu_n}(k_1\cdots k_n)=\sqrt{\pi}\left\{\sqrt{n+1}\int_{C_+}d\Omega_0(k)\times[-\hat{f}^\rho(k)f_\rho{}^\nu(k)]\Psi^{(n+1)}_{\nu\mu_1\cdots\mu_n}(kk_1\cdots k_n)+\frac{1}{\sqrt{n}}\sum_{j=1}^n\hat{f}^\nu(-k_j)f_{\nu\mu_j}(k)\times\Psi^{(n-1)}_{\mu_1\cdots\hat{\mu}_j\cdots\mu_n}(k_1\cdots\hat{k}_j\cdots k_n)\right\}. \quad (2.62)$$

If it is added to $A(0;f)$, we get a field which has the two-point function coinciding with (2.58) because

$$\langle\Psi_0,B(f)A(0;g)\Psi_0\rangle=-\pi\int_{C_+}d\Omega_0(k)\hat{f}^\rho(k)f_{\rho\nu}(k)\hat{g}^\nu(-k),$$
$$\langle\Psi_0,A(0;f)B(g)\Psi_0\rangle=-\pi\int_{C_+}d\Omega_0(k)\hat{f}^\rho(k)f_{\nu\rho}(k)\hat{g}^\nu(-k),$$
$$\langle\Psi_0,B(f)B(g)\Psi_0\rangle=-\pi\int_{C_+}d\Omega_0(k)\hat{f}^\rho(k)f_\rho{}^\sigma(k)f_{\nu\sigma}(k)\hat{g}^\nu(-k).$$

This operator is defined and bounded on $\oplus_{n=0}^N\mathcal{H}^{(n)}$ for all N and satisfies $B(f)^+=B(f)^-$ there. The condition (2.59) guarantees that the expressions $\hat{f}^\rho(k)f_\rho{}^\nu(k)$ and $\hat{f}^\rho(-k)f_\rho{}^\nu(k)$ which appear in (2.62) are, when restricted to C_+, the mantle of the future light cone, functions in $L^2(d\Omega_0(k))$.

If we define a closed subspace of $\mathcal{H}$ by the requirement

$$[\partial_\mu B^{\mu(-)}(f)+\partial_\mu A^{\mu(-)}(M,f)]\Psi=0,$$

it defines the same subspace $\mathcal{H}'$ as (2.29), by virtue of (2.60).

Notice that $A_\mu(0;x)$ and $A_\mu(0;x)+B_\mu(x)$ do not yield the same electromagnetic field operator in general. However, the field operators they yield have identical matrix elements in $\mathcal{H}'$, so the correspondence $A_\mu(0;x)\to A_\mu(0;x)+B_\mu(x)$ is a special gauge transformation.

We can now prove the equivalence of $A_\mu(x)$ and $A_\mu(0;x)+B_\mu(x)$. We define a linear operator V as mapping Ψ_{02} onto Ψ_{01} and $\mathcal{P}(A_2(f))\Psi_{02}$ onto $\mathcal{P}(A(0;f)+B(f))\Psi_{01}$. By the above arguments V is a scalar product preserving bijection. When $f_{\mu\nu}=F_\nu(k)k_\mu$, one defines an operator $\chi(x)$

$$(\chi(f)\Psi^{(n)}_{\mu_1\cdots\mu_n}(k_1\cdots k_n)=i\sqrt{\pi}\left\{\sqrt{n+1}\int_{C_+}d\Omega_0(k)\hat{f}(k)F^\nu(k)\Psi^{(n+1)}_{\nu\mu_1\cdots\mu_n}(kk_1\cdots k_n)+\frac{1}{\sqrt{n}}\sum_{j=0}^n F_{\mu_j}(k)\hat{f}(-k)\Psi^{(n-1)}_{\mu_1\cdots\hat{\mu}_j\cdots\mu_n}(k_1\cdots\hat{k}_j\cdots k_n)\right\}$$

and one easily verifies that the field $A_\mu(0;x)-\partial_\mu\chi(x)$ has the same two-point function as the field $A_{2\mu}(x)$. The argument then proceeds along the same pattern as before, for the field $B_\mu(x)$.

Remark: We conjecture that every special gauge transformation from the Gupta–Bleuler gauge is of the form given in this proposition.

Examples of gauges obtained by the special gauge transformation (2.62)

1. The Kallen-Rollnik-Stech-Nunnemann gauge $A_\mu(M,x)$[28–30]

Here

$$\langle\Psi_0,A_\mu(x)A_\nu(y)\Psi_0\rangle=-[g_{\mu\nu}+2M\partial_\mu\partial_\nu]\frac{1}{i}D^{(+)}(x-y). \quad (2.63)$$

As already remarked in (2.40), this gauge is obtained by setting

$$F_\mu(k)=Mk_\mu,\quad f_{\mu\nu}(k)=Mk_\mu k_\nu. \quad (2.64)$$

This is the only special gauge transformation which preserves locality and covariance.

2. The Evans-Fulton gauge[31]

Here

$$\langle\Psi_0,A_\mu(x)A_\nu(y)\Psi_0\rangle=-(g_{\mu\nu}+\partial_\mu\partial_\nu\partial^{-2}-n_\mu\partial_\nu\partial^{-1}-n_\nu\partial_\mu\partial^{-1})\frac{1}{i}D^{(+)}(x-y) \quad (2.65)$$

where n is a time like vector, $\partial=n_\mu\partial^\mu$, and ∂^{-1} is defined through its Fourier transform. This gauge is obtained by setting

$$F_\mu(k)=-n_\mu/n\cdot k. \quad (2.66)$$

Condition (2.61) is satisfied because n is timelike and k is lightlike, so $n\cdot k\neq 0$ on C_+ for $k\neq 0$ and $F_\mu(k)g(k)$ and $F_\mu(k)k_\nu g(k)\in L^2(d\Omega_0(k))$ for each $g(x)\in\mathcal{S}(\mathbb{R}^4)$.

3. The Valatin gauge[32]

Here

$$\langle\Psi_0,A_\mu(x)A_\nu(y)\Psi_0\rangle=-(g_{\mu\nu}+2\partial_\mu\partial_\nu\partial^{-2}-n_\mu\partial_\nu\partial^{-1}-n_\nu\partial_\mu\partial^{-1})\frac{1}{i}D^{(+)}(x-y). \quad (2.67)$$

This gauge is obtained by setting

$$F_\mu(k)=-\left(\frac{n_\mu}{n\cdot k}+\frac{k_\mu}{(n\cdot k)^2}\right). \quad (2.68)$$

4. Coulomb type gauge

We choose

$$f_{\mu\nu}(k)=-n_\mu n_\nu-\frac{(k_\mu-n_\mu n\cdot k)}{n\cdot k}\frac{(k_\nu-n_\nu n\cdot k)}{nk}, \quad (2.69)$$

where n_ν is a timelike 4-vector for which, for simplicity, we choose the form $n_\mu\equiv(1,0,0,0)$. Then by the construction described in the proof of Proposition 2.2, we obtain a field $A^c_\mu(x)$ such that

$$\mathrm{div}\mathbf{A}^c=0, \quad (2.70)$$
$$A^c_0=0, \quad (2.71)$$
$$\partial_\mu A^\mu_c=0, \quad (2.72)$$

as operator equations. Moreover, the two-point function for A^c_μ is

$$\langle\Psi_0, A^c_\mu(x)A^c_\nu(y)\Psi_0\rangle = \begin{cases} 0 & \text{if } \mu=0 \text{ or } \nu=0 \\ \left(\delta_{ij}-\dfrac{\partial_i\partial_j}{\Delta}\right)\dfrac{1}{i}D^{(+)}(x-y), & i,j=1,2,3. \end{cases} \tag{2.73}$$

It is not difficult to recognize that A^c_μ is essentially the Coulomb gauge potential defined in the Gupta–Bleuler gauge. The vectors of the set $\underline{L}$, obtained by applying polynomials in $A^c_\mu(f)$ to the vacuum, have components $\Psi^{(n)}=\{\Psi^{(n)}_{\mu_1\cdots\mu_n}(k_1\cdots k_n)\}$ such that

$$k_j^{\mu_j}\Psi^{(n)}_{\mu_1\cdots\mu_j\cdots\mu_n}(k_1\cdots k_n)=0, \tag{2.74}$$

$$k_j^i\Psi^{(n)}_{\mu_1\cdots i\cdots\mu_n}=0, \tag{2.75}$$

$$\Psi^{(n)}_{\mu_1\cdots\mu_n}=0 \quad \text{if any } \mu_j=0. \tag{2.76}$$

Thus $\underline{L}\subset\mathcal{H}'$, $A^c(f)\underline{\mathcal{H}}'\subset\mathcal{H}'$, $A^c(f)\mathcal{H}''=0$. The Hilbert space closure $L^c=\overline{\underline{L}}$ is a subspace of $\mathcal{H}'$ on which the sesquilinear form $\langle\cdot,\cdot\rangle$ is positive definite. In agreement with Theorems 1.1 and 1.2, A^c_μ is a nonlocal and noncovariant operator. Moreover, A^c_μ is not irreducible in $\mathcal{H}$. The restriction of the theory to L^c gives exactly the Coulomb gauge.

Finally, it is worth remarking that the gauge transformation induced by (2.69) cannot be written in the form (2.7) and that

$$\partial_\mu F_c^{\mu\nu}\neq\partial_\mu F^{\mu\nu}. \tag{2.77}$$

The Landau gauge for the free field

A crucial test of the adequacy of Definitions 2.1 and 2.2 of gauge and generalized gauge transformation is the description of a Laudau gauge. It and gauges related to it by special gauge transformations are very important in practical applications. We make the following definition[33]:

Definition 2.4: Let $G_{\mu\nu}$ be the two-point distribution of the vector potential

$$G_{\mu\nu}(x,y)=\langle\Psi_0, A_\mu(x)A_\nu(y)\Psi_0\rangle. \tag{2.78}$$

$G_{\mu\nu}$ satisfies the *spectral condition* if

$$\text{supp}[\hat{G}_{\mu\nu}(k,l)]\subset\overline{V}_+\times\overline{V}_-, \tag{2.79}$$

i.e., if the Fourier transform $\hat{G}_{\mu\nu}(k,l)$ vanishes for k outside the future cone and l outside the past cone.

$G_{\mu\nu}$ is called *covariant* if

$$\Lambda_\mu{}^\kappa\Lambda_\nu{}^\lambda G_{\kappa\lambda}(x,y)=G_{\mu\nu}(\Lambda x+a,\Lambda y+a) \tag{2.80}$$

for all Poincaré transformations $\{a,\Lambda\}$. It is called *transverse* if

$$\frac{\partial}{\partial x_\mu}G_{\mu\nu}(x,y)=0=\frac{\partial}{\partial y_\nu}G_{\mu\nu}(x,y). \tag{2.81}$$

A gauge for the free electromagnetic field is called a *Laudau gauge* if it is covariant in the sense that the vector potential satisfies

$$U(a,\Lambda)A_\mu(x)U(a,\Lambda)^{-1}=\Lambda^\nu{}_\mu A_\nu(\Lambda x+a)$$

and transverse in the sense that

$$\partial_\mu A_\mu(x)=0$$

and has a $G_{\mu\nu}$ satisfying the spectral condition.

In a Laudau gauge the $G_{\mu\nu}$ is automatically covariant so it depends only on the difference variable $G_{\mu\nu}(x,y)=G_{\mu\nu}(x-y)$, and the spectral condition reduces to $\text{supp}\hat{G}_{\mu\nu}(k)\subset\overline{V}_+$.

The form of the two-point distribution in a covariant gauge satisfying the spectral condition is fixed by the following lemma which we state as generally as we know how so that it is applicable to coupled as well as free fields. Notice that the hypotheses of the lemma do not include the free wave equation for A_μ nor the transverseness condition (2.81). Furthermore, the spectral condition is a little different from (2.79), and one can draw no conclusion about the support of the Fourier transform $\hat{G}$, of the invariant distribution G, occurring in (2.85). With the spectral condition (2.79), both $\hat{F}$ and $\hat{G}$ would have support in $\overline{V}_+$.

Proposition 2.3: Let $A_\mu(x)$ be defined as an operator-valued distribution in a Hilbert space $\mathcal{H}$, which is equipped with a sesqui-linear form $\langle\cdot,\cdot\rangle$.

Suppose

(i) In $\mathcal{H}$ there is a continuous representation of the Poincaré group $\{a,\Lambda\}\to U(a,\Lambda)$ such that

$$U(a,\Lambda)A_\mu(x)U(a,\Lambda)^{-1}=\Lambda^\nu{}_\mu A_\nu(\Lambda x+a) \tag{2.82}$$

and $U(a,\Lambda)$ is unitary with respect to $\langle\cdot,\cdot\rangle$.

(ii) There exists a unique vector Ψ_0 in $\mathcal{H}$ invariant under $U(a,\Lambda)$.

(iii) If $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$ then for all $\Phi\in\mathcal{H}$ the Fourier transform of $\langle\Phi, F_{\mu\nu}(x)\Psi_0\rangle$ has support contained in $\overline{V}_+$.

(iv) PCT symmetry holds, i.e., there exists an antilinear operator Θ, antiunitary with respect to $\langle\cdot,\cdot\rangle$, such that

$$\Theta\Psi_0=\Psi_0 \tag{2.83}$$

and

$$\Theta A_\mu(x)\Theta^{-1}=-A_\mu(-x). \tag{2.84}$$

Then

$$\langle\Psi_0, A_\mu(x)A_\nu(y)\Psi_0\rangle=g_{\mu\nu}F(y-x)+\partial_\mu\partial_\nu G(y-x) \tag{2.85}$$

where F and G are Lorentz invariant distributions. The Fourier transform of F has support in $\overline{V}_+$.

Proof: The analysis runs parallel in part to that carried out in Refs. 9 and 11. The first step considers the two-point functions

$$G_{\kappa\lambda\mu}(y-x)=\langle\Psi_0, A_\kappa(x)F_{\lambda\mu}(y)\Psi_0\rangle.$$

By virtue of the spectral condition (iii), the $G_{\kappa\lambda\mu}(x)$ can be written as boundary values of functions holomorphic in the variable $z=x+i\eta$ in the tube $\mathcal{T}_+$ which consists of all z such that $\eta\in V_+$. The holomorphic functions will be denoted $G_{\kappa\lambda\mu}(z)$. The same theorem that asserts the existence of $G_{\kappa\lambda\mu}(z)$ says that as z approaches the boundary of $\mathcal{T}_+$, the absolute values $|G_{\kappa\lambda\mu}(z)|$ grow at worst as a power of the Euclidean distance to the boundary.[34]

The relativistic transformation properties (i) and (ii) imply that the $G_{\kappa\lambda\mu}$ satisfy

$$\Lambda_\nu{}^\kappa \Lambda_\rho{}^\lambda \Lambda_\sigma{}^\mu G_{\kappa\lambda\mu}(x) = G_{\nu\rho\sigma}(\Lambda x),$$

an equation that also holds for the holomorphic functions throughout $\mathcal{T}_+$. There is a standard theorem on covariant families of holomorphic functions,[35] which assures us that the $G_{\kappa\lambda\mu}(z)$ can be expanded in the standard covariants

$$z_\kappa z_\lambda z_\mu, g_{\kappa\lambda} z_\mu, g_{\kappa\mu} z_\lambda, g_{\lambda\mu} z_\kappa, \epsilon_{\kappa\lambda\mu\rho} z^\rho \tag{2.86}$$

with uniquely determined coefficients that are holomorphic functions of z^2 for $z \in \mathcal{T}_+$, bounded by a power of $|z^2|$ as z approaches the boundary of $\mathcal{T}_+$. Because of the antisymmetry of $G_{\kappa\lambda\mu}$ in the indices $\lambda\mu$ only the coefficients of antisymmetric combinations of the (2.86) are nonzero and, therefore,

$$G_{\kappa\lambda\mu}(z) = (g_{\kappa\mu} z_\lambda - g_{\kappa\lambda} z_\mu) G(z^2) + \epsilon_{\kappa\lambda\mu\rho} z^\rho G_1(z^2). \tag{2.87}$$

By using a standard trick, one can rewrite this with the differentiation $\partial/\partial z^\sigma$ replacing z_σ:

$$G_{\kappa\lambda\mu}(z) = \left(g_{\kappa\mu} \frac{\partial}{\partial z^\lambda} - g_{\kappa\lambda} \frac{\partial}{\partial z^\mu} \right) F(z^2) + \epsilon_{\kappa\lambda\mu\rho} \frac{\partial}{\partial z_\rho} F_1(z^2) \tag{2.88}$$

where $F(z^2)$ is defined by

$$F(z^2) = \tfrac{1}{2} \int_{z_0^2}^{z^2} G(\zeta)\, d\zeta \tag{2.89}$$

and analogously for F_1. The path of integration goes from some arbitrary fixed point z_0^2 to z^2 without touching the positive real axis; such a path can be constructed using vectors in $\mathcal{T}_+$. Passing to the boundary by the limit process $\lim_{\lambda\to 0_+} G_{\kappa\lambda\mu}(x + i\lambda\eta)$ with $\eta \in V_+$, one obtains from representation (2.88) of $G_{\kappa\lambda\mu}(z)$ a representation

$$G_{\kappa\lambda\mu}(x) = (g_{\kappa\mu}\partial_\lambda - g_{\kappa\lambda}\partial_\mu) F(x) + \epsilon_{\kappa\lambda\mu}{}^\rho \partial_\rho F_1(x) \tag{2.90}$$

where F and F_1 are Lorentz invariant tempered distributions.

Now $F_{\kappa\lambda} = \partial_\kappa A_\lambda - \partial_\lambda A_\kappa$, so $\partial^\kappa \epsilon_\kappa{}^{\lambda\mu\nu} F_{\mu\nu} = 0$ is an operator identity. It implies

$$\partial^\kappa \epsilon_\kappa{}^{\lambda\mu\nu} G_{\rho\mu\nu}(x) = 0. \tag{2.91}$$

When the right-hand side of (2.90) is inserted into (2.91) only the term in F_1 survives. It yields

$$(\Box g_{\lambda\mu} - \partial_\lambda\partial_\mu) F_1(x) = 0. \tag{2.92}$$

Thus F_1 is a constant and does not contribute to (2.90). [To see this take the trace of (2.92) to get $\Box F_1 = 0$, and, therefore, $\partial_\lambda\partial_\mu F_1 = 0$. Then note that $\partial_\lambda\partial_\mu F_1 = 0$ implies that F_1 is a constant plus a linear function of x. However, the linear function must vanish since F_1 is Lorentz invariant.] Thus,

$$G_{\kappa\lambda\mu}(x) = (g_{\kappa\mu}\partial_\lambda - g_{\kappa\lambda}\partial_\mu) F(x). \tag{2.93}$$

Next consider the two-point distribution of the vector potential, writing

$$\langle \Psi_0, A_\mu(x) A_\nu(y) \Psi_0 \rangle = g_{\mu\nu} F(y-x) + H_{\mu\nu}(y-x). \tag{2.94}$$

By taking the curl of (2.94) in its second argument, we get an alternative expression for $G_{\kappa\lambda\mu}(x)$

$$G_{\kappa\lambda\mu}(x) = (g_{\kappa\mu}\partial_\lambda - g_{\kappa\lambda}\partial_\mu) F(x) + \partial_\lambda H_{\kappa\mu}(x) - \partial_\mu H_{\kappa\lambda}(x). \tag{2.95}$$

Comparison with (2.93) yields

$$\partial_\lambda H_{\kappa\mu} - \partial_\mu H_{\kappa\lambda} = 0. \tag{2.96}$$

A standard theorem of distribution theory now applies; it says that H must be the gradient of a scalar in its second index (see Ref. 36, p. 59)

$$H_{\kappa\lambda}(x) = \partial_\lambda H_\kappa(x) \tag{2.97}$$

where H_κ, $\kappa = 0,1,2,3$ are some tempered distributions.

Finally, PCT symmetry implies

$$\langle \Psi_0, A_\mu(x) A_\nu(y) \Psi_0 \rangle = \langle \Psi_0, A_\nu(-y) A_\mu(-x) \Psi_0 \rangle, \tag{2.98}$$

so

$$\partial_\lambda H_\kappa(x) = \partial_\kappa H_\lambda(x) \tag{2.99}$$

and that implies by the same theorem of distribution theory

$$H_\kappa(x) = \partial_\kappa G(x). \tag{2.100}$$

This completes the proof of the formula (2.85). It remains to show that G may be chosen Lorentz invariant without affecting (2.85).

The Lorentz transformation properties of the vector potential assumed in (i) and those of the vacuum assumed in (ii) imply, when combined with the Lorentz invariance of F proved above

$$(\partial_\mu\partial_\nu G)(x) = \Lambda^\kappa{}_\mu \Lambda^\lambda{}_\nu (\partial_\kappa\partial_\lambda G)(\Lambda x),$$

i.e.,

$$\partial_\mu\partial_\nu [G(x) - G(\Lambda x)] = 0. \tag{2.101}$$

Thus, $G(x)$ differs from $G(\Lambda x)$ at worst by a constant plus a linear term in x

$$G(x) - G(\Lambda x) = c(\Lambda) + d(\Lambda)^\mu x_\mu. \tag{2.102}$$

Comparing this with

$$G(\Lambda x) - G(\Lambda_1 \Lambda x) = c(\Lambda_1) + d(\Lambda_1)^\mu (\Lambda x)_\mu \tag{2.103}$$

and

$$G(x) - G(\Lambda_1\Lambda x) = c(\Lambda_1\Lambda) + d(\Lambda_1\Lambda)^\mu x_\mu, \tag{2.104}$$

we see

$$c(\Lambda_1\Lambda) = c(\Lambda) + c(\Lambda_1), \tag{2.105}$$

$$d(\Lambda_1\Lambda) = \Lambda^{-1} d(\Lambda_1) + d(\Lambda). \tag{2.106}$$

$c(\Lambda)$ and $d(\Lambda)$ are infinitely differentiable functions of Λ. [To see this, one first smears (2.102) with a test function $\phi(x)$ such that $\int x^\mu \phi(x) = 0$, $\mu = 0,1,2,3$, and $\int \phi(x)\, dx \neq 0$. Since in the resulting identity

$$G(\phi) - G(\Lambda\phi) = c(\Lambda) \int \phi(x)\, d^4x$$

the left-hand side is infinitely differentiable in Λ, so is $c(\Lambda)$. A similar argument works for $d(\Lambda)$.]

Since the restricted Lorentz group has no nontrivial one-dimensional representations $c(\Lambda) = 0$ for $\Lambda \in L_+^\uparrow$. For I_s, because $I_s^2 = 1$, we have $0 = 2c(I_s)$; so $c(\Lambda) = 0$ for $\Lambda \in L^\uparrow$.

As far as $d(\Lambda)$ is concerned, there are nontrivial solutions; for example,

$$d(\Lambda) = n - \Lambda^{-1} n \tag{2.107}$$

where n is any real vector. However, a term of this form can be eliminated by subtracting $n^\mu x_\mu$ from $G(x)$, an operation that does not affect the validity of (2.85).

That the only nontrivial solutions of (2.106) are of the form (2.107) is well known.[37] Nevertheless, for the convenience of the reader we have given a direct proof in Appendix A. This completes the proof of Proposition 2.3.

If instead of the spectral condition (iii) of Proposition 2.3, we imposed the stronger spectral condition (2.79), we could conclude that the support of the Fourier transform $\hat{G}$ of G lies in the future light cone $\bar{V}_+$.

There is one evident arbitrariness in the functions F and G occurring in Proposition 2.3. They may be altered by the transformation

$$F(x)\to F(x)+c, \quad G(x)\to G(x)-\tfrac{1}{2}cx^2 \tag{2.108}$$

where c is any real constant, without affecting the validity of (2.85). This will be used in Proposition 2.4 to reduce F to standard form.

Now we turn to the consequences of the free Maxwell equations and the condition that A_μ be transverse.

Proposition 2.4: If a quantization of the free Maxwell equations according to Definition 2.1 is given, then the field operator $F_{\kappa\lambda}$ satisfies

$$\langle\Psi_0, \partial^\kappa F_{\kappa\lambda}(x)F_{\mu\nu}(y)\Psi_0\rangle=0 \tag{2.109}$$

which, if the gauge is covariant, implies that the F of (2.85) satisfies

$$F(-x)=-c_1 i^{-1}D^{(+)}(x)+c_3 \tag{2.110}$$

where c_1 and c_3 are some real constants and c_1 is positive. Without loss of generality c_3 may be taken zero. With the conventional normalization of the electromagnetic field $c_1=1$.

If, in addition, the gauge is transverse, then the other invariant function G in (2.85) satisfies

$$G(-x)=c_1 i^{-1}D^{(+)}(x). \tag{2.111}$$

The most general solution of this equation is

$$G(-x)=\frac{c_1}{(4\pi)^2}[\log|x^2|-i\pi\,\mathrm{sgn}x^0\theta(x^2)]+G_{\mathrm{Hom}} \tag{2.112}$$

where G_{Hom} is an arbitrary invariant solution of

$$\Box G_{\mathrm{Hom}}=0. \tag{2.113}$$

Consequently, in a Laudau gauge the free vector potential, which satisfies the spectral condition (2.79) by Definition 2.4, has a two-point function

$$\begin{aligned}\langle\Psi_0, A_\mu(x)A_\nu(y)\Psi_0\rangle^L &= -[g_{\mu\nu}-\partial_\mu\partial_\nu(\Box^{-1}-2M)]i^{-1}D^{(+)}(x-y)\\ &= -g_{\mu\nu}i^{-1}D^{(+)}(x-y)+\partial_\mu\partial_\nu G(y-x)\end{aligned} \tag{2.114}$$

with

$$G(-x)=\frac{1}{(4\pi)^2}[\log|x^2|-i\pi\,\mathrm{sgn}x^0\theta(x^2)]-2Mi^{-1}D^{(+)}(x)$$

where c_1 has been chosen equal to one.

Proof: By assumption (d.iii) of Definition 2.1, $\Psi_0\in\mathcal{H}'$; by assumption (d.ii), $F_{\mu\nu}(f)\Psi_0\in\mathcal{H}'$, and so from (2.44), (2.109) follows.

Now assuming the gauge is covariant, the hypotheses of Proposition 2.3 hold and so the standard form (2.85) for $\langle\Psi_0, A_\mu(x)A_\nu(y)\Psi_0\rangle$ may be inserted in (2.109). After some algebraic reduction the result is $(g_{\lambda\nu}\partial_\mu-g_{\lambda\mu}\partial_\nu)\Box F=0$ which implies

$$\Box F=c_2 \tag{2.115}$$

where c_2 is some real constant. The polynomial $\frac{1}{2}c_2x^2$ is a particular solution of this equation. Thus the general solution for $F(-x)$ is $\frac{1}{2}c_2x^2$ plus an invariant solution of the homogeneous equation whose Fourier transform has support in $\bar{V}_+$. Every solution of the homogeneous wave equation with these properties is of the form

$$-c_1 i^{-1}D^{(+)}(x)+c_3. \tag{2.116}$$

(For a proof, see for example, Ref. 38, p. 45.) Thus

$$F(-x)=-c_1 i^{-1}D^{(+)}(x)+\tfrac{1}{2}c_2x^2+c_3. \tag{2.117}$$

Since in a transverse gauge, (2.85) implies

$$\partial_\nu[F(x)+\Box G(x)]=0,$$

this form (2.117) of F implies

$$\begin{aligned}\Box G(-x)&=-F(-x)+c_4\\ &=c_1 i^{-1}D^{(+)}(x)-\tfrac{1}{2}c_2x^2-c_3+c_4.\end{aligned} \tag{2.118}$$

To see that c_2 must be zero and c_1 positive, note that from (2.93)

$$\begin{aligned}&\langle\Psi_0, F_{\kappa\lambda}(x)F_{\mu\nu}(y)\Psi_0\rangle\\ &=-(g_{\lambda\nu}\partial_\kappa\partial_\mu+g_{\kappa\mu}\partial_\lambda\partial_\nu-g_{\lambda\mu}\partial_\kappa\partial_\nu-g_{\kappa\nu}\partial_\lambda\partial_\mu)\frac{c_1}{i}D^{(+)}(x-y)\\ &\quad+(g_{\kappa\mu}g_{\lambda\nu}-g_{\kappa\nu}g_{\lambda\mu})c_2.\end{aligned}$$

When the left-hand side is regarded as a scalar product in $\mathcal{H}_{\mathrm{phys}}$, one can use the SNAG theorem for the space—time translation group to analyze the states of zero momentum.[39] If E_0 is the projection on the zero-momentum states, one has, if one uses the uniqueness of the vacuum,

$$\begin{aligned}\langle\Psi_0, F_{\kappa\lambda}(x)E_0F_{\mu\nu}(y)\Psi_0\rangle&=\langle\Psi_0, F_{\kappa\lambda}(x)\Psi_0\rangle\langle\Psi_0, F_{\mu\nu}(y)\Psi_0\rangle\\ &=(g_{\kappa\mu}g_{\lambda\nu}-g_{\kappa\nu}g_{\lambda\mu})c_2.\end{aligned}$$

But on grounds of Poincaré invariance alone $\langle\Psi_0, F_{\kappa\lambda}(x)\Psi_0\rangle=0$. Thus, $c_2=0$. At the same time, because $\langle F_{\kappa\lambda}(f)\Psi_0, F_{\kappa\lambda}(f)\Psi_0\rangle\geq 0$, the scalar product being positive on $\mathcal{H}_{\mathrm{phys}}$, we have $c_1\geq 0$.

If we accept for the moment that

$$H(x)=\frac{c_1}{(4\pi)^2}[\log|x^2|-i\pi\,\mathrm{sgn}(x^0)\theta(x^2)]+G_{\mathrm{Hom}} \tag{2.119}$$

is the general solution of

$$\Box H(x)=c_1 i^{-1}D^{(+)}(x), \tag{2.120}$$

then the general solution of (2.117) is clearly

$$G(-x)=H(x)+\tfrac{1}{2}(c_4-c_3)x^2, \tag{2.121}$$

the contributions to F and G from the terms proportional to c_3 exactly cancel when they are combined in (2.85). It is for this reason that we may take $c_3=0$ without loss of generality. The term proportional to c_4 yields the constant $g_{\mu\nu}c_4$ when inserted in (2.85). Because Poincaré invariance implies $\langle\Psi_0, A_\mu(x)\Psi_0\rangle=0$, we conclude

$c_4 = 0$; but the argument has to be different from that leading to $c_2 = 0$, because $A_\mu(f)\Psi_0$ is not, in general, a vector of $\mathcal{H}'$ and the translation operator is only unitary with respect to the indefinite form $\langle\cdot,\cdot\rangle$ in $\mathcal{H}$. In fact, since the argument is much more easily given after we have constructed the Hilbert space $\mathcal{H}^{(1)}$ in the Laudau gauge, we delay the proof until that point.

There are many ways to derive the formula (2.119) for $H(x)$. One instructive procedure is the following. Note that since

$$(\Box + m^2)\Delta^{(+)}(m^2, x) = 0,$$

$$(\Box + m^2)\frac{\partial}{\partial m^2}\Delta^{(+)}(m^2, x) = -\Delta^{(+)}(m^2, x). \tag{2.122}$$

This suggests that $H(x)$ might be defined as

$$i\frac{\partial}{\partial m^2}\Delta^{(+)}(m^2, x)\Big|_{m^2=0} \tag{2.123}$$

However, a direct evaluation shows this quantity is infinite. Nevertheless, the infinity in (2.123) is independent of x, so its derivatives are finite and well defined. To see this in detail, it is convenient to work with the analytic function whose boundary value is $\Delta^{(+)}(m^2, x)$,

$$\Delta^{(+)}(m^2, z) = \frac{m^2}{8\pi}\frac{H_1^{(1)}(im\sqrt{-z^2})}{im\sqrt{-z^2}} \tag{2.124}$$

where $z = x + iy$. Now[40]

$$\frac{H_1^{(1)}(mi\sqrt{-z^2})}{mi\sqrt{-z^2}} = \frac{J_1(mi\sqrt{-z^2})}{mi\sqrt{-z^2}} + \frac{Y_1(mi\sqrt{-z^2})}{m\sqrt{-z^2}} \tag{2.125}$$

and

$$\frac{Y_1(mi\sqrt{-z^2})}{mi\sqrt{-z^2}} = -\frac{2}{\pi m^2 z^2} + \frac{1}{2\pi}\sum_{r=0}^{\infty}\frac{(m^2z^2/4)^r}{r!\,(r+1)!} \times[\log(m^2z^2/4) - \psi(r+1) - \psi(r+2)]. \tag{2.126}$$

Thus, $\Delta^{(+)}(m^2, z)$ has the singular part

$$-\frac{i}{4\pi^2 z^2} + \frac{m^2 i}{16\pi^2}\sum_{r=0}^{\infty}\left[\frac{(m^2z^2/4)^r}{r!\,(r+1)!}\right]\log\left[\frac{m^2z^2}{4}\right]. \tag{2.127}$$

This analytic function has a well-defined limit $-i[4\pi^2 z^2]^{-1}$ as $m^2 \to 0+$, but its derivative with respect to m^2 does not, since $\partial/\partial m^2\,(m^2\log m^2) \to \infty$ as $m^2 \to 0+$. The deletion of $(m^2 i/16\pi^2)\log m^2$, a term independent of z^2, leaves an expression whose derivative with respect to m^2 does have a limit as $m^2 \to 0+$. Thus, one can define a finite part of (2.123) as

$$-\frac{1}{16\pi^2}\log(-z^2) + \text{const} \tag{2.128}$$

which leads directly to the formula (2.119) for $H(x)$. This completes the proof of Proposition 2.4 except for the argument that $c_4 = 0$, which is to come later.

We now turn to the construction of Фок space representations for Laudau gauge. As compared with Gupta–Bleuler gauge two-point functions, the Laudau gauge two-point functions are more singular. Whereas the Gupta–Bleuler gauge two-point functions involve the integral of test functions $\hat{f}^\mu(k)$ and their components along k, $k_\mu \hat{f}^\mu(k)$, over the light cone C_+, the Landau gauge two-point functions also involve derivatives $\partial/\partial k^\mu \hat{f}^\nu(k)$. Consequently, if we attempt to use a Laudau gauge two-point function $\langle A(f)\Psi_0, A(g)\Psi_0\rangle^L$ to define a sesquilinear form on the Hilbert space $\mathcal{H}^{(1)}$ defined in (2.18), it will be unbounded. Explicitly, by Fourier transforming (2.114), we obtain

$$\langle\Psi_0, A(f)A(g)\Psi_0\rangle^L = -\pi\int d\Omega_0(k)\Big[\hat{f}^\mu(k)[g_{\mu\nu} - 2Mk_\mu k_\nu]\hat{g}^\nu(-k) - \frac{1}{4}\left(\frac{\partial}{\partial k^\nu}k_\mu + \frac{\partial}{\partial k^\mu}k_\nu\right)[\hat{f}^\mu(k)\hat{g}^\nu(-k)]\Big]. \tag{2.129}$$

Here we have a choice. We can either alter the definition of the scalar product $(\cdot,\cdot)$ so that it contains appropriate derivatives and (2.129) becomes bounded relative to it, or, alternatively, accept the form as only densely defined, but show that the basic constructions can still be carried out. A mitigating circumstance for the latter possibility is that the form (2.129) reduces to the Gupta–Bleuler form when it is restricted to f and g such that $k^\mu \hat{f}_\mu(k) = 0$ or $k^\mu \hat{g}_\mu(-k) = 0$ for k in a neighborhood of C_+. We will work out the former possibility in detail.

We want to introduce single particle wavefunctions whose components involve both a vector valued function $\hat{f}^\mu(k)$ and such quantities as $k^\mu\,\partial/\partial k^\nu \hat{f}^\nu(k)$ so that (2.129) can be written as a bilinear form in the components. Notice that (2.129) can be rewritten

$$\langle\Psi_0, A(f)A(g)\Psi_0\rangle = \pi\int d\Omega_0(k)\{\hat{f}_\mu(k)[-g^{\mu\nu} + 2Mk^\mu k^\nu]g_\nu(-k) - [\hat{F}^\mu(k)\hat{g}_\mu(-k) + \hat{f}^\mu(k)\hat{G}_\mu(-k)]\}, \tag{2.130}$$

where

$$\hat{F}_\mu(k) = \frac{1}{4}\left(\frac{\partial}{\partial k^\mu}k_\nu + k_\mu\frac{\partial}{\partial k^\nu}\right)\hat{f}^\nu(k),$$
$$\hat{G}_\mu(-k) = \frac{1}{4}\left(\frac{\partial}{\partial k^\mu}k_\nu + k_\mu\frac{\partial}{\partial k^\nu}\right)\hat{g}^\nu(-k). \tag{2.131}$$

This suggests that one introduce the eight component wavefunction

$$\Phi_\beta(k) = \sqrt{\pi}\begin{Bmatrix}\hat{f}_\mu \\ \hat{F}_\mu(k)\end{Bmatrix}, \quad \beta = 0, 1, \ldots, 7, \quad \mu = 0, 1, 2, 3, \tag{2.132}$$

the scalar product

$$(\Phi, \Psi)^{(1)} = \int_{C_+} d\Omega_0(k)\sum_{\alpha=0}^{7}\overline{\Phi_\alpha(k)}\Psi_\alpha(k), \tag{2.133}$$

and the sesquilinear form

$$\langle\Phi, \Psi\rangle^{(1)} = \int_{C_+} d\Omega_0(k)\sum_{\alpha,\beta=0}^{7}\overline{\Phi_\alpha(k)}\eta_{\alpha\beta}\Psi_\beta(k) \tag{2.134}$$

with η the 8×8 matrix

$$\eta = \left\{\begin{array}{c|c} -g_{\mu\nu} & g_{\mu\nu} \\ \hline g_{\mu\nu} & 0\end{array}\right\}. \tag{2.135}$$

Let $\mathcal{H}^{(1)}$ be the Hilbert space obtained by completion in the metric $(\cdot,\cdot)^{(1)}$ defined by (2.133), starting with the set of vectors given by (2.132), with $\hat{f}^\mu \in \mathcal{S}(\mathbb{R}^4)$. Then $\langle\cdot,\cdot\rangle^{(1)}$ is a bounded sesquilinear form densely defined on $\mathcal{H}^{(1)}$ which can be extended by continuity to all of $\mathcal{H}^{(1)}$. In fact, $\langle\Phi, \Psi\rangle^{(1)} = (\Phi, \eta\Psi)^{(1)}$.

If $\partial_\mu f^\mu(x) = 0$, then the first four components of the corresponding Φ satisfy $\sum_{\beta=0}^{3} k^{\beta}\Phi_{\beta}(k)=0$ while the last four are of the form $k_{\beta-4}h(k)$, $\beta=4,5,6,7$ and the same is true of the Φ's obtained as limits of these. This suggests that the closed linear subspace that consists of such vectors be defined as $\mathcal{H}^{(1)\prime}$. Clearly, $\langle\cdot,\cdot\rangle$ is nonnegative on $\mathcal{H}^{(1)\prime}$. The subspace $\mathcal{H}^{(1)\prime\prime}$ consists of those vectors that satisfy in addition $\Phi_{\beta}=k_{\beta}h_1(k)$, $\beta=0,1,2,3$.

There is a representation of the Poincaré group induced in $\mathcal{H}^{(1)}$ by the transformation law of the fields:

$$(U^{(1)}(a,\Lambda)\Phi)_{\beta}(k)=\exp(ik\cdot a)\det\Lambda\sum_{\gamma=0}^{3}\Lambda_{\beta}{}^{\gamma}\times\begin{cases}\Phi_{\gamma}(\Lambda^{-1}k), & \Lambda\in L^{\uparrow},\\ \overline{\Phi_{\gamma}(-\Lambda^{-1}k)}, & \Lambda\in L^{\downarrow},\end{cases}\quad\beta=0,1,2,3;$$

$$(U^{(1)}(a,\Lambda)\Phi)_{\beta}(k)=\exp(ik\cdot a)\det\Lambda\times\begin{cases}\sum_{\gamma=0}^{3}\Big(\Lambda_{\beta-4}{}^{\gamma}\Phi_{\gamma}(\Lambda^{-1}k)+\frac{i}{4}(a_{\beta-4}k_{\gamma}+k_{\beta-4}a_{\gamma})\times\sum_{\delta=0}^{3}\Lambda^{\gamma\delta}\Phi_{\delta}(\Lambda^{-1}k)\Big), & \Lambda\in L^{\uparrow},\\ \sum_{\gamma=0}^{3}\Big(\Lambda_{\beta-4}{}^{\gamma}\overline{\Phi_{\gamma}(-\Lambda^{-1}k)}+\frac{i}{4}(a_{\beta-4}k_{\gamma}+k_{\beta-4}a_{\gamma})\times\sum_{\delta=0}^{3}\Lambda^{\gamma\delta}\overline{\Phi_{\delta}(-\Lambda^{-1}k)}\Big), & \Lambda\in L^{\downarrow},\end{cases}$$
$$\beta=4,5,6,7. \tag{2.136}$$

This representation has been so constructed that it leaves the form $\langle\cdot,\cdot\rangle^{(1)}$ invariant.

The Hilbert space $\mathcal{H}$ is the symmetric Фок space $\mathcal{F}_s(\mathcal{H}^{(1)})$ built over $\mathcal{H}^{(1)}$, i.e., the Hilbert space defined by (2.23)–(2.26) with the old $\mathcal{H}^{(1)}$ replaced by the $\mathcal{H}^{(1)}$ just defined. The scalar product is defined by (2.27) with the indices now running from 0 to 7 instead of 0 to 3. The indefinite sesquilinear form is defined by

$$\langle\Phi,\Psi\rangle=\overline{\Phi^{(0)}}\Psi^{(0)}+\sum_{n=0}^{\infty}\sum_{\alpha_1\cdots\alpha_n}\sum_{\beta_1\cdots\beta_n}\int\cdots\int\Big(\prod_{j=1}^{n}d\Omega_0(k_j)\eta_{\alpha_j\beta_j}\Big)\times\overline{\Phi^{(n)}_{\alpha_1\cdots\alpha_n}(k_1\cdots k_n)}\Psi^{(n)}_{\beta_1\cdots\beta_n}(k_1\cdots k_n). \tag{2.137}$$

We may write $\langle\Phi,\Psi\rangle=(\Phi,\eta\Psi)$ where, with a slight abuse of notation, η stands for the Hermitian operator

$$(\eta\Psi)^{(n)}=\eta^{\otimes n}\Psi^{(n)},\quad n=1,2,\cdots,$$

which lets the matrix η act on each of the matrix indices of $\Psi^{(n)}$. It is worth noting that η has an inverse that is a bounded operator on $\mathcal{H}$, because the matrix η has the inverse

$$\begin{Bmatrix}0 & g^{\mu\nu}\\ g^{\mu\nu} & g^{\mu\nu}\end{Bmatrix}.$$

The subspace $\mathcal{H}'$ is defined by

$$\sum_{\beta_j=0}^{3}k_j^{\beta_j}\Phi^{(n)}_{\beta_1\cdots\beta_n}(k_1\cdots k_n)=0 \tag{2.138}$$

and

$$\Phi^{(n)}_{\beta_1\cdots\beta_n}(k_1\cdots k_n)=k_{j(\beta_j-4)}\Phi'^{(n)}_{\beta_1\cdots\widehat{\beta_j}\cdots\beta_n}(k_1\cdots\widehat{k_j}\cdots k_n),\quad\beta_j=4,5,6,7, \tag{2.139}$$

for some j and therefore by the symmetry for all j. The elements of $\mathcal{H}''$ satisfy (2.139) for $\beta_j=0,1,2,3$ with the subscript β_{j-4} replaced by β_j.

The representation $U^{(1)}$ defined above in $\mathcal{H}^{(1)}$ by (2.136) induces a representation in $\mathcal{H}$ which is unitary relative to the form $\langle\cdot,\cdot\rangle$.

The vector potential operator is defined by analogy with (2.31)–(2.35),

$$A^L(f)=a(\Pi_+^L(M)f)+a^*(\Pi_-^L(M)f), \tag{2.140}$$

with the annihilation operator a defined by

$$(a(\chi)\Psi)^{(n)}_{\alpha_1\cdots\alpha_n}(k_1\cdots k_n)=\sqrt{n+1}\int d\Omega_0(k)\sum_{\alpha,\beta=0}^{7}\chi_{\alpha}(k)\eta_{\alpha\beta}\Psi^{(n+1)}_{\beta\alpha_1\cdots\alpha_n}(kk_1\cdots k_n). \tag{2.141}$$

The creation operator a^* is defined by

$$(a^*(\Phi)\Psi)^{(n)}_{\alpha_1\cdots\alpha_n}(k_1\cdots k_n)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Phi_{\alpha_j}(k_j)\Psi^{(n-1)}_{\alpha_1\cdots\widehat{\alpha_j}\cdots\alpha_n}(k_1\cdots\widehat{k_j}\cdots k_n) \tag{2.142}$$

The Landau gauge mappings $\Pi_\pm^L(M)$ are defined by

$$(\Pi_\pm^L(M)f)_{\mu}(k)=\sqrt{\pi}\begin{cases}\hat{f}_{\mu}(\pm k),\\ \hat{f}_{\mu}(\pm k)+Mk_{\mu}k_{\nu}\hat{f}^{\nu}(\pm k).\end{cases} \tag{2.143}$$

It is easy to check that, with this definition of $A(f)$ and the vacuum state $\Psi_0=\{1,0,\cdots\}$, the two-point function $\langle\Psi_0,A(f)A(g)\Psi_0\rangle$ is the Landau gauge two-point function (2.129), and $A^L(f)^*=A^L(\bar{f})$. Furthermore, the transformation law

$$U(a,\Lambda)A^L(f)U(a,\Lambda)^{-1}=A^L(\{a,\Lambda\}f) \tag{2.144}$$

holds with

$$(\{a,\Lambda\}f)_{\mu}(x)=\det\Lambda\Lambda^{\nu}{}_{\mu}f_{\nu}(\Lambda^{-1}(a,\Lambda)). \tag{2.145}$$

The Landau gauge vector potential satisfies the Lorentz condition

$$\partial^{\mu}A_{\mu}^L(x)=0 \tag{2.146}$$

as an operator identity in $\mathcal{H}$. The verification is easy from the definition [insert the test function $f^{\mu}(x)=\partial^{\mu}h(x)$] provided one recognizes the identity

$$\int d\Omega_0(k)k^{\mu}\frac{\partial}{\partial k^{\mu}}\hat{h}(k)=-2\int d\Omega_0(k)\hat{h}(k). \tag{2.147}$$

The validity of (2.147) follows from the identity

$$z^{\mu}\frac{\partial}{\partial z^{\mu}}D^{(+)}(z)=-2D^{(+)}(z), \tag{2.148}$$

valid for z in the future tube, which is easily checked from the explicit form of $D^{(+)}(z)$

$$D^{(+)}(z)=-i[4\pi^2z^2]^{-1}. \tag{2.149}$$

The electromagnetic field computed from the Landau gauge vector potential (2.140) satisfies

$$\partial^{\mu}F_{\mu\nu}^L(x)=\partial_{\nu}B(x) \tag{2.150}$$

where the local scalar field $B(x)$ is chosen so that

$$(\Box A^L)(f)=A^L(\Box f)=-B(\partial\cdot f) \tag{2.151}$$

and therefore for any scalar test function g,

$$B(g)=a(\Pi_+'^Lg)+a^*(\Pi_+'^Lg) \tag{2.152}$$

with

$$(\Pi_{\pm}^{\prime L} g)_\alpha(k) = \sqrt{\pi}\begin{Bmatrix} 0 \\ \mp 2ik_\mu \hat{g}(\pm k)\end{Bmatrix}. \tag{2.153}$$

The appearance of $B(x)$ is a standard feature of quantizations of the electromagnetic field in the Landau gauge starting from an action principle.[17,41]

There is one aspect of the Hilbert space $\mathcal{H}$ just constructed in the Landau gauge which is essentially more complicated than in a Gupta–Bleuler gauge. That is the characterization of the local behavior of the components Φ_μ on the light cone C_+. In a Gupta–Bleuler gauge they are simply locally square integrable with respect to the measure $d\Omega_0(k)$. However, in the Landau gauge the initial set of wavefunctions from which the general element of $\mathcal{H}$ is obtained by completion, involves components $k_\mu\, \partial/\partial k^\nu \hat{f}^\nu(k) + \partial/\partial k^\mu\, (k_\nu \hat{f}^\nu(k))$ and it appears difficult to characterize the restrictions on the local behavior of the $\Phi_\beta(k)$, $\beta = 0,1,2,3$, which result from the requirement that the components involving derivatives converge in the mean in $L^2(d\Omega_0)$. It is possible to construct an alternative formalism in which the wavefunction has many more components but with simply stated local regularity properties (and $\langle\cdot,\cdot\rangle$ is still bounded). One simply puts all thirty-eight quantities $|\Phi^\mu|^2$, $|k^\mu\Phi^\nu|^2$, $|\partial/\partial k^\mu\, \Phi^\nu|^2$ into the norm defining $\mathcal{H}$. The only delicacy is that the derivatives $\partial/\partial k^0 \Phi^\mu$ need not converge to a function which is a derivative. We omit the details. This completes the construction of the operator formulation of the Landau gauge.

Now we return to the proof of $c_4 = 0$ in (2.118). We use the experience gained in the above reconstruction of $\mathcal{H}$ and A_μ given the two-point function (2.129) as a guide to an analogous construction in the presence of c_4. We write

$$\begin{aligned}(\Phi,\Psi)^{(1)} &= \mathrm{rhs}(2.133) + |c_4| \sum_{\mu=0}^{3} \hat{f}_\mu(0)\hat{g}_\mu(0),\\ \langle\Phi,\Psi\rangle^{(1)} &= \mathrm{rhs}(2.134) + c_4 \hat{f}^\mu(0)\hat{g}_\mu(0).\end{aligned} \tag{2.154}$$

The $\Phi o\kappa$ space $\mathcal{H}$ is constructed just as before but with the altered scalar product $(\cdot,\cdot)^{(1)}$ and $\langle\cdot,\cdot\rangle^{(1)}$. We have $\langle\Phi,\Psi\rangle = (\Phi,\eta\Psi)$ where η is the old η together with multiplication by $\operatorname{sgn} c_4$ in the new zero momentum term. Thus the new η also has an inverse. Next for each test function $g\in\mathcal{S}(\mathbb{R}^4)$ we define a linear functional F_g:

$$\begin{aligned}&F_g(\Psi_0) = 0, \quad F_g(A(f)\Psi_0) = c_4\overline{\hat{g}_\mu(0)}\hat{f}^\mu(0),\\ &F_g(:A(f_1)\cdots A(f_n):\Psi_0) = 0, \quad n>1,\end{aligned} \tag{2.155}$$

F_g is clearly uniformly bounded on the dense set of $\mathcal{H}$ spanned by $:A^L(f_1)\cdots A^L(f_n):\Psi_0$, $n = 0,1,2,\cdots$. Thus, there exists a vector Φ_g such that

$$F_g(\Psi) = (\Phi_g, \Psi) = (\eta^{-1}\Phi_g, \eta\Psi) = \langle\eta^{-1}\Phi_g, \Psi\rangle. \tag{2.156}$$

Notice that

$$F_{\{a,1\}g}(A(\{a,1\}f)\Psi_0) = F_{\{a,1\}g}(U(a,1)A(f)\Psi_0) = F_g(A(f)\Psi_0),$$

so

$$U(a,1)^{-1}\eta^{-1}\Phi_{\{a,1\}g} = \eta^{-1}\Phi_g. \tag{2.157}$$

But because F_g only depends on the zero-momentum value of $\hat{g}$, $F_{\{a,1\}g} = F_g$; so $\Phi_{\{a,1\}g} = \Phi_g$. Thus, $\eta^{-1}\Phi_g$ is invariant under the translation group and, if there is to be only one translation invariant state Ψ_0 in $\mathcal{H}$, the vector $\eta^{-1}\Phi_g$ must be proportioned to Ψ_0, say $\eta^{-1}\Phi_g = \lambda_g\Psi_0$. Thus,

$$F_g(A(f)\Psi_0) = c_4\overline{\hat{g}^\mu(0)}\hat{f}_\mu(0) = \overline{\lambda_g}\langle\Psi_0, A(f)\Psi_0\rangle \tag{2.158}$$

and since by the Lorentz invariance of Ψ_0, $\langle\Psi_0, A_\mu(x)\Psi_0\rangle = 0$, we conclude that $c_4 = 0$.

There is only one point of this argument that deserves further comment. We have constructed one particular Landau gauge yielding the two-point function (2.114) with the extra term $g_{\mu\nu}c_4$ and have verified within it that uniqueness of the vacuum requires $c_4 = 0$. How do we know there is not another gauge yielding the same two-point function within which c_4 does not have to be zero? The answer is that the gauge is unique up to isomorphism, i.e.,

Proposition 2.5: If $\{A_\mu, \mathcal{H}_1, \langle\cdot,\cdot\rangle_1, \mathcal{H}_1'\}$ and $\{A_{2\mu}, \mathcal{H}_2, \langle\cdot,\cdot\rangle, \mathcal{H}_2'\}$ are gauges for the free electromagnetic field that yield the same two-point function, there exists a mapping V of $\mathcal{H}_1$ onto $\mathcal{H}_2$ which satisfies

$$\langle V\Phi, V\Psi\rangle_2 = \langle\Phi,\Psi\rangle_1 \tag{2.159}$$

for all $\Phi,\Psi\in\mathcal{H}_1$

$$V\mathcal{H}_1' = \mathcal{H}_2', \quad VA_{1\mu}(x)V^{-1} = A_{2\mu}(x). \tag{2.160}$$

The proof of this proposition runs parallel to that for the standard result of the general theory of quantized fields that asserts the unitary equivalence of two theories given the equality of their vacuum expectation values.[14]

We have already remarked in connection with the Gupta–Bleuler and Coulomb gauges that, in general, a gauge transformation cannot be realized as a special gauge transformation of the form (2.7). That this statement applies in particular to Landau and Gupta–Bleuler gauges is the content of the following proposition.

Proposition 2.6: There is no special gauge transformation of the form (2.7) leading from a Gupta–Bleuler to a Landau gauge for the free electromagnetic field.

Proof: If such a transformation exists, it is possible to realize the Landau gauge vector potential $A_\mu{}^L$ as an operator valued distribution in the same Hilbert space as the Gupta–Bleuler vector potential A_μ. Let Ψ_0 be the vacuum of the Gupta–Bleuler gauge. Then

$$\begin{aligned}&\langle\Psi_0, \partial^\mu F^L_{\mu\nu}(x)A_\lambda(y)\Psi_0\rangle\\ &\quad= \langle\Psi_0, [\Box A^L_\nu(x) - \partial_\nu\partial^\lambda A^L_\lambda(x)]A_\lambda(y)\Psi_0\rangle\\ &\quad= \Box\langle\Psi_0, A^L_\nu(x)A_\lambda(y)\Psi_0\rangle = \langle\Psi_0, A^L_\nu(x)\Box A_\lambda(y)\Psi_0\rangle = 0\end{aligned} \tag{2.161}$$

where in the second step the transverseness of A^L_μ has been used and in the last the fact that the Gupta–Bleuler potential satisfies the waveequation.

On the other hand, for the Gupta–Bleuler gauge

$$\langle\Psi_0, \partial^\mu F_{\mu\nu}(x)A_\lambda(y)\Psi_0\rangle = -\langle\Psi_0, \partial_\nu(\partial^\mu A_\mu)(x)A_\lambda(y)\Psi_0\rangle \neq 0. \tag{2.162}$$

Thus,

$$\partial^\mu F^L_{\mu\nu} \neq \partial^\mu F_{\mu\nu}, \tag{2.163}$$

so A^L_μ and A_μ cannot be related by a special gauge transformation of the form (2.7).

Starting from the Landau gauge one can obtain by special gauge transformation a variety of other useful gauges.

Proposition 2.7: Any gauge for the free electromagnetic field in which the two-point function has the form

$$\langle\Psi_0, A_\mu(f)A_\nu(g)\Psi_0\rangle = \pi\int d\Omega_0(k)\exp[-ik\cdot(x-y)] \times\Big\{\hat f^\mu(k)[-g_{\mu\nu}+2Mk_\mu k_\nu]\hat g^\nu(-k) + \frac{\alpha}{4}\Big(\frac{\partial}{\partial k^\mu}k_\nu+\frac{\partial}{\partial k^\nu}k_\mu\Big)[\hat f^\mu(k)\hat g^\nu(-k)]\Big\}, \qquad (2.164)$$

α and M being real constants, is obtainable from Landau gauge by special gauge transformation.

Proof: The proof may be carried out by explicit construction of the field A_μ in the Hilbert space, $\mathcal{H}$ of a Landau gauge with $M=0$, as displayed in (2.131)—(2.143):

$$A(f)=a(\Pi_+(\alpha,M)f)+a^+(\Pi_-(\alpha,M)f), \qquad (2.165)$$

where a and a^+ are the annihilation and creation operators defined in $\mathcal{H}$ by (2.141). The mappings $\Pi_\pm(\alpha,M)$ are defined by

$$(\Pi_+(\alpha,M)f)_B(k)=\sqrt{\pi}\begin{Bmatrix}[g_{\mu\nu}-\rho k_\mu k_\nu]\hat f^\nu(k)\\ \sigma\Big\{\Big(\frac{\partial}{\partial k^\mu}k_\nu+k_\mu\frac{\partial}{\partial k^\nu}\Big)+\tau k_\mu k_\nu\Big\}\hat f^\nu(k)\end{Bmatrix},$$

$$(\Pi_-(\alpha,M)f)_B(k)=\sqrt{\pi}\begin{Bmatrix}[g_{\mu\nu}-\rho k_\mu k_\nu]\hat f^\nu(-k)\\ \Big\{\sigma\Big(k_\mu\frac{\partial}{\partial k^\nu}+\frac{\partial}{\partial k^\mu}k_\nu\Big)+\tau k_\mu k_\nu\Big\}\hat f^\nu(-k)\end{Bmatrix}, \qquad (2.166)$$

where $\sigma=\alpha$ and $M=\rho+\tau+\rho\sigma$. It is an easy calculation to check that with this definition of A_μ one gets the two-point function (2.164), provide one remembers the identity (2.147).

Examples:

1. The Landau gauge: $\alpha=1$.
2. The Fried—Yennie gauge[42]: $\alpha=-2$ $M=0$. Here
$$\langle\Psi_0, A_\mu(x)A_\nu(y)\Psi_0\rangle^{FY}=-[g_{\mu\nu}+2\partial_\mu\partial_\nu\Box^{-1}]i^{-1}D^{(+)}(x-y). \qquad (2.167)$$
3. Reducible Gupta—Bleuler type gauges: $\alpha=0$.

For $\alpha=0$, one gets a vector potential A_μ having the two-point function of Gupta—Bleuler gauge. However, the vacuum is not a cyclic vector for this vector potential. The situation is analogous to that for the Coulomb type gauges [described at the end of our discussion of the Gupta—Bleuler gauge, Example 4, after (2.77)].

This concludes our discussion of the free electromagnetic field in the Landau gauges. We believe it shows that these gauges can be treated smoothly and with mathematical precision within the indefinite metric formalism.

Gauges and gauge transformations in the presence of charges

The presence of charges gives rise to new technical difficulties in the discussion of gauge transformations. The transformation law (2.6) of a charged field $\psi(x)\to\exp[iq\chi(x)]\psi(x)$ involves the product of an exponential of a field and another field. Such an expression in general does not define a field; to give it a sense one must use special prescriptions and argue their effectiveness from the special properties of χ and ψ. An example of the complications in question is the problem of defining the exponential. If χ is a free scalar field, $:\exp iq\chi:(x)$ defined as a Wick ordered power series is not a tempered field but is a Jaffe field.[43] If χ is not a free field, the discussion of the exponential involves hypothetical properties of χ. Attempts to avoid discussing such properties by dealing always with the in and out fields (which are free fields) run up against the difficulty that in quantum electrodynamics, because of infrared phenomena, the conventional in and out fields belonging to charge-carrying fields do not exist, and collision theory is only now beginning to reach the stage where an appropriate substitute for them can be located.[44–46]

We have no doubt that a really satisfactory discussion of gauge transformations in quantum electrodynamics must resolve these difficulties. However, at the present stage of development of field theory, it seems sensible to accept less demanding tests of the adequacy of a formalism. If a formulation of a gauge theory makes sense to all orders in an expansion in the renormalized coupling constant, it would seem reasonable to accept it as a basis for investigations independent of perturbation theory. We do so in the following.

The most important new feature of the coupled theory is the presence of a nontrivial electromagnetic current. If one follows the ideas of Gupta—Bleuler, part of the indefinite metric formalism is essentially the same as for the free electromagnetic field. There is still a big Hilbert space $\mathcal{H}$ equipped with a sesquilinear form $\langle\cdot,\cdot\rangle$ and a subspace $\mathcal{H}'$ on which the form is nonnegative. As before, $\mathcal{H}_{\rm phys}=\mathcal{H}'/\mathcal{H}''$ where $\mathcal{H}''$ is the subspace of $\mathcal{H}'$ consisting of vectors Φ of zero length $\langle\Phi,\Phi\rangle=0$. However, here the free Maxwell's equations are replaced by

$$\partial_\mu F^{\mu\nu}=j^\nu+\mathcal{A}^\nu, \quad \partial_\mu{}^*F^{\mu\nu}=0, \qquad (2.168)$$

where j^μ is the electromagnetic current and $\mathcal{A}^\nu\mathcal{H}'\subset\mathcal{H}''$. As before, $\mathcal{A}^\nu$ takes different forms in the different gauges: $\mathcal{A}^\nu=-\partial^\nu\partial_\mu A^\mu$ in the Gupta—Bleuler gauges, and $\partial^\nu B$ in the Landau gauge. The different gauges are now labeled $\{A_\mu,\psi_i,\mathcal{H},\langle\cdot,\cdot\rangle,\mathcal{H}'\}$ where the ψ_i are a set of fields in $\mathcal{H}$ such that the vacuum Ψ_0 is cyclic for the ψ_i and A_μ together. The specification of the dynamics of these theories is completed by the equations of motion of the charged fields and the prescription for the source terms as functions of the charged field and A_μ. For example, in the quantum electrodynamics of a charged spin $\frac12$ field ψ

$$(-i\gamma^\mu\partial_\mu+m)\psi(x)=f(x) \qquad (2.169)$$

where j and f are functions of ψ and A.

For spin $\frac12$ quantum electrodynamics, the renormalized Green's functions have been defined to all orders in renormalized perturbation theory in the Gupta—Bleuler gauge with $M=0$, and are consistent with (2.168) and (2.169). The crucial point in the proof is to carry out the renormalization in such a way that there are no residual infrared divergences and so that Ward's identities are satisfied. The problem is posed and partly solved in Ref. 47; the solution is completed in Ref. 16. An alternative solution is provided by Ref. 48 which uses the 't Hooft—Veltman renormalization method.[49] Although an explicit published proof is not known to the present authors, the same arguments ought to be extendable to the quantum electrodynamics of spin $\frac{1}{2}$ in the other local covariant gauges.

On the other hand, for the Coulomb gauge, the Green's functions are not expected to be tempered distributions, since the connection between the matter field for the Coulomb gauge ψ_c and that for say a Gupta—Bleuler gauge ψ is formally (see Ref. 17, p. 128)

$$\psi_c(x) = \exp\{ie[(-\Delta)^{-1}\nabla\cdot\mathbf{A}](x)\}\psi(x). \tag{2.170}$$

One could enlarge the framework to consider fields in one of Jaffe's classes and then presumably treat the Coulomb gauge along with the local covariant gauges; we regard that as an eminently reasonable proposal. However, that is of little use to us here since we will need explicit locality and covariance and the Coulomb gauge has neither property.

The main conclusion of this discussion is that the evidence of perturbation theory supports the view that the indefinite metric formalism also works for coupled electromagnetic and charged fields. Thus, we make the following definition.

Definition 2.5: A *gauge* for the coupled Maxwell equations is specified by

(a) operator valued distributions: A_μ, the vector potential; j_μ, the electromagnetic current; and ψ_i, the other fields of the gauge in a Hilbert space $\mathcal{H}$;

(b) a representation U of the Poincaré group in $\mathcal{H}$;

(c) a sesquilinear form $\langle\cdot,\cdot\rangle$ on $\mathcal{H}$ with respect to which U is unitary;

(d) a distinguished subspace $\mathcal{H}'\in\mathcal{H}$ such that

(i) The restriction of the sesquilinear form $\langle\cdot,\cdot\rangle$ to $\mathcal{H}'$ is bounded and nonnegative

$$\langle\Psi,\Psi\rangle\geq 0 \quad \text{for } \Psi\in\mathcal{H}'.$$

(ii) There is a common dense domain $D\subset\mathcal{H}'$ for all local observables such that

$$F_{\mu\nu}(f)D\subset D, \quad j_\mu(f)D\subset D \tag{2.171}$$

for all $f\in\mathcal{S}$. Here $F_{\mu\nu}(x)=\partial_\mu A_\nu(x)-\partial_\nu A_\mu(x)$, and $F_{\mu\nu}(f)=\int F_{\mu\nu}(x)f(x)\,d^4x$, and $j_\mu(f)=\int d^4x f(x)j_\mu(x)$ are local fields satisfying

$$\begin{aligned}&\langle\Phi,[\partial_\mu F^{\mu\nu}(f)-j^\nu(f)]\Psi\rangle=0,\\ &\langle\Phi,\partial_\mu{}^*F^{\mu\nu}(f)\Psi\rangle=0\end{aligned} \tag{2.172}$$

for all $\Phi,\Psi\in\mathcal{H}'$ with Ψ in the domain of $\partial^\lambda F^{\mu\nu}(f)$ and $j(f)$.

(iii) The representation U leaves D invariant and therefore also leaves invariant the subset D'' of D consisting of those vectors of zero length. $\mathcal{H}''$ consists of the vectors in $\mathcal{H}'$ of zero length $\langle\Phi,\Phi\rangle=0$. $\mathcal{H}_{\text{phys}}=\mathcal{H}'/\mathcal{H}''$. There exists a unique vector Ψ_0, called the vacuum invariant under the translation subgroup of the Poincaré group. The vector Ψ_0 is invariant under the whole Poincaré group, lies in $\mathcal{H}'$, and is cyclic for the fields A_μ, j_μ, and ψ_i.

(iv) For all $\Phi\in\mathcal{H}'$, the Fourier transforms of $\langle\Phi,F_{\mu\nu}(x)\Psi_0\rangle$ and $\langle\Phi,j_\mu(x)\Psi_0\rangle$ have support contained in the closure of the future light cone $\bar{V}_+$.

Just as for the free Maxwell equations there is the stronger spectral condition.

(iv′) For all $\Phi\in\mathcal{H}'$, the Fourier transform of $\langle\Phi,A_\mu(x)A_\nu(y)\Psi_0\rangle$ and of $\langle\Phi,\psi_i(x)\psi_j(y)\Psi_0\rangle$ has support contained in $\bar{V}_+$ in the momenta conjugate to $(x+y)/2$ and $x-y$.

Remark: We have defined $\mathcal{H}_{\text{phys}}$ as the quotient space $\mathcal{H}'/\mathcal{H}''$. So defined it is guaranteed to be a pre-Hilbert space but not, on general grounds, a Hilbert space, i.e., it may not be complete. If it is not complete, we should define $\mathcal{H}_{\text{phys}}$ as $\overline{\mathcal{H}'/\mathcal{H}''}$, the completion. For the free field case it is not difficult to show that, in fact, $\mathcal{H}'/\mathcal{H}''$ is complete. For the coupled electromagnetic field the question is open. In the following the completion will not be indicated and will be discussed only when its consequences are significant.

An important aspect of j_μ is its relation to infinitesimal gauge transformations. Specifically, what is assumed here is (a) j^μ is conserved: $\partial^\mu j_\mu=0$. (This implies $\partial_\nu\mathcal{J}^\nu=0$ since $\partial_\mu\partial_\nu F^{\mu\nu}=0$ by the antisymmetry of $F_{\mu\nu}$. Of course, in concrete cases the conservation law of j_μ will be a consequence of the equations for the matter field.) (b) The corresponding charge Q, defined as a quadratic form by the limit $Q=\lim_{R\to\infty}Q_R$,

$$Q_R=\int\int f_d(x^0)f_R(|\mathbf{x}|)j^0(x^0,\mathbf{x})\,d^3x \tag{2.173}$$

with f_d and f_R positive, of compact support, and satisfying

$$\int f_d(x^0)\,dx^0=1, \quad f_R(|\mathbf{x}|)=1 \quad \text{for } |\mathbf{x}|\leq R \tag{2.174}$$

is the infinitesimal generator of gauge transformations of the first kind. That is, the quadratic form Q defines a unique self-adjoint operator, also denoted Q, and a field ψ carrying a charge q satisfies

$$\exp(i\alpha Q)\psi\exp(-i\alpha Q)=\exp(i\alpha q)\psi. \tag{2.175}$$

According to the usual ideas of quantum field theory, the domain of Q_R as a quadratic form includes all vectors obtained from the vacuum by application of polynomials in the local fields of the theory smeared with test functions of rapid decrease. Thus, if Φ and Ψ are any two vectors from that domain the assumed convergence is

$$\lim_{R\to\infty}\langle\Phi,Q_R\Psi\rangle=\langle\Phi,Q\Psi\rangle. \tag{2.176}$$

The infinitesimal form of (2.168) is

$$[Q,\psi]=q\psi, \tag{2.177}$$

which under our assumptions implies

$$\lim_{R\to\infty}\langle\Phi,[Q_R,\psi]\Psi\rangle=q\langle\Phi,\psi\Psi\rangle. \tag{2.178}$$

The relations between the one-parameter group $\exp(i\tau Q)$

and the limit expressions (2.176) and (2.178) for its infinitesimal generator have been studied intensively.[50–52] We will not attempt to summarize the results of those studies. However, we want to emphasize that they are all consistent with a mathematical framework for the theory generalizing that proposed by Haag and Kastler. In it there is a field algebra $\mathfrak{F}(\mathcal{O})$ attached to each bounded open set $\mathcal{O}$. It is the algebra of bounded operators in $\mathcal{H}$ generated by bounded functions of the smeared vector potential A_μ and charged fields ψ. The charge defines an automorphism of these field algebras

$$\alpha_\tau(A) = \exp(i\tau Q) A \exp(-i\tau Q). \tag{2.179}$$

The next step in the procedure is to distinguish a subalgebra $\mathcal{A}(\mathcal{O})$ of $\mathfrak{F}(\mathcal{O})$ that should play the role of Haag and Kastler's local algebra of observables. Here the indefinite metric complicates the situation. We expect the state space of the theory to be $\mathcal{H}_{\text{phys}} = \mathcal{H}'/\mathcal{H}''$ acting as state vectors for some algebra $\tilde{\mathcal{A}}(\mathcal{O})$ of operators acting in $\mathcal{H}_{\text{phys}}$. Thus, we have to connect $\mathfrak{A}(\mathcal{O})$ with $\tilde{\mathfrak{A}}(\mathcal{O})$. As we will see in the next section there are distinguished operators of $\mathfrak{F}(\mathcal{O})$ with a property we will call gauge independence. Each such operator determines a unique $\tilde{A}$ in $\mathcal{H}_{\text{phys}}$. These are candidates for observables, but the set of all gauge independent operators in $\mathfrak{F}(\mathcal{O})$ is not a candidate for $\mathfrak{A}(\mathcal{O})$ because it does not form a subalgebra of $\mathfrak{F}(\mathcal{O})$ and the mapping $A \to \tilde{A}$ is not an algebraic homomorphism. We will also define a stronger property called weak gauge invariance such that the weakly gauge invariant operators do form an algebra and to that extent constitute a candidate for $\mathfrak{A}(\mathcal{O})$. However, $A \to \tilde{A}$ is not an algebraic homomorphism for A weakly gauge invariant, so we cannot take the $\tilde{A}$ arising from weakly gauge invariant A as constituting $\tilde{\mathcal{A}}(\mathcal{O})$. We will give yet another more restrictive definition of a property called gauge invariance. The gauge invariant operators form an algebra and $A \to \tilde{A}$ is an algebraic homomorphism from it into an algebra of operators in $\mathcal{H}_{\text{phys}}$. The gauge invariant operators therefore yield candidates for $\mathfrak{A}(\mathcal{O})$ and $\tilde{\mathfrak{A}}(\mathcal{O})$.

Clearly, we have offered no definitive choice for $\mathfrak{A}(\mathcal{O})$ and $\tilde{\mathfrak{A}}(\mathcal{O})$. The main point is that, for the proof of the charge superselection rule, it does not matter what $\mathfrak{A}(\mathcal{O})$ is, it does not even have to be an algebra provided that every element of $\mathfrak{A}(\mathcal{O})$ has a property that we will call gauge independence. We will prove in the next section that under plausible assumptions every observable must arise from a gauge independent operator. We will, in any case, refer to $\tilde{\mathfrak{F}}(\mathcal{O})$ as the generalized Haag–Kastler field algebra.

It should be emphasized that gauge transformations in the wide sense of Definition 2.2 do *not*, in general, define automorphisms of the field algebra. In fact, in such gauges as the Coulomb gauge the basic local commutativity property that plays such an important role in Haag–Kastler theory does not hold. The question which gauge transformations do define automorphisms is open and interesting.

The existence of local covariant gauges that permit the construction of a generalized Haag–Kastler algebra $\mathfrak{F}(\mathcal{O})$ provides candidates for the subalgebra $\mathcal{A}(\mathcal{O})$, and therefrom, via the mapping $A \to \tilde{A}$, the algebra $\tilde{\mathcal{A}}(\mathcal{O})$. If the mapping is an algebraic homomorphism, $\tilde{\mathcal{A}}(\mathcal{O})$ will satisfy the requirements of relativistic invariance and local commutativity. So far no general construction has been found for the inverse process: given the $\tilde{\mathfrak{A}}(\mathcal{O})$ for a theory of charged particles to reconstruct $(\mathcal{O})$ and an appropriate gauge $\{A_\mu, \psi_i, \mathcal{H}, \langle\cdot,\cdot\rangle, \mathcal{H}'\}$ although there is no indication it cannot be done. The main trouble appears to be that the detailed dynamics of the theory in question must play a much larger role in the construction than in the standard reconstruction theorems of the general theory of quantized fields and new ideas appear to be necessary to carry through the program of Doplicher, Haag, and Roberts.

In the proof of the charge superselection rule in the following section Lemma 2.1 will be needed.

Lemma 2.1[6,7]: Let the continuous one parameter unitary group defined by the charge operator Q be $\exp(i\alpha Q)$. (The spectrum of Q consists of integer multiples ne of the magnitude e of the charge on the electron.) An element of the generalized Haag–Kastler field algebra $\mathfrak{F}(\mathcal{O})$ is said to carry charge q if it satisfies (2.175). Then $\mathfrak{F}(\mathcal{O})$ is generated by those of its elements that carry definite charge.

The idea of the proof is that associated with any $A \in \mathfrak{F}(\mathcal{O})$ there is a family of operators

$$A_n = \frac{1}{2\pi}\int_0^{2\pi} \exp(-ine\tau)\alpha_\tau(A)\,d\tau \tag{2.180}$$

which carry charge ne. A_n will lie in $\mathfrak{F}(\mathcal{O})$ because $\mathfrak{F}(\mathcal{O})$ is assumed weakly closed. Furthermore, we can recover $\alpha_\tau(A)$ as the weak limit

$$\alpha_\tau(A) = \lim_{N\to\infty} \sum_{-N}^{N} \exp(ine\tau) A_n \tag{2.181}$$

and, in particular, $A = \alpha_0(A)$.

Notice that the compactness of the gauge group of the first kind plays an essential role in this argument. (The compactness forces the invariant measure on the group to be finite and the spectrum of charge to be discrete. The fact that the gauge group is not only compact but a one-parameter group forces the spectrum to consist of multiples of some fixed charge.[51]

Clearly, the content of the above lemma remains true if one replaces $\mathfrak{F}(\mathcal{O})$ by any of its subalgebras, say $\mathfrak{A}(\mathcal{O})$, provided it is weakly closed.

By virtue of this lemma if one wants to prove

$$\langle\Phi, \alpha_\tau(A)\Psi\rangle = \langle\Phi, A\Psi\rangle, \quad \Phi, \Psi \in \mathcal{H}' \tag{2.182}$$

for all $A \in \mathfrak{A}(\mathcal{O})$ it suffices to prove it for A carrying charge q.

We are going to interpret the charge superselection rule as the validity of (2.182) for all quasilocal A, and that deserves some comment. First of all, notice that the validity of (2.182) for all $A \in \mathfrak{A}(\mathcal{O})$ and all bounded open sets $\mathcal{O}$, implies it for all quasilocal $A \in \mathfrak{A}$, i.e., all A that can be obtained as norm limits of sequences of A_n lying in some $\mathfrak{A}(\mathcal{O}_n)$ ($\lim_{n\to\infty}\|A_n - A\| = 0$ implies $\lim_{n\to\infty}\|\alpha_\tau(A_n) - \alpha_\tau(A)\| = 0$). Thus, it makes no difference whether the charge superselection rule is stated as Q

commutes with all local observables or Q commutes with all quasilocal observables. Second, it is a strong physical assumption that the only observables that matter are local or quasilocal. We adopt this as a natural *definition* within the framework of local relativistic quantum theory but recognize the possibility that while the ideas of Haag and Kastler seem natural and general to us there may be alternatives more favored by Nature.

It should be mentioned that the terminology used to describe observables in local relativistic quantum theory has not been fixed. One speaks also of global observables which are such quantities as charge, baryon, number, and lepton numbers. We take no position on this conflict in terminology but point out that the adjunction of such global observables to the quasilocal observables in no way changes the charge superselection rule as long as these global observables commute with each other as well as with all quasilocal observables. Such global observables are an analog in our context of the elements of the algebra of observables at infinity associated with a representation of the quasilocal algebra by Dobrushin[53] and Lanford and Ruelle.[54] This description accords with the so-called hypothesis of commutative superselection rules.[55]

There is a related question of terminology in the labeling of superselection sectors. In their systematic theory, Doplicher, Haag, and Roberts consider a general gauge group (the gauge group is the set of automorphisms of the field algebra that carries the field algebra of each bounded region into itself and leaves each observable fixed). They label the superselection sectors by the unitary equivalence classes of irreducible representations of the gauge group. Thus, for isospin I where the gauge group is isomorphic to $SU(2)$, the sectors are labeled by an angular momentum quantum number or the eigenvalues of I^2. The three noncommuting components of isospin I_1, I_2, I_3 are nontrivially represented in every sector save that of isospin 0. At first sight this definition of superselection sector would seem to conflict with the one quoted at the beginning of the present paper because I_1, I_2, and I_3 commute with all observables and therefore apparently define superselection rules according to the old definition. On the other hand, I_1, I_2, and I_3 clearly do not and cannot take on definite values in a superselection sector. The discrepancy is resolved if one recognizes that the subspace of the Hilbert space in which DHR realize the gauge group in one of its irreducible representations is a highly redundant description of a superselection sector. Two vectors of such a subspace may give the same expectation values for all quasilocal observables and thus be physically equivalent. With the DHR convention for labeling sectors, it is these equivalence classes of vector states that constitute the elements of a superselection sector; in the present case, they form a Hilbert space $\mathcal{H}_{I(I+1)}$. If one considers

$$\bigoplus_{I=0,1/2,1,\cdots} \mathcal{H}_{I(I+1)}$$

the direct sum of the spaces $\mathcal{H}_{I(I+1)}$, one has a formalism very similar to that arising from the hypothesis of commutative superselection rules, with the additional complication that the sectors are labeled by the irreducible representations of a noncommutative (group) rather than those of a commutative group. There is no room in this Hilbert space for I_1, I_2, I_3. On the other hand, the redundant description in which I_1, I_2, I_3 are nontrivially represented seems much better adapted to quantum field theory.

Although it is possible to maintain the view that $\oplus_I \mathcal{H}_{I(I+1)}$ is the natural Hilbert space to describe states and that, therefore, the hypothesis of commutative superselection rules holds in this example, it appears more natural to alter the definition within the DRH formalism and define it to mean that the gauge group is commutative. With this altered definition the hypothesis of commutative superselection rules does not hold in this example.

Finally, it should be remarked that the quantum electrodynamics of massless charged fermions in two-dimensional space—time is an exactly soluble model illustrating the discussion of this section,[56] but showing features presumably not possessed by quantum electrodynamics in higher dimensions. In any local covariant gauge the electromagnetic current exists and defines a charge with associated charge sectors in $\mathcal{H}$. However, $\mathcal{H}'$ lies entirely in the zero charge sector so $\mathcal{H}_{\text{phys}}$ does not contain any states of charge different from zero, and the fermions have disappeared from the theory (this has been proposed as a mechanism for hiding quarks[57]). As one would expect from these statements about $\mathcal{H}_{\text{phys}}$, in the Coulomb gauge there is no charge operator: Q is defined as a quadratic form by (2.166) but the corresponding symmetry is broken. It is a striking feature of the model that in a local covariant gauge it yields a unique vacuum and no symmetry breaking but $\mathcal{H}'$ is entirely contained in the states of zero charge, while in another gauge, the Coulomb gauge, the same physical situation is described in such a way that the conservation law of charge appears as a broken symmetry in the sense that the form Q does not define an operator and therefore cannot be the infinitesimal generator of a unitary group.

Although quantum electrodynamics in two-dimensional space—time has these peculiar features, it should be emphasized that in four-dimensional space—time the renormalized perturbation series for the Green's functions in Gupta—Bleuler gauges supports the view that for massive fermions in four dimensions assumptions (a) and (b) [see (2.173)—(2.176)] are valid. See Ref. 17 and 58 for relevant discussion.

Strict gauge invariance, gauge invariance, weak gauge invariance and gauge independence

Symanzik has pointed out the usefulness of distinguishing three notions of gauge invariance in the indefinite metric formalism.[17] We add a fourth which we call gauge independence. They are given in order of increasing restriction in the following definition.

Definition 2.6: An operator B, mapping $\mathcal{H}$ into $\mathcal{H}$, is *gauge independent* if

$$\langle \Phi, B\Psi\rangle = \langle \Phi + \chi_1, B(\Psi + \chi_2)\rangle \tag{2.183}$$

for all $\Phi, \Psi \in \mathcal{H}'$ and any $\chi_1, \chi_2 \in \mathcal{H}''$. In other words, the matrix elements $\langle\Phi, B\Psi\rangle$ for $\Phi, \Psi\in\mathcal{H}'$ depend only on the equivalence classes $[\Phi], [\Psi]\in\mathcal{H}_{\rm phys}$.

An operator B, mapping $\mathcal{H}$ into $\mathcal{H}$, is *weakly gauge invariant* if it and its adjoint leave $\mathcal{H}''$ invariant:

$$B\mathcal{H}''\subset\mathcal{H}'', \quad B^*\mathcal{H}''\subset\mathcal{H}''. \tag{2.184}$$

An operator B, mapping $\mathcal{H}$ into $\mathcal{H}$, is *gauge invariant* if it and its adjoint leave $\mathcal{H}'$ invariant

$$B\mathcal{H}'\subset\mathcal{H}', \quad B^*\mathcal{H}'\subset\mathcal{H}'. \tag{2.185}$$

An operator B, mapping $\mathcal{H}$ into $\mathcal{H}$, is *strictly gauge invariant* if it is gauge invariant and commutes with $\mathcal{A}_\mu$

$$[B, \mathcal{A}_\mu]=0. \tag{2.186}$$

This definition has a straightforward extension to operators defined only on appropriate dense subsets which will be elaborated in detail only when it is needed.

We are going to prove the chain of implications (2.183) $\Leftarrow$ (2.184) $\Leftarrow$ (2.185). For that we need an elementary lemma.

Lemma 2.2: Let $\langle\cdot,\cdot\rangle$ be a sesquilinear form everywhere defined and nonnegative on a complex vector space $\mathcal{H}'$. Let $\mathcal{H}''$ be the subset of $\mathcal{H}'$ consisting of those vectors Ψ having zero length

$$\langle\Psi, \Psi\rangle = 0.$$

Then $\Phi\in\mathcal{H}''$ if and only if

$$\langle\chi, \Phi\rangle = 0$$

for all vectors $\chi\in\mathcal{H}'$.

Proof: Schwarz's inequality is valid in $\mathcal{H}'$ (by a standard argument valid even when $\langle\cdot,\cdot\rangle$ is not strictly positive[59]), so

$$|\langle\chi, \Phi\rangle|^2\leq\langle\chi,\chi\rangle\langle\Phi,\Phi\rangle.$$

Thus, $\langle\Phi,\Phi\rangle=0$ implies $\langle\chi,\Phi\rangle=0$ for all $\chi\in\mathcal{H}'$. Conversely, take $\chi=\Phi$. ■

If B is weakly gauge invariant, and $\chi_1, \chi_2\in\mathcal{H}''$, then $B\chi_2$, $B^*\chi_1\in\mathcal{H}''$ and, therefore, by Lemma 2.2

$$\langle\Phi+\chi_1, B(\Psi+\chi_2)\rangle=\langle\Phi+\chi_1, B\Psi\rangle=\langle B^*(\Phi+\chi_1),\Psi\rangle=\langle\Phi, B\Psi\rangle$$

for every $\Phi, \Psi\in\mathcal{H}'$. Thus, if B is weakly gauge invariant, it is gauge independent.

If B is gauge invariant, then $B\mathcal{H}'\subset\mathcal{H}'$ and $B^*\mathcal{H}'\subset\mathcal{H}'$, so we have, for $\chi\in\mathcal{H}'$ and $\Phi\in\mathcal{H}''$,

$$\langle\chi, B\Phi\rangle=\langle B^*\chi, \Phi\rangle=0.$$

The second equality holds because $B^*\chi\in\mathcal{H}'$ and $\Phi\in\mathcal{H}''$. The vanishing of $\langle\chi, B\Phi\rangle$ for all $\chi\in\mathcal{H}'$ implies $B\Phi\in\mathcal{H}''$. Thus every gauge invariant operator is weakly gauge invariant.

By definition, every strictly gauge invariant operator is gauge invariant. In a large class of local covariant gauges, one can prove the stronger statement: If C is a local operator, $[C, \mathcal{A}_\mu]=0$ implies that C is gauge invariant. The class of gauges in question is that for which there is a local scalar field B such that $\mathcal{A}_\nu=\partial_\nu B$, and $\Psi\in\mathcal{H}'$ is characterized by $B^{(-)}\Psi=0$. It includes all Gupta–Bleuler and Landau gauges. Because $\partial_\nu\mathcal{A}^\nu=0$, B satisfies the wave equations, $\Box B=0$. By a standard argument, one can split the field B into positive and negative parts:

$$B(x)=B^{(+)}(x)+B^{(-)}(x), \tag{2.187}$$

where

$$\operatorname{supp}\hat{B}^{(+)}(k)\subset\bar{V}_+ \quad\text{and}\quad \operatorname{supp}\hat{B}^{(-)}(k)\subset\bar{V}_-.$$

Then $[C, \mathcal{A}_\nu]=0$ implies $[C, B^{(+)}]=0=[C, B^{(-)}]$ and, consequently, if $\Psi\in\mathcal{H}'$,

$$B^{(-)}C\Psi=CB^{(-)}\Psi=0.$$

Therefore, $C\Psi\in\mathcal{H}'$, and C is gauge invariant.

A rough idea of the distinction between strict gauge invariance and gauge invariance may be expressed as follows: Strictly gauge invariant operators are functions of A_μ, j_μ, and the other fields that do not change when $A_\mu\to A_\mu+\partial_\mu\chi$, $\psi_j(x)\to\exp[iq_j\chi(x)]\psi_j(x)$ where χ is any smooth real-valued function, while gauge invariant operators may change but only by the addition of an operator mapping $\mathcal{H}'$ into $\mathcal{H}''$. Examples of operators which are strictly gauge invariant are $F_{\mu\nu}$ and j_μ. An example of an operator which is gauge invariant but not strictly gauge invariant is the energy momentum tensor $Ⓗ_{\mu\nu}$. An example of an operator which is gauge independent but not weakly gauge invariant is, in a free Gupta–Bleuler gauge, $\partial_\mu A^\mu(f)A(g)$. It satisfies (2.183), but does not map $\mathcal{H}''$ into itself. An example of a weakly gauge invariant operator which is not gauge invariant is, in a free Gupta–Bleuler gauge,

$$(B\Psi)^{(n)}_{\mu_1\cdots\mu_n}=n_{\mu_1}\cdots n_{\mu_n}b^{\nu_1}(k_1)\cdots b^{\nu_n}(k_n)\Psi^{(n)}_{\nu_1\cdots\nu_n}(k_1\cdots k_n) \tag{2.188}$$

where $k_\mu b^\mu(k)=0$, $b\in\mathcal{S}(\mathbb{R}^4)$ and $k_\mu n^\mu\neq 0$. For further information and references see Ref. 17. For the present discussion of the charge superselection rule, it is mainly gauge independence that is significant.

The following elementary lemma will be used in the proof of the charge superselection rule.

Lemma 2.3: If B is gauge independent, then

$$\langle\Phi, B\Psi\rangle=0=\langle\Phi, B^*\Psi\rangle$$

for every $\Phi\in\mathcal{H}'$ and $\Psi\in\mathcal{H}''$.

Proof: It suffices to remark that the matrix elements depend on Ψ only through $[\Psi]$, which contains the zero vector.

Associated with a gauge invariant operator B in $\mathcal{H}$, there is a uniquely determined operator $\tilde{B}$ in $\mathcal{H}'/\mathcal{H}''$ as indicated in the diagram

$$\begin{tikzcd} \mathcal{H}' \arrow[r, "B"] \arrow[d] & \mathcal{H}' \arrow[d] \\ \mathcal{H}'/\mathcal{H}'' \arrow[r, "\tilde{B}"] & \mathcal{H}'/\mathcal{H}'' \end{tikzcd} \tag{2.189}$$

where the vertical arrows indicate the mapping $\Phi\to[\Phi]$. The crucial point is that if $\chi\in\mathcal{H}''$, $B(\Phi+\chi)=B\Phi+B\chi$, where $B\chi\in\mathcal{H}''$, so we can define $\tilde{B}[\Phi]=[B\Phi]$. $\tilde{B}$ is clearly linear and $\tilde{B}^*=(\tilde{B})^*$. Furthermore, $BC=\tilde{B}\tilde{C}$, $(B+C)=\tilde{B}+\tilde{C}$ and $\lambda B=\lambda\tilde{B}$ so $B\to\tilde{B}$ is a homomorphism of the gauge invariant operators of $\mathcal{H}$, which form an algebra, into the algebra of operators of $\mathcal{H}_{\rm phys}=\mathcal{H}'/\mathcal{H}''$.

If B is weakly gauge invariant or only gauge independent, the construction of $\tilde{B}$ requires alteration. Given B defined on $\mathcal{H}$, one can consider the sesquilinear form $\langle\Phi, B\Psi\rangle$ restricted to $\Phi, \Psi \in \mathcal{H}'$. If B is gauge independent, this form actually depends only on $[\Phi]$ and $[\Psi]$ and therefore defines a sesquilinear form on $\mathcal{H}'/\mathcal{H}''$, which we denote by $\tilde{B}([\Phi],[\Psi])$. We would like to conclude that there exists an operator $\tilde{B}$ in $\mathcal{H}'/\mathcal{H}''$ such that

$$\tilde{B}([\Phi],[\Psi]) = ([\Phi], \tilde{B}[\Psi])_{\mathcal{H}'/\mathcal{H}''}.$$

For this to hold, it is sufficient that $\tilde{B}([\Phi],[\Psi])$ be separately continuous in $[\Phi]$ and $[\Psi]$. The situation is quite different depending on the completeness of $\mathcal{H}'/\mathcal{H}''$. If $\mathcal{H}'/\mathcal{H}''$ is complete its topology can be defined in two different but equivalent ways. On the one hand, it has a topology induced by the seminorm $\sqrt{\langle\Phi,\Phi\rangle}$. On the other hand, $\mathcal{H}'/\mathcal{H}''$ as the quotient of two closed subspaces of $\mathcal{H}$ has a topology induced by the norm $\sqrt{(\Phi,\Phi)}$ in $\mathcal{H}$. Since $\langle\Phi, B\Psi\rangle = (\Phi, \eta B\Psi)$ is evidently separately continuous in the latter description, $\tilde{B}([\Phi],[\Psi])$ is separately continuous in the former. If $\mathcal{H}'/\mathcal{H}''$ is not complete little can be said without further information on $\langle\cdot,\cdot\rangle$ and the spaces. Apparently, there is no general argument that guarantees the existence of $\tilde{B}$ in this case. Of course, if $\tilde{B}$ does exist, it can be extended by continuity to all of $\overline{\mathcal{H}'/\mathcal{H}''}$.

We continue the analysis assuming $\mathcal{H}'/\mathcal{H}''$ is complete. Clearly the mapping $B \to \tilde{B}$ is linear and $(\tilde{B})^+ = \tilde{B}^+$. However, the preservation of algebraic properties is not assured. To begin with, if B and C are gauge independent, in general BC will not be, so the gauge independent operators do not form an algebra. If B and C are weakly gauge invariant, their product is also, since it too leaves $\mathcal{H}''$ invariant. Thus, the weakly gauge invariant operators form an algebra. Furthermore, the operation $B \to \tilde{B}$ is defined on the algebra of weakly gauge invariant operators. Unfortunately, we cannot in general say that $\widetilde{BC} = \tilde{B}\tilde{C}$ because C need not carry $\mathcal{H}'$ into $\mathcal{H}'$. On the other hand, if B is weakly gauge invariant and C is gauge invariant, we have $\widetilde{BC} = \tilde{B}\tilde{C}$ and $\widetilde{CB} = \tilde{C}\tilde{B}$ because if $\Phi, \Psi \in \mathcal{H}'$

$$\langle\Phi, BC\Psi\rangle = \langle\Phi, B(C\Psi)\rangle = \langle[\Phi], \tilde{B}[C\Psi]\rangle_{\mathcal{H}'/\mathcal{H}''}$$

since $C\Psi \in \mathcal{H}'$. The gauge invariance of C implies $[C\Psi] = \tilde{C}[\Psi]$, so

$$\begin{aligned}\langle\Phi, BC\Psi\rangle &= \langle[\Phi], \tilde{B}\tilde{C}[\Psi]\rangle_{\mathcal{H}'/\mathcal{H}''}\\ &= \langle[\Phi], \widetilde{BC}[\Psi]\rangle_{\mathcal{H}'/\mathcal{H}''}\end{aligned}$$

The argument for $\widetilde{CB} = \tilde{C}\tilde{B}$ follows by passing to adjoints.

In summary, even under the assumption that $\mathcal{H}'/\mathcal{H}''$ is complete, it appears that although the mapping $A \to \tilde{A}$ is defined for every gauge independent A we cannot guarantee it to be an algebraic homomorphism unless A is gauge invariant, even though the set of weakly gauge invariant operators do form an algebra.

We now turn to a closer examination of the physical meaning of gauge independence. Our notation $\mathcal{H}_{\text{phys}} = \mathcal{H}'/\mathcal{H}''$ indicates the received wisdom that physically realizable states should be described by vectors in $\mathcal{H}_{\text{phys}}$ and therefore observables should be described by self-adjoint operators mapping $\mathcal{H}_{\text{phys}}$ into itself. As we have seen, a gauge independent self-adjoint operator A in $\mathcal{H}'$ gives rise to a unique self-adjoint operator in A in $\mathcal{H}_{\text{phys}}$. We ask, "Does it ever make sense to regard a nongauge independent self-adjoint operator in $\mathcal{H}$ as an observable?" We are going to answer this question in the negative under a plausible additional assumption on $\mathcal{H}'$.

There are four essential constituents of a description of observations in quantum theory: *observables*, *physically realizable states*, *expectation values of observables in states*, and *transition probabilities between states*. If $\underline{A}$ is an observable and $\underline{\Psi}$ is a physically realizable state, we write $E(\underline{A},\underline{\Psi})$, a real number, for the expectation value of $\underline{A}$ in $\underline{\Psi}$. If $\underline{\Psi_1}$ and $\underline{\Psi_2}$ are two physically realizable states, there is an associated transition probability $T(\underline{\Psi_1},\underline{\Psi_2})$, a real positive number ≤ 1. The internal consistency of the description requires that

(a) Two observables $\underline{A_1}$ and $\underline{A_2}$ are equal if and only if

$$E(\underline{A_1},\underline{\Psi}) = E(\underline{A_2},\underline{\Psi})$$

for every physically realizable state $\underline{\Psi}$.

(b) Two physically realizable states $\underline{\Psi_1}$ and $\underline{\Psi_2}$ are equal if and only if

$$E(\underline{A},\underline{\Psi_1}) = E(\underline{A},\underline{\Psi_2})$$

for every observable $\underline{A}$.

(c) Two physically realizable states $\underline{\Psi_1}$ and $\underline{\Psi_2}$ are equal if and only if

$$T(\underline{\Psi_1},\underline{\chi}) = T(\underline{\Psi_2},\underline{\chi})$$

for every physically realizable state $\underline{\chi}$.

This general description is somewhat redundant since the information contained in the expectation values of all observables can also be expressed in terms of transition probabilities and conversely. [For example, in the usual Hilbert space formalism where $\underline{\Psi}$ stands for a unit ray, i.e., a unit vector up to a phase factor, we have $T(\underline{\Psi_1},\Psi_2) = E(P_{\Psi_1},\underline{\Psi_2}) = |(\Psi_1,\Psi_2)|^2$ where P_{Ψ_1} is the projection operator onto Ψ_1.]

Now suppose we are given an indefinite metric formalism with its triple of subspaces $\mathcal{H}$, $\mathcal{H}'$, and $\mathcal{H}''$ and its Hermitian sesquilinear form $\langle\cdot,\cdot\rangle$ positive on $\mathcal{H}'$. We assume:

(d) There is a subset $\Sigma \subset \mathcal{H}'$ of vectors Ψ normalized to 1, $\langle\Psi,\Psi\rangle = 1$, which determine corresponding states $\underline{\Psi}$. Similarly, there is a family $\mathcal{O}$ of operators acting in $\mathcal{H}$ and Hermitian relative to $\langle\cdot,\cdot\rangle$; the operators A in $\mathcal{O}$ determine observables $\underline{A}$. The mappings $\Psi \to \underline{\Psi}$ and $A \to \underline{A}$ are such that all physically realizable states $\underline{\Psi}$ arise from vectors $\Psi \in \Sigma$ and all observables $\underline{A}$ arise from operators A in $\mathcal{O}$. The expectation value and transition probability satisfy

$$E(\underline{A},\underline{\Psi}) = \langle\Psi, A\Psi\rangle, \tag{2.190}$$

$$T(\underline{\Psi_1},\underline{\Psi_2}) = |\langle\Psi_1,\Psi_2\rangle|^2. \tag{2.191}$$

The transition probabilities defined by (2.191) have the property that they depend only on equivalence classes in $\mathcal{H}'/\mathcal{H}''$:

$$T(\underline{\Psi_1}, \underline{\Psi_2}) = T(\underline{\Psi_1 + \chi_1}, \underline{\Psi_2 + \chi_2})$$

for all $\chi_1, \chi_2 \in \mathcal{H}''$, since by Lemma 2, 2 $\langle \chi_1, \Psi_2 \rangle = \langle \psi_1, \chi_2 \rangle = 0$. If we insist on the principle that physical statements expressible in terms of transition probabilities be also expressible in terms of expectation values of observables we have to require:

(e) If $\underline{\Psi}$ is a physically realizable state arising from a vector Ψ, then $\Psi + \chi$ for any $\chi \in \mathcal{H}''$ is also a vector giving rise to the same physically realizable state and

$$E(\underline{A}, \underline{\Psi}) = E(\underline{A}, \underline{\Psi + \chi}) \tag{2.192}$$

for all observables $\underline{A}$.

In this general setting the answer to the question of the gauge properties of observables in straightforward.

Proposition 2.8: In a quantum mechanical theory using an indefinite metric formalism satisfying (a), (b), (c), (d), (e), every operator A in $\mathcal{H}$ that gives rise to an observable $\underline{A}$ is gauge independent.

Proof: If (2.192) holds for every physically realizable $\underline{\Psi}$ and every $\chi \in \mathcal{H}''$, then

$$\langle \Psi, A\Psi \rangle = \langle \Psi + \lambda\chi, A(\Psi + \lambda\chi) \rangle = \langle \Psi, A\Psi \rangle + |\lambda|^2 \langle \chi, A\chi \rangle + 2\,\mathrm{Re}\lambda \langle \Psi, A\chi \rangle$$

for all complex λ, so

$$\langle \Psi, A\chi \rangle = 0. \tag{2.193}$$

If the Ψ are dense in $\mathcal{H}'$, this condition implies (2.183). On the other hand, if superselection rules operate in the theory there will be a set of orthogonal subspaces of $\mathcal{H}'$ spanning the whole space and the Ψ will be dense in each of these so one can again recover (2.193) for all $\Psi \in \mathcal{H}'$ and hence again the gauge independence of A.

3. STATEMENT AND PROOF OF THE CHARGE SUPERSELECTION RULE

The crux of the argument is contained in the following proposition.

Proposition 3.1: Let $\mathfrak{F}(\mathcal{O})$ be the local algebra of bounded operators associated with the bounded region $\mathcal{O}$ of space—time (the generalized Haag—Kastler field algebra) in the Hilbert space $\mathcal{H}$. Suppose that in $\mathcal{H}$ one has the basic structures of an indefinite metric formalism for the electromagnetic field, so that one can give meaning to the statement that an element of $\mathfrak{F}(\mathcal{O})$ carries charge q. Then each element $A \in \mathfrak{F}(\mathcal{O})$ that is gauge independent and carries charge q, either carries zero charge or has zero matrix elements $\langle \Phi, A\Psi \rangle$ for all vectors $\Phi, \Psi \in \mathcal{H}'$.

Proof: Suppose $A \in \mathfrak{F}(\mathcal{O})$ and

$$\lim_{R \to \infty} \langle \Phi, [Q_R, A]\Psi \rangle = q \langle \Phi, A\Psi \rangle$$

for every $\Phi, \Psi \in D$, the dense domain of Definition 2.5. For sufficiently large $|\mathbf{y}|$, $\partial_i F^{i0}(y^0, \mathbf{y})$ commutes with A. Thus for sufficiently large R, the right-hand side of

$$\langle \Phi, [Q_R, A]\Psi \rangle = \int\int d^3y\, dy^0 f_d(y^0) f_R(|\mathbf{y}|) \langle \Phi, [j^0(y^0, \mathbf{y}), A]\Psi \rangle$$

is

$$\int\int d^3y\, dy^0 f_d(y^0) f_R(|\mathbf{y}|) \langle \Phi, [(j^0(y^0, \mathbf{y}) - \partial_i F^{i0}(y^0, \mathbf{y})), A]\Psi \rangle$$

$$\begin{aligned} &= -\int\int d^3 dy^0 f_d(y^0) f_R(|\mathbf{y}|) \langle \Phi, [\mathcal{A}^0(y^0, \mathbf{y}), A]\Psi \rangle \\ &= -\langle \Phi, [\mathcal{A}^0(f_d f_R), A]\Psi \rangle \\ &= -\langle \mathcal{A}^0(f_d f_R)\Phi, A\Psi \rangle + \langle \Phi, A\mathcal{A}^0(f_d f_R)\Psi \rangle. \end{aligned}$$

Since $\mathcal{A}^0(f_d f_R)\mathcal{H}' \subset \mathcal{H}''$ and A is gauge independent, both of these last terms vanish by Lemma 2.3. Thus

$$q \langle \Phi, A\Psi \rangle = 0$$

for all $\Phi, \Psi \in D$, and we conclude that either $q = 0$ or all matrix elements of A between states of $\mathcal{H}'$ vanish.

An operator A such that all its matrix elements between vectors of $\mathcal{H}'$ vanish clearly describes a trivial observable since the $\tilde{}$ mapping yields $\tilde{A} = 0$ in $\mathcal{H}_{\text{phys}}$. Moreover, two local operators A and B differing by a physically trivial operator $[\langle \Phi, (A - B)\Psi \rangle = 0$ for all $\Phi, \Psi \in \mathcal{H}']$ describe the same observable since $\tilde{A} = \tilde{B}$. We will say, in this case, that A is equal to B modulo a physically trivial operator.

With this preparation we can now establish the main result of the paper.

Theorem 3.1 (The Charge Superselection Rule): In quantum electrodynamics in a local covariant gauge satisfying the hypothesis of Proposition 2.8, every quasilocal observable modulo a physically trivial operator commutes with $\exp(i\alpha Q)$ where Q is the electric charge and α is any real number.

Proof: We reduce the theorem to the preceding Proposition 3.1, using the fact the quasilocal observables must described by gauge independent operators by Proposition 2.8. We argue as we did just after Lemma 2.1, that if every gauge independent element of each of the generalized Haag—Kastler algebras $\mathfrak{F}(\mathcal{O})$ modulo a physically trivial operator commutes with $\exp(i\alpha Q)$, then so do their norm limits and therefore so does every gauge independent quasilocal observable. We notice that if A is a gauge independent element of $\mathfrak{F}(\mathcal{O})$ so are its constituents that carry definite charge, described in Lemma 2.1. Moreover, all its constituents carrying nonzero charge have zero matrix elements between vectors of $\mathcal{H}'$ according to Proposition 3.1 and therefore it differs from its zero constituent A_0 by a physically trivial operator. Clearly, A_0 commutes with $\exp(i\alpha Q)$, and the theorem is proved.

It is clear from the above discussion that locality and dynamics (Maxwell's equations) play an essential role in the proof of the charge superselection rule. It is not difficult to see that they are crucial for the argument. For example, the charge carrying fields introduced by Mandelstam,[60]

$$\Phi(x) = \phi(x) \exp[-ie \int_{-\infty}^{x} d\xi^\mu A^\mu(\xi)],$$

are (at least formally) strictly gauge invariant, as Mandelstam has shown. $\Phi(x)$ does not provide a counterexample to the above theorem because it is not local. This shows that the assumption of locality cannot be dispensed with in this proof. The role played by Maxwell's equations in the derivation makes it clear that the charge superselection rule is not a kinematical property, following from purely group theoretical considerations. It arises because the charge is coupled to a massless field $F_{\mu\nu}$ through the divergence $\partial_\mu F^{\mu\nu}$. More explicitly, it is the fact that

$$j^\nu = \partial_\mu F^{\mu\nu} - \mathcal{A}^\nu$$

where $\mathcal{A}^\nu$ is not observable that allows us to conclude $\langle \Phi, [Q_R, A]\Psi\rangle = 0$ in the above argument. Maxwell's equations are crucial in the deduction.

It is useful at this point to go back to the question why the theorem has not been proved directly in the Coulomb gauge or in $\mathcal{H}'/\mathcal{H}''$. One might think that a much simpler proof could be obtained by using only the triplet ($R \equiv$ [set of gauge invariant operators], $F_{\mu\nu}, j_\nu$), since this would allow us to work directly in $\mathcal{H}'/\mathcal{H}''$ rather than in $\mathcal{H}$. The main objection is that one should then exhibit a convincing proof that observables must be not only gauge independent but also gauge invariant. Even if the result seems plausible, providing a proof does not seem to be trivial and this justifies our giving the argument in $\mathcal{H}$, where gauge independence of observables can be established, rather than in $\mathcal{H}'/\mathcal{H}''$.

There is another deeper reason why the argument does not work in $\mathcal{H}'/\mathcal{H}''$. The crux of the proof of charge superselection rule is that a *global* observable like the electric charge Q can be obtained as a limit of local operator $Q_R = \int j^0(x) f_R(x)\, dx$. It is not obvious on general grounds that this has to be true in one gauge and not in another. From perturbation theory one learns that this is true in the Gupta—Bleuler gauge since, at each order of perturbation theory, for any local field A

$$[Q, A] = \lim_{R\to\infty} [Q_R, A]. \tag{3.1}$$

Since local fields are irreducible one is allowed to conclude that $Q = \lim_{R\to\infty} Q_R$, in the sense that gauge transformations of the first kind admit a local generator in the Gupta—Bleuler gauge.

In order to have an irreducible set of operators in the Coulomb gauge, or in $\mathcal{H}'/\mathcal{H}''$, one must include charged fields and for them

$$[\hat{Q}, \hat{\psi}] \neq \lim_R [\hat{Q}_R, \hat{\psi}]$$

as one learns from perturbation theory (1.23). This is not surprising because the property that gauge transformations of the first kind admit a local generator is usually dependent on the gauge. An example is provided by two-dimensional QED, where a local generator exists in the indefinite metric gauge but not in the Coulomb gauge.

The validity of (3.1) in $\mathcal{H}$ does not allow one to conclude that the same equation holds for the operators, $\hat{Q}$, $\hat{Q}_R$, and $\hat{A}$ in $\mathcal{H}'/\mathcal{H}''$, i.e., that

$$[\hat{Q}, \hat{A}] = \lim_{R\to\infty} [\hat{Q}_R, \hat{A}]. \tag{3.2}$$

Equation (3.2) follows from (3.1) if A is gauge invariant. But the set of gauge invariant operators does not form an irreducible set of operators in $\mathcal{H}'/\mathcal{H}''$ and one cannot conclude $[\hat{Q}, \cdot\,] = \lim_R [\hat{Q}_R, \cdot\,]$. In fact, for charged fields $\hat{\psi}$ this equation does not hold in perturbation theory. More generally, if A is a local operator in

$$[Q, A] = [\mathcal{A}_R, A], \quad R \text{ sufficiently large,}$$

whereas, in general,

$$[\hat{Q}, \hat{A}] \neq [\hat{\mathcal{A}}_R, \hat{A}] = 0,$$

since $\hat{\mathcal{A}}_R = 0$. Thus, relations valid in $\mathcal{H}$ do not carry over to $\mathcal{H}'/\mathcal{H}''$ in general.

4. SPECULATIONS ON THE BARYON AND LEPTON SUPERSELECTION RULES

To complete the paper, we want to discuss the possibility of an analogous treatment of other superselection rules, in particular, the baryon and lepton superselection rules. It is natural to look for such an explanation in the possibility that there are gauge fields associated with the strong and weak interactions, a perennial idea that is currently under extensive study.

As a preliminary we ask the question, "What is the difference between a conservation law arising from a symmetry of the theory under a (finite) Lie group and one arising from symmetry under an infinite Lie group or gauge group?" For example, the invariance of a Lagrangian under a $U(1)$ group (gauge invariance of the first kind) implies the existence of a current J^μ which is conserved, $\partial_\mu J^\mu = 0$, and whose charge, $\int d^3x\, J^0$, is the infinitesimal generator of the representation of the $U(1)$ group. If the Lagrangian is adjusted so as to be invariant under a local group which extends the $U(1)$ group, does the theory acquire new conserved currents or new restrictions on J^μ? This question can be put another way: If we are given a strictly conserved charge which is the infinitesimal generator of a $U(1)$ group, is there any obstruction to obtaining an extension to an associated local gauge group?

The standard answers to these questions appear to be: In going from a $U(1)$ gauge invariance to a corresponding local gauge invariance one acquires no new conservation laws. (See, for example, Ref. 61, especially p. 1557, and Ref. 62, especially p. 1083, or, for a general and pedagogical account, Ref. 63.) On the other hand, local gauge invariance does imply an additional restriction on the structure of the current: It must be the divergence of an antisymmetric tensor. The resulting special structure of the current is responsible for a variety of soft pion theorems, etc.[63] What we have to add to this is the remark that *conserved currents arising from local gauge invariance give rise to superselection rules if the gauge invariance is not broken.*

Just as for quantum electrodynamics, the arguments come in two stages. One has first to convince oneself that an indefinite metric formalism is necessary in order that one should have a local and physically interesting theory. Then within the indefinite metric formalism one proves the existence of superselection rules.

To be concrete, we consider the Yang—Mills theory[64] in which the isospin group is extended to a local gauge group. In one formulation of this theory, the basic equations are

$$\partial_\mu \mathbf{f}_{\mu\nu} = \mathbf{J}_\nu, \quad \mathbf{J}_\nu = \mathbf{j}_\nu + 2\epsilon \mathbf{b}^\mu \times \mathbf{f}_{\mu\nu} \tag{4.1}$$

the analog of the first of the Maxwell equations (1.13). Of course, in the full dress quantum field theory the term involving $\mathbf{b}^\mu \times \mathbf{f}_{\mu\nu}$ in the current $\mathbf{J}_\nu$ has to be defined by some limiting procedure, just as the constituent contributions to $\mathbf{J}_\nu$ do, when they are expressed as nonlinear functions of the other field variables. The triplet of conserved currents $\mathbf{J}_\nu$ is related to a triplet of charges

$$\mathbf{Q}_R = \int \mathbf{J}_0(x^0, \mathbf{x}) f_R(|\mathbf{x}|) f_d(x^0)\, d^4x \tag{4.2}$$

which are generators of local gauge transformations.

The arguments used in the proof of Theorem 1.3 now imply.

Proposition 4.1:

(a) In any local theory in which Eq. (4.1) holds, the charges Q_R^i generate the identity transformation of the local fields, and therefore cannot be associated with any quantity of physical interest.

(b) In any local theory in which there are local fields ψ of nonvanishing Q^i charge

$$\lim_{R\to\infty} [Q_R^i, \psi(f)] = q^i \psi(f), \tag{4.3}$$

the field equation (4.1) can hold only in a subspace $\mathcal{H}'$ of the Hilbert space $\mathcal{H}$ of states:

$$\langle \Phi, (\partial^\mu \mathbf{f}_{\mu\nu} - \mathbf{J}_\nu)\Psi\rangle = 0 \tag{4.4}$$

for $\Phi, \Psi \in \mathcal{H}'$. Here $\langle\cdot,\cdot\rangle$ must be indefinite on $\mathcal{H}$. On $\mathcal{H}$, the difference

$$\partial^\mu \mathbf{f}_{\mu\nu} - \mathbf{J}_\nu = \mathcal{A}_\nu \tag{4.5}$$

does not vanish identically, but satisfies

$$\mathcal{A}_\nu \mathcal{H}' \subset \mathcal{H}''. \tag{4.6}$$

The proof of (a) is an immediate consequence of locality and Gauss theorem, the argument being that of the proof of Proposition 3.1 with the additional simplification that the analog of $\mathcal{A}^0$ vanishes. The proof of (b) runs precisely parallel to that in Ref. 13 for quantum electrodynamics. [Recall that the essential constituents of that proof were that the Maxwell equation holds in the sense that matrix elements $\langle \Phi, (\partial_\mu F^{\mu\nu} - j^\nu)\Psi\rangle = 0$ for a dense set of vectors Φ, Ψ in $\mathcal{H}'$, that $\langle\cdot,\cdot\rangle$ be nonnegative on $\mathcal{H}'$ and nondegenerate on $\mathcal{H}$, and that the automorphism induced by the charge $Q_R = \int d^4x\, j^0(x^0, \mathbf{x}) f_R \times (|\mathbf{x}|) f_d(x^0)$ be nontrivial. Then it follows that $\mathcal{A}^\nu \Phi$ cannot be zero for all Φ in the dense set, but that it must be vector of zero length and, consequently. $\langle\cdot,\cdot\rangle$ cannot be nonnegative and nondegenerate on all of $\mathcal{H}$.]

When Yang and Mills quantized their theory, in fact, they adopted an alternative formulation of their equations in which $\mathcal{A}_\nu$ is not zero, being $-\partial_\nu \partial_\mu \mathbf{b}^\mu$, and most later authors have followed them in this. (See, for example, Refs. 61 and 65.) What we want to insist on is that if these procedures are carried out consistently they *always* yield a formalism with indefinite metric.

Once the formalism for Yang–Mills theory with gauges specified by the analog of Definition 2.1 is accepted, one can define strict gauge invariance, gauge invariance, weak gauge invariance, and gauge independence in precise analogy with Definition 2.6 and prove an analog of Theorem 3.1.

Proposition 4.2: In Yang–Mills gauge theories, the charges Q^i, if they are the infinitesimal generators of symmetries, are not only conserved, they also generate superselection rules. If a charge Q^i corresponds to a spontaneously broken symmetry it does not define a conserved quantity and thus does not define a superselection rule.

There are several remarks to be made on the significance of this result. First, a notational matter: Theories for which conserved noncommuting Q^i exist do not provide examples of noncommutative superselection rules. As explained in Sec. 2 after Lemma 2.1, we follow Ref. 5; the superselection sectors will be labeled by the unitary equivalence classes of the representations of the group generated by the conserved Q^i. Second, the theorem offers no clue whether a given Q^i will define a conserved quantity or not. The question whether a symmetry is spontaneously broken or not is a deep dynamical problem about which we have nothing to say. However, this theorem does provide a precise answer to the questions posed above. Conserved currents of gauge theories of the Yang–Mills type do have an additional property: When their charges are conserved they define superselection rules. Third, there is reason to believe that these arguments sketched for Yang–Mills theories can be extended to general gauge theories. It has been proved by Utiyama[66] that a Lagrangian invariant under local gauge transformations gives rise to equations of the form

$$J_\nu^\alpha = \partial^\mu F_{\mu\nu}^\alpha, \quad F_{\mu\nu}^\alpha = -F_{\nu\mu}^\alpha. \tag{4.7}$$

The conserved currents J^α give rise to charges for which Proposition 4.1 can again be proved. Fourth, among the gauge theories that come under Utiyama's results is the theory of the gravitational field. In it there is a family of conserved currents, the energy momentum tensor $\Theta_{\mu\nu}$, and the angular momentum density $\mathcal{M}_{\kappa;\lambda\mu}$. As was pointed out to us in correspondence with Deser, the argument of this paper apparently generalizes to yield a superselection rule for energy, momentum, and angular momentum. (See Ref. 62 for a discussion of Gauss law for these cases.) However, the very foundation of local quantum theory for theories of quantized gravitation needs further study[67] and so we defer discussion of this problem.

Now we turn to the baryon and lepton superselection rules. Here there are two experimental facts to be explained: the superselection rule itself and the fact that there is no long range force associated with the "charges", baryon number, and lepton numbers. [Experiment can, of course, provide only approximate support for such absolute statements[68] and it is not inconsistent to regard baryon conservation (and, of course, lepton conservation) as approximate.[69] We consider the conservation laws here as absolute.]

The absence of long-range forces associated with the baryon number is an old problem,[70] to which a number of interesting solutions have been proposed. (See, for example, Ref. 71.) Here we will consider a recent proposal[68] which exploits the Higgs mechanism of local gauge theories: The massless gauge boson whose exchange would give rise to the long-range force is converted to massive particles by the interaction to the accompaniment of spontaneous breaking of the local gauge invariance. The natural question is how one can exploit this mechanism to solve the second problem without at the same time losing the conservation law of baryon number because of the breaking of gauge symmetry. Pais proposed a solution to this problem in which there is an auxiliary non-Hermitian scalar field ϕ, a conserved baryon current J^μ, and a conserved ϕ current J^μ_ϕ. The baryon number Q associated with J^μ is assumed conserved, while that Q_ϕ associated with J^μ_ϕ is not. What we are going to show is that a slight generalization of the above analysis makes it applicable to Pais' model and leads to the conclusion that the model also has a baryon number superselection rule.

Variation of the classical Largangian which is Pais' starting point, with respect to the vector boson field, yields the equation

$$\partial_\mu G^{\mu\nu} = J^\nu + J^\nu_\phi \tag{4.8}$$

which is supposed to be an operator identity connecting local fields in the quantized version of the theory. Just as before we want to argue that an indefinite metric formalism in which the Eq. (4.8) is replaced by

$$\partial_\mu G^{\mu\nu} = J^\nu + J^\nu_\phi + \mathcal{A}^\nu \tag{4.9}$$

where $\mathcal{A}^\nu \mathcal{H}' \subset \mathcal{H}''$ is unavoidable if J^ν is to be a conserved current generating a conservation law for Q (here the baryon number). The extra complication is that J^ν_ϕ, while conserved $\partial_\mu J^\mu_\phi = 0$, is supposed *not* to generate a conserved quantity Q_ϕ; just that fact can be exploited to yield the proof. For consider the action of Q_R on a local field, $A \in \mathfrak{F}(\mathcal{O})$

$$(\Phi, [Q_R, A]\Psi) = -(\Phi, [Q_{\phi R}, A]\Psi) + (\Phi, [\partial_i F^{i0}(f_R f_d), A]\Psi) \tag{4.10}$$

where Φ, Ψ run over a dense domain of localized states in $\mathcal{H}$. For R sufficiently large the last term vanishes and we see that Q_ϕ has the same action on local fields as $-Q$. Since by assumption the latter action is unitarily implementable, so must be the former, a contradiction. Notice that for this argument to be valid Q_ϕ need only exist as a densely defined sesquilinear form; that always holds under our assumptions. Thus, (4.8) must be interpreted in a weaker sense. If we require it to hold in the sense of

$$\langle \Phi, (\partial_\mu G^{\mu\nu} - J^\nu - J^\nu_\phi)\Psi\rangle = 0 \tag{4.11}$$

on some dense subset of $\mathcal{H}'$ a closed subspace the Hilbert space $\mathcal{H}$, we conclude as in quantum electrodynamics that $\mathcal{H}'$ cannot be $\mathcal{H}$ and must contain vectors of zero length in the form $\langle\cdot,\cdot\rangle$. Those vectors form a subspace $\mathcal{H}''$. Furthermore, $\mathcal{A}^\nu \mathcal{H}' \subset \mathcal{H}''$ just as in Propositions 1.3 and 4.1.

Having accepted the equation of motion (4.9), we have an analog of (4.10):

$$\begin{aligned}\langle \Phi, [Q_R, A]\Psi\rangle = &-\langle \Phi, [Q_{\phi R} + \mathcal{A}^0(f_R f_d), A]\Psi\rangle \\ &+ \langle \Phi, [\partial_i F^{i0}(f_R f_d), A]\Psi\rangle.\end{aligned} \tag{4.12}$$

Again the last term vanishes for sufficiently large R. Now we restrict our attention to Φ, Ψ states in $\mathcal{H}'$, and A a local gauge independent operator. Then by the argument of the proof of Proposition 3.1 the terms containing $\mathcal{A}^0$ drop out and we are again left with

$$\lim_{R\to\infty}\langle \Phi, [Q_R, A]\Psi\rangle = -\lim_{R\to\infty}\langle \Phi, [Q_{\phi R}, A]\Psi\rangle; \tag{4.13}$$

this time, however, with A local and gauge independent, and Φ, Ψ is a dense set of states in $\mathcal{H}'$. Now the assumption that Q is a conserved observable quantity implies that the left-hand side of (4.13) is actually dependent on Φ and Ψ only through $[\Phi]$ and $[\Psi]$ (to see this recall that the algebra of observables is generated by elements of definite charge) and defines the infinitesimal form of an automorphism of the observables $\tilde{A}$ in $\mathcal{H}_{\text{phys}}$. Thus the same is true of the right-hand side. This is only compatible with the assumption that Q_ϕ defines a broken symmetry if the automorphism is trivial. Then baryon number carrying local fields carry baryon number zero and the baryon number superselection rule can be established as in the proof of Theorem 3.1. Thus, if Pais' model has solutions realizing the Higgs mechanism the baryon number Q defines a superselection rule.

There is a sense in which Pais' model resolves the problem of the absence of long-range forces associated with baryon charge by definition. In it the baryon current J_μ itself is not the source of a gauge field; rather it is $J_\mu + J_{\phi\mu}$. The Lagrangian possesses gauge invariance of the first kind and not of the second kind with respect to phase changes of the baryon field. Nevertheless, as the above argument shows, the broken gauge invariance of the second kind of $J_\mu + J_{\phi\mu}$ implies a superselection rule for the baryon charge.

The above discussion may perhaps be regarded more as an object lesson than as a general theorem. It makes plausible that there is a generic explanation of the baryon (and similarly the lepton) superselection rule in a class of gauge theories even if partial breaking of the gauge invariance occurs.

ACKNOWLEDGMENTS

While on leave from I. N. F. N. one of us (F. S.) received hospitality and financial support from Princeton University where most of this work was done. The other (A. W.) thanks Dr. Carl Kaysen for the hospitality of Institute for Advanced Study. Comments from V. Bargmann, S. Deser, R. Haag, and J. Swieca, which helped us in the writing and rewriting of this paper, are gratefully acknowledged.

APPENDIX A: SOLUTION OF THE EQUATION $d(\Lambda_1\Lambda) = \Lambda^{-1}d(\Lambda_1) + d(\Lambda)$

Wigner already gave a proof in 1939 (Ref. 37) that all solutions of the indicated functional equation are of the form

$$d(\Lambda) = (1 - \Lambda^{-1})n \tag{A1}$$

where n is some fixed vector. We found the following alternate proof which may make up in simplicity for what it lacks in priority.

Let G_a be the subgroup of the restricted Lorentz group $L^\uparrow_+$ consisting of all $\Lambda \in L^\uparrow_+$ such that

$$\Lambda a = a \tag{A2}$$

where a is some real timelike vector. We average the defining equation

$$d(\Lambda_1\Lambda) = \Lambda^{-1}d(\Lambda_1) + d(\Lambda) \tag{A3}$$

over G_a in the variable Λ_1 to obtain

$$\int_{G_a} d\mu(\Lambda_1)d(\Lambda_1\,\Lambda) = \Lambda^{-1}\int_{G_a} d\mu(\Lambda_1)d(\Lambda_1) + d(\Lambda) \tag{A4}$$

where $d\mu(\Lambda_1)$ is the invariant measure on G_a normalized so that $\int_{G_a} d\mu(\Lambda_1) = 1$. Since $d\mu$ is invariant, $d\mu(\Lambda_1) = d\mu(\Lambda_1\Lambda)$; and, therefore, denoting $\int_{G_a} d\mu(\Lambda_1)d(\Lambda_1)$ by n, we have (A1) at least for all $\Lambda \in G_a$ and with an n which can depend on a.

How uniquely is n determined? Clearly, adding any multiple of a to n does not affect (A1). Furthermore, that is all the arbitrariness in n, since

$$(1-\Lambda^{-1})n = (1-\Lambda^{-1})n'$$

for all $\Lambda \in G_a$ implies $n - n'$ is a multiple of a.

Next let a_1 and a_2 be two noncollinear timelike vectors. The subgroup $G_{a_1} \cap G_{a_2}$ consists of Lorentz transformations leaving a_1 and a_2 fixed and acting on the two-dimensional spacelike plane of vectors orthogonal to a_1 and a_2. If n_1 and n_2 are the n vectors determined as above for G_{a_1} and G_{a_2}

$$(1-\Lambda^{-1})n_1 = (1-\Lambda^{-1})n_2$$

for all $\Lambda \in G_{a_1} \cap G_{a_2}$. That implies that $n_1 - n_2$ lies in the plane spanned by a_1 and a_2,

$$n_1 - n_2 = \alpha a_1 + \beta a_2.$$

The requirement that $\alpha = 0 = \beta$ fixes the arbitrariness in n_1 and n_2 uniquely. Thus, there exists a uniquely determined n so that (A1) holds for all $\Lambda \in G_{a_1}$ and all $\Lambda \in G_{a_2}$.

It remains to show that with this same n (A1) holds for all $\Lambda \in L_+^\uparrow$. What follows immediately from (A3) is

$$\begin{aligned} d(\Lambda_1\Lambda) &= \Lambda^{-1}(1-\Lambda^{-1})n + (1-\Lambda^{-1})n \\ &= [1-(\Lambda_1\Lambda)^{-1}]n \end{aligned}$$

for all Λ_1 and Λ with $\Lambda_1 \in G_{a_1}$ and $\Lambda \in G_{a_2}$. By repeating this argument by induction, we obtain that (A1) holds for all Λ in the group generated by G_{a_1} and G_{a_2}. It remains to argue that the only subgroup of $L_+^\uparrow$ that contains both G_{a_1} and G_{a_2} is $L_+^\uparrow$ itself. This statement is an immediate consequence of the fact pointed out to us by Bargmann that if Λ is any element of $L_+^\uparrow$ not in $SO(3)$, the subgroup of $L_+^\uparrow$ generated by Λ and $SO(3)$ together is all of $L_+^\uparrow$.[72]

However, let us, for completeness, offer a proof, also suggested by Bargmann. It is convenient to work with the covering groups and therefore to show that if A is any element of $SL(2,\mathbb{C})$ not in $SU(2)$ and G is the smallest subgroup of $SL(2,\mathbb{C})$ containing A and $SU(2)$ then $G = SL(2,\mathbb{C})$. Clearly, if this result can be established the desired statement for $SO(3)$ and $L_+^\uparrow$ follows.

Now it suffices to show that G contains all positive diagonal matrices of determinant 1:

$$D_\eta = \begin{pmatrix} \eta & 0 \\ 0 & \eta^{-1} \end{pmatrix}, \quad 1 \le \eta < \infty,$$

because by the polar decomposition every $B \in SL(2,\mathbb{C})$ can be written $B = UH$ where $U \in SU(2)$ and H is positive, Hermitian, and of determinant 1, and H in turn can be written $H = U_1 D_\eta U_1^{-1}$ for some $U_1 \in SU(2)$ and some η with $1 \le \eta < \infty$. Thus the general element B of $SL(2,\mathbb{C})$ is a product of the element $UU_1 \in SU(2)$, D_η, and $U_1^{-1} \in SU(2)$. Notice further that the pair $\{\eta, \eta^{-1}\}$ is uniquely determined by B and, in fact, can be computed from the formula

$$\begin{aligned} \operatorname{tr}(B^*B) &= \operatorname{tr}([U_1 D_\eta^* U_1 U^*][UU_1 D_\eta U_1^*]) \\ &= \operatorname{tr}(D_\eta^2) = \eta^2 + \eta^{-2}. \end{aligned}$$

Thus, all that has to be shown is that G contains elements for which $\operatorname{tr}(B^*B)$ takes every value in the interval $2 \le \operatorname{tr}(B^*B) < \infty$.

By the same reasoning we may as well assume the element A is diagonal and Hermitian, say D_d. Consider, then, the element

$$A_\theta = \begin{Bmatrix} \cos\theta - \sin\theta \\ \sin\theta \cos\theta \end{Bmatrix}$$

in $SU(2)$, and its transform $V_{\theta,n}$ by D_d^n

$$V_{\theta,n} = D_d^n A_\theta D_d^{-n} = \begin{Bmatrix} \cos\theta & -d^{2n}\sin\theta \\ d^{-2n}\sin\theta & \cos\theta \end{Bmatrix}.$$

For it

$$\operatorname{tr}(V_{\theta,n}^* V_{\theta,n}) = \sum_{j,k=1}^{2} |V_{\theta,njk}|^2 = 2\cos^2\theta + (d^{2n} + d^{-2n})\sin^2\theta$$

which takes the value 2 for $\theta = 0$, and $(d^{2n} + d^{-2n})$ for $\theta = 2\pi$. Since it is continuous in θ it takes every value in between. Since n can be chosen as large as one likes, the trace takes every value in the interval $2 \le \operatorname{tr}(V_{\theta,n}^* V_{\theta,n}) < \infty$ for suitably chosen θ and n and the theorem is proved.

*Research supported by Air Force Office of Scientific Research under Contract F44620-71-C-0108.

[1]G.C. Wick, A.S. Wightman, and E.P. Wigner, Phys. Rev. 88, 101 (1952).
[2]R.M.F. Houtappel, H. van Dam, and E. Wigner, Rev. Mod. Phys. 37, 595 (1965), especially p. 622.
[3]R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964).
[4]R. Kadison, "Lectures on Operator Algebras" in *Cargèse Lectures in Theoretical Physics Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967), pp. 41—82, especially pp. 64-5.
[5]S. Doplicher, Commun. Math. Phys. 1, 1 (1965); S. Doplicher, I.M. Singer, and T. Regge, Commun. Math. Phys. 7, 51 (1968).
[6]S. Doplicher, R. Haag, and J. Roberts, Commun. Math. Phys. 13, 1 (1969).
[7]S. Doplicher, R. Haag, and J. Roberts, Commun. Math. Phys. 15, 173 (1969).
[8]S. Doplicher, R. Haag, and J. Roberts, Commun. Math. Phys. 23, 199 (1971).
[9]F. Strocchi, Phys. Rev. 162, 1429 (1967); F. Strocchi, "Locality and Covariance in QED and Gravitation General Proof of Gupta—Bleuler Type Formulations," Lectures at Boulder Summer Institute, August 1971.
[10]F. Strocchi, Phys. Rev. 166, 1302 (1968).
[11]F. Strocchi, Phys. Rev. D 2, 2334 (1970).
[12]J. Schwinger, Phys. Rev. 115, 721 (1959).
[13]R. Ferrari, L. Picasso, and F. Strocchi, "Some Remarks on Local Operators in Quantum Electrodynamics," Commun. Math. Phys. 35, 25 (1974).
[14]L. Gårding and A.S. Wightman, Arkiv för Fysik 28, 129 (1964). Here one finds a framework on which the present paper rests. For a more detailed exposition of the results

of the general theory of quantized fields see, for example, R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[15]F.J. Dyson, Phys. Rev. **82**, 428 (1951); **83**, 608 (1951).
[16]P. Blanchard and R. Seneor, Ann. Inst. Henri Poincaré (to be published).
[17]K. Symanzik, "Lectures on Lagrangian Quantum Field Theory," Desy T-71/1.
[18]R. Haag, Ann. der Phys. 7, 29 (1963).
[19]A. Swieca, "Goldstone's Theorem and Related Topics" *Cargèse Lectures in Physics*, edited by D. Kastler (Gordon and Breach, New York, 1970), Vol. 4, pp. 215—234.
[20]D. Sharp, Phys. Rev. **165**, 1867 (1968), especially pp. 1876, 7.
[21]W. Ehrenburg and R. Siday, Proc. Phys. Soc. Lond. B **62**, 8 (1949).
[22]Y. Aharanov and D. Bohm, Phys. Rev. **130**, 1625 (1963).
[23]B. de Witt, Phys. Rev. **125**, 2189 (1962).
[24]M. Peshkin, I. Talmi, and L.J. Tassie, Ann. Phys. **12**, 426 (1961).
[25]R.F. Dashen and D. Sharp, Phys. Rev. **165**, 1857 (1968).
[26]P. Jordan and W. Pauli, Z. Phys. **47**, 151 (1928).
[27]S. Gupta, Proc. Phys. Soc. Lond. A **63**, 681 (1950).
[28]K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).
[29]G. Källén, Helv. Phys. Acta **25**, 417 (1952).
[30]R. Rollnick, B. Stech, and E. Nunnemann, Z. Phys. **159**, 482 (1960).
[31]L.E. Evans and T. Fulton, Nucl. Phys. **21**, 492 (1960).
[32]J. Valatin, Dansk. Vid. Selsk. Mat.-fys. Medd. **26**, No. 13 (1951).
[33]L.D. Landau and I.M. Khalatnikov, J. Exp. Theor. Phys. **29**, 89 (1955) [Sov. Phys. JETP **2**, 69 (1956)].
[34]J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. **6**, 77 (1967).
[35]K. Hepp, Helv. Phys. Acta **36**, 355 (1963).
[36]L. Schwartz, *Théorie des Distributions I* (Hermann, Paris, 1957), p. 59.
[37]E.P. Wigner, Ann. Math. **40**, 149 (1939), especially pp. 174—6.
[38]L. Gårding and J. Lions, Nuovo Cimento Suppl. **14**, 9 (1959), especially p. 45.
[39]M.A. Naimark, *Normed Rings* (Groningen, Noordhoff, 1964), Chap. VIII.
[40]G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1962), pp. 62, 64, 73.
[41]N. Nakanishi, Prog. Theor. Phys. **35**, 111 (1966).
[42]H.M. Fried and D. Yennie, Phys. Rev. **112**, 1391 (1958).
[43]A. Jaffe, Phys. Rev. **158**, 1454 (1967).
[44]P. Blanchard, Commun. Math. Phys. **15**, 156 (1969).
[45]P.P. Kulish and L.D. Faddeev, Teor. Mat. Fiz. **4**, 153 (1970) [Theor. Math. Phys. **4**, 745 (1970)].
[46]D. Zwanziger, Phys. Rev. Lett. **30**, 934 (1973).
[47]K. Hepp, "Renormalization Theory," in *Statistical Mechanics and Quantum Field Theory*, Les Houches 1970, edited by C. De Witt and R. Stora (Gordon and Breach, New York, 1971), pp. 429—500.
[48]E. Speer, *Renormalization and Ward Identities Using Complex Space-Time Dimension*, Rutgers University preprint.
[49]G. 't Hooft and M. Veltman, *Regularization and Renormalization of Gauge Fields* Nucl. Phys. **B44**, 189-213 (1972); *Combinations of Gauge Fields* Nucl. Phys. **B50**, 318-353 (1972).
[50]H. Reeh, *On Goldstone's Theorem* Fortschritte der Physik **15**, 681-704 (1967).
[51]C.N. Yang, *Charge Quantization, Compactness of the Gauge Group, and Flux Quantization* Phys. Rev. **D1**, 2360 (1970).
[52]C.A. Orzalesi, Rev. Mod. Phys. **42**, 381 (1970).
[53]R.L. Dobrushin, Teorija Ver. i ee Prim. **13**, 201 (1968) [Theory of Probability and Its Applications **13**, 197 (1968)].
[54]O.E. Lanford and D. Ruelle, Commun. Math. Phys. **13**, 194 (1969).
[55]A.S. Wightman, Nuovo Cimento Suppl. **14**, 81 (1959), especially p. 84.
[56]J. Lowenstein and J. Swieca, Ann. Phys. **68**, 172 (1971).
[57]A. Casher, J. Kogut, and L. Susskind, Phys. Rev. Lett. **31**, 792 (1973).
[58]J. Swieca, "Goldstone's Theorem and Related Topics," in *Cargèse Lectures in Physics*, edited by D. Kastler (Gordon and Breach, New York, 1970), Vol. 4, pp. 215—33.
[59]N. Bourbaki, *Eléments de Mathématique, Fasc. XVIII, Livre V, Espaces vectoriels topologiques* (Hermann, Paris, 1967), Chap. V., p. 127.
[60]S. Mandelstam, Ann. Phys. **19**, 1 (1962).
[61]T.W. Kibble, Phys. Rev. **155**, 1554 (1967).
[62]S. Deser, Amer. J. Phys. **40**, 1082 (1972).
[63]S. Weinberg, "Dynamic and Algebraic Symmetries," in *Lectures on Elementary Particles and Quantum Field Theory*, 1970 *Brandeis University Summer Institute in Theoretical Physics*, edited by S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, Mass., 1970), Vol. 7, pp. 283—393.
[64]C.N. Yang and R. Mills, Phys. Rev. **96**, 191 (1954).
[65]J. Schwinger, Phys. Rev. **130**, 402 (1962).
[66]R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[67]F.J. Dyson, Bull. Amer. Math. Soc. **78**, 635 (1972).
[68]A. Pais, Phys. Rev. **D8**, 1844 (1973).
[69]J. Pati and A. Salam, Phys. Rev. Lett. **31**, 661 (1973).
[70]T.D. Lee and C.N. Yang, Phys. Rev. **98**, 1501 (1955).
[71]J. Schwinger, Phys. Rev. **125**, 397 (1962); **128**, 2425 (1962).
[72]R. Brauer, Arch Ration. Mech. Anal. **18**, 97 (1965).

Perturbation theory of a nonideal Bose gas. II*

I. RamaRao

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Franz Mohling

Department of Physics, University of Colorado, Boulder, Colorado 80302
(Received 19 June 1974)

With the pair Hamiltonian model as the starting point, perturbation theory calculations are performed for the radial distribution function and the structure factor through second order in the two-body interaction potential for a many-boson system below the Bose–Einstein transition temperature. For the special case of a dilute hard-sphere Bose gas at $T=0°$ K, the structure factor is calculated in the low-momentum limit keeping terms through second order in the hard-sphere gas parameter $(na^3)^{1/2}$, where n is the gas density and a is the hard-sphere diameter. The resulting expression for the structure factor is shown to satisfy the Feynman–Bijl relation explicitly.

1. INTRODUCTION

In a previous paper[1] perturbation theory calculations were performed for the average energy and momentum distribution of a degenerate Bose system (a Bose system below the Bose-Einstein transition temperature), through second order in the two-body interaction potential using the pair Hamiltonian model as a starting point. In the present work we perform similar calculations for correlation functions; in particular, the radial distribution function and the structure factor for a degenerate Bose system are both calculated. We specialize these results to the case of a dilute hard sphere Bose gas at $T=0°$K, using the procedure outlined in Appendix F of I. The resulting expression for the structure factor disagrees in part with an earlier result of T. T. Wu.[2] In the low-momentum limit we obtain an explicit expression for the structure factor, valid through second order in the hard sphere gas parameter $(na^3)^{1/2}$, where n is the density and a is the hard sphere diameter.

The Feynman—Bijl relation[3] which relates $S_{\text{in}}(q)$, the inelastic part of the structure factor $S(q)$, to the low-lying excitations of a degenerate Bose system is given by

$$S_{\text{in}}(q)=\frac{\hbar^2}{2m}\cdot\frac{q^2}{W_{\text{exc}}(q)}\quad(q\to 0). \tag{1.1}$$

The result for the structure factor obtained in this paper is shown to satisfy (1.1) for the dilute hard sphere Bose gas (DHSBG) provided we identify $W_{\text{exc}}(q)$ with the quasi-particle energy $\epsilon_+(q)$ obtained in I [see Eq. (I. F25)]. We shall use units $\hbar=2m=1$ throughout this paper.

2. RADIAL DISTRIBUTION FUNCTION AND STRUCTURE FACTOR

We begin by defining the one- and two-particle correlation functions $P_1(r_1)$ and $P_2(r_1,r_2)$, where r_1 and r_2 label the coordinates of particles 1 and 2. Using Fock space notation, these are

$$P_1(r_1)=n^{-1}\langle\psi^+(r_1)\psi(r_1)\rangle, \tag{2.1}$$

$$P_2(r_1,r_2)=n^{-2}\langle\psi^+(r_1)\psi^+(r_2)\psi(r_2)\psi(r_1)\rangle, \tag{2.2}$$

where

$$\psi(r)=\Omega^{-1/2}\sum_k a_k e^{ik\cdot r}, \tag{2.3}$$

Ω being the volume. The operators a_k in (2.3) are the annihilation operators of free bosons, and these together with the Hermitian conjugate operators a_k^+ satisfy the Bose commutation relations. $P_1(r_1)$ gives the normalized probability for finding one particle at the position r_1 in the N-particle system. Similarly, $P_2(r_1,r_2)$ gives the probability for finding one particle at r_1 and another at r_2 in the N-particle system. For an isotropic system in the infinite-volume limit[4] it can be shown that

$$P_1(r_1)\equiv 1, \tag{2.4}$$

$$P_2(r_1,r_2)=P_2(|r_1-r_2|)\equiv P_2(r), \tag{2.5}$$

$$r=|r_1-r_2|, \tag{2.6}$$

where $P_2(r)$, known as the radial distribution function, now represents the probability that two particles in the N-boson system are separated by a distance $r=|r_1-r_2|$. In the infinite-volume limit, using (2.3) and momentum conservation, we can rewrite Eq. (2.2) for $P_2(r)$ at $T=0°$K as

$$P_2(r)=(n\Omega)^{-2}\sum_{\substack{k_1k_2k_3k_4\\(k_1+k_2=k_3+k_4)}}\exp[-i(k_1-k_4)\cdot r]\times\frac{\langle\Psi\,|\,a_{k_1}^+\,a_{k_2}^+a_{k_3}a_{k_4}\,|\,\Psi\rangle}{\langle\Psi\,|\,\Psi\rangle} \tag{2.7}$$

where $|\Psi\rangle$ is the ground state vector of the many-boson system. The Fourier transform of $P_2(r)$, known as the structure factor $S(q)$, is given by

$$S(q)=1+n\int P_2(r)\,e^{iq\cdot r}\,d^3r. \tag{2.8}$$

Interest in this quantity occurs especially, because experiments using neutron and x-ray diffraction give direct information about $S(q)$, and hence about $P_2(r)$.[5] We shall be interested here only in the inelastic part of $S(q)$, i.e.,

$$\begin{aligned}S_{\text{in}}(q)&=S(q)-n(2\pi)^3\delta^{(3)}(q)\\&=1+n\int[P_2(r)-1]\,e^{iq\cdot r}\,d^3r.\end{aligned} \tag{2.9}$$

In this section, we shall give formal calculations of $P_2(r)$ and $S_{\text{in}}(q)$, starting from (2.7) and (2.9), respectively, to second order in the two-body interaction potential for a many-boson system below the Bose—Einstein transition temperature T_λ. The procedure for the perturbation theory calculations using the pair Hamiltonian model as the starting point has already been outlined in Sec. 2 of I, and we shall not repeat these details here.

As in I, we assume that a degenerate Bose system at rest is characterized by macroscopic occupation of the zero-momentum state. According to Sec. 2 of I we must consider separately the two possibilities for k_i; namely, $k_i = 0$ and $k_i = p_i \neq 0$ and make the Bogoliubov approximation [see (I. 2. 2) and following] $a_0 \sim a_0^+ \sim N_0^{1/2}$. Here N_0 is the number of particles in the zero-momentum state. We then obtain the following expression for $P_2(r)$ from (2. 7):

$$P_2(r) = 1 + P_{2A}(r) + F_{11}(r)F_{11}(-r) + F_{20}(r)F_{02}(r) + 2\xi F_{12}(r) + F_{22}(r) + \tilde{F}_{22}(r), \tag{2.10}$$

where

$$\xi = x/n \quad \text{with } x = N_0/\Omega, \tag{2.11}$$

denotes the fraction of particles in the zero-momentum state. The functions introduced in (2. 10) are

$$P_{2A}(r) = \xi(n\Omega)^{-1}\sum_p e^{ip\cdot r}B(p), \tag{2.12}$$

with

$$B(p) = \langle\Psi|(2a_p^+a_p + a_p^+a_{-p}^+ + a_pa_{-p})|\Psi\rangle\langle\Psi|\Psi\rangle^{-1} \tag{2.13}$$

$$F_{11}(r) = (n\Omega)^{-1}\sum_p \langle n(p)\rangle e^{ip\cdot r}, \tag{2.14}$$

$$F_{02}(r) = (n\Omega)^{-1}\sum_p e^{ip\cdot r}\langle\Psi|a_pa_{-p}|\Psi\rangle\langle\Psi|\Psi\rangle^{-1}, \tag{2.15}$$

$$F_{20}(r) = F_{02}^+(-r), \tag{2.16}$$

$$F_{12}(r) = (n\Omega)^{-1}N_0^{-1/2}\sum_{\substack{p_2p_3p_4\\(p_2=p_3+p_4)}} e^{ip_4\cdot r}\langle\Psi|a_{p_2}^+a_{p_3}a_{p_4} + a_{p_4}^+a_{p_3}^+a_{p_2})|\Psi\rangle\langle\Psi|\Psi\rangle^{-1} \tag{2.17}$$

$$\begin{aligned}\tilde{F}_{22}(r) = (n\Omega)^{-2}\sum_{p_1p_2}\{1 + \exp[-i(p_1-p_2)\cdot r]\} \\ \times[\langle\Psi|a_{p_1}^+a_{p_2}^+a_{p_1}a_{p_2}|\Psi\rangle\langle\Psi|\Psi\rangle^{-1} \\ - \langle n(p_1)\rangle\langle n(p_2)\rangle], \\ + (n\Omega)^{-2}\sum_{p_1p_2}\exp[-i(p_1+p_2)\cdot r] \\ \times\{\langle\Psi|a_{p_1}^+a_{-p_1}^+a_{p_2}a_{-p_2}|\Psi\rangle\langle\Psi|\Psi\rangle^{-1} \\ - \langle a_{p_1}^+a_{-p_1}^+\rangle\langle a_{p_2}a_{-p}\rangle_2\},\end{aligned} \tag{2.18}$$

$$F_{22}(r) = (n\Omega)^{-2}\sum_{\substack{p_1p_2p_3p_4\\(p_1+p_2=p_3+p_4)}}\exp[-i(p_1-p_4)\cdot r] \times\langle\Psi|a_{p_1}^+a_{p_2}^+a_{p_3}a_{p_4}|\Psi\rangle\langle\Psi|\Psi\rangle^{-1}. \tag{2.19}$$

In equations (2. 12)–(2. 19), $\langle n(p)\rangle$ is the momentum distribution, already calculated in I and given there by (5. 10).

The calculation of $B(q)$ and the $F_{ij}(r)$ to second order in the (real) two-body interaction matrix element $\langle k_1k_2|V^{(s)}|k_3k_4\rangle$ [see Eq. (I. 2. 2)] now proceeds as follows: First we transform to a quasiparticle representation by using the Bogoliubov transformation (I. B13), which was used to diagonalize the pair Hamiltonian (I. 2. 11) (see Appendix B of I). Then we substitute the expression for the state vector $|\Psi\rangle$, as given by (I. 2. 22)–(I. 2. 30) to second order in V. The perturbation theory calculations then proceed exactly as in I, and we summarize only the results here. For the quantity $B(q)$ of (2. 13) we find

$$B(q) = -1 + [1-\alpha_-(q)][1+\alpha_-(q)]^{-1}[1+2n_+(q)] + S_{11}(q)/\xi, \tag{2.20}$$

where $\alpha_-(q)$ is given by (I. 4. 3) and

$$\begin{aligned}S_{11}(q) = -\xi[1-\alpha_-(q)_0][1+\alpha_-(q)_0]^{-1} \\ \times x\Omega\sum_{p_2p_3}\sum_{ijk=\pm}[f_i(q)_0n_i(q)] \\ \times[jf_j(p_2)_0n_j(p_2)]\cdot[kf_k(p_3)_0n_k(p_3)] \\ \times|A_{ijk}(qp_2p_3)|^2 \\ \times[\epsilon_i(q)_0 + \epsilon_j(p_2)_0 + \epsilon_k(p_3)_0]^{-2} \\ - \tfrac{1}{3}\xi[1-\alpha_-(q)_0][1+\alpha_-(q)_0]^{-1}\sum_{p_2p_3p_4}\sum_{ijkl=\pm} \\ \times[f_i(q)_0n_i(q)] \\ \times[jf_j(p_2)_0n_j(p_2)]\cdot[kf_k(p_3)_0n_k(p_3)][if_l(p_4)_0n_l(p_4)] \\ \times|A_{ijkl}(qp_2p_3p_4)|^2\cdot[\epsilon_i(q)_0 + \epsilon_j(p_2)_0 + \epsilon_k(p_3)_0 + \epsilon_l(p_4)_0]^{-2} \\ - \xi\alpha_-(q)_0[1+\alpha_-(q)_0]^{-2}[2\epsilon_+(q)_0]^{-1} \\ \times\Big(x\Omega\sum_{rjk=\pm} r\sum_{p_2p_3}[jf_j(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)] \\ \times A_{rjk}^{(T)}(qp_2p_3)A_{-rjk}(qp_2p_3) \\ \times[\epsilon_r(q)_0 + \epsilon_j(p_2)_0 + \epsilon_k(p_3)_0]^{-1} \\ + \frac{1}{3}\sum_{rjkl=\pm}\sum_{p_2p_3p_4} r[jf_j(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)] \\ \times[lf_l(p_4)_0n_l(p_4)]A_{rjkl}^{(T)}(qp_2p_3p_4)A_{-rjkl}(qp_2p_3p_4) \\ \times[\epsilon_r(q)_0 + \epsilon_j(p_2)_0 + \epsilon_k(p_3)_0 + \epsilon_l(p_4)_0]^{-1}\Big).\end{aligned} \tag{2.21}$$

Here $\epsilon_+(q)_0$ is the quasiparticle energy (I. 2. 17) obtained by diagonalizing the pair-Hamiltonian model. The (real) quantities $\alpha_{-i}(p)_0$, $f_i(p)_0$, $A_{ijk}(p_1p_2p_3)$, and $A_{ijkl}(p_1p_2p_3p_4)$, with $\sum_i p_i = 0$ in the A functions, are given by Eqs. (I. 2. 20), (I. 2. 21), (I. C. 12), and (I. C. 14), respectively. Finally, the quantities $n_\pm(p)$ in Eq. (2. 21) are the eigenvalues of the quasiparticle occupation-number operators $\xi_p^+\xi_p$, $-\xi_p\xi_p^+$, which were introduced via the Bogoliubov transformation [see Eqs. (I. B13) and (I. 2. 15)]. Thus

$$\begin{aligned}n_+(p) = \langle\xi_p^+\xi_p\rangle_0, \\ n_-(p) = -\langle\xi_p\xi_p^+\rangle = -[1+n_+(p)].\end{aligned} \tag{2.22}$$

Strictly speaking, we must set $n_+ = 0$ and $n_- = -1$ in Eqs. (2. 20)–(2. 22), since we consider only the limit $T = 0$ °K in this paper. However, we shall continue to retain the functional dependence of various quantities on $n_\pm(p)$, because the $T \neq 0$ °K results for correlation functions are obtained by setting $n_+ = [\exp(\beta\epsilon_+) - 1]^{-1}$ and $n_- = -[1+n_+]$, where $\beta = (\kappa T)^{-1}$. This procedure is justified in Appendices A and B of I.

The calculation of $F_{20}(r)$ and $F_{02}(r)$ is facilitated by defining the quantity

$$\eta(p) = \langle\Psi|a_p^+a_{-p}^+ + a_pa_{-p}|\Psi\rangle\langle\Psi|\Psi\rangle^{-1}. \tag{2.23}$$

We can then rewrite $F_{20}(r)$ and $F_{02}(r)$ in terms of $\eta(p)$, for Hermitian $V^{(s)}$, as

$$F_{20}(r) = F_{02}(r) = (2n\Omega)^{-1}\sum_p e^{ip\cdot r}\eta(p). \tag{2.24}$$

It is easy to show that the quantity $\eta(p)$ is related to

$B(p)$ of Eq. (2.13) by

$$\eta(p)=B(p)-2\langle n(p)\rangle . \tag{2.25}$$

Upon substituting Eqs. (2.20) and (I.5.10) into (2.25) and simplifying, we obtain the following expression for $\eta(p)$:

$$\eta(p)=-2f_+(p)\alpha_-(p)[1+2n_+(p)]+A_{20}(p), \tag{2.26}$$

$$\begin{aligned}A_{20}(p)=&2x\Omega f_+(p)_0\alpha_-(p)_0\sum_{p_2p_3}\sum_{ijk}[f_i(p)_0n_i(p)][jf_j(p_2)_0n_j(p_2)]\\&\times[kf_k(p_3)_0n_k(p_3)]|A_{ijk}(pp_2p_3)|^2[\epsilon_i(p)_0+\epsilon_j(p_2)_0\\&+\epsilon_k(p_3)_0]^{-2}+\tfrac{2}{3}f_+(p)_0\alpha_-(p)_0\sum_{p_2p_3p_4}\sum_{ijkl}[f_i(p)_0n_i(p)]\\&\times[jf_j(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)][lf_l(p_4)_0n_l(p_4)]\\&\times|A_{ijkl}(pp_2p_3p_4)|^2[\epsilon_i(p)_0+\epsilon_j(p_2)_0+\epsilon_k(p_3)_0+\epsilon_l(p_4)_0]^{-2}\\&+\tfrac{1}{2}x\Omega f_+(p)_0\alpha_-(p)_0[f_+(p)_0+f_-(p)_0]\\&\times\sum_{p_2p_3}\sum_{ijk}[jf_j(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)]\\&\times A^{(T)}_{ijk}(pp_2p_3)A_{-ijk}(pp_2p_3)[\epsilon_i(p)_0+\epsilon_j(p_2)_0+\epsilon_k(p_3)_0]^{-1}\\&\times[\epsilon_j(p_2)_0+\epsilon_k(p_3)_0-\epsilon_i(p)_0]^{-1}\\&+\tfrac{1}{6}f_+(p)_0\alpha_-(p)_0[f_+(p)_0+f_-(p)_0]\\&\times\sum_{p_2p_3p_4}\sum_{jkl}[jf_j(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)][lf_l(p_4)_0n_l(p_4)]\\&\times A^{(T)}_{ijkl}(pp_2p_3p_4)\,A_{-ijkl}(pp_2p_3p_4)\\&\times[\epsilon_j(p_2)_0+\epsilon_k(p_3)_0+\epsilon_l(p_4)_0+\epsilon_i(p)_0]^{-1}\cdot[\epsilon_j(p_2)_0+\epsilon_k(p_3)_0\\&+\epsilon_l(p_4)_0-\epsilon_i(p)_0]^{-1}.\end{aligned} \tag{2.27}$$

The quantities $F_{12}(r)$ and $F_{22}(r)$ are given similarly, to $O(V)$ only, by

$$F_{12}(r)=(2\xi n\Omega)^{-1}\sum_q e^{-iq\cdot r}S_{12}(q), \tag{2.28}$$

$$F_{22}(r)=(n\Omega)^{-1}\sum_q e^{-iq\cdot r}S_{22}(q), \tag{2.29}$$

where

$$\begin{aligned}S_{12}(q)=&-4\xi\sum_{p_2p_3}\sum_{ijk}[if_i(q)_0n_i(q)][jf_j(p_2)_0n_j(p_2)]\\&\times[kf_k(p_3)_0n_k(p_3)][\alpha_{-k}(p_3)_0\alpha_{-i}(q)_0-\alpha_{-j}(p_2)_0]\\&\times A_{jki}(p_2p_3-q)[\epsilon_j(p_2)_0+\epsilon_k(p_3)_0+\epsilon_i(q)_0]^{-1},\end{aligned} \tag{2.30}$$

$$\begin{aligned}S_{22}(q)=&2(n\Omega)^{-1}\sum_{p_1p_2p_3p_4}\delta_{q,p_1-p_4}\sum_{ijkl}[if_i(p_1)_0\alpha_{-i}(p_1)_0n_i(p_1)]\\&\times[jf_j(p_2)_0\alpha_{-j}(p_2)_0n_j(p_2)][kf_k(p_3)_0n_k(p_3)]\\&\times[lf_l(p_4)_0n_l(p_4)]A_{ijkl}(p_1p_2p_3p_4)\\&\times[\epsilon_i(p_1)_0+\epsilon_j(p_2)_0+\epsilon_k(p_3)_0+\epsilon_l(p_4)_0]^{-1}.\end{aligned} \tag{2.31}$$

Fortunately, it is not necessary for us to include $O(V^2)$ terms in these last two quantities, to the order of interest required in Sec. 3 for the DHSBG calculation of the structure factor. The quantity $\widetilde{F}_{22}(r)$ can be shown to be $O(V^3)$; hence we shall not calculate this term in this paper. These equations complete the formal calculation of the radial distribution function $P_2(r)$.

The structure factor is obtained by substituting Eqs. (2.10)–(2.31) into (2.9). We then obtain

$$\begin{aligned}S_{in}(q)=&\xi[1+2n_+(q)][1-\alpha_-(q)][1+\alpha_-(q)]^{-1}+S_2(q)\\&+S_{11}(q)+S_{12}(q)+S_{22}(q)+\widetilde{S}_{22}(q)+O(V^2),\end{aligned} \tag{2.32}$$

where $S_{11}(q)$, $S_{12}(q)$, and $S_{22}(q)$ are given by Eqs. (2.21), (2.30), and (2.31), respectively, and

$$S_2(q)=(1-\xi)+(n\Omega)^{-1}\sum_p[\langle n(p)\rangle\langle n(p+q)\rangle+\tfrac{1}{4}\eta(p)\eta(p+q)], \tag{2.33}$$

$$\widetilde{S}_{22}(q)=n\int d^3r\,e^{iq\cdot r}\,\widetilde{F}_{22}(r)=O(V^3). \tag{2.34}$$

3. DHSBG CALCULATION OF STRUCTURE FACTOR

In this section we shall derive an explicit expression for the structure factor for the model system of a dilute hard sphere Bose gas. In Appendix F of I we have outlined in detail how to obtain DHSBG results from corresponding perturbation theory results. Upon applying this procedure to the results of Sec. 2, it is straightforward to deduce the following expression for the structure factor at $T=0\,^\circ$K (which corresponds to $n_+=0$ and $n_-=-1$):

$$\begin{aligned}S_{in}(q)=&\xi[1-\alpha_-(q)][1+\alpha_-(q)]^{-1}+S_2(q)+S_{11}(q)\\&+S_{12}(q)+0(na^3).\end{aligned} \tag{3.1}$$

In obtaining (3.1) from (2.32) we have neglected the quantity $S_{22}(q)$ of (2.31), because it involves summation over three independent momenta. The reason is explained following Eq. (F21b) of I. For the same reason we may immediately neglect $\widetilde{S}_{22}(q)$ and those parts of $S_{11}(q)$, $\alpha_-(q)$, $\langle n(p)\rangle$ and $\eta(p)$ which involve three or more independent momenta.

The quantity ξ in Eq. (3.1), which denotes the fraction of particles in the zero-momentum state, has already been evaluated in I. From (I.F8) we have

$$\xi=1-\tfrac{8}{3}(n\xi^3a^3/\pi)^{1/2}+O(na^3). \tag{3.2}$$

We also need a second-order expression for the quantity $[1-\alpha_-(p)][1+\alpha_-(p)]^{-1}$ and hence for $\alpha_-(p)$. Here we appeal to Eq. (I.D7) for $\delta\alpha_-(p)$. But we must also refer to Eqs. (I.D2) and (I.D4), with $\Delta_{1i}(p)_2\rightarrow\Delta_{1i}(p)_2-\delta g$, because the quantity $\alpha_{-i}(p)_0$ of (I.F6) is only an approximation to the corresponding quantity of (I.2.20). Thus, the first-order energies in Eqs. (I.F19) and (I.F20) also contribute to the expression for $\delta\alpha_-(p)$. Similar considerations apply to the calculation of $\delta\epsilon_+(q)$ in Eq. (3.10) below. With these remarks in mind, and with the aid of other equations from Appendix F of I, we obtain

$$\alpha_-(p)=\alpha_-(p)_0+\delta\alpha_-(p), \tag{3.3}$$

$$\alpha_-(p)_0=1+2y^2-2y(y^2+1)^{1/2}, \tag{3.4}$$

$$\begin{aligned}\delta\alpha_-(p)=&\tfrac{1}{2}W^2\xi[2\epsilon_+(p)_0]^{-1}(n\Omega)^{-1}\sum_{p_2p_3}f_+(p_2)_0f_+(p_3)_0\\&\times Q(pp_2p_3)R(pp_2p_3)[(\epsilon_+(p_2)_0+\epsilon_+(p_3)_0+\epsilon_+(p)_0)^{-1}\\&+(\epsilon_+(p_2)_0+\epsilon_+(p_3)_0-\epsilon_+(p)_0)^{-1}]\\&-[2\epsilon_+(p)_0]^{-1}\tfrac{1}{2}W(n\Omega)^{-1}(1+\alpha_-(p)_0)^2\sum_{p_2}f_+(p_2)_0\alpha_-(p_2)_0\\&+[2\epsilon_+(p)_0]^{-1}\tfrac{1}{8}W^2\xi(n\Omega)^{-1}(1-\alpha_-(p)_0)^2\sum_{p_2}[\omega(p_2)-g_0]^{-1},\end{aligned} \tag{3.5}$$

$$\begin{aligned}Q(p_1p_2p_3)=&[\alpha_-(p_1)_0+\alpha_-(p_2)_0+\alpha_-(p_3)_0\\&-\alpha_-(p_1)_0\alpha_-(p_2)_0-\alpha_-(p_1)_0\alpha_-(p_3)_0\\&-\alpha_-(p_2)_0\alpha_-(p_3)_0],\end{aligned} \tag{3.6}$$

$$R(p_1p_2p_3)=[1-\alpha_-(p_2)_0-\alpha_-(p_3)_0+\alpha_-(p_1)_0\alpha_-(p_2)_0 + \alpha_-(p_1)_0\alpha_-(p_3)_0-\alpha_-(p_1)_0\alpha_-(p_2)_0\alpha_-(p_3)_0], \quad (3.7)$$

where $y=p/p_0$, with $p_0^2=W\xi=16\pi an\xi$, and $p_1+p_2+p_3=0$ in the functions Q and R.

We can now expand the quantity $[1-\alpha_-(p)][1+\alpha_-(p)]^{-1}$ to obtain

$$\xi[1-\alpha_-(p)][1+\alpha_-(p)]^{-1} \approx \xi[1-\alpha_-(p)_0][1+\alpha_-(p)_0]^{-1}-2\xi[1+\alpha_-(p)_0]^{-2}\delta\alpha_-(p) = \xi q^2/\epsilon_+(q)_0-2\xi[1+\alpha_-(p)_0]^{-2}\delta\alpha_- . \quad (3.8)$$

For later convenience we rewrite this last equation as follows:

$$\xi[1-\alpha_-(q)][1+\alpha_-(q)]^{-1} \approx \xi q^2/\epsilon_+(q)+\xi q^2[\delta\epsilon_+(q)/\epsilon_+^2(q)]-2\xi[1+\alpha_-(q)_0]^{-2}\delta\alpha_-(q) = \xi q^2/\epsilon_+(q)+\xi[H_A(q)/\epsilon_+(q)_0]+O(na^3), \quad (3.9)$$

where

$$H_A(q)=[1-\alpha_-(q)_0][1+\alpha_-(q)_0]^{-1}\delta\epsilon_+(q) - 2\epsilon_+(q)_0[1+\alpha_-(q)_0]^{-2}\delta\alpha_-(q). \quad (3.10)$$

After substituting Eqs. (I. 4. 30), (I. 4. 31), and (3. 5) into (3. 10), and recalling the remarks above (3. 3) before simplifying, we obtain for $H_A(q)$ at $T=0\,°\mathrm{K}$ the result

$$H_A(q)=-\tfrac{1}{2}W^2\xi(n\Omega)^{-1}[1+\alpha_-(q)_0]^{-1}\sum_{p_2p_3}f_+(p_2)_0f_+(p_3)_0 \times[1-\alpha(p_2)_0\alpha(p_3)_0]\{Q(qp_2p_3)[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0+\epsilon_+(q)_0]^{-1} + R(qp_2p_3)[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0-\epsilon_+(q)_0]^{-1}\} + W(n\Omega)^{-1}\sum_{p_2}f_+(p_2)_0\alpha_-(p_2)_0. \quad (3.11)$$

The last term in (3. 11) occurs effectively as a subtraction term, as explained following (I. F16).

We now give DHSBG limits for the quantities $S_{11}(q)$, $S_{12}(q)$, and $S_2(q)$ of Eqs. (2. 21), (2. 30), and (2. 33), respectively, following the procedure outlined in Appendix F of I. We find

$$S_{11}(q)=S_{11}(q)_A+S_{11}(q)_B, \quad (3.12)$$

where

$$S_{11}(q)_A=(W\xi)^2[1+\alpha_-(q)_0]^{-2}(n\Omega)^{-1}\sum_{p_2p_3}f_+(p_2)_0f_+(p_3)_0 \times[Q(qp_2p_3)]^2[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0+\epsilon_+(q)_0]^{-2}, \quad (3.13)$$

$$S_{11}(q)_B=(W\xi)^2\alpha_-(q)_0(n\Omega)^{-1}[1+\alpha_-(q)_0]^{-2} \times\sum_{p_2p_3}f_+(p_2)_0f_+(p_3)_0Q(qp_2p_3)R(qp_2p_3) \times[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0+\epsilon_+(q)_0]^{-1} \times[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0-\epsilon_+(q)_0]^{-1}, \quad (3.14)$$

$$S_{12}(q)=-4W\xi(n\Omega)^{-1}[1-\alpha_-^2(q)_0]^{-1}\sum_{p_2p_3}[f_+(p_2)_0f_+(p_3)_0] \times Q(qp_2p_3)[\alpha_-(p_2)_0-\alpha_-(p_3)_0\alpha_-(q)_0] \times[\epsilon_+(p_2)_0+\epsilon_+(p_3)_0+\epsilon_+(q)_0]^{-1}, \quad (3.15)$$

$$S_2(q)=(n\Omega)^{-1}\sum_p\alpha_-(p)_0[\alpha_-(p)_0+\alpha_-(p+q)_0] \times f_+(p)_0f_+(p+q)_0, \quad (3.16)$$

where $W=16\pi an$. These results disagree in part with corresponding results obtained by T. T. Wu.[6]

In the low-momentum limit the quantities, $H_A(q)$, $S_2(q)$, $S_{11}(q)$, and $S_{12}(q)$ can be evaluated approximately. We obtain the limiting expressions

$$H_A(q)=-\tfrac{4}{3}W(n\xi^3a^3/\pi)^{1/2}[y^2+O(y^4)], \quad (3.17)$$

$$S_2(q)=2(n\xi^3a^3/\pi)^{1/2}[\pi-y+O(y^2)], \quad (3.18)$$

$$S_{11}(q)=2(n\xi^3a^3/\pi)^{1/2}[\pi-3y+O(y^2)], \quad (3.19)$$

$$S_{12}(q)=-4(n\xi^3a^3/\pi)^{1/2}[\pi-3y+O(y^2\ln y)]. \quad (3.20)$$

Finally, upon substituting Eqs. (3. 9), (3. 10), and (3. 17)—(3. 20) into (3. 1), and using (3. 2), we obtain the following expression for $S_{in}(q)$ in the low-momentum limit:

$$S_{in}(q)=q^2/W_{\mathrm{exc}}(q) \quad (q\to 0), \quad (3.21)$$

where

$$W_{\mathrm{exc}}(q)=\epsilon_+(q)\{1+O(na^3)+O(y\ln y)\}, \quad (3.22)$$

and $\epsilon_+(q)$ is given by Eq. (I. F25). Thus we have demonstrated, to second order in the DHSBG parameter $(na^3)^{1/2}$, that our expression for the structure factor satisfies the Feynman—Bijl relation (1. 1) with $W_{\mathrm{exc}}(q)=\epsilon_+(q)$. It is important to observe, however, that the formal expressions for $S_{in}(q)$ and $\epsilon_i(q)$, given in Sec. 2 and by Eqs. (I. 4. 23)—(I. 4. 31), will not satisfy the Feynman—Bijl relation in general, but only in the low-momentum limit.

The general validity of the Feynman—Bijl relation in the low-momentum limit has been proved earlier by many authors.[7] Therefore, the explicit verification of this relation in the present paper means that our formal expressions for the structure factor can be used as a consistency check on other calculations of this quantity for a degenerate Bose system.

4. SUMMARY

The radial distribution function and the structure factor have been calculated for a degenerate Bose system to second order in perturbation theory using the pair Hamiltonian model as a starting point. These results will constitute an important check on any realistic calculations of the radial distribution function and the structure factor. The special case of a dilute hard sphere Bose gas has been investigated in detail at $T=0\,°\mathrm{K}$. In the low-momentum limit our explicit expression for the structure factor satisfies the Feynman—Bijl relation.

*Based in part on a thesis submitted by one of us (I.R.R.) to the Department of Physics, University of Colorado, in partial fulfillment of the requirements for the Ph. D. degree (1969).

[1]I. RamaRao and F. Mohling, J. Math. Phys. 12, 1631 (1971); hereafter referred to as I. Equation (n) of this paper will be referred to as (I. n).
[2]T. T. Wu, Phys. Rev. 115, 1390 (1959).
[3]R. P. Feynman, Phys. Rev. 94, 262 (1954).
[4]L. Van Hove and K. W. McVoy, Nucl. Phys. 33, 468 (1962).
[5]D. G. Henshaw, Phys. Rev. 119, 9 (1960).
[6]See Sec. 7C of Reference 2.
[7]See, e.g., K. Huang and A. Klein, Ann. Phys. 30, 203 (1964); and J. Gavoret and P. Nozieres, Ann. Phys. (N.Y.) 28, 349 (1964).

A scalar-tensor theory of gravitation

K. A. Dunn

Department of Mathematics, Dalhousie University, Halifax, Nova Scotia, Canada
(Received 29 November 1973; final revised manuscript received 9 May 1974)

A scalar–tensor theory of gravitation is constructed using a non-Riemannian geometry in which both the metric tensor and the scalar function have an unambiguous geometric interpretation. The scalar function is introduced by defining a linear connection with nonvanishing torsion. The field equations of the theory, and the Lagrangian from which they are derived, are identical to those given by Dicke in an alternate formulation of the Brans–Dicke theory. By using the static spherically symmetric solution to the field equations it is found that, with a proper choice of parameter, this theory agrees with experimental results in the three classical tests of a gravitational theory.

I. INTRODUCTION

Probably the major difference between the general theory of relativity and the classical theory of gravitation is the geometrization of the gravitational field. More precisely, space—time is described as a four-dimensional Riemannian manifold in which the components g_{ij} of the Riemannian metric completely describe the gravitational field.

Motivated by ideas of Mach, Brans and Dicke[1] introduced an alternate theory of gravitation involving a scalar function as well as the metric tensor. This scalar—tensor theory is not purely geometrical however, as the scalar field is introduced in a rather ad hoc manner into the Riemannian manifold.

Several attempts have been made to cast a scalar—tensor theory of gravitation in a wider geometrical context. Brans and Dicke[1] observed in their work the formal connection between their theory and that of Jordan[2] which uses a five-dimensional manifold. Peters[3] has shown that the scalar field of Dicke can be geometrized using the techniques of geometrodynamics[4]; unfortunately, this method restricts consideration only to source-free regions of space—time. Ross[5] has constructed a scalar—tensor theory of gravitation using the Weyl formulation of Riemannian geometry, and Dunn and Sen[6] have introduced a scalar—tensor theory modeled on a modification of Riemannian geometry suggested by Lyra.

In the present work we introduce a geometry which differs from the usual Riemannian geometry in that its linear connection has nonvanishing torsion defined in terms of a scalar function. In this way both the metric tensor and the scalar field have a well-defined geometric meaning in the spirit of general relativity. In Sec. II, we define the geometry and describe some of its relevant properties. In Sec. III, we formulate a scalar—tensor theory of gravitation whose field equations are identical to those given by Dicke[7] in an alternate presentation of the Brans—Dicke theory. In Sec. IV, the static spherically symmetric solution to the field equations is found and used to compute the values of red shift, deflection of light, and perihelion advance.

II. DEFINITION OF THE GEOMETRY

Let M be a four-dimensional C^∞ manifold; $\chi(M)$ denotes the Lie algebra of C^∞ vector fields on M and $C^\infty(M)$ the ring of C^∞ functions on M. We suppose that M is endowed with a nonsingular metric, that is, a second-order symmetric covariant tensor field g such that, at every point $m \in M$, the induced form g_m on the tangent space at m is nondegenerate and is given in a local coordinate system by components g_{ij}.

We also require a linear connection ∇ on M which allows one to compare vectors at different points of M. Recall[8] that a linear connection ∇ is a mapping $\nabla: \chi(M)\times\chi(M)\to\chi(M)$, usually written $(X, Y)\to\nabla_X Y$ which satisfies

(i) $\nabla_X(Y+Z)=\nabla_X Y+\nabla_X Z$

(ii) $\nabla_{fX+gY} Z=f\nabla_X Z+g\nabla_Y Z$

(iii) $\nabla_X fY=X(f)\,Y+f\nabla_X Y$

where $f,\ g\in C^\infty(M)$ and $X, Y, Z\in\chi(M)$.

Choosing a local coordinate system (x^i) with basis vectors $\{e_i=\partial/\partial x^i\}$, the components of the connection are given by $\nabla_{e_i} e_j=\Gamma^k_{ji} e_k$. The torsion of ∇ is the mapping $TOR_\nabla(X, Y)=\nabla_X Y-\nabla_Y X-[X, Y]$ and is given in a local coordinate system by $TOR_\nabla(e_i, e_j) = (\Gamma^k_{ji}-\Gamma^k_{ij})e_k$.

The linear connection ∇ enables one to define a general covariant derivative of any tensor field on M with respect to a vector field X. This derivative preserves the tensor type and in particular the covariant derivative of the metric tensor $\nabla_X g$ is again a second-order covariant tensor field defined by

$$\nabla_X g(Y, Z)=X(g(Y, Z))-g(\nabla_X Y, Z)-g(Y, \nabla_X Z), \quad X, Y, Z\in\chi(M). \tag{2.1}$$

By a geometry we will mean a manifold M endowed with a metric g and a connection ∇. The geometry of interest for the scalar—tensor theory is given by the following.

Proposition 1: Given a metric g and a scalar function $\lambda\in C^\infty(M)$ which vanishes nowhere on M, there exists a unique connection ∇ on M satisfying

$$\nabla_Z g(X, Y)=0, \tag{2.2}$$

$$TOR_\nabla(X, Y)=(k/\lambda)\,X(\lambda)\,Y-(k/\lambda)\,Y(\lambda)\,X \tag{2.3}$$

for all $X, Y, Z\in\chi(M)$, k a constant.

Proof[9]: Notice that if $k=0$ or λ is a constant the connection is metric preserving and torsion-free; i.e., (2.2) and (2.3) reduce to the usual definition of the Riemannian connection. The proof of the general case follows by expanding (2.2) and (2.3) in a local coordinate system using (2.1) and exhibiting the unique connection components

$$\Gamma^r_{ij} = \left\{ {r \atop ij} \right\} + (k/\lambda)\, g_{ij}\lambda^{,r} - (k/\lambda)\,\delta^r_j\,\lambda_{,i} \tag{2.4}$$

where $\left\{ {r \atop ij} \right\}$ are the Christoffel symbols of the metric g_{ij}.

We call a geometry defined by a metric g and the unique linear connection (2.4) a scalar—tensor geometry. Such a geometry is completely determined given the metric, the scalar function λ, and the constant k.

The curvature tensor for a connection ∇ is defined by

$$K(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, \qquad X, Y, Z \in \chi(M). \tag{2.5}$$

The curvature tensor of the scalar—tensor geometry is given in a local coordinate system by

$$\begin{aligned} K^s_{lij} = R^s_{lij} &+ k\left(g_{lj}\frac{\lambda^{,s}}{\lambda}\right)_{;i} - k\left(g_{li}\frac{\lambda^{,s}}{\lambda}\right)_{;j} - k\left(\delta^s_j\frac{\lambda_{,l}}{\lambda}\right)_{;i} + k\left(\delta^s_i\frac{\lambda_{,l}}{\lambda}\right)_{;j} \\ &+ k^2 g_{lj}g_{ri}\frac{\lambda^{,r}\lambda^{,s}}{\lambda^2} - k^2 g_{li}g_{rj}\frac{\lambda^{,r}\lambda^{,s}}{\lambda^2} \\ &- k^2 g_{lj}\delta^s_i\frac{\lambda_{,r}\lambda^{,r}}{\lambda^2} + k^2 g_{li}\delta^s_j\frac{\lambda_{,r}\lambda^{,r}}{\lambda^2} \\ &+ k^2\,\delta^s_i\frac{\lambda_{,l}\lambda_{,j}}{\lambda^2} - k^2\,\delta^s_j\frac{\lambda_{,l}\lambda_{,i}}{\lambda^2} \end{aligned} \tag{2.6}$$

where R^s_{lij} is the Riemannian curvature tensor defined by the Christoffel symbols of the metric g and a semicolon denotes covariant differentiation with respect to these Christoffel symbols.

Proposition 2: If the tensor $\bar{K}$ of covariant degree 4 is defined by $\bar{K}(X,Y,Z,W) = g(X, K(Z,W)Y)$; $X, Y, Z, W \in \chi(M)$, then the following relations are true for the scalar—tensor curvature tensor:

(a) $K(X,Y)Z + K(Z,X)Y + K(Y,Z)X = 0$,

(b) $\bar{K}(X,Y,Z,W) = -\bar{K}(Y,X,Z,W)$,

(c) $\bar{K}(X,Y,Z,W) = -\bar{K}(X,Y,W,Z)$,

(d) $\bar{K}(X,Y,Z,W) = \bar{K}(Z,W,X,Y)$,

i.e., the curvature tensor K^r_{lij} has exactly the same algebraic symmetries as the Riemannian curvature tensor R^r_{lij}.

Proof: For (a), use the Jacobi identity, Eq. (2.3) and compute. (b), (c), and (d) follow as in the Riemannian case.[8]

Because of the symmetries of the curvature tensor K^r_{lij}, we can contract to form a unique (up to sign) symmetric covariant tensor of degree two, $K_{li} = K^s_{lsi}$, and hence a unique curvature scalar

$$K = g^{li}K_{li} = R - 6k^2\frac{\lambda_{,r}\lambda^{,r}}{\lambda^2} + \frac{6k}{\sqrt{-g}}\frac{\partial}{\partial x^s}(\log\lambda)^{,s} \tag{2.7}$$

where R is the curvature scalar defined by the metric g and we assume $\lambda > 0$.

The geodesic equations for the scalar—tensor geometry, i.e., those curves whose tangent vectors X satisfy $\nabla_X X = 0$, are given in a local coordinate system by

$$\frac{d^2x^i}{ds^2} + \left\{ {i \atop jk} \right\}\frac{dx^j}{ds}\frac{dx^k}{ds} + k\frac{\lambda^{,i}}{\lambda} = g_{rs}\frac{dx^r}{ds}\frac{dx^s}{ds} - \frac{k}{\lambda}\lambda_{,j}\frac{dx^j}{ds}\frac{dx^i}{ds} = 0. \tag{2.8}$$

For timelike geodesics we may choose the parameter s such that $g_{ij}\,dx^i/ds\; dx^j/ds = 1$ along the trajectory. For null geodesics, the reparametrization $t = \int_0^s \lambda^k(s)\,ds$ reduces (2.8) to the form

$$\frac{d^2x^i}{dt^2} + \left\{ {i \atop jk} \right\}\frac{dx^j}{dt}\frac{dx^k}{dt} = 0 \tag{2.9}$$

which are the same trajectories as the null geodesics of the Riemannian geometry determined by g.

III. A SCALAR TENSOR THEORY OF GRAVITATION

We choose the scalar—tensor geometry determined by metric g, scalar λ, and constant k as the geometric framework for a scalar—tensor theory of gravitation.

The vacuum field equations of general relativity can be obtained from the variational principle

$$\delta\int R\sqrt{-g}\,d^4x = 0 \tag{3.1}$$

where R is the curvature scalar of the metric g. In our scalar—tensor geometry, this variational principle becomes

$$\delta\int K\sqrt{-g}\,d^4x = 0 \tag{3.2}$$

where K is given by (2.7). On variation of (3.2) with respect to g_{ij} and λ, the vacuum field equations are

$$R_{ij} - \tfrac{1}{2}g_{ij}R = 6k^2\left(\frac{\lambda_{,i}\lambda_{,j}}{\lambda^2} - \tfrac{1}{2}g_{ij}\frac{\lambda_{,r}\lambda^{,r}}{\lambda^2}\right), \tag{3.3}$$

$$\frac{\partial}{\partial x^s}(\lambda_{,j}\sqrt{-g}\,g^{is}) - \frac{\lambda_{,s}\lambda_{,i}}{\lambda}g^{si}\sqrt{-g} = 0. \tag{3.4}$$

Several things should be noted about these equations. First, the variation principle (3.2) and hence the field equations (3.3) and (3.4) are identical in the vacuum case to those given by Dicke[7] in the second formulation of the Brans—Dicke theory. Second, although the curvature tensor (2.6) of the scalar—tensor geometry does not satisfy the second Bianchi identity,[8] the invariance of the integral (3.2) under arbitrary coordinate transformations yields conservation laws as in the relativistic case.[10] Finally, we have considered here only regions of space—time with zero charge and mass densities. As in general relativity, we can generalize our Lagrangian to include such terms. The precise form of this matter Lagrangian goes beyond the geometric considerations of this work, however, and we shall limit ourselves to regions where charge and mass densities vanish.

Consider now the solution to the field equations (3.3) and (3.4) for the static spherically symmetric field about a point mass. We express the line element in isotropic form

$$ds^2 = e^{2\alpha}dt^2 - e^{2\beta}[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] \tag{3.5}$$

where α and β are functions of r only. The general vacuum solution for $k \neq 0$ is given by

$$g_{00} = e^{2\alpha} = \phi_0 e^{2\alpha_0} \left[\frac{1 - B/r}{1 + B/r}\right]^{(2+C)/p},$$

$$-g_{11} = e^{2\beta} = \phi_0 e^{2\beta_0} (1 + B/r)^4 \left[\frac{1 - B/r}{1 + B/r}\right]^{(2p-C-2)/p} \tag{3.6}$$

$$\lambda = \phi_0 \left[\frac{1 - B/r}{1 + B/r}\right]^{C/p} \tag{3.7}$$

where

$$p = [(3k^2 + \tfrac{1}{4}) C^2 + C + 1]^{1/2} \tag{3.8}$$

and α_0, β_0, ϕ_0, B, and C are arbitrary constants. This solution is also valid for $k = 0$ if we restrict the possible values of the constant C to $C \neq -2$. In this case the geometry defined by (2.2) and (2.3) is Riemannian and the metric (3.6) is the Schwarzschild solution of general relativity.

IV. TESTS OF THE THEORY

In order to compare the theoretical predictions of the scalar—tensor theory with experimental results we must specify the arbitrary constants in (3.6). We assume, therefore, that the solution (3.6) is asymptotically flat; i.e., as $r \to \infty$, $g_{ij} \to \eta_{ij}$, where η_{ij} is the Minkowski metric, and also that the weak-field limit of the timelike geodesics (2.8) correspond to Newtonian theory[11] for a single central pointmass M. This second assumption yields

$$1 - 2GM/rc^2 = g_{00} + \ln\lambda^{2k} \tag{4.1}$$

where G is the gravitational constant. These two conditions specify the constants as follows:

$$\phi_0 = 1, \quad \alpha_0 = \beta_0 = 0,$$
$$2B\left(\frac{2+C}{p}\right) + \frac{4kCB}{p} = \frac{2GM}{c^2}. \tag{4.2}$$

We can now examine the predictions of the scalar—tensor theory with respect to the three classical tests.

The gravitational red shift is determined by g_{00} to first order in $1/r$. To obtain agreement with experimental results, using (3.6), we must have

$$B[(2 + C)/p] = G'M/c^2 \tag{4.3}$$

where G' is the gravitational constant measured experimentally.

The deflection of light is determined, not by g_{00} alone, but from the ratio g_{11}/g_{00}. It is easily shown using first-order terms in (3.6) that the light deflection computed from this theory is

$$\delta\theta = (G'/G) \times (\text{general relativity result}). \tag{4.4}$$

Finally, the advance of the perihelion of a planetary orbit requires g_{00} to second order in $1/r$ and g_{11} to first order. The result of this calculation is that the perihelion rotation rate of a planetary orbit is

$$[\tfrac{2}{3}(G'/G) + \tfrac{1}{3}] \times (\text{general relativity result}). \tag{4.5}$$

We have said nothing concerning the possible values of the constant k introduced in the definition of the scalar—tensor geometry. Combining (4.2) and (4.3), we see that k and C must satisfy

$$1 + 2kC/(2 + C) = G/G'. \tag{4.6}$$

For $G \neq G'$, (4.6) gives C as a well-defined function of k as long as $k \neq \frac{1}{2}(G/G' - 1)$. For $G = G'$ we must have $k = 0$ or $C = 0$, both of which imply that the solution (3.6) reduces to the Schwarzschild solution of general relativity. No other restrictions are placed on k by the assumptions made in this work.

V. CONCLUSION

In this work we have introduced a scalar—tensor theory of gravitation in which both the metric tensor and the scalar function have an unambiguous geometric interpretation. The vacuum field equations, and the Lagrangian from which they are derived, are identical to those presented by Dicke in an alternate formulation of the Brans—Dicke theory. These two theories are not the same however. For example, the geodesics (2.8) differ from the equations of motion of test particles in Dicke's theory, and also the scalar functions enter the two theories in quite different manners.

A viable theory of gravitation is one which satisfies three criteria[11]: self-consistency, completeness, and agreement with past experiment. The scalar—tensor theory as presented here cannot be considered a viable gravitational model. For one reason, it is not complete since we have given no rules to specify the matter Lagrangian; furthermore, the introduction of torsion could have observational consequences in other physical applications (only gravitation has been considered here). However, the three classical tests of red shift, light deflection, and perihelian advance can be accommodated by the adjustment of the parameter k and hence the theory is worthy of further examination.

*Supported in part by the National Research Council of Canada.

[1]C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).

[2]P. Jordan, *Schwerkraft und Weltall* (Vieweg and Sohn, Braunschweig, 1955).

[3]P.C. Peters, J. Math. Phys. **10**, 1029 (1969).

[4]J.A. Wheeler, *Geometrodynamics* (Academic, New York 1962).

[5]D.K. Ross, Phys. Rev. D **5**, 284 (1972).

[6]D.K. Sen and K.A. Dunn, J. Math. Phys. **12**, 578 (1971).

[7]R.H. Dicke, Phys. Rev. **125**, 2163 (1962).

[8]N.J. Hicks, *Notes on Differential Geomtry* (Van Nostrand, Princeton, N.J., 1965).

[9]This is a special case of a more general result on connections stated in Hicks (p. 104). A proof can be found in D.K. Sen and J.R. Vanstone, J. Math. Phys. **13**, 990 (1970).

[10]A. Schild, *Relativity Theory Astrophysics*, American Mathematical Society, Lectures in Applied Mathematics (AMS, Providence, R.I.), Vol. 8.

[11]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

The uniqueness of the Bertotti–Robinson electromagnetic universe*

N. Tariq and B. O. J. Tupper

Department of Mathematics, University of New Brunswick, Fredericton, N. B., Canada
(Received 16 November 1973; revised manuscript received 27 March 1974)

By employing the spin coefficient formalism of Newman and Penrose a direct proof is obtained that the Bertotti–Robinson electromagnetic universe is the only conformally flat solution of the source-free Einstein–Maxwell equations for nonnull fields.

1. INTRODUCTION

Bertotti[1] and Robinson[2] have found a solution of the Einstein–Maxwell equations in the absence of sources. The space–time solution, which can be written in the form

$$ds^2=(e^2/r^2)(c^2\,dt-dr^2-r^2\,d\theta^2-r^2\sin^2\theta\,d\psi^2), \tag{1.1}$$

where e is a constant, is conformally flat and the electromagnetic field is nonnull. This solution was also found by Lovelock[3] and the interpretation of the solution has been discussed by Dolan.[4]

Cahen and Leroy[5] have claimed that the general conformally flat solution of the Einstein–Maxwell equations for nonnull fields is of the form

$$ds^2=Kr^2\,du^2+2du\,dr-2(1+\tfrac{1}{2}Kz\bar{z})^{-2}dz\,d\bar{z}, \tag{1.2}$$

where z is a complex coordinate and K is the constant curvature of the surfaces $u=r=\text{const}$. When $K>0$, this solution is precisely (1.1). The case when $K<0$, which Cahen and Leroy apparently regarded as a possibility, can be dismissed on the grounds that the energy condition is violated, as we show in the Appendix to this article. The proof given by Cahen and Leroy that (1.2), and hence (1.1), is the general solution to the problem is not satisfactory since they apply a limiting process to type N solutions of the Einstein–Maxwell equations and appear to neglect the possibility of conformally flat solutions being derived by applying a similar limiting process to type D solutions; the existence of this possibility is clear from the Penrose diagram.[6] Furthermore, it has not been proved that *all* conformally flat solutions are necessarily obtained by the limiting process.

In this article we give a direct proof that (1.1) is indeed the unique conformally flat solution of the source-free Einstein–Maxwell equations for a nonnull electromagnetic field. The method of solution is via the spin coefficient formalism of Newman and Penrose,[7] and we lean on the calculations described by Newman, Tamburino, and Unti[8] and Kinnersley.[9] This method has the advantage of automatically eliminating the case $K<0$ in (1.2) and leads only to the positive curvature form (1.1).

2. NOTATION AND EQUATIONS

A tetrad system of null vectors $(l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$, where l^μ, n^μ are real and m^μ, $\bar{m}^\mu$ are complex conjugate vectors, is defined by the relations

$$l_\mu n^\mu=-m_\mu\bar{m}^\mu=1$$

with all other contractions zero.

If $F_{\mu\nu}$ is the electromagnetic tensor then the three "Maxwell scalars" are defined by

$$\phi_0\equiv F_{\mu\nu}l^\mu m^\nu,\quad \phi_2\equiv F_{\mu\nu}\bar{m}^\mu n^\nu,$$

$$\phi_1\equiv\tfrac{1}{2}F_{\mu\nu}(l^\mu n^\nu+\bar{m}^\mu m^\nu)$$

and for a nonnull field the tetrad can be chosen[10] so that $\phi_0=\phi_2=0$, $\phi_1\equiv\phi\neq 0$. In this case l^μ, n^μ are the principal null vectors of the electromagnetic field.

Dolan[4] showed that when the condition for conformal flatness is used, i.e., the vanishing of the five Weyl scalars, eight of the spin coefficients vanish. The four remaining nonzero spin coefficients are

$$\epsilon=\tfrac{1}{2}(l_{\mu;\nu}n^\mu l^\nu-m_{\mu;\nu}\bar{m}^\mu l^\nu),$$
$$\gamma=\tfrac{1}{2}(l_{\mu;\nu}n^\mu n^\nu-m_{\mu;\nu}\bar{m}^\mu n^\nu),$$
$$\alpha=\tfrac{1}{2}(l_{\mu;\nu}n^\mu\bar{m}^\nu-m_{\mu;\nu}\bar{m}^\mu\bar{m}^\nu),$$
$$\beta=\tfrac{1}{2}(l_{\mu;\nu}n^\mu m^\nu-m_{\mu;\nu}\bar{m}^\mu m^\nu).$$

As a result it is found that ϕ is a constant since the Maxwell equations become

$$D\phi=\Delta\phi=\delta\phi=\bar{\delta}\phi=0,$$

where D, Δ, δ, $\bar{\delta}$ are differential operators defined by

$$D\phi=\phi_{;\mu}l^\mu,\quad \Delta\phi=\phi_{;\mu}n^\mu,$$
$$\delta\phi=\phi_{;\mu}m^\mu,\quad \bar{\delta}\phi=\phi_{;\mu}\bar{m}^\mu.$$

By a suitable choice of units the Einstein–Maxwell field equations may be written in the form

$$\Phi_{AB}=\phi_A\bar{\phi}_B,$$

where Φ_{AB} are the complex tetrad components of the Ricci tensor and A, B take the values 0, 1, 2. In the case under consideration here it follows that the only nonzero component of Φ_{AB} is $\Phi_{11}=\phi\bar{\phi}$.

The nontrivial Newman-Penrose equations are

$$\begin{aligned}
&D\alpha-\bar{\delta}\epsilon=(\bar{\epsilon}-2\epsilon)\alpha-\bar{\beta}\epsilon,\\
&D\beta-\delta\epsilon=-\bar{\epsilon}\beta-\bar{\alpha}\epsilon,\\
&D\gamma-\Delta\epsilon=-(\epsilon+\bar{\epsilon})\gamma-(\gamma+\bar{\gamma})\epsilon+\phi\bar{\phi},\\
&\delta\alpha-\bar{\delta}\beta=\alpha\bar{\alpha}+\beta\bar{\beta}-2\alpha\beta+\phi\bar{\phi},\\
&\delta\gamma-\Delta\beta=-(\bar{\alpha}+\beta)\gamma-\beta(\gamma-\bar{\gamma}),\\
&\Delta\alpha-\bar{\delta}\gamma=\alpha\bar{\gamma}+\bar{\beta}\gamma.
\end{aligned} \tag{2.1}$$

3. SIMPLIFICATION OF THE EQUATIONS

In order to preserve l^μ, n^μ as the principal null vectors of the electromagnetic field, the available tetrad freedom is confined to rotations of the form

$$l^\mu \to R l^\mu,$$
$$n^\mu \to R^{-1} n^\mu, \qquad (3.1)$$
$$m^\mu \to e^{iS} m^\mu,$$

where $R>0$, S are real functions. By a combination of these rotations we can choose either $\epsilon=0$ or $\gamma=0$ but not both. We will choose $\epsilon=0$, and this is preserved under the rotations (3.1) provided that $DR=DS=0$. We may also choose $\bar\alpha+\beta=\tau=0$, and this is preserved if $\delta R=0$.

Following Newman and Penrose,[7] we choose coordinates such that $l^\mu=\delta_2{}^\mu$ and $x^2=r$ is an affine parameter along l^μ. The tetrad components are

$$l^\mu=(0,\ 1,\ 0,\ 0),$$
$$n^\mu=(1,\ U,\ X^3,\ X^4),$$
$$m^\mu=(0,\ \omega,\ \xi^3,\ \xi^4)$$

so that, writing $x^1=u$, the differential operators are

$$D=\frac{\partial}{\partial r},$$
$$\Delta=U\frac{\partial}{\partial r}+\frac{\partial}{\partial u}+X^i\frac{\partial}{\partial x^i},$$
$$\delta=\omega\frac{\partial}{\partial r}+\xi^i\frac{\partial}{\partial x^i},$$

where $i=3,\ 4$.

Equations (2.1) now take the form

$$D\alpha=0, \qquad (3.2a)$$
$$D\gamma=\phi\bar\phi, \qquad (3.2b)$$
$$\delta\alpha+\bar\delta\bar\alpha=4\alpha\bar\alpha+\phi\bar\phi, \qquad (3.2c)$$
$$\delta\gamma+\Delta\bar\alpha=\bar\alpha(\gamma-\bar\gamma), \qquad (3.2d)$$
$$\delta(\gamma+\bar\gamma)=0, \qquad (3.2e)$$

and, in addition, we obtain the following equations for the quantities $U,\ \omega,\ X^i,\ \xi^i$:

$$DU=-(\gamma+\bar\gamma), \qquad (3.3a)$$
$$D\omega=0, \qquad (3.3b)$$
$$DX^i=0, \qquad (3.3c)$$
$$D\xi^i=0, \qquad (3.3d)$$
$$\delta U-\Delta\omega=-(\gamma-\bar\gamma)\omega, \qquad (3.3e)$$
$$\delta X^i-\Delta\xi^i=-(\gamma-\bar\gamma)\xi^i, \qquad (3.3f)$$
$$\bar\delta\omega-\delta\bar\omega=2\alpha\omega-2\bar\alpha\bar\omega, \qquad (3.3g)$$
$$\bar\delta\xi^i-\delta\bar\xi^i=2\alpha\xi^i-2\bar\alpha\bar\xi^i. \qquad (3.3h)$$

The commutation relations, which are

$$\Delta D-D\Delta=(\gamma+\bar\gamma)D,$$
$$\delta D-D\delta=0,$$
$$\delta\Delta-\Delta\delta=-(\gamma-\bar\gamma)\delta,$$
$$\delta\bar\delta-\bar\delta\delta=-2\alpha\delta+2\bar\alpha\bar\delta,$$

give no further information.

Since ϕ is a constant, we put $\phi\bar\phi=\frac12 e^{-2}$, where e is a nonzero real constant, and integrate the radial equations (3.2a,b) and (3.3a, b, c) to obtain

$$\alpha=\alpha^\circ, \qquad (3.4a)$$
$$\gamma=\gamma^\circ+\tfrac12 e^{-2}r, \qquad (3.4b)$$
$$U=U^\circ-(\gamma^\circ+\bar\gamma^\circ)r-\tfrac12 e^{-2}r^2, \qquad (3.4c)$$
$$\omega=\omega^\circ, \qquad (3.4d)$$
$$X^i=X^{\circ i}, \qquad (3.4e)$$
$$\xi^i=\xi^{\circ i}, \qquad (3.4f)$$

where the superscript $^\circ$ indicates independence of r.

From (3.2e) and (3.4c) we calculate

$$\delta U=\xi^{\circ i}U^\circ{}_{,i}-\omega^\circ(\gamma^\circ+\bar\gamma^\circ)+2e^{-2}r\omega^\circ,$$

but from (3.3e), (3.4b, d) we see that δU is independent of r. It follows that $\omega^\circ=0$ and the nonradial equations become

$$\xi^{\circ i}\alpha^\circ{}_{,i}+\bar\xi^{\circ i}\bar\alpha^\circ{}_{,i}=4\alpha^\circ\bar\alpha^\circ+\tfrac12 e^{-2}, \qquad (3.5a)$$
$$\xi^{\circ i}\gamma^\circ{}_{,i}+\bar\alpha^\circ{}_{,1}+X^{\circ i}\bar\alpha^\circ{}_{,i}=\bar\alpha^\circ(\gamma^\circ-\bar\gamma^\circ), \qquad (3.5b)$$
$$\xi^{\circ i}(\gamma^\circ+\bar\gamma^\circ)_{,i}=0, \qquad (3.5c)$$
$$\xi^{\circ i}U^\circ{}_{,i}=0, \qquad (3.5d)$$
$$\xi^{\circ j}X^{\circ i}{}_{,j}-\xi^{\circ i}{}_{,1}-X^{\circ j}\xi^{\circ i}{}_{,j}=-(\gamma^\circ-\bar\gamma^\circ)\xi^{\circ i}, \qquad (3.5e)$$
$$\bar\xi^{\circ j}\xi^{\circ i}{}_{,j}-\xi^{\circ j}\bar\xi^{\circ i}{}_{,j}=2\alpha^\circ\xi^{\circ i}-2\bar\alpha^\circ\bar\xi^{\circ i}. \qquad (3.5f)$$

Following an argument used by Kinnersley,[9] we can eliminate U° and $\gamma^\circ+\bar\gamma^\circ$ by means of a combined coordinate transformation and tetrad rotation of the form

$$l^\mu\to R(x^1)l^\mu,$$
$$n^\mu\to R^{-1}(x^1)n^\mu,$$
$$x^1\to\int_0^{x^1}R(v)\,dv,$$
$$r\to rR^{-1}(x^1)+U^\circ f(x^1).$$

Since U° and $\bar\gamma^\circ$ are, from (3.5c, d), arbitrary functions of x^1, we can find functions R, f such that the new U° and $\gamma^\circ+\bar\gamma^\circ$ are zero provided that sufficient assumptions of continuity are made. As a result of this we have

$$U=-\tfrac12 e^{-2}r^2. \qquad (3.6)$$

Using the fact that $\gamma^\circ+\bar\gamma^\circ=0$ a tetrad rotation of the form

$$m^\mu\to e^{iS}m^\mu$$

can be used to set

$$\gamma^\circ=0 \qquad (3.7)$$

since both S and γ° are independent of r. In order to preserve (3.7), future rotations must satisfy $\Delta S=0$.

The coordinate transformations

$$x^3\to\theta^3(x^1,\ x^3,\ x^4),$$
$$x^4\to\theta^4(x^1,\ x^3,\ x^4)$$

can be used to set

$$X^{\circ 3}=X^{\circ 4}=0, \qquad (3.8)$$

and, from Eqs. (3.5b, e) this implies that

$$\xi^{oi}{}_{,1}=0, \tag{3.9a}$$

$$\alpha^{\circ}{}_{,1}=0. \tag{3.9b}$$

As a result of Eq. (3.8) the requirement that future rotations must satisfy $DS=\Delta S=0$ implies that S must be independent of x^1 and r. Since α° and ξ^{o3} are both functions of x^3 and x^4 only, we can now show that by a rotation of the form $m^\mu \to e^{iS}m^\mu$, where $S=S(x^3,\ x^4)$, it is possible to make both α° and ξ^{o3} real. Under this rotation α° and ξ^{o3} become

$$\alpha^{\circ\prime}=e^{-iS}(\alpha^\circ+\tfrac{1}{2}i\bar{\delta}S),$$
$$\xi^{o3\prime}=e^{iS}\xi^{o3}.$$

For $\alpha^{\circ\prime}$, $\xi^{o3\prime}$ to be real, we have

$$e^{-iS}(\alpha^\circ+\tfrac{1}{2}i\bar{\xi}^{oj}S_{,j})-e^{iS}(\bar{\alpha}^\circ-\tfrac{1}{2}i\xi^{oj}S_{,j})=0, \tag{3.10a}$$

$$e^{iS}\xi^{o3}-e^{-iS}\bar{\xi}^{o3}=0. \tag{3.10b}$$

Differentiating the last equation with respect to x^j $(j=3,\ 4)$, we find

$$iS_{,j}(e^{iS}\xi^{o3}+e^{-iS}\bar{\xi}^{o3})+e^{iS}\xi^{o3}{}_{,j}-e^{-iS}\bar{\xi}^{o3}{}_{,j}=0,$$

and, substituting for $S_{,j}$ in (3.10a), we obtain

$$e^{-iS}[2\alpha^\circ(e^{iS}\xi^{o3}+e^{-iS}\bar{\xi}^{o3})-\bar{\xi}^{oj}(e^{iS}\xi^{o3}{}_{,j}-e^{-iS}\bar{\xi}^{o3}{}_{,j})]$$
$$=e^{iS}[2\bar{\alpha}^\circ(e^{-iS}\bar{\xi}^{o3}+e^{iS}\xi^{o3})+\xi^{oj}(e^{iS}\xi^{o3}{}_{,j}-e^{-iS}\bar{\xi}^{o3}{}_{,j})].$$

Using Eq. (3.5f) to simplify, this becomes

$$e^{2iS}(\xi^{oj}\xi^{o3}{}_{,j}+2\bar{\alpha}^\circ\xi^{o3})=e^{-2iS}(\bar{\xi}^{oj}\bar{\xi}^{o3}{}_{,j}+2\alpha^\circ\bar{\xi}^{o3}).$$

Substituting for e^{2iS} from (3.10b) and using (3.5f) again, we obtain finally

$$(\bar{\xi}^{o3}\xi^{oj}-\xi^{o3}\bar{\xi}^{oj})(\xi^{o3}\bar{\xi}^{o3})_{,j}=0,$$

$$\text{i.e.,}\ (\bar{\xi}^{o3}\xi^{o4}-\xi^{o3}\bar{\xi}^{o4})(\xi^{o3}\bar{\xi}^{o3})_{,4}=0 \tag{3.11}$$

The quantity $(\bar{\xi}^{o3}\xi^{o4}-\xi^{o3}\bar{\xi}^{o4})$ is nonzero since it is proportional to det $g^{\mu\nu}$ and hence the rotation will achieve the desired effect if $(\xi^{o3}\bar{\xi}^{o3})$ is independent of x^4. This condition can be satisfied since we still have the coordinate freedom expressed by the transformations

$$x^3\to f^3(x^3,\ x^4), \tag{3.12a}$$

$$x^4\to f^4(x^3,\ x^4). \tag{3.12b}$$

By using the transformation (3.12a) we can choose the coordinate x^3 so that $(\xi^{o3}\bar{\xi}^{o3})$ is a function of x^3 only. It follows that $\xi^{o3\prime}$ is also a function of x^3 only and since $\xi^{o3\prime}$ is real we can use the remaining coordinate freedom (3.12b) to make $\xi^{o4\prime}$ imaginary.

The possibility exists that Eq. (3.11) may be satisfied by $\xi^{o3}\bar{\xi}^{o3}$, and consequently $\xi^{o3\prime}$, being a constant rather than a function of x^3. Both possibilities are covered by using a coordinate transformation of the form $x^3\to h(x^3)$ to transform

$$\int\frac{dx^3}{\xi^{o3\prime}}\to\sqrt{2}\,ex^3,$$

which is equivalent to taking

$$\xi^{o3\prime}=(\sqrt{2}\,e)^{-1} \tag{3.13}$$

in the new coordinate system.

Discarding the primes and using the facts that α° is real, ξ^{o4} is imaginary and γ°, ξ^{o3} are given by (3.7) and (3.13), Eqs. (3.5a, f) lead to

$$(\sqrt{2}\,e)^{-1}\alpha^\circ{}_{,3}=4\alpha^{o2}+\tfrac{1}{2}e^{-2}, \tag{3.14a}$$

$$(\sqrt{2}\,e)^{-1}\xi^{o4}{}_{,3}=2\alpha^\circ\xi^{o4}. \tag{3.14b}$$

4. THE SOLUTION

From Eqs. (3.14a, b) together with equations (3.6), (3.7), (3.9a, b), and (3.13) the complete solution is

$$\begin{aligned}&\xi^{o3}=(\sqrt{2}\,e)^{-1},\\&\xi^{o4}=i(\sqrt{2}\,e)^{-1}A\sec(x^3-B),\\&U=-\tfrac{1}{2}e^{-2}r^2,\\&\gamma=\tfrac{1}{2}e^{-2}r,\\&\alpha=(2\sqrt{2}\,e)^{-1}\tan(x^3-B),\end{aligned} \tag{4.1}$$

where A, B are functions of x^4 only. Using the fact that the metric tensor is given by

$$g^{\mu\nu}=l^\mu n^\nu+l^\nu n^\mu-m^\mu\bar{m}^\nu-m^\nu\bar{m}^\mu,$$

the matric of the space—time solution is found to be

$$ds^2=e^{-2}r^2du^2+2du\,dr-e^2(dx^3)^2$$
$$-e^2A^{-2}\cos^2(x^3-B)(dx^4)^2.$$

We can put $A=1$ by a redefinition of the x^4 coordinate and, by calculating the components of the curvature tensor, we note that the two-dimensional space with metric

$$e^2(dx^3)^2+e^2\cos^2(x^3-B)(dx^4)^2$$

is a space of constant positive curvature so that there exists a coordinate system in which the metric takes the form[11]

$$e^2(dx^3)^2+e^2\cos^2x^3(dx^4)^2,$$

i.e., in which $B=0$. Hence the most general conformally flat solution of the source-free Einstein—Maxwell equations for nonnull fields is

$$ds^2=e^{-2}r^2du^2+2du\,dr-e^2(dx^3)^2-e^2\cos^2x^3(dx^4)^2.$$

The coordinate transformations

$$r=e^2/r',\quad u=t+r',\quad x^3=\theta-\pi/2,\quad x^4=\psi$$

convert this metric into the form (1.1), which is thus the unique solution of this type.

ACKNOWLEDGMENTS

We wish to express our thanks to G.F.R. Ellis, C.W. Kilmister, and particularly R.G. McLenaghan for helpful comments and discussions. We would also like to thank the entire relativity group at the University of Waterloo for enduring a talk by one of us (BOJT) on the subject of this article.

APPENDIX

The energy condition appearing in the Rainich conditions is

$$T_{\mu\nu}v^\mu v^\nu>0,$$

where $T_{\mu\nu}$ is the electromagnetic energy tensor and v^μ is an arbitrary timelike vector. In view of the field equations this can be written in the form

$$R_{\mu\nu}v^\mu v^\nu < 0. \tag{A1}$$

Now rewrite the metric (1.2) by defining coordinates $\xi(=x^3)$ and $\eta(=x^4)$ such that $z=\xi+i\eta$, i.e.,

$$ds^2 = Kr^2 du^2 + 2du\,dr - 2(1+\tfrac{1}{2}K\rho^2)^{-2}(d\xi^2 + d\eta^2),$$

where $\rho^2=\xi^2+\eta^2$. The nonzero components of the Ricci tensor are

$$R_{11} = -K^2r^2,$$

$$R_{12} = -K,$$

$$R_{33} = R_{44} = -2K(1+\tfrac{1}{2}K\rho^2)^{-2}.$$

The condition (A1) becomes

$$R_{\mu\nu}v^\mu v^\nu = -K^2r^2(v^1)^2 - 2Kv^1v^2 - 2K(1+\tfrac{1}{2}K\rho^2)^{-2}[(v^3)^2+(v^4)^2] < 0. \tag{A2}$$

Since v^μ is a timelike vector, $v_\mu v^\mu$ is positive, i.e.,

$$g_{\mu\nu}v^\mu v^\nu = Kr^2(v^1)^2 + 2v^1v^2 - 2(1+\tfrac{1}{2}K\rho^2)^{-2}[(v^3)^2+(v^4)^2] = p^2 \tag{A3}$$

From Eqs. (A2) and (A3) we find that the energy condition becomes

$$R_{\mu\nu}v^\mu v^\nu = -K(p^2+q^2) < 0, \tag{A4}$$

where

$$q^2 = 4(1+\tfrac{1}{2}K\rho^2)^{-2}[(v^3)^2+(v^4)^2].$$

The energy condition (A4) is satisfied only when $K>0$. Thus the case when $K<0$ may be dismissed.

*Research supported in part by the National Research Council of Canada through operating Grant A 7589.

[1]B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[2]I. Robinson, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astr. Phys. **7**, 351 (1959).
[3]D. Lovelock, Comm. Math. Phys. **5**, 257 (1967).
[4]P. Dolan, Comm. Math. Phys. **9**, 161 (1968).
[5]M. Cahen and J. Leroy, Bull. Acad. Roy. Belg. **51**, 1179 (1965); J. Math. Mech. **16**, 501 (1966).
[6]See, for example, the article by F.A.E. Pirani in A. Trautman, F.A.E. Pirani, and H. Bondi, *Lectures in General Relativity* (Prentic-Hall, Englewood Cliffs, N.J., 1965).
[7]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[8]E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[9]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[10]G.C. Debney and J.D. Zund, Tensor, N.S. **22**, 333 (1971).
[11]T.J. Willmore, *An Introduction to Differential Geometry* (Clarendon, Oxford, 1959).

Physical axiomatics

A. R. Marlow

Department of Physics, Spring Hill College, Mobile, Alabama 36608
(Received 30 May 1973)

An axiomatic mathematical structure is presented in terms of which arbitrary mathematical models of a physical system can be rigorously formulated and studied. Physical systems, observables, and states all arise in a natural way from the primitive notion of a physical operation. In particular, the structure is independent of the special mathematical completion defined by the quantum mechanical model, since no lattice theoretical or Hilbert space assumptions are used. The precise relationship between the present structure and the usual quantum mechanical model is investigated in a succeeding paper.

Recent work in the foundations of quantum mechanics has led to a variety of axiom sets for the quantum mechanical model of a physical system,[1] each with the basic aim of elucidating the logical foundations of the model in order to determine which elements are physically natural and which are basically *a priori*, and hence possibly dispensable. The present paper in a sense continues Mackey's work in this area,[2] since we use four of Mackey's axioms substantially intact, with only the modifications necessary to guarantee independence from the quantum mechanical model.[3] These axioms together with five others define a logical and mathematical structure which we believe to be both natural and even inevitable for arbitrary models of actually realizable physical systems; because of the independence of the resulting structure from the quantum mechanical model, we have a rigorous basis for the formulation and study of general mathematical models of physical systems, whether classical, quantum, or of some new type. In particular, since systems, observables, *and* states are all given a natural physical interpretation within the structure in terms of the primitive notion of a physical operation, we have a means of clarifying such troublesome questions as simultaneous observability, the meaning of physical indeterminacy principles, the collapse of a state under observation, and the possibility of hidden variables. In a succeeding paper we investigate the relation of the present structure to the standard quantum mechanical model. We will neither need nor use in the succeeding paper all of the axioms formulated in this paper, but we feel it best to state them now, since they are natural physical requirements and future work can be built on them.

In this and the following paper we will use standard mathematical and logical notation throughout, and we will signal the end of a formal statement of our axiomatic structure with the symbol □; the letters A, D, T, L, and C will stand for "axiom," "definition," "theorem," "lemma" and "corollary," respectively. In general, square brackets will be used to separate off independent segments of a symbolic logic or mathematical statement, while curly brackets will be used in defining and referring to sets.

1. THE PHYSICAL UNIVERSE

Following the axiomatic method, we will need a universe of discourse, consisting of certain undefined elements together with certain elements already defined in terms of existing axiomatic structures; thus we will feel free to introduce any purely mathematical elements into our structure at need and without apology, but physical elements will have to be introduced as undefined terms, whose formal interpretation will be given by succeeding axioms, or as terms defined using only previously introduced elements. We will always have in mind a physical interpretation for our axioms and definitions, and we will give this interpretation as we go along, but it will play no formal part in our axiomatic structure; the purpose of this motivating interpretation is simply to guarantee the physical realizability of our axioms to ensure that we are not playing logical games in a physical vacuum. The basic undefined notion we will need is that of a physical operation; from this we will construct systems, states, and observables. Intuitively, we can say that what we mean by an operation is simply any action or "interference" in our environment which observers can actually perform in practice in a clearly communicable way. This latter qualification is intended to imply that to each operation there corresponds a list of practical instructions for performing the operation.[4] We will also want to include from the start the mathematical theory of probability measures in such a way that the measures will distinguish between different operations. Finally, we want to include the possibility of combining two operations in a given order to produce a third operation. With this preamble we can now state formally:

U. Our universe of discourse will consist of a set Φ together with an associative composition law

$$*:\Phi\times\Phi\to\Phi$$
$$(\phi,\phi')\to\phi'*\phi$$

and a mapping

$$\pi:\Phi\times\mathcal{B}\to[0,1]$$
$$(\phi,B)\to\pi_\phi(B),$$

where $\mathcal{B}$ is the class of all Borel sets[5] in the real line R, and $[0,1]$ is the closed unit interval in R. Finally, given any two operations ϕ,ϕ', we have $\phi=\phi'$ if and only if $\pi_{\phi*\phi''}=\pi_{\phi'*\phi''}$, for all ϕ'' in Φ.

Elements $\phi,\phi',\ldots$ of the set Φ will be called *operations* and the mapping π will be called the *probability mapping* of our universe. The motivation behind the introduction of associativity is the obvious fact that if three operations ϕ,ϕ',ϕ'' can be performed successively at all (i.e., if the corresponding sets of operational prescriptions can be taken together in the required order as a single actualizable set of prescriptions), then the two alternative groupings $(\phi''*\phi')*\phi$ and $\phi''*(\phi'*\phi)$, are not operationally distinguishable.

But now it is necessary to take explicit account of the fact that two arbitrary sets of operational prescriptions, taken together in a certain order, do not always yield a new set of prescriptions which can actually be performed, as well as the fact that it is possible to do nothing at all. Also, we want to introduce left inverses for operations. All of this will be taken care of by our first axiom:

A-1. There exist two elements of Φ, specifically, ι and θ, such that, $\forall \phi \in \Phi$:

$$\iota * \phi = \phi * \iota = \phi \text{ and } \theta * \phi = \phi * \theta = \theta.$$

Also, $\forall \phi \in \phi - \{\theta\}$, $\exists \phi^{-1}$ such that:

$$\phi^{-1} * \phi = \iota. \qquad \square$$

We will call ι the *identity operation*, and θ the *impossible operation*. The operation ι is interpreted as the operation of "doing nothing," while an equation of the form $\phi' * \phi = \theta$ is interpreted as meaning that some prescription involved in the performance of ϕ' cannot be fulfilled by reason of the previous fulfillment of the prescriptions for ϕ.[6] The introduction of the left inverse elements is motivated by the general physical requirement of repeatability of experiments and measurements —there must be some operation, even if it is only a "gedanken" operation (e.g., time reversal for astronomical measurements), by which we can in principle return to the original initial conditions and then perform again our experiment or measurement.

In the presence of A-1, we can describe our set Φ mathematically as a semigroup (that is, a set with an associative composition law) with a two-sided identity element and a two-sided null element (the impossible operation) and, in addition, a left inverse for every element of the semigroup. From this we have our first theorem:

T-1. The set $\Phi - \{\theta\}$ is a group. $\square$

Proof of T-1. The only thing we need to check is that the left inverses serve also as right inverses for every element of $\Phi - \{\theta\}$. This is seen from the following line of reasoning:

$$\varphi * \varphi^{-1} = (\varphi^{-1})^{-1} * (\varphi^{-1} * \varphi) * \varphi^{-1} = \iota. \qquad \text{QED}$$

From simple and standard algebraic considerations, we now have C-1. The operations ι, θ, and the two-sided inverses of the group $\Phi - \{\theta\}$ are uniquely defined. $\square$

We have introduced the mapping π into our structure in order to be able to select the elements necessary for the construction of physical systems, but before doing this we will need some definitions and further axioms.

D-1. Let $\mathfrak{M}$ be the set of all $\phi \in \Phi$ such that the mapping $\pi_\phi : \mathcal{B} \to [0, 1]$ is a probability measure on $\mathcal{B}$, and let Φ_0 be the set of all $\phi \in \Phi$ such that π_ϕ is the zero measure on $\mathcal{B}$. $\square$

If $\emptyset$ represents the empty set, we can now state our second axiom:

A-2. $\mathfrak{M} \neq \emptyset$; $\iota, \theta \in \Phi_0$; and $\Phi = \mathfrak{M} \cup \Phi_0$. $\square$

The elements of the set $\mathfrak{M}$ will be called the *measurements* of our universe; intuitively, we can describe a measurement $\phi \in \mathfrak{M}$ as an operation such that each performance p_ϕ of ϕ determines a single real number r in a well defined way, and repeated performances p_ϕ, p'_ϕ, p''_ϕ, yield results $r, r', r'' \in R$ in accordance with a frequency distribution specified by the probability measure π_ϕ. We thus give the natural interpretation to the number $\pi_\phi(B)$ as the probability that the result r of an arbitrary performance p_ϕ of ϕ will lie in the Borel set B. A-2 also guarantees that there will be measurements in our formal structure, and makes the mapping π the means of distinguishing measurements from all other operations; A-2 also expresses axiomatically the fact that the identity operation and the impossible operation are not measurements.

For our next axiom we will need to take over from measure theory the notion of *absolute continuity:* a measure π', in symbols, $\pi \ll \pi'$, if and only if $[\pi'(B) = 0] \Rightarrow [\pi(B) = 0, \forall B \in \mathcal{B}]$. We can now state:

A-3. For all pairs $(\phi, \phi') \in \mathfrak{M} \times \Phi$,

(1) $\pi_{\phi*\phi'} \ll \pi_\phi$

and

(2) $[\phi * \phi' \neq \theta] \to [\phi * \phi' \in \mathfrak{M}]$. $\square$

Our axiom A-3 needs a bit more justification than the previous axioms; basically, it expresses our intention (1) to make the mappings π_ϕ express *total* probabilities, as well as our intuition (2) that if a measurement ϕ is possible after some preparatory operation ϕ' (i.e., if $\phi * \phi' \neq \theta$), then $\phi * \phi'$ should also be a measurement, in general distinct from ϕ, but related to it by (1) above. For example, if ϕ is a measurement and B a Borel set such that $\pi_\phi(B) = 0$, then there should be no probability, from any "source" whatsoever, that the result of a performance of $\phi * \phi'$ will lie in B, no matter what preparatory operation ϕ' we use; hence, $\pi_{\phi*\phi'}(B)$ should also be zero.[7] By (2) above, the only instance in which $\phi * \phi'$ is not a measurement is had when it is the impossible operation, and even then (1) is obeyed, since the zero measure is absolutely continuous with respect to every measure. Thus, in the presence of A-3, the measurements become a very restricted type of operation, since each measurement must have in its set of operational prescriptions sufficient safeguards to guarantee that a unique probability measure will be associated with the measurement. It is our belief, however, that all of modern physics is predicated on the basis of the actual existence of measurements as we have defined them, and that the mappings π_ϕ are sufficient to express all physical data.

It does not seem clear physically, however, that every operation which determines a number on each performance is a measurement—it seems quite conceivable that an operation can determine numbers without a probability distribution. We want our structure to make allowance for this, and so we introduce:

D-2. For each $\phi \in \Phi$, let $Z(\phi)$ be the union of all open intervals $I \subset R$ such that $\pi_\phi(I) = 0$, and let $N(\phi)$ be the set defined by:

$$N(\phi) = \bigcup_{\phi' \in \Phi} (R - Z(\phi * \phi')) = R - \bigcap_{\phi' \in \Phi} Z(\phi * \phi').$$

Then we define the set $\mathfrak{N}$ by the prescription: $\mathfrak{N} = \{\phi \in \Phi : N(\phi) \neq \emptyset\}$. $\square$

We will call $N(\phi)$ the *numerical set* of the operation ϕ, and $\mathfrak{N}$ the set of *numerical operations* of our universe. Thus the numerical operations are precisely the operations with nonempty numerical sets; we also have the following theorem, characterizing the relationship between numerical operations and measurements:

T-2. $\mathfrak{N} = \{\phi \in \Phi : \exists\, \phi' \in \Phi, \text{with } \phi * \phi' \in \mathfrak{M}\}$. $\square$

Proof of T-2. We have $[N(\phi) \neq \emptyset] \Longleftrightarrow [\exists\, \phi' \in \Phi : R \not\subset Z(\phi * \phi')]$. But $[R \not\subset Z(\phi * \phi')] \Longleftrightarrow [\pi_{\phi*\phi'} \neq 0]$, and by A-2 $[\pi_{\phi*\phi'} \neq 0] \Longleftrightarrow [\phi * \phi' \in \mathfrak{M}]$. Therefore, $[N(\phi) \neq \emptyset] \Longleftrightarrow [\exists\, \phi' \in \Phi : \phi * \phi' \in \mathfrak{M}]$. QED

Since for all $\phi \in \mathfrak{M}$ we have $\phi * \iota \in \mathfrak{M}$, we conclude that the set of measurements is a subset of the set of numerical operations; we leave open the possibility that, in physical reality, $\mathfrak{M}$ is actually equal to $\mathfrak{N}$. As indicated earlier, we interpret $\mathfrak{N}$ as including, besides the measurements, all those operations which determine a number on each performance, but do not do so in accordance with a probability distribution. From now on, we will use the Greek letters $\alpha, \beta, \ldots$ to symbolize the numerical operations (and hence, the measurements also), in order to have a convenient way of distinguishing them from other operations.

It will also be useful now to introduce two other classes of operations:

D-3. Let $\mathfrak{Q} \in \mathfrak{N}$ be the set $\mathfrak{Q} = \{\alpha \in \mathfrak{N}: N(\alpha) \subset \{0, 1\}\}$, and let $\mathfrak{D} \subset \mathfrak{Q}$ be the set $\mathfrak{D} = \{\alpha \in \mathfrak{N}: N(\alpha) = \{1\}\}$. □

We call elements of the set $\mathfrak{Q}$ *questioning operations* and elements of $\mathfrak{D}$ *determinations*. For questioning operations we will use the symbols $q, q', q'', \ldots$; because they have only the numbers 1 and 0 in their numerical sets, we give them the obvious interpretation of operations which ask physical questions which have yes–no answers. The determinations, on the other hand, form the subset of those operations in $\mathfrak{Q}$ which have only the answer "yes."

Our final axiom in this section is essentially Mackey's axiom III, extended, however, to include numerical operations.[8]

A-4. Let $\mathfrak{F}$ be the set of real-valued Borel functions on R.[5] Then to each pair $(\alpha, f) \in \mathfrak{N} \times \mathfrak{F}$ there corresponds a unique element $f(\alpha) \in \mathfrak{N}$ such that, $\forall(\phi, B) \in \Phi \times \mathfrak{B}$,

(1) $[f(\alpha)*\phi \in \mathfrak{N}] \longleftrightarrow [\alpha*\phi \in \mathfrak{N}]$

(2) $[\alpha*\phi \in \mathfrak{N}] \rightarrow [f(\alpha)*\phi = f(\alpha*\phi)]$

(3) $\pi_{f(\alpha)}(B) = \pi_\alpha(f^{-1}(B))$. □

The motivation behind A-4 is the same as that given by Mackey—physically, the operation $f(\alpha)$ results from α as follows: whatever we do to perform α, we perform $f(\alpha)$ by simply applying the function f to the numerical result obtained in a performance of α. Several theorems can be proven now.

T-3. $\forall(\alpha, f) \in \mathfrak{N} \times \mathfrak{F}, [f(\alpha) \in \mathfrak{M}] \Longleftrightarrow [\alpha \in \mathfrak{M}]$. □

Proof of T-3. $\forall(\alpha, f), \pi_{f(\alpha)}(R) = \pi_\alpha(R)$, by A-4. Hence, $[\pi_{f(\alpha)} = 0] \Longleftrightarrow [\pi_\alpha = 0]$, and T-3 follows from A-2. QED

T-4. $\forall(\alpha, f) \in \mathfrak{N} \times \mathfrak{F}, [f(\alpha)*\phi \in \mathfrak{M}] \longleftrightarrow [\alpha*\phi \in \mathfrak{M}]$. □

Proof of T-4. By A-4, $[f(\alpha)*\phi \in \mathfrak{M}] \Rightarrow [\alpha*\phi \in \mathfrak{N}]$, and $[\alpha*\phi \in \mathfrak{N}] \rightarrow [f(\alpha)*\phi = f(\alpha*\phi)]$. Hence, $[f(\alpha)*\phi \in \mathfrak{M}] \rightarrow [f(\alpha*\phi) \in \mathfrak{M}]$ and so by T-3 we have (1) $[f(\alpha)*\phi \in \mathfrak{M}] \Rightarrow [\alpha*\phi \in \mathfrak{M}]$. Similarly, by T-3, $[\alpha*\phi \in \mathfrak{M}] \Rightarrow [f(\alpha*\phi) \in \mathfrak{M}]$ and by A-4 $[\alpha*\phi \in \mathfrak{M}] \Rightarrow [f(\alpha*\phi) = f(\alpha)*\phi]$. Therefore, we have (2) $[\alpha*\phi \in \mathfrak{M}] \Rightarrow [f(\alpha)*\phi \in \mathfrak{M}]$, and T-4 is equivalent to the conjunction of (1) and (2). QED

T-5. The sets $\mathfrak{Q}$ and $\mathfrak{D}$ are not empty. □

Proof of T-5. For all $B \in \mathfrak{B}$, the characteristic function Q_B [defined on the real line by the prescription: $Q_B(r) = 1, r \in B$; $Q_B(r) = 0, r \in B$] is a Borel function, and for any open interval I such that $\{0, 1\} \not\subset I$ we obviously have $Q_B^{-1}(I) = \emptyset$. Thus the numerical set of $Q_B(\alpha)$, for $\alpha \in \mathfrak{N}$, must be a subset of $\{0, 1\}$, and so for all numerical operations α we have $Q_B(\alpha) \in \mathfrak{Q}, \forall B \in \mathfrak{B}$. In particular, for $\alpha \in \mathfrak{M}$, the measurement $Q_R(\alpha)$ is a determination since its numerical set can only be $\{1\}$. Since by A-2 $\mathfrak{M}$ is a nonempty subset of $\mathfrak{N}$, we see that $\mathfrak{Q}$ and $\mathfrak{D}$ are nonempty. QED

To conclude this section, we will call any formal structure U satisfying axioms A-1, through A-4 a *physical universe*.

2. PHYSICAL SYSTEMS

We are now in a position to construct physical systems within a physical universe. To begin, for any questioning operation $q \in \mathfrak{Q}$ we construct the two sets:

D-4. $\Phi_q = \{\phi \in \Phi: q*\phi \in \mathfrak{D}\}$; $\mathfrak{N}_q = \{\alpha \in \mathfrak{N}: \alpha*\phi \in \mathfrak{M}, \forall \phi \in \Phi_q\}$. □

Now from our description U of the physical universe, and the material in footnote 4, Mackey's axioms I and II follow immediately as theorems of our present structure, i.e., to each pair $(\alpha, \phi) \in \mathfrak{N}_q \times \Phi_q$ there corresponds uniquely a probability measure $\pi_{\alpha*\phi}$. Mackey's axiom III also follows from our present structure, and so after stating our definition of a physical system, we will state these three axioms as theorems of our structure.[8]

D-5. A structure $\sum = \{\mathcal{O}_\Sigma, \mathcal{S}_\Sigma\}$ consisting of two sets will be called a *physical system* if and only if there exists a questioning operation q with $\Phi_q \neq \emptyset$ such that $\mathcal{O}_\Sigma = \mathfrak{N}_q$ and $\mathcal{S}_\Sigma = \Phi_q$. In this case we will call q a *defining operation* of the system, $\mathcal{S}_\Sigma$ the set of *physical states* of the system, and $\mathcal{O}_\Sigma$ the set of *physical observables* of the system. □

The qualification "physical" in D-5 has been put in for later convenience, when it will be necessary to distinguish the physically constructed elements of a system from purely mathematically constructed elements in some particular mathematical model of the system. When this distinction is not necessary we will simply refer to the observables and states of the system. We can now state three theorems valid for arbitrary physical systems; these simply summarize Mackey's axioms I-III. From our definitions it should be obvious that only T-8 needs any explicit verification.

T-6. $\forall(\alpha, \phi) \in \mathcal{O}_\Sigma \times \mathcal{S}_\Sigma, \pi_{\alpha*\phi}$ is a probability measure. □

T-7. $\forall(\alpha, \beta) \in \mathcal{O}_\Sigma \times \mathcal{O}_\Sigma, [\alpha = \beta] \Longleftrightarrow [\pi_{\alpha*\phi} = \pi_{\beta*\phi}, \forall \phi \in \mathcal{S}_\Sigma]$;

$\forall(\phi, \psi) \mathcal{S}_\Sigma \times \mathcal{S}_\Sigma, [\phi = \psi] \Longleftrightarrow [\pi_{\alpha*\phi} = \pi_{\alpha*\psi}, \forall \alpha \in \mathcal{O}_\Sigma]$. □

T-8. $\forall(\alpha, f) \in \mathcal{O}_\Sigma \times \mathfrak{F}$, there exists a unique element $f(\alpha) \in \mathcal{O}_\Sigma$ such that $\pi_{f(\alpha)*\phi}(B) = \pi_{\alpha*\phi}(f^{-1}(B)), \forall(\phi, B) \in \mathcal{S}_\Sigma \times \mathfrak{B}$. □

Proof of T-8. From a simple application of A-4 and T-4 it is clear that there exist elements of $\mathcal{O}_\Sigma$ which satisfy the conditions of T-8. It then follows immediately from T-7 that there can be only one such element for any pair (α, f). QED

The interpretations of the elements introduced in the last three definitions can be given quite simply. The defining operation of a system is the means of verifying the actual fulfilment of the necessary and sufficient conditions for the presence of the system; this requires of course that we have a good operational definition of the system, i.e., that we know precisely under what experimental conditions we will say that the system has been actualized. Conceptually the defining operation is simple, but in practice it may be a conjunction of a number of subsidiary questioning operations. For example, in the case of some macroscopic systems it may involve only checking to see whether the right experimental arrangement is present and in working condition, while for microscopic systems it will generally involve also detailed questioning operations regarding the proper values of mass, charge, spin, etc. In any case, we say a certain system $\sum$ is present if and only if we are in circumstances such that the defining operation q has the answer "yes" with certainty. The elements of the set Φ_q are then precisely those preparatory operations ϕ which produce various examples of circumstances such that the questioning operation q has 1 for its only possible result (i.e., such that $q*\phi \in \mathfrak{D}$), and so they are all the operations which render the system present in its various possible configurations or states; the set $\mathfrak{M}_q$ then consists of all those numerical operations α which result in measurements $\alpha*\phi$ when they are performed after any of the preparatory operations $\phi \in \Phi_q$.

Several remarks are in order at this point. We have purposely omitted any requirement to the effect that a measurement $\alpha*\phi$ should itself correspond to a state of the system being considered (i.e., we do not impose the condition $q*\alpha*\phi \in \mathfrak{D}$, for q the defining operation of a system and $\alpha \in \mathfrak{M}_q$, $\phi \in \Phi_q$—in fact we do not even require that the operation $q*\alpha*\phi$ be possible); this is realistic since many actual measurements, for example those involving high-energy scattering processes, may destroy the initial system under study, so that the system must be prepared again if another measurement is to be performed on it. We also avoid any assumption regarding the "collapse of the state" after measurement, along with the difficulties and confusions to which this assumption can lead.[9] In this same line of thought, indeterminacy principles have a very precise interpretation in the present context and will arise in the standard way from any quantum mechanical model resulting from our structure. Such a principle, holding, say, between two observables α and β on a given system $\sum$, makes no statement about the possibility or impossibility of the operations $\alpha*\beta$ or $\beta*\alpha$. Instead, the principle states a relationship between measurements of the form $\alpha*\phi$ or $\beta*\phi$, implying that there exists no preparation ϕ of $\sum$ which will make the repeated performances of both measurements determinate, although such determinacy can be had for either one or the other of the two measurements, provided we use a suitably chosen state (an "eigenpreparation" for one of the observables) and are willing to suffer indeterminacy in the results of the other measurement. Finally, it should be noted that the definition we have given of a system is general enough to include all the physical models so far proposed, whether for classical or quantum systems; there seems to be no reason to believe that it cannot be applied to more general systems as long as they are subject to quantitative measurement (e.g., biological systems, psychological systems, sociological systems), but we will use the terms "system" and "physical system" in reference to the structure defined above.

We now formally incorporate into our structure Mackey's axioms IV–VI (and the relevant definitions) as our axioms 5–7, with only the required notational changes.[10] Consequently, all of the theorems and results which depend only on Mackey's first six axioms will be available in our structure.

A-5. Let $\phi_i, i = 1, 2, \ldots$, be a finite or infinite sequence of states of a system $\sum$, and let λ_i be a corresponding sequence of real numbers such that $0 < \lambda_i \leq 1$ and $\sum_i \lambda_i = 1$. Then there exists a $\phi \in \mathcal{S}_\Sigma$ such that, $\forall \alpha \in \mathcal{O}_\Sigma$ $\pi_{\alpha*\phi} = \sum_i \lambda_i \pi_{\alpha*\phi_i}$. We will denote this state by the symbol $\sum_i \lambda_i \phi_i$. □

D-7. For any system $\sum$ let $\mathcal{S}_\Sigma^m$ be the set of all states ϕ which are convex linear combinations $\sum_i \lambda_i \phi$ of states $\phi_i \neq \phi$, and let $\mathcal{S}_\Sigma^p$ be the set $\mathcal{S}_\Sigma - \mathcal{S}_\Sigma^m$. □

We will call the states in $\mathcal{S}_\Sigma^m$ *mixed states* of the system and those in $\mathcal{S}_\Sigma^p$ *pure states*. Obviously, the motivation behind A-5 is the introduction of statistical ensembles of states into our structure.

With our next definition we introduce the important set $\mathcal{Q}_\Sigma$ of *questions* on a system $\sum$.

D-8. For any system $\sum$, let $\mathcal{Q}_\Sigma$ be the subset of observables defined by $\mathcal{Q}_\Sigma = \{\alpha \in \mathcal{O}_\Sigma : \pi_{\alpha*\phi}(\{0, 1\}) = 1, \forall \phi \in \mathcal{S}_\Sigma\}$. □

That $\mathcal{Q}_\Sigma$ is not empty can be seen from the same reasoning that led to T-5: for any $\alpha \in \mathcal{O}_\Sigma$ the observables $Q_B(\alpha)$ corresponding to the characteristic functions of the sets $B \in \mathfrak{B}$ are elements of $\mathcal{Q}_\Sigma$. In particular, the defining operation q of $\sum$ corresponds to a unique question in $\mathcal{Q}_\Sigma$ which we will general symbolize as "1" since it is clearly also the Borel function $Q_R = 1$ applied to any observable $\alpha \in \mathcal{O}_\Sigma$. A question of the form $Q_B(\alpha)$ has an obvious interpretation as the observable which yields the result "1" whenever a performance of the observable α yields a result in the set B, and yields the result "0" otherwise; in this sense it is the yes-no question: "Did the measurement of α lead to a result in B?" Now for arbitrary $q \in \mathcal{Q}_\Sigma$ the Borel function $n: r \to 1 - r, r \in R$, defines by T-8 a new observable $n(q)$, which we will also write as $1 - q$; this observable is clearly the question whose answer is "yes" if and only if the answer to q is "no," and for $q = Q_B(\alpha)$, $n(q)$ corresponds to the question "Did the performance of α lead to a result in $R - B$?"

There is a natural partial ordering on the set $\mathcal{Q}_\Sigma$, but to specify this we need the definition of the *mean value* of an observable on a state:

D-9. Let $m_\phi(\alpha) = \int_R r d\pi_{\alpha*\phi}(r)$ for all pairs $(\alpha, \phi) \in \mathcal{O}_\Sigma \times \mathcal{S}_\Sigma$ such that the integral on the right exists. □

For a question $q \in \mathcal{Q}_\Sigma$ the mean value always exists and is easily evaluated since for any $\phi \in \mathcal{S}_\Sigma$ we have

$$m_\phi(q) = 0 \cdot \pi_{q*\phi}(\{0\}) + 1 \cdot \pi_{q*\phi}(\{1\}) = \pi_{q*\phi}(\{1\})$$

and, since $\pi_{q*\phi}(\{0\}) + \pi_{q*\phi}(\{1\}) = \pi_{q*\phi}(\{0, 1\}) = 1$, it can be seen that the set of mean values $m_\phi(q)$, $\phi \in \mathcal{S}_\Sigma$, completely characterizes the question q. We now define the partial ordering on $\mathcal{Q}_\Sigma$ by:

D-10. $\forall q, q' \in \mathcal{Q}_\Sigma, [q \leq q'] \Longleftrightarrow [m_\phi(q) \leq m_\phi(q'), \forall \phi \in \mathcal{S}_\Sigma]$. □

We also introduce Mackey's useful relationship of *disjointness* between questions:

D-11. $\forall q, q' \in \mathcal{Q}_\Sigma, [q \perp q'] \Longleftrightarrow [q \leq n(q')]$. □[11]

We obviously have $q \perp n(q), \forall q \in \mathcal{Q}_\Sigma$, and $m_\phi(q) + m_\phi(n(q)) = 1, \forall \phi \in \mathcal{S}_\Sigma$; we also have for all disjoint pairs q, q', $m_\phi(q) + m_\phi(q') \geq 1, \forall \phi \in \mathcal{S}_\Sigma$. This suggests the possibility of defining an operation of summation for arbitrary families q_i of pairwise disjoint questions, but this notion will only be meaningful if $\sum_i m_\phi(q_i) \leq 1, \forall \phi \in \mathcal{S}_\Sigma$, and pairwise disjointness alone is not sufficient to guarantee this latter inequality. Nevertheless, we will include Mackey's axiom V as our

A-6. Let q_i be any pairwise disjoint family of questions on a physical system $\sum$. Then there exists in $\mathcal{Q}_\Sigma$ a question q (in symbols, $q = \sum_i q_i$) such that, $\forall \phi \in \mathcal{S}_\Sigma$, $m_\phi(q) = \sum_i m_\phi(q_i)$. □

This axiom obviously needs a clear motivation and interpretation. We assume it primarily because in the most important cases—i.e., when $q_i = Q_{B_i}(\alpha)$ for some observable α and some pairwise disjoint sequence B_i of Borel sets—it is evidently true and we have a precise physical prescription for asking the question q on any state ϕ: we simply perform α on ϕ and see if the resulting number lies in the Borel set $U_i B_i$; if so we have gotten the answer 1 for q; if not we have gotten the answer 0. The remainder of the motivation for A-6 lies in a closer look at the partial ordering in terms of which disjointness is defined. Basically, A-6 expresses our intention of interpreting $q \leq q'$ physically as implying more than the simple probabilistic relation (D-10) which defines it mathematically: we interpret $q \leq q'$ tomean that q' depends intrinsically on q in such a way that q is a partial measurement of q' in the sense that any individual performance of q which yields the result 1 can be taken as a simultaneous performance of q' yielding the result 1. A careful working out of the implications of this interpretation shows that A-6 is fundamentally an assumption concerning the number of states in a physical system—we intend to include enough states in our structure (even if they have to be purely mathematically constructed states and not physical states) so that if q' does not depend on q in the way we have specified there will exist some state ϕ for the system such that $m_\phi(q') < m_\phi(q)$. With this interpretation we also see that $q \perp q'$ means that q and q' cannot have simultaneous "yes" answers, and that a question of the form $\sum_i q_i$ may be interpreted physically as the question whose answer is "yes" if and only if at least (and therefore at most) one of the questions q_i has the answer "yes." We note that, in virtue of T-7, the summations introduced by **A-6** are uniquely defined questions.

In terms of disjointness and disjoint sums we can now define the notion of a *question-valued measure* which will be needed for our next axiom.

D-12. For a physical system $\sum$, we will call a mapping $Q : \mathcal{B} \to \mathcal{Q}_\Sigma : B \to Q_B$ a question-valued measure if and only if:

(1) $[B \cap B' = \emptyset] \Rightarrow [Q_B \perp Q_{B'}]$;

(2) For any sequence B_i of pairwise disjoint sets, $[B = U_i B_i] \Rightarrow [Q_B = \sum_i Q_{B_i}]$;

(3) $Q_R = 1$ and $Q_0 = 0$.

In condition (3) above, the questions designated 1 and 0 are, of course, to be interpreted as the corresponding constant Borel functions. Now for any observable α it is clear that the correspondence $B \to Q_B(\alpha)$ is a question-valued measure, and conversely it is easily seen that the question-valued measures have all the formal properties of observables. This motivates our axiom.

A-7. For any question-valued measure Q on a system $\sum$ there exists an observable α such that $Q_B = Q_B(\alpha)$, $\forall B \in \mathcal{B}$. □

As Mackey shows,[12] the correspondence between observables and question-valued measures on a system is bijective, and each state $\phi \in \mathcal{S}_\Sigma$ is completely specified by the function $m_\phi : \mathcal{Q}_\Sigma \to [0, 1] : q \to m_\phi(q)$. These functions have the following properties: (a) if q_i is any pairwise disjoint family of questions, then $m_\phi(\sum_i q_i) = \sum_i m_\phi(q_i)$; (b) $0 \leq m_\phi(q) \leq 1, \forall q \in \mathcal{Q}_\Sigma$; (c) $m_\phi(0) = 0$ and $m_\phi(1) = 1$. Such a mapping m satisfying (a), (b), and (c) on an orthocomplemented partially ordered set $\mathcal{Q}$ (on $\mathcal{Q}_\Sigma$ the correspondence $q \to n(q)$ defines the orthocomplementation) is called a *probability measure* on $\mathcal{Q}$. Since, by A-5, the full set of states is specified by the subset $\mathcal{S}_\Sigma^p$ of pure states, we see that a physical system $\sum$ is completely equivalent to the triplet $\{\mathcal{Q}_\Sigma, \mathcal{S}_\Sigma^p, m\}$, where m is the mean value function restricted to $\mathcal{Q}_\Sigma \times \mathcal{S}_\Sigma^p$—the full set of observables is defined by A-7 as the set of question-valued measures in $\mathcal{Q}_\Sigma$ and the probability mapping π is recovered from the prescription $\pi_{\alpha*\phi}(B) = m_\phi(Q_B(\alpha))$, where $B \to Q_B(\alpha)$ is the question corresponding to α. It will often be convenient to consider a physical system $\sum$ simply to be the corresponding structure $\{\mathcal{Q}_\Sigma, \mathcal{S}_\Sigma^p, m\}$.

Now any mathematical model of a physical system has as its main purpose the computation of the elements contained in the set of numbers defined by:

D-13. For a physical system $\sum$ let D_Σ (*the data set of* $\sum$) be given by $D_\Sigma \equiv \{m_\phi(q) : \phi \in \mathcal{S}_\Sigma^p, q \in \mathcal{Q}_\Sigma\}$. □

It is obvious, however, that for such a computation to take place we must complete our physical structure, by the addition of further mathematically defined elements, into a mathematical structure provided with a computational rule for deriving the numbers in D_Σ. Even without such a complete mathematical structure, though, we can already construct a number of useful definitions. For example, in analogy with the definition of the numerical set of an operation (D-2) we can define the spectrum $\sigma(\alpha)$ of an observable α on a system $\sum$: let $Z(\alpha)$ be the null set of α [i.e., the union of all open intervals I in R such that $\pi_{\alpha*\phi}(I) = 0$ for all states ϕ] and let $\sigma(\alpha) = R - Z(\alpha)$; the *point spectrum* will then be the subset $\sigma_p(\alpha) = \{r \in R : Q_{\{r\}}(\alpha) \neq 0\}$. We have then as an immediate result the fact that the spectrum of any observable is always a closed nonempty subset of R. If we let the *norm* $\|\alpha\|$ of an observable be the least number $a \in [0, \infty]$ such that $|r| \leq a, \forall r \in \sigma(\alpha)$, and that an observable is bounded if it has a finite norm, then it is easy to show that the mean value of a bounded observable is always finite on any state. We can say two observables α and α' are *compatible* or *simultaneously observable* (in symbols, $\alpha \leftrightarrow \alpha'$) if there exists an observable β and two Borel functions f and f' such that $\alpha = f(\beta)$ and $\alpha' = f'(\beta)$; for questions q and q' this obviously implies that $q \leftrightarrow q'$ if and only if there exists an observable β and two Borel sets B and B' such that $q = Q_B(\beta)$ and $q' = Q'_B(\beta)$.[13] This notion should be distinguished carefully from a similar relationship, that of *simultaneous determinability*, which can be said to hold between questions q and q' if there exists a state ϕ such that both $q*\phi$ and $q'*\phi$ are determinations.

It is evident that these definitions have been motivated by the usual formulation of the quantum mechanical model of a physical system in terms of the mathematics of Hilbert space, although we have formally used only probability theory and our axioms; this can be understood as an indication that the quantum mechanical model for a physical system is a very natural mathematical completion of a physical system. In a succeeding paper we will investigate this possibility.[14]

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Boston University and Spring Hill College for their hospitality during the writing of this paper, to Dr. George W. Mackey for a stimulating discussion, and to Dr. Abner Shimony for helpful comments and suggestions.

[1]Cf. for example, C. Piron, Helv. Phys. Acta, 37, 439 (1964); M. Guenin. J. Math. Phys. 7, 271 (1966). This latter paper contains a rather extensive set of references to previous work in axiomatic quantum theory. The paper most similar in spirit to the present one is the work of S. Gudder, J. Math. Phys. 8, 1848 (1967); however, we do not make any assumptions concerning the existence of coordinate and momentum observables. If Gudder's axioms for these observables are added to the present set of axioms then his results will be valid in the resulting model. In the paper following this one we investigate the relationship of a physical system to the quantum mechanical model without these additional assumptions. As Gudder points out, the basic difficulty with previous axiomatic theories is that they assume an orthocomplemented atomic lattice structure for the set of physical questions, while it is not at all physically evident that the set of questions forms any sort of lattice, much less an atomic one. The axiomatic structure of the present paper leaves open the possibility that a physical system may well be described by such a lattice theoretical model, but it also leaves open the possibility that a physical system may be better describable in terms of quite different mathematical structures.

[2]G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).

[3]Mackey's axioms III–VI are essentially our axioms 4–7. In particular, we do not use Mackey's axioms VII and VIII (the specifically quantum mechanical axioms).

[4]As in all statements of the motivation behind an axiomatic system, we have to avoid pushing our description of physical operations too far. An analogous situation arises in the case of axiomatic geometries versus physical reality–the formal elements "point", "line", "circle" are meaningful in their axiomatic context, but when we try to describe rigorously the physical notions from which they arose we meet innumerable difficulties. The best we can hope for is that our formal system mirrors enough of reality to make it useful.

[5]For those readers unfamiliar with the notions of Borel sets and Borel functions, it is sufficient to say that these classes contain all the sets and functions that physics will need, excluding only some rather "weird" mathematically constructable sets and functions.

[6]Our intuitive description of operations breaks down somewhat at this point (in the sense of footnote 4), since the "artificial" operations introduced by A-1 for mathematical convenience do not have any unique sets of operational prescriptions.

[7]Considering the converse statement "If an operation ϕ does not satisfy A–3, then ϕ is not a measurement", demonstrates the true nature of A–3 as a defining specification of the restricted class of objects we are going to call "measurements." This class is more restricted than the ordinary usage of the term, which would roughly correspond to "A measurement is any operation which results in a number every time it is performed." This broader class of "numerical operations" will be defined in the following paragraph and used in the construction of systems, observables and states, but the strict notion of measurement as given in the text will remain basic to our considerations since it is needed even to define rigorously the notion of a numerical operation.

[8]Reference 2, p. 63.

[9]For some of these difficulties, cf. H. Margenau, Phil. Sci. 30, 1, 138 (1963).

[10]Reference 2, pp. 63–66. We generalize Mackey's axiom V slightly so that it does not assume implicitly that only a countable set of questions can be pairwise disjoint; since we do not use Mackey's Hilbert space axioms or theorems, this modification will have no effect on our present considerations. When we examine the relationship of the present structure to quantum mechanical models, the modification will allow for the possibility that a quantum mechanical model for a given system may have to be realized in terms of nonseparable Hilbert space structures. It will be seen then that the clarification of the physical nature of a state given in this paper will result in a clarification of the physical significance of the separability requirement ordinarily placed on quantum mechanical models.

[11]Reference 2, p. 64.

[12]Reference 2, pp. 66, 67. Our axiom A-6, of course, implies that any given state ϕ has only a countable set of nonzero mean values with respect to any particular pairwise disjoint family of questions; this is not the same as requiring that any such family of questions be countable.

[13]We chose this definition of simultaneous observability because it or its equivalent forms (cf. Ref. 2, pp. 70, 71, and the note on p. 137) seem to be the only known ways to avoid the obscurity inherent in the usual physical descriptions of the notion–in particular, the definition given above does not conflict with relativistic principles, since it does not refer to any form of temporal simultaneity.

[14]For papers related to this paper, by Gudder, Pool, and Mielnik, cf. Trans. Am. Math. Soc. 119, 428 (1965); Pac. J. Math. 19, 81 (1966); Comm. Math. Phys. 9, 118 (1968); Comm. Math. Phys. 9, 55 (1968); Comm. Math. Phys. 15, 1 (1969).

Twistors and the conformal group

Frederick S. Klotz

Department of Applied Mathematics, Trinity College, Dublin, Ireland
(Received 1 March 1973)

The connected symmetry group $SU(2, 2)$ of twistor space ($\mathcal{T}$), a four-dimensional complex manifold with metric

$$ds_{\mathcal{T}}{}^2 \equiv dX^0 d\overline{X^2} + dX^2 d\overline{X^0} + dX^1 d\overline{X^3} + dX^3 d\overline{X^1},$$

and the connected symmetry group $O_0(2, 4)$ of conformal space ($\mathcal{C}$), a six-dimensional real manifold with metric

$$ds_{\mathcal{C}}{}^2 \equiv (dX^0)^2 - (dX^1)^2 - (dX^2)^2 - (dX^3)^2 - (dX^4)^2 + (dX^5)^2,$$

are 4:1 and 2:1 homomorphic, respectively, to the restricted conformal group in compactified Minkowski space ($\mathfrak{M}$). We obtain explicit realizations for these homomorphisms and explore the invariant geometrical relationships they imply between $\mathcal{T}$, $\mathcal{C}$, and $\mathfrak{M}$. As an application of the twistor formalism, we show that every continuous conformal transformation has a unique decomposition as the product of a Lorentz transformation, a translation, an acceleration, a dilation, and one of four special conformal transformations.

I. INTRODUCTION

Twistors may briefly be characterized as the tensors of the group $SU(2, 2)$. Now $SU(2, 2)$ is $2:1$ homomorphic to $O_0(2, 4)$, the connected component of $O(2, 4)$, and $O_0(2, 4)$ in turn is $2:1$ homomorphic to $\mathfrak{C}$, the 15-parameter restricted conformal group in compactified Minkowski space ($\mathfrak{M}$). Consequently twistors of rank one form a basis for a fourfold representation of $\mathfrak{C}$ (see Fig. 1). More generally, any finite-dimensional representation of $\mathfrak{C}$ is equivalent to a direct sum of twistor representations. Thus we can think of twistors as the spinors appropriate to $\mathfrak{M}$.[1,2] They provide, as Penrose[3,4] has shown, a natural means for formulating conformally invariant relationships in $\mathfrak{M}$.

$SU(2, 2)$ is the connected component of the symmetry group of twistor space ($\mathcal{T}$), a four-dimensional complex manifold with metric

$$ds^2_{\mathcal{T}} \equiv dX^0 d\overline{X^2} + dX^2 d\overline{X^0} + dX^1 d\overline{X^3} + dX^3 d\overline{X^1}.$$

$O_0(2, 4)$ is the connected component of the symmetry group of conformal space ($\mathcal{C}$), a six-dimensional real manifold with metric

$$ds^2_{\mathcal{C}} = (dX^0)^2 - (dX^1)^2 - (dX^2)^2 - (dX^3)^2 - (dX^4)^2 + (dX^5)^2. \tag{1}$$

The homomorphisms of figure one link the abstract spaces $\mathcal{T}$ and $\mathcal{C}$ with the space $\mathfrak{M}$ of direct physical interest. It is the purpose of this paper to derive explicit formulas for these homomorphisms and to examine their implications.

We begin in Sec. II with a study of conformal transformations in Minkowski space. The construction of conformal space leads naturally to an explicit realization of the homomorphism between $O(2, 4)$ and the conformal group, which is also available elsewhere.[5,6] In Sec. III, after a brief review of twistor algebra, we build up the theory necessary to work out the homomorphism between $SU(2, 2)$ and $O_0(2, 4)$ in a fully covariant manner. In the course of this work there emerges a general formula for arbitrary products of Dirac γ matrices. This formula is the conformally covariant version of the standard rules for multiplication of γ matrices given by Macfarlane.[7] With the aid of these preliminary results we obtain the explicit homomorphism between $SU(2, 2)$ and $O_0(2, 4)$. Finally, in Sec. IV, we complete the circle of relationships indicated in Fig. 1 by working out the homomorphism between $SU(2, 2)$ and $\mathfrak{C}$. As expected, we get the usual homomorphism between $SL(2, C)$ and the restricted Lorentz group by specializing our results appropriately. Sec. V contains a few simple applications of the formalism of Secs. II–IV. Working in $\mathcal{T}$, we decompose an arbitrary continuous conformal transformation into the product of a Lorentz transformation, a translation, an acceleration, a dilation, and one of four special conformal transformations. We also point out several of the conformally invariant geometrical relationships between $\mathcal{T}$, $\mathcal{C}$, and $\mathfrak{M}$ which our work implies. We conclude in Sec. VI with a few remarks on the significance of the abstract spaces $\mathcal{T}$ and $\mathcal{C}$ for physics.

In the course of the work outlined above it will be necessary to introduce four distinct kinds of indices: conformal space indices (range: $0, \ldots, 5$), spinor indices (range: $0, 1$), twistor indices (range: $0, \ldots, 3$), and Minkowski space indices (range: $0, \ldots, 3$). In order to avoid confusion between different types of indices, we adopt the following conventions. Upper case Latin indices attached to upper case symbols, e.g., X^A, I^A, represent conformal space indices; upper case Latin indices attached to lower case symbols, e.g., ξ^A, $x^{AA'}$, represent spinor indices. Greek indices attached to upper case symbols are twistor indices, e.g., L^α, X^α, $\sum^A_{\alpha\beta}$; and Greek indices attached to lower case symbols are Minkowski space indices, e.g., x^μ, $\tau^{AA'}_\mu$. The only exceptions to these rules will be the universal use of δ for the Kronecker delta and ϵ for the Levi-Civita alternating tensor, where no confusion as to the nature of the indices is likely to occur.

II. $O(2,4)$ AND THE CONFORMAL GROUP

Conformal transformations in Minkowski space preserve the form of the metric

$$ds^2 \equiv \sigma^2 ds^2_{\mathfrak{M}},$$

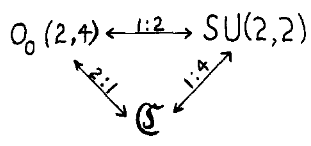

Fig. 1. Homomorphisms between SU(2, 2), $O_0(2, 4)$, and $\mathfrak{C}$.

where ds^2_M is the Minkowski metric

$$ds^2_M \equiv \eta_{\mu\nu} dx^\mu dx^\nu,$$

and $\sigma(x)$ is some unspecified conformal factor. The invariant metric ds^2 may be written in the form

$$ds^2 = \frac{\partial K^A}{\partial x^\mu} \frac{\partial K_A}{\partial x^\nu} dx^\mu dx^\nu \tag{2}$$

where

$$K^A(x) \equiv \sigma(x)\left(x^\alpha, \frac{1}{\sqrt{2}}\left[1+\frac{x^2}{2}\right], \frac{1}{\sqrt{2}}\left[1-\frac{x^2}{2}\right]\right), \tag{3}$$

$$K_A \equiv \Lambda_{AB} K^B,$$

$$\Lambda_{AB} \equiv \mathrm{diag}(+1,-1,-1,-1,-1,+1).$$

By (3), $K^A(x)$ is a null vector in $\mathcal{C}$:

$$K^2 \equiv \Lambda_{AB} K^A K^B = 0.$$

As x^μ varies, $K^A(x)$ sweeps out a four-dimensional submanifold of $\mathcal{C}$ with metric ds^2 related to the Minkowski space metric ds^2_M by the conformal factor σ^2. If we allow arbitrary σ, (3) maps points of Minkowski space (M) into pairs of null directions (corresponding to $\sigma > 0$ or $\sigma < 0$) in $\mathcal{C}$.

With the aid of (3) we can work out the explicit conection between symmetry transformations in $\mathcal{C}$ and conformal transformations in M. The linear transformation

$$X^{A\prime} = C^A{}_B X^B,$$

$$C^A{}_R C^B{}_S A_{AB} = A_{RS}$$

which leaves invariant the conformal space metric (1), induces a transformation in Minkowski space via

$$K^{A\prime}(x) = K^A(x') = C^A{}_B K^B(x). \tag{4}$$

From (2),

$$ds'^2 = \sigma'^2 ds'^2_M = \sigma^2 ds^2_M = ds^2$$

so that

$$ds'^2_M = \Omega^2 ds^2_M,$$

where

$$\sigma' = \Omega^{-1}\sigma.$$

Consequently (4) associates with every element of $O(2,4)$ a conformal transformation in Minkowski space. For those $C \in O(2,4)$ of the form $C = \exp\{\Gamma\}$, we can write

$$\Gamma^{RS} = -\Gamma^{SR} = \hat{\Gamma}^{RS} + 2\hat{A}^{[R}O^{S]} + 2\hat{B}^{[R}I^{S]} + 2\omega I^{[R}O^{S]}.$$

Here $\hat{A}^R$, $\hat{B}^R$, and $\hat{\Gamma}^{RS}$ are projected forms of conformal space tensors,

$$\hat{A}^R \equiv (a^\rho, 0, 0),$$

$$\hat{B}^R \equiv (b^\rho, 0, 0),$$

$$\hat{\Gamma}^R{}_S = \begin{bmatrix} \gamma^\rho{}_\sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$O^R \equiv (1/\sqrt{2})(0,0,0,0,1,1),$$

$$I^R \equiv (1/\sqrt{2})(0,0,0,0,-1,1).$$

Now the (restricted) Lorentz, translation, acceleration, and dilation subgroups of $\mathfrak{C}$ are characterized in $\mathcal{C}$ as follows:

Lorentz transformations: $\hat{A}^R = \hat{B}^R = \omega = 0 \Rightarrow$

$$x^{\mu\prime} = \xi^\mu{}_\nu x^\nu,$$

$$\xi^\mu{}_\nu = (\exp\gamma)^\mu{}_\nu.$$

Translations: $\hat{\Gamma}^{RS} = \hat{A}^R = \omega = 0 \Rightarrow$

$$x^{\mu\prime} = x^\mu + b^\mu.$$

Accelerations: $\hat{\Gamma}^{RS} = \hat{B}^R = \omega = 0 \Rightarrow$

$$x^{\mu\prime} = \Omega[x^\mu - a^\mu(x^2/2)],$$

$$\Omega \equiv (1 - a\cdot x + a^2x^2/4)^{-1}.$$

Dilations: $\hat{\Gamma}^{RS} = \hat{A}^R = \hat{B}^R = 0 \Rightarrow$

$$x^{\mu\prime} = \Omega x^\mu,$$

$$\Omega \equiv e^\omega.$$

There are in fact two matrices $C_\pm = \pm C \in O(2,4)$ for every conformal transformation in M, since the two null directions $\pm K^A(x)$ in $\mathcal{C}$ correspond to the same point x^μ in M. That is, $O(2,4)$ is 2 : 1 homomorphic to the conformal group. This homomorphism has also been obtained by Grgin.[5]

While every point in M defines a pair of null directions in $\mathcal{C}$ via (3), the converse is not true. There exist some null directions in $\mathcal{C}$ for which there are no corresponding points in M, namely those of the form

$$N^A \equiv \left(a^\alpha, -\frac{b}{\sqrt{2}}, \frac{b}{\sqrt{2}}\right),$$

with

$$a^2 \equiv a_\mu a^\mu = 0 \Rightarrow N^2 \equiv N_A N^A = 0.$$

To give these null directions a Minkowski space interpretation, we associate them with *sets* of points x^μ which satisfy

$$N_A K^A(x) = 0 \Rightarrow a_\mu x^\mu + b = 0.$$

For $a^\mu \neq 0$, $a^2 = 0$, this is the equation of a null hyperplane in M. In order to complete the conformally invariant correspondence of (3) between $\mathcal{C}$ and M, we must add to M points at infinity identified with the N^A, so that all null directions in $\mathcal{C}$ have an equivalent interpretation. That is, we must add to M a point at infinity for each null hyperplane of M (the point where the parallel null generators of the hyperplane intersect) plus one additional point I at infinity for the null direction $N^A = bI^A$, with $a^\alpha = 0$ (the point where the null lines at infinity intersect). The result of this procedure is the conformal compactification $\mathfrak{M}$ of M. The behavior of 0^A and I^A under conformal transformations reflects the transformation properties of the origin and timelike (spacelike) infinity, the points of $\mathfrak{M}$ to which, by (3), they correspond:

Lorentz Transformations	*Translations*	*Accelerations*	*Dilations*
$O^{R\prime} = O^R$	$O^{R\prime} = O^R + \hat{B}^R - \frac{1}{2}\hat{B}^2 I^R$	$O^{R\prime} = O^R$	$O^{R\prime} = \Omega^{-1}O^R$
$I^{R\prime} = I^R$	$I^{R\prime} = I^R$	$I^{R\prime} = I^R + \hat{A}^R - \frac{1}{2}\hat{A}^2 O^R$	$I^{R\prime} = \Omega I^R$

III. $SU(2,2)$ AND $O_0(2,4)$

Having worked out explicitly the relationship between $O(2,4)$ and the conformal group in (compactified) Minkowski space, we move on to consider the relationship between $SU(2,2)$ and $O_0(2,4)$. For this purpose we must first review a few pertinent details of twistor algebra. We shall employ the notation of Ref. 3 as far as possible.

Twistors Z^α of valence $\binom{1}{0}$ form a basis for a realization of $SU(2,2)$ in terms of 4×4 complex matrices $T^\alpha{}_\beta$:

$$Z^{\alpha\prime} = T^\alpha{}_\beta Z^\beta,$$
$$T^\alpha{}_\rho \overline{T^\beta{}_\sigma} \Lambda_{\alpha\beta} = \Lambda_{\rho\sigma},$$
$$\det T^\alpha{}_\beta = 1,$$

where

$$\Lambda_{\alpha\beta} \equiv \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \Rightarrow \text{Signature}\,(\Lambda_{\alpha\beta}) = (+,+,-,-).$$

The transformation law for more general contravariant twistors is

$$Z^{\alpha\beta\cdots\kappa\prime} = T^\alpha{}_\mu T^\beta{}_\nu \cdots T^\kappa{}_\rho Z^{\mu\nu\cdots\rho}.$$

In order to produce an invariant scalar product, we lower twistor indices as follows:

$$\overline{Z}_{\alpha\beta\cdots\kappa} \equiv \Lambda_{\alpha\mu}\Lambda_{\beta\nu}\cdots\Lambda_{\kappa\rho}\overline{Z^{\mu\nu\cdots\rho}} \Rightarrow \overline{Z}'_{\alpha\beta\cdots\kappa} Z^{\alpha\beta\cdots\kappa\prime} = \overline{Z}_{\alpha\beta\cdots\kappa} Z^{\alpha\beta\cdots\kappa}$$

We can also use the four-index permutation symbol $\epsilon^{\mu\nu\rho\sigma}$ to raise and lower antisymmetric index pairs, in this case without complex conjugation:

$$Z_{\alpha\beta} = \tfrac{1}{2}\epsilon_{\alpha\beta\mu\nu} Z^{\mu\nu} \Longleftrightarrow Z^{\mu\nu} = \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma} Z_{\rho\sigma}.$$

We say that an antisymmetric twistor $Z^{\alpha\beta}$ is real if $Z^{\alpha\beta} = \overline{Z}^{\alpha\beta}$. This rather eccentric definition of reality has an invariant meaning in $\mathcal{T}$,

$$Z^{\alpha\beta} = \overline{Z}^{\alpha\beta} \Rightarrow Z^{\alpha\beta\prime} = \overline{Z}^{\alpha\beta\prime},$$

which the ordinary definition of reality lacks.

To obtain a 2:1 mapping of $SU(2,2)$ onto $O_0(2,4)$, we must associate rank-one conformal space vectors Z^A, whose transformation law is linear in $C^A{}_B$, with rank two twistors $Z^{\alpha\beta}$, whose transformation law is bilinear in $T^\alpha{}_\beta$, so that $\pm T^\alpha{}_\beta \Leftrightarrow C^A{}_B$.[8] The only rank two twistors uniquely determined by six independent real parameters—the number supplied by the Z^A—are real antisymmetric twistors $Z^{\alpha\beta} = \overline{Z}^{\alpha\beta} = -Z^{\beta\alpha}$. Accordingly we assume a one-to-one correspondence between Z^A and $Z^{\alpha\beta}$,

$$Z^A = \Sigma^A_{\alpha\beta} Z^{\alpha\beta}, \tag{5}$$

which preserves the scalar product

$$Z^2 = Z^A Z_A = Z^{\alpha\beta} Z_{\alpha\beta}.$$

One can easily show that these assumptions require

(i) $\Sigma^A_{\alpha\beta} = -\Sigma^A_{\beta\alpha}$,

(ii) $\Sigma^A_{\alpha\beta} = \overline{\Sigma}^A_{\alpha\beta}$,

(iii) $\Sigma^A_{\alpha\beta}\Sigma^{\alpha\beta}_B = \delta^A_B$,

(iv) $\Sigma^A_{\alpha\beta}\Sigma^{\rho\sigma}_A = \tfrac{1}{2}(\delta^\rho_\alpha\delta^\sigma_\beta - \delta^\rho_\beta\delta^\sigma_\alpha)$.

We shall take (i)–(iv) as the defining properties of the $\Sigma^A_{\alpha\beta}$. These quantities, which provide the basic link between $\mathcal{T}$ and $\mathfrak{C}$, are analogs of the Pauli matrices $\tau^{AB\prime}_\mu$, which relate spinors to Minkowski space tensors.

If we now perform a coordinate transformation in $\mathcal{T}$,

$$Z^{\alpha\beta\prime} = T^\alpha{}_\rho T^\beta{}_\sigma Z^{\rho\sigma} = T^\alpha{}_\rho T^\beta{}_\sigma \Sigma^{\rho\sigma}{}_A Z^A,$$

and set

$$Z^{\alpha\beta\prime} = \Sigma^{\alpha\beta}_A Z^{A\prime} = \Sigma^{\alpha\beta}_A C^A{}_B Z^B,$$

we obtain

$$C^A{}_B = \Sigma^A_{\alpha\beta}\Sigma^{\rho\sigma}_B T^\sigma{}_\rho T^\beta{}_\sigma. \tag{6}$$

This gives the conformal space transformation $C^A{}_B$ associated with the twistor space transformations $T^\alpha_{\pm\beta} \equiv \pm T^\alpha{}_\beta$. One can check explicitly that, given any $T^\alpha{}_\beta \in SU(2,2)$ and any set of $\Sigma^A_{\alpha\beta}$ satisfying (i)-(iv), the matrix $C^A{}_B$ constructed according to the prescription (6) belongs to $O_0(2,4)$. Furthermore,

$$C(T_1)C(T_2) = C(T_1T_2).$$

Thus we have indeed secured the required homomorphism.

It is useful to extract from (6) an expression for T in terms of C. As a first step in this direction, we solve (6) for T in the case of infinitesimal transformations. In this case the nonlinearity of (6) presents no difficulties. If we write

$$T^\alpha{}_\beta = \delta^\alpha_\beta - i\epsilon G^\alpha{}_\beta,$$
$$C^A{}_B = \delta^A_B + \epsilon\Gamma^A{}_B,$$

with $G^\alpha{}_\beta$ tracefree and Hermitian (in the $\mathcal{T}$-invariant sense) and Γ^{AB} antisymmetric,

$$G^\alpha{}_\beta = \overline{G}_\beta{}^\alpha,$$
$$G^\alpha{}_\alpha = 0,$$
$$\Gamma^{AB} = -\Gamma^{BA},$$

we have

$$G^\alpha{}_\beta = \frac{i}{2}\Sigma_{AB}{}^\alpha{}_\beta\Gamma^{AB}, \tag{7}$$

$$\Gamma_{AB} = i\Sigma_{AB}{}^\alpha{}_\beta G^\beta{}_\alpha, \tag{8}$$

where

$$\Sigma_{AB}{}^\alpha{}_\beta \equiv 2[\Sigma^{\gamma\alpha}_A\Sigma_{B\gamma\beta} - \tfrac{1}{4}\Lambda_{AB}\delta^\alpha_\beta]. \tag{9}$$

The $\Sigma_{AB}{}^\alpha{}_\beta$ possess a number of interesting properties of their own, which follow directly from properties (i)–(iv) of the $\Sigma^A_{\alpha\beta}$:

(I). $\Sigma_{AB}{}^\alpha{}_\alpha = 0$,

(II). $\Sigma_{AB}{}^\alpha{}_\beta = -\Sigma_{BA}{}^\alpha{}_\beta = -\overline{\Sigma}_{AB\beta}{}^\alpha$,

(III). $[\Sigma_{AB}, \Sigma_{CD}] = \Lambda_{BC}\Sigma_{AD} - \Lambda_{BD}\Sigma_{AC} + \Lambda_{AD}\Sigma_{BC} - \Lambda_{AC}\Sigma_{BD}$ } Lie algebra for $O(2,4)$,

(IV). $\Sigma^\alpha{}_\beta, \Sigma^\rho{}_\sigma = \Sigma^\alpha{}_\sigma\delta^\rho_\beta - \Sigma^\rho{}_\beta\delta^\alpha$ } Lie algebra for $SU(2,2)$,

(V). $$\Sigma_{AB}\Sigma_{CD} = \frac{1}{4}\Lambda_{AD}\Lambda_{BC} - \frac{1}{4}\Lambda_{AC}\Lambda_{BD} + \frac{1}{2}\Lambda_{BC}\Sigma_{AD} - \frac{1}{2}\Lambda_{BD}\Sigma_{AC} + \frac{1}{2}\Lambda_{AD}\Sigma_{BC} - \frac{1}{2}\Lambda_{AC}\Sigma_{BD} \mp \frac{i}{4}\epsilon_{ABCDEF}\Sigma^{EF}.$$

In (III)–(V) we have employed matrix notation. The $\sum_{AB}$, with the twistor indices suppressed, represent 4×4 tracefree Hermitian matrices, and the $\sum^{\alpha}{}_{\beta}$, with the conformal space indices suppressed, represent 6×6 antisymmetric matrices. Without loss of generality we may write

$$\sum_{AB} = -\frac{i}{2}\begin{bmatrix} \sigma_{\mu\nu} & \sigma_{\mu} & \pm\gamma_{\mu} \\ -\sigma_{\mu} & 0 & \pm\gamma_5 \\ \mp\gamma_{\mu} & \mp\gamma_5 & 0 \end{bmatrix}. \tag{10}$$

By (I), (II), and (V) the $\gamma_{\mu}, \gamma_5, \sigma_{\mu}$, and $\sigma_{\mu\nu}$ which appear above must have all the defining properties of Dirac matrices. In particular,

$$\gamma_{\mu}\gamma_{\nu} + \gamma_{\nu}\gamma_{\mu} = 2\eta_{\mu\nu}, \quad \sigma_{\mu} = \frac{i}{2}[\gamma_{\mu}, \gamma_5],$$
$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3, \quad \sigma_{\mu\nu} = \frac{i}{2}[\gamma_{\mu}, \gamma_{\nu}].$$

Multiplication rules for more complicated products of γ matrices, such as those given by Macfarlane,[7] may be obtained either directly from (V) or by iteration.

The ambiguity of sign above is connected with the fact that, given an irreducible representation of $SU(2,2)$ by matrices $T^{\alpha}{}_{\beta}$, the matrices $\overline{T^{\alpha}{}_{\beta}}$ constitute a second irreducible representation of $SU(2,2)$ unitarily inequivalent to the first. According to (7), the second representation is obtained from the first by replacing $\sum_{AB}$ with $\overline{\sum_{AB}}$. The $\overline{\sum_{AB}}$ satisfy the same fundamental relations (I)–(V) as the $\sum_{AB}$, except that the sign of the last term in (V) is reversed. As derived, (V) allows for either representation by leaving the sign of that term arbitrary. Of course, no ambiguity occurs in the commutation relations (III) and (IV), as these reflect the structure of the groups themselves, without reference to their representations.

By (6) and (9) we can write for $C(T)$

$$C_{AB}(T) = \frac{1}{2}\mathrm{Tr}(\textstyle\sum_{AC}T)\,\mathrm{Tr}(\sum^{C}{}_{B}T) + \frac{1}{2}\mathrm{Tr}(\sum_{AB}T^2) - \frac{1}{2}\mathrm{Tr}\,T\,\mathrm{Tr}(\sum_{AB}T) + \frac{1}{8}\Lambda_{AB}[(\mathrm{Tr}\,T)^2 - 2\,\mathrm{Tr}\,T^2].$$

We now employ a covariant technique due to Macfarlane,[9] modified to apply in the present circumstances, to solve (6) for T. We first require the orthogonality relation

$$\textstyle\sum^{AB\alpha}{}_{\beta}\sum_{AB}{}^{\rho}{}_{\sigma} = \frac{1}{2}\,\delta^{\alpha}_{\beta}\delta^{\rho}_{\sigma} - 2\,\delta^{\alpha}_{\sigma}\delta^{\rho}_{\beta},$$

which follows from (i)–(iv) and (9). This result, when multiplied on both sides by $T^{\mu}{}_{\alpha}\overline{T}^{\beta}{}_{\rho}$ gives, in matrix notation,

$$2T(\mathrm{Tr}\,T^{-1}) = \frac{1}{2} - T\textstyle\sum^{AB}T^{-1}\sum_{AB},$$

where

$$(T^{-1})^{\alpha}{}_{\beta} \equiv \overline{T}^{\alpha}{}_{\beta}.$$

We may evaluate the last term on the right by means of (6):

$$T\textstyle\sum^{AB}T^{-1} = C^{RA}C^{SB}\sum_{RS}.$$

With the help of (V) we find

$$2T(\mathrm{Tr}\,T^{-1}) = \tfrac{1}{2} + \frac{1}{4}(\mathrm{Tr}\,C)^2 - \frac{1}{4}\mathrm{Tr}\,C^2 + (\mathrm{Tr}\,C) \times C_{AB}\textstyle\sum^{AB} - C^2_{AB}\sum^{AB} \mp \frac{i}{4}\epsilon_{ABCDEF} \times C^{ABC}C^{CD}\sum^{EF}. \tag{11}$$

Now by the unitarity of T,

$$\mathrm{Tr}\,T^{-1} = \overline{\mathrm{Tr}\,T}. \tag{12}$$

Taking the trace of (11) and inserting (12) we get

$$\mathrm{Tr}\,T = e^{i\theta}\sqrt{2}\left[\frac{1}{2} + \frac{1}{4}(\mathrm{Tr}\,C)^2 - \frac{1}{4}\mathrm{Tr}\,C^2\right],$$

which implies

$$T(\pm C) = e^{i\theta}\left\{\frac{2 + (\mathrm{Tr}\,C)^2 - \mathrm{Tr}\,C^2 + 4(\mathrm{Tr}\,C)C_{AB}\sum^{AB} - 4C^2_{AB}\sum^{AB} \mp i\epsilon_{ABCDEF}C^{ABC}C^{CD}\sum^{EF}}{4\sqrt{2}\,[2 + (\mathrm{Tr}\,C)^2 - \mathrm{Tr}\,C^2]^{1/2}}\right\}.$$

θ is determined by the condition $\det T = 1$ to be

$$\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}.$$

Thus if we set

$$T(C) = \frac{2 + (\mathrm{Tr}\,C)^2 - \mathrm{Tr}\,C^2 + 4(\mathrm{Tr}\,C)C_{AB}\sum^{AB} - 4C^2_{AB}\sum^{AB} \mp i\epsilon_{ABCDEF}C^{ABC}C^{CD}\sum^{EF}}{4\sqrt{2}\,[2 + (\mathrm{Tr}\,C)^2 - \mathrm{Tr}\,C^2]^{1/2}}, \tag{13}$$

we may write

$$T_{\pm}(C) = \pm\,T(C), \tag{14}$$

$$T_{\pm}(-C) = \pm\,iT(C). \tag{15}$$

Equations (13) and (14) establish the 2 : 1 homomorphism between $SU(2,2)$ and $O_0(2,4)$; while (13)–(15), coupled with the results of Sec. II, give the 4 : 1 homomorphism between $SU(2,2)$ and $\mathfrak{C}$. [The general formula (13) gives an indeterminate result when the denominator vanishes. In practice, however, this case presents no difficulty, since for those $C \in O_0(2,4)$ with $2 + (\mathrm{Tr}\,C)^2 - \mathrm{Tr}\,C^2 = 0$, we can define $T(C)$ by continuity.]

IV. $SU(2,2)$ AND $\mathfrak{C}$

We now study the 4 : 1 homomorphism between $SU(2,2)$ and $\mathfrak{C}$ in greater detail. In doing so it will be convenient to introduce a specific set of $\sum^{A}_{\alpha\beta}$, namely

$$\textstyle\sum^{M}_{\alpha\beta} \equiv \frac{i}{\sqrt{2}}\begin{bmatrix} 0 & -\tau^{\mu}{}_{A}{}^{B'} \\ \rho^{\mu A'}{}_{B} & 0 \end{bmatrix} \quad (M = \mu),$$
$$\textstyle\sum^{4}_{\alpha\beta} \equiv \frac{1}{2}\begin{bmatrix} \epsilon_{AB} & 0 \\ 0 & -\epsilon^{A'B'} \end{bmatrix}, \quad \sum^{5}_{\alpha\beta} \equiv \frac{1}{2}\begin{bmatrix} \epsilon_{AB} & 0 \\ 0 & \epsilon^{A'B'} \end{bmatrix}. \tag{16}$$

Here the τ_{μ} are the ordinary 2×2 Pauli matrices,

$$\tau_0 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \tau_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\tau_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \tau_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and

$$\rho_{\mu A'B} = \overline{\tau_{\mu AB'}} = \epsilon_{AC}\epsilon_{B'D'}\overline{\tau_{\mu}^{CD'}}.$$

When inserted into (9) and (10), (16) implies a particular representation for the Dirac γ matrices:

$$\gamma_\mu = \sqrt{2}\begin{bmatrix} 0 & \tau_\mu^{AB'} \\ \rho_{\mu A'B} & 0 \end{bmatrix}.$$

For simplicity we have chosen the top sign in (V). With these conventions, we list below the twistor space transformations which correspond to the Lorentz, translation, acceleration, and dilation subgroups of $\mathfrak{C}$.

Lorentz Transformations:

$$T(C) = \frac{2(\mathrm{Tr}\,\xi + 1) + \frac{1}{2}(\mathrm{Tr}\,\xi)^2 - \frac{1}{2}\mathrm{Tr}\,\xi^2 - i(\mathrm{Tr}\,\xi + 2)\xi_{\mu\nu}\sigma^{\mu\nu} + i\xi^2_{\mu\nu}\sigma^{\mu\nu} - \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\xi^{\mu\nu}\xi^{\rho\sigma}\gamma^5}{4[2(\mathrm{Tr}\,\xi + 1) + \frac{1}{2}(\mathrm{Tr}\,\xi)^2 - \frac{1}{2}\mathrm{Tr}\,\xi^2]^{1/2}},$$

$$T(C) = \begin{bmatrix} \xi^A{}_B & 0 \\ 0 & -\bar{\xi}_{A'}{}^{B'} \end{bmatrix}, \tag{17}$$

$$\xi^A{}_B \equiv \frac{\xi_{\mu\nu}\tau^{\mu AC'}\rho^\nu_{C'B}}{[\xi_{\alpha\delta}\,\xi_{\kappa\lambda}\tau^\alpha\rho^\beta\tau^\kappa\rho^\lambda]^{1/2}}, = \frac{2\xi_{\mu\nu}\tau^{\mu AC'}\rho^\nu_{C'B}}{[(\mathrm{Tr}\,\xi)^2 - \mathrm{Tr}\,\xi^2 + 4 + i\epsilon_{\alpha\beta\kappa\lambda}\xi^{\alpha\beta}\xi^{\kappa\lambda}]^{1/2}} \in SL(2, C);$$

Translations:

$$T(C) = \mathrm{I} - \frac{i}{2\sqrt{2}}[b^\mu\gamma_\mu(\mathrm{I} + i\gamma^5)].$$

$$T(C) = \begin{bmatrix} \mathrm{I}^A{}_B & 0 \\ -ib_{A'B} & \mathrm{I}_{A'}{}^{B'} \end{bmatrix} \quad b_{A'B} \equiv b_\mu\rho^\mu_{A'B}; \tag{18}$$

Accelerations:

$$T(C) = I - \frac{i}{2\sqrt{2}}[a^\mu\gamma_\mu(\mathrm{I} - i\gamma^5)],$$

$$T(C) = \begin{bmatrix} \mathrm{I}^A{}_B - ia^{AB'} \\ 0 & \mathrm{I}_{A'}{}^{B'} \end{bmatrix} \quad a^{AB'} \equiv a^\mu\tau_\mu^{AB'}; \tag{19}$$

Dilations:

$$T(C) = \cosh\frac{\omega}{2} - i\gamma^5\sinh\frac{\omega}{2},$$

$$T(C) = \begin{bmatrix} \Omega^{-1/2}I^A{}_B & 0 \\ 0 & \Omega^{1/2}\mathrm{I}_{A'}{}^{B'} \end{bmatrix}, \Omega \equiv e^\omega \tag{20}$$

We see that (13) and (14) yield, as a special case, the 2 : 1 homomorphism between elements of $SL(2, C)$ and the restricted Lorentz group.[9] This relationship, as expressed in (17), is responsible for the existence of a spinor decomposition of twistors.[3] If we set

$$L^\alpha \equiv (\lambda^A, \mu_{A'}),$$

then

$$L^{\alpha'} = T(\xi)^\alpha{}_\beta L^\beta$$

gives the correct transformation law for the spinors λ^A and $\mu_{A'}$,

$$\tilde{\lambda}^A = \xi^A{}_B\lambda^B,$$

$$\tilde{\mu}_{A'} = -\bar{\xi}_{A'}{}^{B'}\mu_{B'},$$

so long as T corresponds to a pure Lorentz transformation. For general $T(C)$, however, this is no longer true, and as a result no conformally invariant identification of twistors with spinor pairs is possible.

V. APPLICATIONS

As an application of the above work, we show in this section that every continuous conformal transformation may be represented as the product of a Lorentz transformation, a translation, an acceleration, a dilation, and one of the following four special conformal transformations: (a) the identity transformation $x^{\mu'} = x^\mu$; (b) the inversion

$$t' = -\frac{2t}{x_\mu x^\mu}, \quad x^a = \frac{2x^a}{x_\mu x^\mu} \quad (a = 1, 2, 3);$$

(c), (d) the transformations

$$t' = -\frac{1 - \frac{1}{2}x_\mu x^\mu}{t \pm z}, \quad x' = \frac{-\sqrt{2}x}{t \pm z},$$

$$z = \mp\frac{1 + \frac{1}{2}x_\mu x^\mu}{t \pm z}, \quad y' = \frac{-\sqrt{2}y}{t \pm z}.$$

The twistor space analog of this decomposition is, for $T(C) \in SU(2, 2)$,

$$T(C) = T(L)T(T)T(A)T(D)T_s,$$

where $T(L)$, $T(T)$, $T(A)$, and $T(D)$ have the forms of (17), (18), (19), and (20), respectively, and $T_s \in SU(2, 2)$ ($s = 1, 2, 3, 4$) is one of the four transformation matrices shown below:

$$T_1 = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathrm{I} \end{bmatrix}, \quad T_2 = \begin{bmatrix} 0 & \mathrm{I} \\ \mathrm{I} & 0 \end{bmatrix},$$

$$T_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ 0 & i & 0 & 0 \end{bmatrix}, \quad T_4 = \begin{bmatrix} 0 & 0 & i & 0 \\ 0 & 1 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

To obtain this decomposition we set

$$T(C) = \begin{bmatrix} \underset{1}{t}{}^A{}_B & \underset{2}{t}{}^{AB'} \\ \underset{3}{t}{}_{A'B} & \underset{4}{t}{}_{A'}{}^{B'} \end{bmatrix}.$$

If $D_1(C) \equiv \det \underset{1}{t}(C) \neq 0$, one can write $T(C) = T(L)T(T)T(A)T(D)$, where

$$T(L) = \begin{bmatrix} \underset{1}{t}{}^A{}_B(\det \underset{1}{t})^{-1/2} & 0 \\ 0 & -\underset{1}{\bar{t}}{}_{A'}{}^{B'}(\det \underset{1}{t})^{-1/2} \end{bmatrix}, \tag{21}$$

$$T(T)=\begin{bmatrix} \mathrm{I}^{A}{}_{B} & 0 \\ -\underset{1}{\bar t}{}^{C'}{}_{A'}\,\underset{3}{t}{}_{C'B}(\det \underset{1}{t})^{-1} & \mathrm{I}_{A'}{}^{B'} \end{bmatrix}, \tag{22}$$

$$T(A)=\begin{bmatrix} \mathrm{I}^{A}{}_{B} & \underset{1}{t}{}^{CA}\underset{2}{t}{}_{C}{}^{B'} \\ 0 & \mathrm{I}_{A'}{}^{B'} \end{bmatrix}, \tag{23}$$

$$T(D)=\begin{bmatrix} (\det \underset{1}{t})^{1/2}\,\mathrm{I}^{A}{}_{B} & 0 \\ 0 & (\det \underset{1}{t})^{-1/2}\,\mathrm{I}_{A'}{}^{B'} \end{bmatrix}. \tag{24}$$

Now $T(C) \in SU(2,2)$ implies that $D_1 = \overline{D}_1$. Furthermore, if $D_1(C) < 0$, then $D_1(-C) = -D_1(C) > 0$ by (15), and so we can always arrange for $(D_1)^{1/2}$ to be real. The unitarity of T then implies that (21), (22), (23), and (24) have all the properties of (17), (18), (19), and (20), respectively.

Suppose next that $D_1 = 0$. When this occurs we construct $\tilde{T} \in SU(2,2)$, with $\det \underset{1}{\tilde{t}} \neq 0$, so that our previous argument applies with T replaced by $\tilde{T}$. If $D_2 \equiv T^0{}_2 T^1{}_3 - T^0{}_3 T^1{}_2 \neq 0$, then

$$\tilde{T}_2 \equiv T T_2^{-1} \Rightarrow \det \underset{1}{\tilde{t}} = D_2 \neq 0.$$

If $D_1 = D_2 = 0$, $D_3 \equiv T^0{}_0 T^1{}_3 - T^1{}_0 T^0{}_3 \neq 0$, then

$$\tilde{T}_3 \equiv T T_3^{-1} \rightarrow \det \underset{1}{\tilde{t}} = D_3 \neq 0.$$

Finally, if $D_1 = D_2 = D_3 = 0$, $T \in SU(2,2)$ implies $D_4 \equiv T^0{}_1 T^1{}_2 - T^0{}_2 T^1{}_1 \neq 0$, and we define

$$\tilde{T}_4 \equiv T T_4^{-1} \Rightarrow \det \underset{1}{\tilde{t}} = i D_4 \neq 0.$$

We now have in general

$$\tilde{T}_s = T T_s^{-1} = T(L)\,T(T)\,T(A)\,T(D),$$

which completes our argument.

It is also possible, with the aid of the formalism of Sec. II–IV, to establish conformally invariant relationships between geometric objects in $\mathcal{T}$, $\mathcal{C}$, and $\mathfrak{M}$. The $\sum^{A}_{\alpha\beta}$ define one such correspondence between real antisymmetric twistors $Z^{\alpha\beta}$ and conformal space vectors Z^A [cf. (5)]. We can further associate the Z^A with sets of points x^μ in (compactified) Minkowski space which satisfy the invariant condition

$$Z_A K^A(x) = 0.$$

This equation defines a hypersphere or a hyperplane in $\mathfrak{M}$, depending on the particular form of Z^A. If $Z^{\alpha\beta} Z_{\alpha\beta} = Z^A Z_A = 0$, the corresponding geometric object in $\mathfrak{M}$ is a null cone (a degenerate hypersphere). The vertex of the cone is the point in $\mathfrak{M}$ associated with the null direction Z^A via (3). Thus real simple twistors $Z^{\alpha\beta}$ correspond to null vectors Z^A in $\mathcal{C}$, and these in turn define points z^μ in $\mathfrak{M}$ (cf. Ref. 3).

A similar series of relationships between $\mathcal{T}$, $\mathcal{C}$, and $\mathfrak{M}$ is based on the $\sum_{AB}{}^{\alpha}{}_{\beta}$. Equations (7) and (8) associate tracefree Hermitian twistors $G^\alpha{}_\beta$ with real antisymmetric conformal space tensors Γ_{AB}. Alternately, given a twistor L^α we can construct a real antisymmetric conformal space tensor $\Gamma_{AB}(L)$ as follows:

$$\Gamma_{AB}(L) \equiv i\bar{L}_\alpha \textstyle\sum_{AB}{}^{\alpha}{}_{\beta} L^\beta. \tag{25}$$

From (25),

$$\Gamma_{AB}(L)\,\Gamma^{BC}(L) = -\frac{1}{4} L^2 \delta^C_A \tag{26}$$

$$\Gamma_{[AB}(L)\,\Gamma_{C]D}(L) = \frac{1}{12} L^2 \epsilon_{ABCDEF} \Gamma^{EF}(L). \tag{27}$$

According to (27), $\Gamma_{AB}(L)$ is simple if and only if L^α is null (10). In this case

$$\Gamma_{AB}(L) = U_{[A} V_{B]},$$

where (26) (with $L^2 = 0$) gives

$$U^2 = V^2 = U \cdot V = 0.$$

The null directions U^A and V^A in $\mathcal{C}$ define points u^μ and v^μ in $\mathfrak{M}$ via (3). The line in $\mathfrak{M}$ joining u^μ and v^μ is null:

$$U(u^\mu) \cdot V(v^\mu) = 0 \Rightarrow (u^\mu - v^\mu)^2 = 0.$$

Thus we conclude that there is a conformally invariant correspondence between null twistors, simple antisymmetric conformal space tensors satisfying (26) (with $L^2 = 0$), and null lines in $\mathfrak{M}$ (cf. Ref. 3).

VI. CONCLUSION

Since $SU(2,2)$ and $O_0(2,4)$ are 4 : 1 and 2 : 1 homomorphic to $\mathfrak{C}$, respectively, every conformally invariant Minkowski space statement must have conformal space and twistor space analogs. In several respects it is easier to deal with the conformal space and twistor space expressions than with the original Minkowski space forms. For one thing, conformal transformations are linear transformations in $\mathcal{T}$ and $\mathcal{C}$, but not in $\mathfrak{M}$. In addition, tensor equations in $\mathcal{T}$ and $\mathcal{C}$ are *manifestly* conformally covariant, while conformal covariance in $\mathfrak{M}$ is not so obvious. Thus the abstract space $\mathcal{T}$ and $\mathcal{C}$ have an important role to play in physics whenever conformal symmetries occur.

ACKNOWLEDGMENT

I am grateful to Dr. Lochlainn O'Raifeartaigh, Director of the Dublin Institute for Advanced Studies, for his help in clarifying and interpreting several aspects of the above work.

[1]E. Cartan, Ann. Ecole Norm. Sup. 31, 263 (1914).
[2]R. Brauer and H. Weyl, Am. J. Math. 57, 425 (1935).
[3]R. Penrose, J. Math. Phys. 8, 345 (1967).
[4]R. Penrose, Int. J. Theor. Phys. 1, 61 (1968).
[5]E. Grgin, J. Math. Phys. 9, 1595 (1968).
[6]W. A. Hepner, Nuovo Cimento 26, 4183 (1962).
[7]A. J. Macfarlane, Commun. Math. Phys. 2, 133 (1966).
[8]I. Segal, Proc. Natl. Acad. Sci. (USA), 57, 194 (1967).
[9]A. J. Macfarlane, J. Math. Phys. 3, 1116 (1962).
[10]J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954).

On the higher order theories of piezoelectric crystal surfaces

M. Cengiz Dökmeci *

Thurston Hall, Cornell University, Ithaca, New York 14850
(Received 8 August 1972)

This paper presents a higher order theory of crystal finite surfaces within the frame of the three-dimensional theory of linear piezoelectricity. First, by modifying Hamilton's principle, a variational theorem is deduced. Then, this theorem together with a method of series expansion is employed to establish the theory in a systematic and consistent manner. The theory consists of a hierarchy of two-dimensional equations of motion, charge equations of electrostatics, initial and boundary conditions, strain-displacement and electric field-electric potential relations, and macroscopic constitutive equations. It governs the extensional and flexural as well as torsional motions of piezoelectric cyrstal shells and plates of uniform thickness. Further, theorems of uniqueness in this theory are presented.

1. INTRODUCTION

Since the discovery of piezoelectricity by the Curie brothers[1] in 1880, the theory of piezoelectric crystals has been well developed and employed.[2,3] With this theory, a large class of applications has been solved in the literature, in particular, following the Second World War. Earlier works have been reviewed in Refs. 2-5. Most of them are concerned with the vibrations of thin rods and planes for both finite and infinite cases. These papers are, in general, restricted to the extremely low-frequency vibrations. For high frequency vibrations of piezoelectric crystals, we mention the works of Mindlin and his co-workers for finite planes, whose investigations have recently been elaborated by Tiersten[6] in a comprehensive monograph, and Dökmeci[7] for finite bars. However, with the exception of some particular problems[8] solved for the case of lower frequency, the vibrations of piezo-electric crystal surfaces are not touched.

Our aim in the present work is to establish a linear theory of piezoelectric crystal finite surfaces, valid for high as well as low frequency vibrations, and to examine the uniqueness of its solution.

In this paper, we set up the theory in the following way. In Sec. 2 we summarize the basic equations of piezoelectric crystals necessary for the subsequent development. Section 3 is devoted to the variational formulation of the field equations and the mixed boundary conditions. Accordingly, a variational theorem is directly deduced from Hamilton's principle as a first step toward the theory. Section 4 is concerned with the geometry and kinematics of a finite surface with no singularities of any type. This surface is considered to describe all types of behavior of thin shells and plates; in fact, all the governing equations are expressed in terms of the quantities referred to this surface. The displacement components and the electric potential are expanded in power series. These expansions imply series distributions for stresses, displacements and electric field. In Sec. 5, by means of the variational theorem together with the assumed electric potential and displacement field, the higher order field equations and natural boundary conditions of piezoelectric crystal shells are consistently established in the same spirit as those of Mindlin and Tiersten. These equations are then supplemented with the appropriate initial conditions and macroscopic constitutive relations. In Sec. 6, we study the uniqueness of solution of the initial mixed boundary value problem described by the governing equations of the theory, and enumerate the conditions to ensure the uniqueness.

Some general conclusions regarding the theory are drawn in the last section.

Notation

In the subsequent development, we use standard space and surface tensors in a Euclidean 3-space of normal coordinates. Latin and Greek indices are used respectively to designate space and surface tensors. Accordingly, Latin indices take the values 1, 2, 3 and Greek indices the values 1, 2. Einstein's summation convention is implied for all the repeated Latin and Greek indices. A comma and a superposed dot denote respectively partial differentiation with respect to the indicated variable and time. Using respectively the space and surface metrics, a semicolon and a stroke stand for covariant total differentiation with respect to the indicated coordinate. Further, we employ a star to designate the prescribed quantities and an overbar to refer to the quantities which are belong to the midsurface.

2. GENERAL EQUATIONS FOR THE LINEAR THEORY OF PIEZOELECTRICITY

Referring the motion of nonpolar continuum to a fixed system of general curvilinear coordinates, the equations of local balance of momenta are

$$\tau^{ij};_i + f^j = \rho b^j \quad \text{in } \mathcal{V} \times [t_0, \infty), \tag{2.1}$$

$$\epsilon_{ijk}\tau^{jk} = 0 \quad \text{in } \mathcal{V} \times [t_0, \infty). \tag{2.2}$$

Here, $\tau^{ij} = \tau^{ji}$ is the spatial (contravariant) components of the stress tensor, ρ the mass density, ϵ_{ijk} the alternating tensor, and f^j and b^j denote the spatial components of the body force and acceleration vectors, respectively. The stress tensor τ^{ij}, across a surface whose unit outward normal vector is $\mathbf{n}$, is related to the traction vector t^i:

$$t^j = n_i \tau^{ij}. \tag{2.3}$$

The charge equation of electrostatics may be written as

$$D^i{}_{;i} = 0 \quad \text{in } \mathcal{V} \times [t_0, \infty), \tag{2.4}$$

where D^i is the component of electric displacement. In Eqs. (2.1)-(2.4), t_0 denotes some prescribed value and/or the initial instant of time t, and $\mathcal{V}$ the volume of the body with its boundary surface $\mathcal{S}$.

The constitutive equations of piezoelectricity are given by

$$\tau^{ij} = C^{ijkl}S_{kl} - C^{kij}E_k, \qquad (2.5)$$

$$D^i = C^{ijk}S_{jk} + C^{ij}E_j, \qquad (2.6)$$

where S_{ij} represents the components of the infinitesimal strain tensor, E_i the components of the quasistatic electric field, and C^{ijkl}, C^{ijk}, and C^{ij} are the components of elastic stiffness, piezoelectric strain constant and dielectric permittivity, respectively. By virtue of Eqs. (2.5) and (2.6), the equations of linear elastodynamics are coupled to the charge equation of electrostatics. The following symmetry relations hold

$$C^{ijkl} = C^{jikl} = C^{klij} = C^{ijlk}, \quad C^{ijk} = C^{ikj}, \quad C^{ij} = C^{ji} \qquad (2.7)$$

for the material coefficients.

The components of electric field and mechanical strain are expressed in terms of the mechanical displacement vector $\mathbf{u}$ and the electric potential φ by

$$S_{ij} = \tfrac{1}{2}(u_{i;j} + u_{j;i}), \qquad (2.8)$$

$$E_i = -\varphi_{,i}. \qquad (2.9)$$

With the help of Eqs. (2.1), (2.4)-(2.6) and (2.8)-(2.9), we readily get the equations

$$f^j + C^{ijkl}u_{k;li} + C^{kij}\varphi_{,ki} - \rho\ddot{u}^j = 0,$$
$$C^{kij}u_{i;jk} - C^{ij}\varphi_{,ij} = 0, \qquad (2.10)$$

which govern the electric potential φ and the displacement $\mathbf{u}$.

Let $\mathcal{S}$, $\mathcal{S}_N$, and $\mathcal{S}_C$, respectively, stand for the entire boundary surface of the body, the portion of $\mathcal{S}$ on which the t^i and/or the surface charge σ are prescribed, and the portion of $\mathcal{S}$ on which the u^i and/or φ are prescribed. Hence, the boundary conditions may be expressed by

$$t^j_* - n_i\tau^{ij} = 0, \quad \sigma_* - n_iD^i = 0 \quad \text{on } \mathcal{S}_N \times [t_0, \infty), \qquad (2.11)$$

$$u^i_* - u^i = 0, \quad \varphi_* - \varphi = 0 \quad \text{on } \mathcal{S}_C \times [t_0, \infty), \qquad (2.12)$$

and

$$\mathcal{S}_N \cap \mathcal{S}_C = 0, \quad \mathcal{S}_N \cup \mathcal{S}_C = \mathcal{S}. \qquad (2.13)$$

The aforementioned equations together with the initial conditions in the volume $\mathcal{V}$, namely,

$$\mathbf{u}(\theta^i, t_0) = \mathbf{v}^*(\theta^i), \quad \dot{\mathbf{u}}(\theta^i, t_0) = \mathbf{w}^*(\theta^i) \text{ in } \mathcal{V}(t_0)^*,$$
$$\mathbf{E}(\theta^i, t_0) = \mathbf{E}^*(\theta^i) \qquad (2.14)$$

completely describe the initial-mixed boundary value problem of interest. (The symbol $\mathcal{V}(t)$ refers to the volume $\mathcal{V}$ at time t.) As subsequently shown, the above initial and boundary conditions are sufficient to ensure a unique solution provided that a positive-definite internal energy function U exists.

Lastly, we define the internal energy density by

$$U = \tfrac{1}{2}\tau^{ij}S_{ij} + \tfrac{1}{2}D^iE_i, \qquad (2.15)$$

the electric enthalpy density H by

$$H = \tfrac{1}{2}\tau^{ij}S_{ij} - \tfrac{1}{2}D^iE_i = U - D^iE_i \qquad (2.16)$$

so that

$$\tau^{ij} = \frac{1}{2}\left(\frac{\partial H}{\partial S_{ij}} + \frac{\partial H}{\partial S_{ji}}\right), \quad D^i = -\frac{\partial H}{\partial E_i} \qquad (2.17)$$

and the kinetic energy density K by

$$K = \tfrac{1}{2}\rho\dot{u}^i\dot{u}_i \qquad (2.18)$$

for future convenience.

Before closing this section, we note that the governing equations of linear piezoelectricity, (2.1)-(2.8), are essentially contained in Ref. 6 but in Cartesian coordinates.

3. VARIATIONAL FORMULATION

For later use, we now proceed to formulate a variational theorem by means of Hamilton's principle.[9] First, we state a generalized version of Hamilton's principle by

$$\delta J = \delta\int_{t_0}^{t_1}\mathcal{L}dt = 0 \qquad (3.1a)$$

with

$$\mathcal{L} = \mathcal{K} - \Omega + \mathcal{H}. \qquad (3.1b)$$

Here, $\mathcal{L}$ is the Lagrangian function, $\mathcal{K}$ the total kinetic energy, Ω the total enthalpy and $\mathcal{H}$ the virtual work due to the external forces. For a piezoelectric body $\mathcal{V} + \mathcal{S}$, having no singularities of any kind and subject to the prescribed surface tractions and charges, $\mathcal{K}$, Ω and $\mathcal{H}$ may be expressed by

$$\mathcal{K} = \int_{\mathcal{V}}Kdv, \quad \Omega = \int_{\mathcal{V}}Hdv,$$
$$\mathcal{H} = \int_{\mathcal{V}}f^iu_idv + \int_{\mathcal{S}_C}[t^i(u_i - u_i^*) + \sigma(\varphi - \varphi_*)]ds$$
$$+ \int_{\mathcal{S}_N}(t^i_*u_i + \sigma_*\varphi)ds. \qquad (3.2)$$

Next, to establish the variational theorem, we carry out the indicated variations in each term of Eq. (3.1). Hence, the variation of the first term is found to be

$$\delta\int_{t_0}^{t_1}\mathcal{K}dt = \delta\int_{t_0}^{t_1}dt\int_{\mathcal{V}}\tfrac{1}{2}\rho\dot{u}^i\dot{u}_idv$$
$$= \int_{\mathcal{V}}[\rho\dot{u}^i\delta u_i]_{t_0}^{t_1}dv - \int_{t_0}^{t_1}dt\int_{\mathcal{V}}\rho\ddot{u}^i\delta u_idv. \qquad (3.3a)$$

This may be written in the form

$$\delta\int_{t_0}^{t_1}\mathcal{K}dt = -\int_{t_0}^{t_1}dt\int_{\mathcal{V}}\rho\ddot{u}^i\delta u_i\,dv \qquad (3.3b)$$

since δu_i vanishes at t_0 and t_1, as is customary in the use of Hamilton's principle.

The variation of the second term is

$$\delta\int_{t_0}^{t_1}\Omega dt = \delta\int_{t_0}^{t_1}dt\int_{\mathcal{V}}H(S_{ij}, E_i)dv$$
$$= \delta\int_{t_0}^{t_1}dt\int_{\mathcal{V}}\left(\frac{\partial H}{\partial S_{ij}}\delta S_{ij} + \frac{\partial H}{\partial E_i}\delta E_i\right)dv. \qquad (3.4a)$$

Considering Eqs. (2.8), (2.13) and (2.17), and using Green's transformation, we finally obtain the following equation.

$$\delta\int_{t_0}^{t_1}\Omega dt = \int_{t_0}^{t_1}dt\int_{\mathcal{S}_N}n_i(\tau^{ij}\delta u_j + D^i\delta\varphi)dS$$
$$-\int_{t_0}^{t_1}dt\int_{\mathcal{V}}[\tau^{ij};i\,\delta u_j + D^i;i\,\delta\varphi]dv. \qquad (3.4b)$$

Lastly, the variation of the third term is of the form

$$\delta\int_{t_0}^{t_1}\mathcal{H}dt = \int_{t_0}^{t_1}dt\int f^i\delta u_idv + \int_{t_0}^{t_1}dt\int (t^i_*\delta u_i + \sigma_*\delta\varphi)dS$$
$$+ \int_{t_0}^{t_1}dt\int_{\mathcal{S}_C}[(u_i - u_i^*)\delta t^i + (\varphi - \varphi_*)\delta\sigma]dS. \qquad (3.5)$$

Substituting Eqs. (3.3)-(3.5) into Eq. (3.1) and combining the volume and surface integrals, we arrive at the equation

$$\delta J = \int_{t_0}^{t_1} \int_{\mathcal{V}} [(\tau^{ij};i + f^j - \rho\ddot{u}^j)\delta u_j + (D^i;i)\delta\varphi]\, dv$$
$$+ \int_{t_0}^{t_1} dt \Big(\int_{\mathcal{S}_N} [(t_*^j - n_i \tau^{ij})\delta u_j + (\sigma_* - n_i D^i)\delta\varphi] dS$$
$$+ \int_{\mathcal{S}_C} [(u_i - u_i^*)\delta t^i + (\varphi - \varphi_*)\delta\sigma] dS \Big) = 0. \qquad (3.6)$$

By the use of the fundamental lemma of the calculus of variations the following theorem is concluded.

Theorem: Give a regular region[10] of space $\mathcal{V} + \mathcal{S}$ with boundary $\mathcal{S}(\mathcal{S}_N \cup \mathcal{S}_C = \mathcal{S})$ in a Euclidean 3-space, and defined all the boundary value problems which admit the functional J has zero first variation for all admissible variations of the field quantities; if and only if $u_i \in C^{(0,2)}, D_i \in C^{(1,0)}, \tau^{ij} \in C^{(1,0)}$, and $\varphi \in C^{(0,0)}$ satisfy Eqs. (2.1), (2.4), (2.11) and (2.12) as appropriate Euler equations. [$C^{(m,n)}$ represents the functions with derivatives of order up to and including (m) and (n) with respect to θ^i and t, respectively.]

Similar theorems have also been discussed, among others, by Tiersten[6,11] and references therein.

4. GEOMETRY OF THE FINITE SURFACE. KINEMATICS

Consider a piezoelectric finite surface embedded in a Euclidean 3-space. Let θ^i denote a system of right-handed geodesic normal coordinates[12] in this space. $\theta^3 = 0$ defines the surface which coincides with the midsurface $\mathcal{A}$ of shell. The θ^1- and θ^2-curve are situated on $\mathcal{A}$. The upper face of the shell $\mathcal{A}_u$ and its lower face $\mathcal{A}_l$ are indicated by the equations $\theta^3 = h$ and $\theta^3 = -h$, respectively. The edge boundary of the shell $\mathcal{S}_e$ is a right cylindrical surface with generators perpendicular to $\mathcal{A}$, and it intersects $\mathcal{A}$ along a Jordan curve $\mathcal{C}$. The metric tensor at any point of the shell space $\mathcal{V}$ is associated with that of $\mathcal{A}$ by the relations

$$g_{\alpha\beta} = \mu_\alpha^\nu \mu_\beta^\lambda a_{\nu\lambda}, \quad g_{\alpha 3} = 0, \quad g_{33} = 1 \qquad (4.1a)$$

with

$$\mu_\beta^\alpha = \delta_\beta^\alpha - \theta^3 b_\beta^\alpha, \qquad (4.1b)$$

where $a_{\alpha\beta}$ and $b_{\alpha\beta}$ denote the covariant components of the first and second fundamental forms of $\mathcal{A}$. Its third fundamental form is defined by

$$C_{\alpha\beta} = b_{\alpha\sigma} b_\beta^\sigma. \qquad (4.2)$$

The elements of volume dv, of Surface dS on $\mathcal{S}$, of area dA on $\mathcal{A}$, and of line ds along $\mathcal{C}$ are of the forms

$$dv = \sqrt{g}\, d\theta^1 d\theta^2 d\theta^3 = dS d\theta^3 = \mu dA d\theta^3,$$
$$n_\alpha dS = \mu \nu_\alpha ds d\theta^3, \quad \mu = |\mu_\beta^\alpha| = (g/a)^{1/2},$$
$$a = |a_{\alpha\beta}|, \quad g = |g_{ij}| \qquad (4.3)$$

in which ν is the unit vector normal to $\mathcal{C}$.

The shifted[13,14] displacement components of a generic point in $\mathcal{V}$ are represented by

$$\bar{u}_i(\theta^j, t) = \sum_{n=0}^{N} P_n(\theta^3) u_i^{(n)}(\theta^\alpha, t) \qquad (4.4)$$

and the electric potential by

$$\varphi(\theta^i, t) = \sum_{n=0}^{N} Q_n(\theta^3) \varphi^{(n)}(\theta^\alpha, t). \qquad (4.5)$$

From the mathematical standpoint, a separation of variables solution is sought for the aforementioned field equations. Therefore, the functions in Eqs. (4.4) and (4.5) are unknown *a priori* and independent functions defined in $\mathcal{V}$. Further, we assume that $\varphi^{(n)}$ and $u_i^{(n)}$ exist and are functions of class $C^{(1,0)}$ and $C^{(2,2)}$, respectively. The functions P_n and Q_n of the form

$$P_n = Q_n = (\theta^3)^n \qquad (4.6)$$

are to be used in this analysis.

In view of Eqs. (2.8), (2.9) and (4.4)-(4.7), we obtain the strain distribution as

$$S_{ij} = \sum_{n=0}^{N} (\theta^3)^n S_{ij}^{(n)}(\theta^\alpha, t) \qquad (4.7a)$$

with

$$S_{\alpha\beta}^{(n)} = \tfrac{1}{2}[u_{\alpha|\beta}^{(n)} + u_{\beta|\alpha}^{(n)} - 2b_{\alpha\beta} u_3^{(n)}$$
$$- (b_\alpha^\nu u_{\nu|\beta}^{(n-1)} + b_\beta^\nu u_{\nu|\alpha}^{(n-1)} - 2C_{\alpha\beta} u_3^{(n-1)})],$$
$$S_{\alpha 3}^{(n)} = \tfrac{1}{2}[(n+1) u_\alpha^{(n+1)} + u_{3,\alpha}^{(n)} - (n-1) b_\alpha^\nu u_\nu^{(n)}], \qquad (4.7b)$$
$$S_{33}^{(n)} = (n+1) u_3^{(n+1)},$$

and the electric field as

$$E_i = \sum_{n=0}^{N} (\theta^3)^n \cdot E_i^{(n)}(\theta^\alpha, t) \qquad (4.8a)$$

with

$$E_\alpha^{(n)} = -\varphi_{,\alpha}^{(n)}, \quad E_3^{(n)} = -(n+1)\varphi^{(n+1)}. \qquad (4.8b)$$

Here, $u_i^{(n)}, \varphi^{(n)}, S_{ij}^{(n)}$ and $E_i^{(n)}$ are henceforth termed the displacement, electric potential, mechanical strain and electric field components of order n, respectively.

5. GOVERNING EQUATIONS OF PIEZOELECTRIC CRYSTAL SURFACES

In this section, the variational theorem (3.6) together with the series expansions (4.4)-(4.6) is applied to establish the macroscopic field equations and the natural boundary conditions of piezoelectric crystal finite surfaces. These equations are then supplemented with the appropriate initial conditions and constitutive equations.

Before proceeding further, we define the stress, body force, electric displacement, and effective load resultants of order n by

$$T_{(n)}^{ij} = T_{(n)}^{ji} = \int_{-h}^{h} \mu \tau^{ij} (\theta^3)^n d\theta^3,$$
$$\{T_{*(n)}^i, D_{(n)}^*\} = \int_{-h}^{h} \mu \{t_*^i, \sigma_*\} (\theta^3)^n d\theta^3,$$
$$\{F_{(n)}^i, D_{(n)}^i\} = \int_{-h}^{h} \mu (\theta^3)^n \{f^i, D^i\} d\theta^3,$$
$$S_{*(n)}^\alpha = T_{*(n)}^\alpha - b_\nu^\alpha T_{*(n+1)}^\nu, \qquad (5.1a)$$
$$\{P_{(n)}^i, R_{(n)}^i\} = [\mu \tau^{3i} (\theta^3)^n]_{|\theta^3 = \{h,-h\}}, \quad S_{(n)} = E_{(n)} - F_{(n)},$$
$$\{E_{(n)}, F_{(n)}\} = [\mu \cdot D^3 (\theta^3)^n]_{|\theta^3 = \{h,-h\}},$$
$$T_{(n)}^\alpha = F_{(n)}^\alpha + P_{(n)}^\alpha - R_{(n)}^\alpha - b_\nu^\alpha (F_{(n+1)}^\nu + P_{(n+1)}^\nu - R_{(n+1)}^\nu),$$
$$T_{(n)}^3 = F_{(n)}^3 + P_{(n)}^3 - R_{(n)}^3,$$

and the acceleration resultants of order n by

$$A_i^{(n)} = \sum_{m=0}^{N} I_{(n+m)} \ddot{\bar{u}}_i^{(m)} \qquad (5.1b)$$

with

$$I_{(n)} = \begin{cases} 2\,[K_0 h^2/(n+3) + 1/(n+1)]h^{n+1} & \text{for } n = 2r \\ -4H_0 h^{n+2}/(n+2) & \text{for } n = 2r+1. \end{cases} \tag{5.1c}$$

Here, H_0 and K_0 are the mean and Gaussian curvatures of the surface $\mathfrak{A}$, and they can be expressed as

$$H_0 = \tfrac{1}{2} b^{\alpha}_{\alpha}, \quad K_0 = |b^{\alpha}_{\beta}|. \tag{5.2}$$

Now, consider first the volume integral in Eq. (3.6), that is,

$$\delta J^1_1 = \int_{t_0}^{t_1} dt \int_{\mathfrak{A}} \int_{\theta^3=-h}^{h} [(\tau^{ij};_i + f^j - \rho \ddot{u}^j)\delta u_j + (D^i;_i)\delta\varphi]\mu\, dA\, d\theta^3. \tag{5.3}$$

Substituting the series expansions (4.4)-(4.6) into this equation, integrating across the thickness and replacing the resultants (5.1), we arrive at the equation

$$\delta J^1_1 = \int_{t_0}^{t_1} dt \int_{\mathfrak{A}} \left(\sum_{n=0}^{N} (V^i_{(n)} \delta u_i^{(n)} + \phi_{(n)} \delta \varphi^{(n)}) \right) dA \tag{5.4a}$$

with

$$V^{\alpha}_{(n)} = (T^{\beta\alpha}_{(n)} - b^{\alpha}_{\nu} T^{\beta\nu}_{(n+1)})_{|\beta} - b^{\alpha}_{\nu} T^{\nu 3}_{(n)} - n(T^{3\alpha}_{(n-1)} - b^{\alpha}_{\nu} T^{3\nu}_{(n)}) + T^{\alpha}_{(n)} - \rho \ddot{A}^{\alpha}_{(n)},$$

$$V^3_{(n)} = T^{\alpha 3}_{(n)|\alpha} + b_{\alpha\beta} T^{\alpha\beta}_{(n)} - C_{\alpha\beta} T^{\alpha\beta}_{(n+1)} - nT^{33}_{(n-1)} + T^3_{(n)} - \rho \ddot{A}^3_{(n)},$$

$$\phi_{(n)} = D^{\alpha}_{(n)|\alpha} - nD^3_{(n-1)} + S_{(n)}, \tag{5.4b}$$

where we have made use of the following identities[14]

$$\begin{aligned} &\mu\mu^{\nu}_{\alpha}\tau^{\alpha\beta}{}_{;\beta} = (\mu\mu^{\nu}_{\lambda}\tau^{\lambda\beta})_{|\beta} - \mu\mu^{\nu}_{\alpha}(\mu^{-1})^{\beta}_{\lambda} b^{\lambda}_{\beta} \tau^{\alpha 3} - \mu b^{\nu}_{\alpha}\tau^{3\alpha}, \\ &\mu\tau^{3\alpha}{}_{;\alpha} = (\mu\tau^{3\alpha})_{|\alpha} + \mu\mu^{\nu}_{\alpha} b_{\nu\beta}\tau^{\alpha\beta} - \mu(\mu^{-1})^{\alpha}_{\nu} b^{\nu}_{\alpha}\tau^{33}, \\ &\mu^{\beta}_{\alpha}\tau^{\alpha 3}{}_{;3} = (\mu^{\beta}_{\alpha}\tau^{\alpha 3})_{,3}, \quad \mu_{,3} = -\mu(\mu^{-1})^{\alpha}_{\beta} b^{\beta}_{\alpha}, \\ &\mu D^{\alpha}{}_{;\alpha} = (\mu D^{\alpha})_{|\alpha} - \mu(\mu^{-1})^{\alpha}_{\beta} b^{\beta}_{\alpha} D^3. \end{aligned} \tag{5.4c}$$

The surface integrals in Eq. (3.6) are

$$\delta J^2_2 = \int_{t_0}^{t_1} dt \oint_{\mathfrak{C}} ds \int_{-h}^{h} \mu[(t^j_* - n_i \tau^{ij})\delta u_j + (\sigma_* - n_i D^i)\delta\varphi]\, d\theta^3 \tag{5.5}$$

and

$$\delta J^3_3 = \int_{t_0}^{t_1} dt \int_{\mathfrak{A}_c} [(u_i - u_i^*)\delta t^i + (\varphi - \varphi_*)\delta\sigma]\, dA, \tag{5.6}$$

where

$$\mathcal{S}_N = \mathcal{S}_e, \quad \mathcal{S}_C = \mathfrak{A}_u \cup \mathfrak{A}_l = \mathfrak{A}_C. \tag{5.7}$$

Carrying out the integrations as in the volume integral, we get the equations

$$\begin{aligned} \delta J^2_2 &= \int_{t_0}^{t_1} dt \oint_{\mathfrak{C}} ds \left(\sum_{n=0}^{N} (V^i_{*(n)} \delta u_i + \phi^*_{(n)} \delta\varphi^{(n)}) \right), \\ \delta J^3_3 &= \int_{t_0}^{t_1} dt \int_{\mathfrak{A}_c} \sum_{n=0}^{N} [(u_i^{(n)} - u_i^{*(n)})\delta t^i + (\varphi^{(n)} - \varphi_*^{(n)})\delta\sigma]\, dA, \end{aligned} \tag{5.8a}$$

where

$$\begin{aligned} V^{\alpha}_{*(n)} &= S^{\alpha}_{*(n)} - \nu_{\beta}(T^{\beta\alpha}_{(n)} - b^{\alpha}_{\nu} T^{\beta\nu}_{(n+1)}), \\ V^3_{*(n)} &= T^3_{*(n)} - \nu_{\alpha} T^{\alpha 3}_{(n)}, \\ \phi^*_{(n)} &= D^*_{(n)} - \nu_{\alpha} D^{\alpha}_{(n)}. \end{aligned} \tag{5.8b}$$

With the help of Eqs. (5.4) and (5.8), the variational equation (3.6) can be written in the form

$$\delta J = \delta J^k_k = 0. \tag{5.9}$$

Setting this equation equal to zero for the arbitrary and independent variations of the displacement, surface tractions and charge, and electric potential components, the hierarchy of the two-dimensional approximate field equations and the corresponding natural boundary conditions are found, and they are given as follows:

$$V^i_{(n)} = 0, \quad \phi_{(n)} = 0 \quad \text{on } \mathfrak{A} \times [t_0, \infty) \tag{5.10a}$$

and

$$\begin{aligned} &V^i_{*(n)} = 0, \quad \phi^*_{(n)} = 0 \quad \text{on } \mathfrak{C} \times [t_0, \infty), \\ &\bar{u}_i^{(n)} - \bar{u}^{*(n)} = 0, \quad \varphi_{(n)} - \varphi^*_{(n)} = 0 \quad \text{on } \mathfrak{A}_c \times [t_0, \infty), \\ &n = 0, 1, 2, \cdots, N. \end{aligned} \tag{5.10b}$$

A set of initial conditions based on Eqs. (2.14) reads as

$$\begin{aligned} &\mathbf{u}^{(n)}(\theta^{\alpha}, t_0) = \mathbf{v}^{*(n)}(\theta^{\alpha}), \quad \dot{\mathbf{u}}^{(n)}(\theta^{\alpha}, t_0) = \mathbf{w}^{*(n)}(\theta^{\alpha}), \\ &\varphi_{(n)}(\theta^{\alpha}, t_0) = \Psi^*_{(n)}(\theta^{\alpha}) \quad \text{on } \mathfrak{A} \times [t_0, \infty). \end{aligned} \tag{5.11}$$

The distributions (4.7) and (4.8) in conjunction with Eqs. (2.5), (2.6) and (5.1) yield the macroscopic constitutive relations

$$\begin{aligned} T^{ij}_{(n)} &= \sum_{m=0}^{N+1} I_{(m+n)}(C^{ijkl} S^{(m)}_{kl} - C^{ijk} E^{(m)}_k), \\ D^i_{(n)} &= \sum_{m=0}^{N+1} I_{(m+n)}(C^{ijk} S^{(m)}_{jk} + C^{ij} E^{(m)}_j). \end{aligned} \tag{5.12}$$

Thus far, a higher order linear theory of piezoelectric crystal surfaces has been established. This consists of the macroscopic field equations (5.8), the strain-displacement and electric field-electric potential relations (4.7) and (4.8), the initial and natural boundary conditions (5.10) and (5.11), and the constitutive equations (5.12).

6. UNIQUENESS OF SOLUTIONS

In the foregoing analysis, an initial-mixed boundary value problem is completely described by the governing equations of piezoelectric crystal finite surfaces. The uniqueness of solutions of this problem is now discussed as in the classical Neumann manner. Accordingly, for the uniqueness of solutions, it is enough to show that the homogeneous problem (homogeneous field equations, homogeneous boundary conditions and so on) has only the trivial solution.

Let E and W, respectively, stand for the kinetic and internal energies per unit area of $\mathfrak{A}$. From Eq. (2.18), we have the rate of the kinetic energy density

$$\dot{E} = \int_{-h}^{h} \dot{K}\mu\, d\theta^3 = \int_{-h}^{h} \delta \ddot{u}^i \dot{u}_i \mu\, d\theta^3. \tag{6.1}$$

Upon use of Eqs. (4.4) and (5.1), this density rate may be written in the form

$$\dot{E} = \sum_{n=0}^{N} \rho \ddot{A}^i_{(n)} \dot{\bar{u}}_i^{(n)}. \tag{6.2}$$

On account of Eqs. (2.5-9) and (2.15), we may express the energy rate $\dot{W}$

$$\dot{W} = \int_{-h}^{h} \mu \dot{U}\, d\theta^3 = \int_{-h}^{h} (\tau^{ij}\dot{u}_{i;j} - D_i \dot{\varphi}_{,i})\mu\, d\theta^3. \tag{6.3}$$

Substituting Eqs. (4.4) and (4.5) into Eq. (6.3) and then integrating, we obtain $\dot{W}$ in terms of the stress and charge resultants (5.1) as follows

$$\dot{W} = \sum_{n=0}^{N} [(T^{\alpha\beta}_{(n)} - b^{\alpha}_{\nu} T^{\nu\beta}_{(n+1)})\dot{\bar{u}}^{(n)}_{\alpha|\beta} + (nT^{\alpha 3}_{(n-1)} + b^{\alpha}_{\nu} T^{3\nu}_{(n)} - nb^{\alpha}_{\nu} T^{3\nu}_{(n)})\dot{\bar{u}}^{(n)}_{\alpha} + (C_{\alpha\beta} T^{\alpha\beta}_{(n+1)} - b_{\alpha\beta} T^{\alpha\beta}_{(n)} + nT^{33}_{(n-1)})\dot{\bar{u}}^{(n)}_{3} + T^{3\alpha}_{(n)} \dot{\bar{u}}^{(n)}_{3,\alpha} - D^{\alpha}_{(n)} \dot{\varphi}^{(n)}_{,\alpha} - nD^{3}_{(n-1)} \dot{\varphi}^{(n)}], \tag{6.4}$$

where the relations between the derivatives of space and surface vectors[14]

$$u_{3|\alpha} = \bar{u}_{3,\alpha} + b^{\lambda}_{\alpha} \bar{u}_{\lambda}, \quad u_{\alpha|3} = \mu^{\nu}_{\alpha} \bar{u}_{\nu,3}, \quad u_{\alpha|\beta} = \mu^{\sigma}_{\alpha}(\bar{u}_{\sigma\|\beta} - b_{\alpha\beta}\bar{u}_3$$

are considered.

Now, consider the following equation for the homogeneous problem,

$$\int_{t_0}^{t_1} dt \int_{\mathfrak{a}} dA \sum_{n=0}^{N} V^{i}_{(n)} \dot{\bar{u}}^{(n)}_{i} = 0, \tag{6.5}$$

where we have used Eqs. (5.10). By applying Green's transformations, and using Eqs. (5.4), (6.2) and (6.4), we arrive at the equation

$$-\int_{t_0}^{t_1} dt \int_{\mathfrak{a}} (\dot{W} + \dot{E}) dA + \int_{t_0}^{t_1} dt \oint_{\mathcal{C}} \nu_{\beta} \sum_{n=0}^{N} (W^{\beta i}_{(n)} \dot{\bar{u}}^{(n)}_{i} - D^{\beta}_{(n)} \dot{\varphi}^{(n)}) ds + \int_{t_0}^{t_1} dt \int_{\mathfrak{a}} \sum_{n=0}^{N} (S^{i}_{(n)} \dot{\bar{u}}^{(n)}_{i} + W_{(n)} \dot{\varphi}^{(n)}) dA = 0 \tag{6.6a}$$

with

$$W^{\beta\alpha}_{(n)} = T^{\beta\alpha}_{(n)} - b^{\alpha}_{\nu} T^{\beta\nu}_{(n+1)}, \quad W^{\beta 3}_{(n)} = T^{\beta 3}_{(n)}, \quad S^{i}_{(n)} = T^{i}_{(n)} - F^{i}_{(n)}, \quad W_{(n)} = -D^{\alpha}_{(n)|\alpha} - nD^{3}_{(n-1)}. \tag{6.6b}$$

This equation can readily be put to the form

$$\mathcal{K}(t_1) + \Sigma(t_1) = \mathcal{K}(t_0) + \Sigma(t_0) + \int_{t_0}^{t_1} dt \left(\oint_{\mathcal{C}} \nu_{\beta} \sum_{n=0}^{N} (W^{\beta i}_{(n)} \dot{\bar{u}}^{(n)}_{i} - D^{\beta}_{(n)} \dot{\varphi}^{(n)}) ds + \int_{\mathfrak{a}} \sum_{n=0}^{N} (S^{i}_{(n)} \dot{\bar{u}}^{(n)}_{i} + W_{(n)} \dot{\varphi}^{(n)}) dA \right) \tag{6.7}$$

in which Eq. (3.2) is taken into account. If the line and area integrals in Eq. (6.7) vanish, e.g., as a result of Eqs. (5.10) and (5.11), then we write down

$$\mathcal{K}(t_1) + \Sigma(t_1) = \mathcal{K}(t_0) + \Sigma(t_0). \tag{6.8}$$

Guided by the usual arguments based on the positive-definiteness of the kinetic and internal energies $\mathcal{K}$ and Σ, we may write

$$\mathcal{K}(t_1) = \Sigma(t_1) = \mathcal{K}(t_0) = \Sigma(t_0) = 0 \tag{6.9}$$

which lead to the uniqueness of solutions. We then state the following theorem.

Theorem: Given a regular region of finite space $\mathcal{V} + \mathcal{S}$ with boundary $\mathcal{S}(\mathcal{S} = \mathcal{S}_N \cup \mathcal{S}_C, \mathcal{S}_N \cap \mathcal{S}_C = 0)$ in a Euclidean 3-space, then there exist at most one-single valued vector functions $\bar{u}_i^{(n)} \in C^{(2,2)}$ and $\varphi^{(n)} \in C^{(1,0)}$ in $\mathcal{V} + \mathcal{S}$ at $t_0 \leq t < \infty$, which satisfy the governing equations of piezoelectric crystal surfaces, i.e., Eqs. (4.7)-(4.8) and (5.10)-(5.12).

Further, it is evident that not only the boundary conditions (5.10) but to prescribe any member of each product in the line and area integrals of (6.7) assures the uniqueness of solutions.

7. CONCLUSION

A rigorous derivation of the theory of piezoelectric crystal finite surfaces has been established on the basis of the linear theory of piezoelectricity. The theory is constructed in a systematic and consistent manner by means of a method of series expansion and a variational theorem deduced from Hamilton's principle. It consists of a hierarchy of two-dimensional approximate equations of motion, charge equation of electrostatics, initial and natural boundary conditions, strain-displacement and electric field-electric potential relations, and macroscopic constitutive equations.

For a piezoelectric crystal plane, our results can be brought in general agreement with those of Mindlin and Tiersten,[6] if the effects of curvature are abrogated in Eqs. (4.7), (5.8), and (5.12), i.e. $b^{\alpha}_{\beta} = 0$. However, this presentation is more general regarding the theorems

In closing, we note that the extension of this theory to Cosserat media[14] and to piezoelectric composites as well as to dielectrics is straightforward. The applications of the theory remain to be exhibited. We deal with some of them in a forthcoming article.[15]

ACKNOWLEDGMENTS

The author wishes to thank Dr. Mg. AlpD who has exhibited some applications of the theory and to acknowledge the financial support provided by the Office of Naval Research.

*Present address: Bakircilar 63, Elazig, Turkey.

[1] J. and P. Curie, Bull. Soc. Min. France 3, 90 (1880); see also *Oeuvres de Pierre Curie* (Gauthier-Villars, Paris, 1908).
[2] W. G. Cady, *Piezoelectricity* (McGraw-Hill, New York, 1946).
[3] P. W. Forsbergh Jr., in *Handbuch der Physik, Dielektrika* (Springer-Verlag, Berlin, 1956), Vol. 17, p. 264.
[4] W. P. Mason, *Piezoelectric Crystals and Their Applications to Ultrasonics* (Van Nostrand, New York, 1950).
[5] *Piezoelectricity* (Post Office Research Station, London, 1957).
[6] H. F. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).
[7] M. C. Dökmeci, "On the oscillations of piezoelectric crystal beams," paper presented at 84th meeting of the Acoustical Society of America, Miami Beach, Nov. 1972. Also, see Int. J. Solids Structures 10, 401 (1974).
[8] D. S. Drumheller and A. Kalnins, J. Acoust. Soc. Am. **47**, 1343 (1970).
[9] C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1949).
[10] O. D. Kellog, *Foundations of Potential Theory* (Ungar, New York, 1929).
[11] H. F. Tiersten, J. Math. Phys. **9**, 1445 (1968).
[12] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1949).
[13] P. M. Naghdi, in *Progress in Solid Mechanics* (North-Holland, Amsterdam, 1963), Vol. 4, p. 1.
[14] M. C. Dökmeci, in *Recent Advances in Engineering Science* (Gordon and Breach, New York, 1970), Vol. 5, pp. 2, 189.
[15] M. C. Dökmeci, "High frequency vibrations of crystal surfaces, plated with electrodes", paper presented at 85th meeting of the Acoustical Society of America, Boston, Mass., April 1973.

On the explicit-time-dependent invariance properties of quantum mechanical systems

G. Burdet, M. Perrin, and P. Sorba

Centre de Physique Théorique, C.N.R.S., 31, chemin J. Aiguier, 13274 Marseille Cedex 2, France
(Received 26 January 1973; revised manuscript received 24 May 1973)

We use the notion of split extension algebra to embody a given invariance algebra of a quantum mechanical system, a realization of which is known in terms of some variables, in a richer invariance algebra expressed in terms of the same variables. By applying this procedure to a free system of particles we show how to obtain the invariance under the Schrödinger algebra and we build a bigger invariance algebra which describes a system of noninteracting particles, for example the asymptotic states in a nonrelativistic scattering problem.

1. INTRODUCTION

As long as we do not take into account an explicit time dependence, the notion of symmetry algebra for a conservative system is defined by the property that all its own generators commute with the Hamiltonian of the problem and furnish constants of motion. In the case of explicitly time-dependent transformations it has been shown[1] that the corresponding generators leave invariant the Hamiltonian of the Schrödinger equation iff:

$$[H, S(t)] = i\,\frac{\partial S(t)}{\partial t}. \tag{1}$$

Lipkin noted that if ψ is an eigenfunction of H, $S\psi$ cannot be an eigenfunction of H with the same eigenvalue [in the case where Eq. (1) is not trivially satisfied]. Hence $S(t)$ satisfying Eq. (1) generates families of states having different energies and provides informations about the energy spectrum of H: a time-dependent symmetry algebra works as a spectrum generating algebra, which increases the own interest of such a concept.

Moreover it is easy to see that all the explicit time derivatives of $S(t)$ satisfying Eq. (1) furnish invariance properties too. In the simplest nontrivial case where $S(t)$ is a linear function of the time, $\partial S/\partial t$ is time-independent and commutes with H. This is the case for the generators of the pure Galilean transformations $S_j(t) \equiv K_j = MR_j - P_j t$ $(j = 1, 2, 3)$ which verify:

$$[H, K_j] = -\,iP_j, \qquad [H, P_j] = 0.$$

Translational invariance generated by P_j is then automatically involved by the requirement of pure Galilean invariance generated by K_j. "It is impossible to consider the former without the latter," said Lipkin, and the constraints brought on the interaction by the invariance under Galilean "boosts" imply those required by the invariance under translations. Then this remark emphasizes the interest to work as far as possible with large invariance algebras.

It appears, from this short discussion, that the concept of explicit time-dependent symmetry makes coexist two mathematical notions: links between some generators of the invariance algebras by integration with respect to time, and correspondingly particular positions of certain subalgebras in the considered invariance algebra.

At this step, it has seemed natural to us to consider the following problem: being given an elementary invariance algebra associated to an Hamiltonian describing a physical system, is it possible to embody this algebra into a bigger invariance one? It is an aim of this paper to answer this question. In fact the hypothesis of the problem can be enlarged and we propose here an algebraic method to embody an elementary invariance algebra for an unspecified Hamiltonian into a larger time-dependent invariance algebra a realization of which can be given in terms of the same observables used for the realization of the starting algebra.

This paper will be divided into three parts:

—In Sec. 2 we discuss the correspondence between invariance properties in the Schrödinger's picture and non-explicit time-dependent observables in the Heisenberg's picture.

—In Sec. 3 we show how the mathematical frame furnished by the notion of derivation algebra can be used to increase a given invariance algebra called a "germ" in our terminology.

—Section 4 deals with two applications of the above method. By choosing as germs two subalgebras of the extended Galilean algebra we generate as invariance algebra the Schrödinger algebra and a bigger one which appears as the smallest invariance algebra able to describe a system of noninteracting particles.

2. TIME-DEPENDENT INVARIANCE ALGEBRA IN THE SCHRÖDINGER AND HEISENBERG PICTURES

It is easy to deduce from Eq. (1) that the commutator $[S_1, S_2] = S_1S_2 - S_2S_1$ for any invariance generators S_1 and S_2 is an invariance generator too, hence the invariance properties of a given Hamiltonian form a Lie algebra. Now Eq. (1) does not imply it is a finite dimensional Lie algebra, but we restrict ourselves to a finite one for mathematical convenience.

Let us therefore consider a finite N-dimensional invariance algebra. In the Schrödinger picture each generator is satisfying the relations

$$[H, S_j(t)] = i\,\frac{\partial S_j}{\partial t} = i\sum_{k=1}^{N}\phi_j^k S_k(t) \quad \forall j = 1, \ldots, N, \tag{2}$$

where the coefficients ϕ_j^k are time-independent.

Such generators possess the following important property: they do not depend explicitly on time in the Heisenberg picture. Indeed in going from the Schrödinger representation to the Heisenberg one, any operator Ω changes according to the rule:

$$\Omega_H = e^{iHt}\,\Omega\, e^{-iHt} \tag{3}$$

and for any generator verifying (1) it is easy to see that

$$\frac{\partial S_H}{\partial t} = e^{iHt}\left(i[H, S] + \partial S/\partial t\right)e^{-iHt} = 0.$$

Consequently, S_H does not depend explicitly on time and, from Eq. (3), takes its value for $t = 0$. We shall write:

$$S_H = S(0) = S_0. \tag{4}$$

Hence the generators of an invariance algebra do not depend explicitly upon time in the Heisenberg picture but do not commute necessarily with H. More precisely the generators S_0 generate an algebra $\mathcal{S}_0$ isomorphic to the starting invariance algebra $\mathcal{S}(t)$ since each S_0 deduces from a corresponding S by means of the automorphism defined in Eq. (3).

Conversely, to any algebra of observables which do not depend explicitly on time corresponds an invariance algebra isomorphic to it in the Schrödinger picture. Indeed the explicit time derivative of $S(t)$ defined by

$$S(t) = e^{-iHt} S_0 e^{iHt} \tag{5}$$

gives exactly the invariance condition (1).

Obviously each coefficient in the development of $S(t)$ in powers of t

$$S(t) = S_0 - it[H, S_0] + \cdots + (-it)^m/m! \,.\, [\underbrace{H[H[H}_{m \text{ times}} \cdots [H, S_0] \cdots] + \cdots \tag{6}$$

is a generator of the algebra $\mathcal{S}_0$. Then it should be noticed that in the case where the development is boundless, the explicit time dependence of the corresponding $S(t)$ is an exponential one. But if there exists m such that

$$[\underbrace{H[H}_{m+1} \cdots [H, S_0] \cdots] = 0,$$

$S(t)$ is a polynomial of highest degree m in t. In this last case the coefficient of t^m is identical in both pictures since it commutes with H.

It is worth noticing that this last results should be obtained directly from the right-hand equation contained in (2) which furnishes a system of N linear differential equations of the first order with constant coefficients when j runs from 1 to N. The resolution of this system determines the explicit time dependence of the $S_j(t)$ and leads to the above properties.

As it was already noticed in the Introduction, chains of elements can be obtained in the algebra $\mathcal{S}(t)$ by successive derivations with respect to the time: $S(t)$, $\partial S(t)/\partial t$, $\partial^2 S(t)/\partial t^2, \cdots$ but also by repeated action of ad(H) if one remembers Eq. (1). Obviously the isomorphism between $\mathcal{S}(t)$ and $\mathcal{S}_0$ insures that ... algebra $\mathcal{S}_0$ but in the Heisenberg picture the chains of elements can only be obtained by repeated action of ad(H). Hence in practice the relationships between generators appear when the whole algebra is decomposed under the action of H.

We conclude this section by remarking that from a mathematical point of view the algebras $\mathcal{S}_0$ and $\mathcal{S}(t)$ are the same algebra depending on t as a real parameter. An interesting consequence is that the Casimir operators of this algebra are time independent, such an example is given in Ref. 2.

3. EXTENSION OF A GIVEN INVARIANCE ALGEBRA

We start with a well-defined invariance algebra: the germ $\mathcal{G}_0$ a realization of which is given in the Heisenberg picture in terms of some noncommuting observables $\omega_j (j = 1, \ldots, n)$ used to describe a quantum mechanical system. We wish to embody this germ into a bigger algebra able to yield new invariance properties in terms of the same variables. Now we insist on the fact that our aim is not to extend the germ by another invariance algebra expressed in terms of other observables which was the case when one tried to extend the Poincare algebra by internal symmetry such as $SU(3)$ for example.

In our approach the Hamiltonian is not given explicitly and we have in mind to exhibit the constraints on H which come from the embedding of the germ into a larger invariance algebra. To solve this problem it is not sufficient to give a realization of a larger algebra chosen in the enveloping algebra $\mathcal{E}(\mathcal{O})$ of the algebra $\mathcal{O}$ generated by the observables ω_j. Indeed in this way one doesn't know the action of the larger algebra on H. In the same mind it is not possible to extend the germ by time integration of some generators, that is to introduce new operators such that $Z_j = \int K_j\, dt = Z_{0j} + MR_j t - \frac{1}{2} P_j t^2$ in the example of the introduction. In this case the commutation relations with H are well defined: $[H, Z_j] = iK_j$ but we have not constructed a bigger invariance algebra since commutators like $[K_j, Z_k]$ and $[P_j, Z_k]$ are not determined. It is then necessary to build an algebraic scheme which provides an algebra $\mathcal{F}$ containing the germ, $\mathcal{G}_0$ and making appear the commutation relations of H with the new introduced generators.

By hypothesis $\mathcal{G}_0$ is a subalgebra of $\mathcal{E}(\mathcal{O})$ and the largest subalgebra of $\mathcal{F}$ which can be set in isomorphism with a subalgebra of $\mathcal{E}(\mathcal{O})$ will be a candidate to the wanted extended algebra $\mathcal{S}_0$. The explicit time realization, i.e., the invariance algebra $\mathcal{S}(t)$ is then obtained by the automorphism defined in Eq. (5).

In order to construct the abovementioned Lie algebra $\mathcal{F}$ we shall use a mathematical tool closely connected to the structure of a given Lie algebra which is the Lie algebra of its automorphisms: the derivation algebra. This notion is very natural and strongly suggested by the fact that in Sec. 2 we have already considered H as a derivation acting on the invariance algebra. On the other hand, it is very simple to extend a Lie algebra by its derivation algebra.[3] Indeed let $\mathcal{A}$ be a finite dimensional Lie algebra and $\mathfrak{D}(\mathcal{A})$ the derivation algebra, it is then possible to construct the split extension of $\mathfrak{D}(\mathcal{A})$ by $\mathcal{A}$, also called the holomorph of $\mathcal{A}$, which is the semidirect sum $\mathcal{A} \,\square\, \mathfrak{D}(\mathcal{A})$.[4] The elements of this new algebra are the ordered pairs (a, d) where a and d belong, respectively, to $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$, and the Lie product is defined by

$$[(a, d), (a', d')] = ([a, a'] + d(a') - d'(a), [d, d']). \tag{7}$$

This construction gives a meaning to the brackets $[d, a']$, indeed:

$$[d, a'] \equiv [(0, d), (a', 0)] = (d(a'), 0) = d(a') \in \mathcal{A}, \tag{8}$$

which is just the action of the derivations on the elements of the initial algebra.

Let us now apply this construction to a given germ $\mathcal{G}_0$. In general the Hamiltonian belongs to the germ and the suggested method is able to make new generators d appear, the action of which on H is not trivial, i.e., $d(H) = S_0$, with $S_0 \in \mathcal{G}_0$ and $S_0 \neq 0$. Therefore they correspond to time integrals of generators of the germ in the Schrödinger picture, but it should be noticed that we have not succeeded to exhibit an extension giving rise to integration of Galilean boosts.

Now it happens that the Hamiltonian doesn't appear alone or accompanied in the derived algebra of the given invariance algebra. It is then possible to exclude H out of the germ and, owing to its action on the invariance algebra, we are sure to identify H in the derivation algebra. We will show in the next section that this proposal is not purely academic but allows us to show some interesting links between some invariance algebras.

4. APPLICATIONS TO A FREE SYSTEM OF PARTICLES

In general the Galilei group is considered as the full kinematical invariance group of the Hamiltonian of a free system, but it has been recently shown by Niederer[5] that a larger group of space–time transformations: the Schrödinger group, leaves invariant the free Schrödinger equation. In fact it has been shown[2] that this embedding of the Galilei group strongly limits the form of the interaction between the constituent particles of the system. We propose to apply the technic above described to generate the Schrödinger algebra and moreover to exhibit a larger invariance algebra which appears as a limit case in the description of a system of particles, since no interaction can take place between them.

We recall that a system of particles can be set in correspondence with an unirrep of a central extension $\tilde{G}$ of the Galilei group, the generator realizations of which, are given, in terms of canonical coordinates $q_j(\mu)$ and $p_j(\mu)$ corresponding to each mass point $m(\mu)$, $\mu = 1, 2, \ldots, N$, by

$$P_j = \sum_{\mu=1}^{N} p_j(\mu), \quad M = \sum_{\mu=1}^{N} m(\mu), \quad J_j = \sum_{\mu} (\mathbf{q}(\mu) \wedge \mathbf{p}(\mu))_j,$$
$$K_j = \sum_{\mu} m(\mu) q_j(\mu) - P_j t = MX_j - P_j t, \tag{9}$$
$$H = \sum_{\mu=1}^{N} \frac{p^2(\mu)}{2m\mu} + V = T + V.$$

The nonzero commutation relations are given by

$$[J_j, J_k] = i\epsilon_{jkl} J_l, \quad [J_j, P_k] = i\epsilon_{jkl} P_l, \quad [J_j, K_k] = i\epsilon_{jkl} K_l,$$
$$[K_j, P_k] = i\delta_{jk} M, \quad [H, K_j] = -iP_j. \tag{10}$$

(A)—Let us choose as a first example the derived extended Galilei group generated by $\{K_j, P_j, J_j, M\}$, $j = 1, 2, 3$, i.e., we take as a germ the corresponding Lie algebra $\tilde{\mathcal{G}}_0'$ in the Heisenberg picture.

As it has been shown in Ref. 6 the derivation algebra $\mathfrak{D}(\tilde{\mathcal{G}}_0')$ contains the subalgebra of derivations isomorphic to the derived nonextended Galilean algebra and four supplementary generators which form between them a subalgebra isomorphic to $\mathcal{G}l(2, \mathbb{R})$. The problem consists now in finding in the envelopping associative algebra $\mathcal{E}(\mathbf{q}(\mu), \mathbf{p}(\mu))$ the largest subalgebra isomorphic to a subalgebra of $\mathcal{F} = \tilde{\mathcal{G}}_0' \,\square\, \mathfrak{D}(\tilde{\mathcal{G}}_0')$ and including $\tilde{\mathcal{G}}_0'$. It is easy to see that we cannot realize simultaneously the whole inner derivation algebra, and if we consider only a subalgebra, the realization coincides with that of the corresponding generators of $\tilde{\mathcal{G}}_0'$ and there is a redundance. Now the subalgebra $\mathcal{G}l(2, \mathbb{R})$ can be decomposed into the direct sum $SU(1,1) \oplus \mathcal{R}_{D_2}$, but the one-parameter subalgebra $\mathcal{R}_{D_2}$ acts as a dilatation on M and obviously cannot be set in correspondence with any element of $\mathcal{E}(\mathbf{q}, \mathbf{p})$ and then must be rejected.

Among the $SU(1,1)$ generators, one of them can be recognized, owing to its action on $\tilde{\mathcal{G}}_0'$, as the Hamiltonian of the system. The others, denoted C_0 and D_0, are given by

$$C_0 = \tfrac{1}{2} \sum_{\mu=1}^{N} m(\mu) \mathbf{q}^2(\mu),$$
$$D_0 = -\tfrac{1}{2} \sum_{j=1}^{3} \sum_{\mu=1}^{N} [q_j(\mu), p_j(\mu)]_+. \tag{11}$$

Their commutation relations are

$$[H, C_0] = iD_0, \quad [H, D_0] = 2iH,$$
$$[C_0, D_0] = -2iC_0. \tag{12}$$

And by using the transformation (5) we obtain the explicit time-dependent generators of the extended invariance algebra in the Schrödinger picture:

$$D = 2Ht + D_0, \quad C = Ht^2 + D_0 t + C_0. \tag{13}$$

In fact this invariance algebra obtained from the derivations of $\tilde{\mathcal{G}}'$ is the Schrödinger algebra $\tilde{\mathcal{S}}_{ch} = \tilde{\mathcal{G}}' \,\square\, SU(1,1)$, the commutation relations of which are given by (10), (12) and following (8) by the nontrivial actions of the derivations on the germ:

$$[D, K_j] = iK_j, \quad [D, P_j] = -iP_j, \quad [C, P_j] = iK_j. \tag{14}$$

(B)—As a second example we will consider a subalgebra in the preceding germ namely the Heisenberg algebra $\mathcal{H}_0$ generated by $\{K_j, P_j, M\}$. The loss of the rotational invariance, i.e., of the semisimple subalgebra generated by $\mathbf{J}$, leads to a richer derivation algebra.[6] Indeed the inner derivations form a six-dimensional Abelian algebra and we obtain a basis of $\mathfrak{D}(\mathcal{H}_0)$ by adding 22 generators which generate a subalgebra isomorphic to $Sp(6, \mathbb{R}) \oplus \mathcal{R}_{D_2}$. For the same reasons as previously, we must neglect in $\mathcal{H}_0 \,\square\, \mathfrak{D}(\mathcal{H}_0)$ the generators corresponding to the inner derivations and to the dilatation of the mass D_2. Finally it is possible to give a realization of the semidirect sum $\mathcal{A}_{s_0} = \mathcal{H}_0 \,\square\, (Sp(6, \mathbb{R}))_0$. Now it appears a new feature: the Hamiltonian is fully determined and cannot contain any interaction: $H = \sum_\mu [p^2(\mu)/2m\mu]$. The realization of a convenient basis of the $(Sp(6, \mathbb{R}))_0$ algebra is given by

$$R_{jk} = \tfrac{1}{2} \sum_{\mu=1}^{N} [p_j(\mu), q_k(\mu)]_+, \quad j, k = 1, 2, 3,$$
$$S_{jk} = \sum_{\mu=1}^{N} m(\mu) q_j(\mu) q_k(\mu) = S_{kj}, \tag{15}$$
$$T_{jk} = \sum_{\mu=1}^{N} \frac{p_j(\mu) p_k(\mu)}{m(\mu)} = T_{kj}.$$

It is easy to see that $\mathcal{A}_{s_0}$ contains the extended Schrödinger algebra and we have the following relations in terms of the generators of the above basis:

$$J_j = -\epsilon_{jkl} R_{kl}, \quad H = \tfrac{1}{2} \sum_{j=1}^{3} T_{jj},$$
$$C_0 = \tfrac{1}{2} \sum_{j=1}^{3} S_{jj}, \quad D_0 = -\sum_{j=1}^{3} R_{jj}. \tag{16}$$

The nonzero commutation relations of $\mathcal{A}_{s_0}$ are given by

$$[R_{jk}, R_{lm}] = i(\delta_{kl} R_{jm} - \delta_{jm} R_{lk}),$$
$$[R_{jk}, T_{lm}] = i(\delta_{kl} T_{jm} + \delta_{km} T_{jl}),$$
$$[R_{jk}, S_{lm}] = -i(\delta_{jl} S_{km} + \delta_{jm} S_{kl}), \tag{17}$$
$$[S_{jk}, T_{lm}] = i(\delta_{jl} R_{mk} + \delta_{jm} R_{lk} + \delta_{kl} R_{mj} + \delta_{km} R_{lj}),$$
$$[R_{kl}, K_{0_j}] = -i\delta_{jk} K_{0_l}, \quad [R_{kl}, P_j] = i\delta_{jl} P_k,$$
$$[S_{kl}, P_j] = i(\delta_{jk} K_{0_l} + \delta_{jl} K_{0_k}),$$
$$[T_{kl}, K_{0_j}] = -i(\delta_{jk} P_l + \delta_{jl} P_k).$$

Owing to the above-mentioned inclusion $\mathcal{S}_{ch_0} \subset \mathcal{A}_{s_0}$ it remains only to give the explicit time dependence of the last 15 generators. Following the method described in Sec. 2 we decompose $\mathfrak{A}_{s_0}$ under the action of H and from (17) firstly we obtain

$$[H, S_{jk}] = -i(R_{jk} + R_{kj}), \quad [H, R_{jk} + R_{kj}] = -2i T_{jk} \quad (j \neq k). \quad (18)$$

Then the transformation (5) furnishes the three following generators:

$$Q_{jk} = T_{jk} t^2 - (R_{jk} + R_{kj})t + S_{jk}, \quad j \neq k, \quad (19)$$

and also the six explicit time derivatives $\dot{Q}_{jk}$ and $\ddot{Q}_{jk}$.

The two last chains correspond to the commutation relations which determine H and which can be written under the convenient but redundant following form:

$$[H, S_{jj} - \tfrac{1}{3}\sum_{k=1}^{3} S_{kk}] = -2i(R_{jj} - \tfrac{1}{3}\sum_k R_{kk}),$$
$$[H, R_{jj} - \tfrac{1}{3}\sum_k R_{kk}] = -2i(T_{jj} - \tfrac{1}{3}\sum_k T_{kk}) = -2i(T_{jj} - \tfrac{1}{3}H) \quad (20)$$

of which we deduce two independent generators

$$Q_{jj} = (T_{jj} - \tfrac{1}{3}\sum_k T_{kk})t^2 - 2(R_{jj} - \tfrac{1}{3}\sum_k R_{kk})t - (S_{jj} - \tfrac{1}{3}\sum S_{kk})_{j=1,2,3} \quad (21)$$

such that

$$\sum_{j=1}^{3} Q_{jj} = 0,$$

and their explicit time derivatives.

In summary we will notice that our approach allowed us to build the $\mathfrak{A}_s$ algebra which is the invariance algebra of a free system of noninteracting particles. As we have seen $\mathfrak{A}_s$ contains the Schrödinger algebra which is the largest algebra of space–time transformations which leaves invariant the free Schrödinger equation; in fact, as it has been shown in Ref. 2, the Schrödinger invariance still allows us to introduce some types of interaction between particles, while the $\mathfrak{A}_s$ algebra, which comes from the derivations of the Heisenberg algebra, does not offer such a possibility and can uniquely describe a system of particles without interaction.

CONCLUSION

The techniques we proposed allows us to construct the following chain of algebras: $\mathcal{K} \subset \tilde{\mathcal{G}} \subset \mathcal{S}_{ch} \subset \mathfrak{A}_s$. In this chain the number of invariance generators increase while the internal interaction melts. As a direct consequence of this property it may be asked what kinds of external fields are compatible with the above invariance algebras? We will treat this subject in a forthcoming paper.

Finally we emphasize the property of the algebra $\mathfrak{A}_s = \mathcal{K} \,\square\, Sp(6, \mathbb{R})$ which really appears as the candidate to describe the asymptotic states in nonrelativistic scattering problems.

[1]H. J. Lipkin in *Nuclear Physics*, edited by C. de Witt and V. Gillet (Gordon and Breach, New York, (1969), p. 644); Y. Dothan, Phys. Rev. D 12, 2944 (1970).
[2]G. Burdet and M. Perrin, Lett. Nuovo Cimento 13, 651 (1972).
[3]Let $\mathfrak{A}$ be a finite dimensional Lie algebra over the field of real numbers. We denote $\mathfrak{D}(\mathfrak{A})$ the Lie algebra of the group of automorphisms of $\mathfrak{A}$ also called the Lie algebra of derivations or derivation algebra of $\mathfrak{A}$. We recall that a derivation of $\mathfrak{A}$ is a linear map d: $\mathfrak{A} \to \mathfrak{A}$ such that: $d[a, a'] = [d(a), a'] + [a, d(a')]$ for $a, a' \in \mathfrak{A}$, and that in $\mathfrak{D}(\mathfrak{A})$ the Lie bracket is defined by: $[d_1, d_2](a) = d_1(d_2(a)) - d_2(d_1(a))$ for $d_1, d_2 \in \mathfrak{D}(\mathfrak{A})$ and $a \in \mathfrak{D}(\mathfrak{A})$.
[4]N. Jacobson, *Lie Algebras* (Interscience, London, 1965).
[5]U. Niederer, Helv. Phys. Acta 45, 802 (1972); see also P. Roman, J. J. Aghassi, R. Santilli, and P. Huddleston, Nuovo Cimento. A 12, 185 (1972).
[6]G. Burdet, M. Perrin, and R. Sorba, "On the Automorphisms of Real Lie algebras," preprint, Marseilles 72/p. 498 (1972).

Erratum: KS-related f–g couples as exact vacuum solutions of Salam's two-tensor theory [J. Math. Phys. 15, 129 (1974)]

Helmuth K. Urbantke

Institut für Theoretische Physik, Universität Wien, Vienna, Austria
(Received 13 August 1974)

Concerning the KS-related solutions to the f–g equations, an additional assumption has been made, but not stated, which restricts the solutions to the class given in the paper. Thus, the abstract is to read as follows: "The exact solutions...are determined for the case that the two tensors differ only by the tensor product of a *principal* null vector field by itself..." What is shown then is that this principal null vector field must be multiple and have vanishing optical scalars. The proof relies on Eq. (18) of the paper which would be different if the null vector field were not a principal one, i.e., if expression (2.26a) of Ref. 9 were nonzero.

