Exact envelope-soliton solutions of a nonlinear wave equation

Ryogo Hirota

RCA Research Laboratories, Inc., P.O. Box 5151, Tokyo International 100-31, Japan
(Received 26 September 1972)

Exact N-envelope-soliton solutions have been obtained for the following nonlinear wave equation, $i\partial\psi/\partial t + i3\alpha|\psi|^2\partial\psi/\partial x + \beta\partial^2\psi/\partial x^2 + i\gamma\partial^3\psi/\partial x^3 + \delta|\psi|^2\psi = 0$, where α, β, γ and δ are real positive constants with the relation $\alpha\beta = \gamma\delta$. In one limit of $\alpha = \gamma = 0$, the equation reduces to the nonlinear Schrödinger equation which describes a plane self-focusing and one-dimensional self-modulation of waves in nonlinear dispersive media. In another limit, $\beta = \delta = 0$, the equation for real Ψ, reduces to the modified Korteweg–de Vries equation. Hence, the solutions reveal the close relation between classical solitons and envelope-solitons.

1. INTRODUCTION

In previous papers, hereafter referred to as I[1], II[2], III[3], we obtained exact N-soliton solutions of various nonlinear wave equations by transforming the nonlinear equations into homogeneous forms of the second degree.

In Paper I we considered the Korteweg-de Vries equation,[4-6] of the following form

$$\frac{\partial u}{\partial t} - 6u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0. \tag{1.1}$$

The equation is transformed using the relation, $u(x,t) = \partial^2 \log f(x,t)/\partial x^2$ into the following

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)\left[\frac{\partial}{\partial t} - \frac{\partial}{\partial t'} + \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^3\right] f(x,t)f(x',t')_{\text{at } t=t',\,x=x'} = 0. \tag{1.2}$$

In Paper II, we considered the modified Korteweg-de Vries equation[7] of the following form

$$\frac{\partial v}{\partial t} + 24v^2\frac{\partial v}{\partial t} + \frac{\partial^3 v}{\partial x^3} = 0. \tag{1.3}$$

The equation is transformed using the relation, $v(x,t) = \partial[\tan^{-1}(g(x,t)/f(x,t))]/\partial x$ into the following coupled equations,

$$\left[\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right) + \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^3\right] g(x,t)f(x',t')_{\text{at } t=t',\,x=x'} = 0, \tag{1.4}$$

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 [f(x,t)f(x',t') + g(x,t)g(x',t')]_{\text{at } t=t',\,x=x'} = 0. \tag{1.5}$$

In Paper III, we considered the sine-Gordon equation[8,9]

$$\frac{\partial^2\varphi}{\partial x^2} - \frac{\partial^2\varphi}{\partial t^2} = \sin\varphi. \tag{1.6}$$

The equation is transformed using the relation $\varphi(x,t) = 4\tan^{-1}[g(x,t)/f(x,t)]$ into the following coupled equations,

$$\left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 - \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^2 - 1\right] f(x,t)g(x',t')_{\text{at } t=t',\,x=x'} = 0, \tag{1.7}$$

$$\left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 - \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^2\right][f(x,t)f(x',t') - g(x,t)g(x',t')]_{\text{at } t=t',\,x=x'} = 0. \tag{1.8}$$

In the present paper we consider the following nonlinear wave equation,

$$i\frac{\partial\psi}{\partial t} + i3\alpha|\psi|^2\frac{\partial\psi}{\partial x} + \beta\frac{\partial^2\psi}{\partial x^2} + i\gamma\frac{\partial^3\psi}{\partial x^3} + \delta|\psi|^2\psi = 0. \tag{1.9}$$

In one limit, the equation reduces to the nonlinear Schrödinger equation[10] which describes plane self-focusing and one-dimensional self-modulation of waves in nonlinear dispersive media. In another limit, the equation for real ψ reduces to the modified Korteweg-de Vries equation.

We shall obtain exact N-envelope-soliton solutions of the equation by using the same method (one based on intuition and experience) as was used in Papers I, II and III.

Zakharov and Shabat[10] have obtained exact N-envelope-soliton solutions of the nonlinear Schrödinger equation by reducing it to the inverse scattering problem for a certain linear differential operator. Recently Wadati[11] obtained exact N-soliton solutions of the modified Korteweg-de Vries equation by using an inverse problem method similar to that of Zakharov and Shabat. However, the inverse problem method is applicable only to equations of the type

$$\frac{\partial u}{\partial t} = S[u], \tag{1.10}$$

where S is a nonlinear operator differential in x. Although our method is rather heuristic in comparison with the inverse problem method, it should be noted that the present method is very flexible and is applicable to various types of nonlinear wave equations.

In succeeding papers we shall obtain exact N-soliton solutions of the following nonlinear wave equations by using the same method:

(i) $$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} - 6\frac{\partial^2 w^2}{\partial x^2} - \frac{\partial^4 w}{\partial x^4} = 0, \tag{1.11}$$

which describes motions of long waves in shallow water under gravity[12] and in a one-dimensional nonlinear lattice,[13]

(ii) $$\frac{\partial^2}{\partial t^2}\log(1 + V_n) = V_{n+1} + V_{n-1} - 2V_n, \tag{1.12}$$

which describes wave propagation in a nonlinear, lumped,

Copyright © 1973 by the American Institute of Physics

ladder-type[14] network and the one-dimensional nonlinear lattice studied by M. Toda,[15] and

$$\text{(iii)}\quad \frac{1}{1+u_n^2}\frac{\partial u_n}{\partial t} = u_{n+1/2} - u_{n+1/2}, \tag{1.13}$$

which describes wave propagation in a nonlinear, lumped, self-dual ladder-type network.

2. EXACT N-ENVELOPE SOLITON SOLUTIONS

We describe exact N-envelope-soliton solutions of the equation

$$i\frac{\partial\psi}{\partial t} + i3\alpha|\psi|^2\frac{\partial\psi}{\partial x} + \beta\frac{\partial^2\psi}{\partial x^2} + i\gamma\frac{\partial^3\psi}{\partial x^2} + \delta|\psi|^2\psi = 0, \tag{2.1}$$

where α, β, γ and δ are positive real constants satisfying the relation $\alpha\beta = \gamma\delta$.

In the limit (i) $\alpha = \gamma = 0$, the equation reduces to the nonlinear Schrödinger equation,[10]

$$i\frac{\partial\psi}{\partial t} + \beta\frac{\partial^2\psi}{\partial x^2} + \delta|\psi|^2\psi = 0. \tag{2.2}$$

In the other limit (ii) $\beta = \delta = 0$, the equation for real ψ reduces to the modified Korteweg-de Vries equation,[2,11]

$$\frac{\partial\psi}{\partial t} + 3\alpha\psi^2\frac{\partial\psi}{\partial x} + \gamma\frac{\partial^3\psi}{\partial x^3} = 0. \tag{2.3}$$

Hence, the present solutions reveal the close relation between classical solitons and envelope-solitons.

In the limit (iii) $\alpha = \gamma = 0$, the equation reduces to the linear partial differential equation

$$i\frac{\partial\psi}{\partial t} + \beta\frac{\partial^2\psi}{\partial x^2} + i\gamma\frac{\partial^3\psi}{\partial x^3} = 0. \tag{2.4}$$

We note that our solutions exclude the case (iv) $\beta = \gamma = 0$.

Exact N-envelope soliton solutions of Eq. (2.1) can be expressed in the following form

$$\psi(x,t) = G(x,t)/F(x,t), \tag{2.5}$$

where

$$F(x,t) = \sum_{\mu=0,1}{}' \exp\left(\sum_{i<j}^{(2N)}\varphi(i,j)\mu_i\mu_j + \sum_{i=1}\mu_i\eta_i\right), \tag{2.6}$$

$$G(x,t) = \sum_{\nu=0,1}{}'' \exp\left(\sum_{i<j}^{(2N)}\varphi(i,j)\nu_i\nu_j + \sum_{i=1}^{2N}\nu_i\eta_i\right), \tag{2.7}$$

$$G(x,t) = \sum_{\nu=0,1}{}''' \exp\left(\sum_{i<j}^{(2N)}\varphi(i,j)\nu_i\nu_j + \sum_{i=1}^{2N}\nu_i\eta_i\right), \tag{2.8}$$

where

$$\eta_i = P_i x - \Omega_i t - \eta_i^0, \tag{2.9}$$

$$\Omega_i = -i\beta P_i^2 + \gamma P_i^3, \quad \text{for } i = 1,2,\ldots,2N, \tag{2.10}$$

and

$$\eta_{i+N} = \eta_i^*, \tag{2.11}$$

$$P_{i+N} = P_i^*, \tag{2.12}$$

$$\Omega_{i+n} = \Omega_i^*, \quad \text{for } i = 1,2,\ldots,N. \tag{2.13}$$

where * implies a complex conjugate, and

$$\varphi(i,j) = \log[\alpha/2\gamma(P_i + P_j)^2],$$

for $i = 1,2,\ldots,N$ and $j = N+1,\ldots,2N$,

or $i = N+1,\ldots,2N$ and $j = 1,2,\ldots,N$, (2.14)

$$\varphi(i,j) = -\log[\alpha/2\gamma(P_i - P_j)^2],$$

for $i = 1,2,\ldots,N$ and $j = 1,2,\ldots,N$,

or $i = N+1,\ldots,2N$ and $j = N+1,\ldots,2N$, (2.15)

where P_i and η_i^0 are complex constants relating respectively to the amplitude and phase of the ith soliton, $\sum'_{\mu=0,1}$ indicates the summation over all possible combinations of $\mu_1 = 0,1, \mu_2 = 0,1,\ldots,\mu_{2N} = 0,1$, under the condition, $\sum_{i=1}^N\mu_i = \sum_{i=1}^N\mu_{i+N}$, $\sum''_{\nu=0,1}$ and $\sum'''_{\nu=0,1}$ indicate the summations over all possible combinations of $\nu_1 = 0,1, \nu_2 = 0,1,\ldots,\nu_{2N} = 0,1$ under the conditions $\sum_{i=1}\nu_i = 1 + \sum_{i=1}^N\nu_{i+N}$ and $1 + \sum_{i=1}^N\nu_i = \sum_{i=1}^N\nu_{i+N}$ respectively, and $\sum_{i<j}^{(2N)}$ indicates the summation over all possible pairs taken from $2N$ elements with the specified condition $i<j$, as indicated. We assume all P_i are different from each other.

As an example, we write forms of F and G for $N = 2$,

$$F(x,t) = 1 + a(1,1^*)\exp(\eta_1 + \eta_1^*) + a(1,2^*)\exp(\eta_1 + \eta_2^*) + a(2,1^*)\exp(\eta_2 + \eta_1^*) + a(2,2^*)\exp(\eta_2 + \eta_2^*) + a(1,2,1^*,2^*)\exp(\eta_1 + \eta_2 + \eta_1^* + \eta_2^*), \tag{2.16}$$

$$G(x,t) = \exp(\eta_1) + \exp(\eta_2) + a(1,2,1^*)\exp(\eta_1 + \eta_2 + \eta_1^*) + a(1,2,2^*)\exp(\eta_1 + \eta_2 + \eta_2^*), \tag{2.17}$$

where

$$a(i,j^*) = (P_i + P_j^*)^{-2}(\alpha/2\gamma), \tag{2.18}$$

$$a(i,j) = (P_i - P_j)^2(2\gamma/\alpha), \tag{2.19}$$

$$a(i^*,j^*) = (P_i - P_j^*)^2(2\gamma/\alpha), \tag{2.20}$$

$$a(i,j,k^*) = a(i,j)a(i,k^*)a(j,k^*), \tag{2.21}$$

and

$$a(i,j,k^*,l^*) = a(i,j)a(i,k^*)a(i,l^*)a(j,k^*)a(j,l^*)a(k^*,l^*). \tag{2.22}$$

It is evident that $\psi(x,t)$, defined by Eqs. (2.5)-(2.15), is a solution of Eq. (2.1) provided that F and G satisfy the following equations:

$$\left[i\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right) + \beta\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 + i\gamma\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^3\right]G(x,t)F(x',t')_{\text{at } x=x',t=t'} = 0 \tag{2.23}$$

$$\gamma\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 F(x,t)F(x',t')_{\text{at } x=x',t=t'} = \alpha G(x,t)G^*(x,t), \tag{2.24}$$

We note that Eq. (2.24) can be written in the following form,

$$|\psi(x,t)|^2 = \frac{\alpha}{2\gamma}\frac{\partial^2}{\partial x^2}\log F(x,t)$$

which corresponds to the relation obtained by Zakharov and Shabat[10];

$$|u(x,t)|^2 = \sqrt{2K}\,\frac{d^2}{dx^2}\ln\det\|A\|.$$

It should be noted that the differential operator in Eq. (2.23),

$$i\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial t'}\right)+\beta\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial x'}\right)^2+i\gamma\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial x'}\right)^3, \tag{2.25}$$

is related to the linear differential operator in Eq. (2.1),

$$i\frac{\partial}{\partial t}+\beta\frac{\partial^2}{\partial x^2}+i\gamma\frac{\partial^3}{\partial x^3}. \tag{2.26}$$

Substituting the expressions for F and G into Eqs. (2.23) and (2.24), we have

$$\sum_{\nu=0,1}''\sum_{\mu=0,1}'\left[i\left(\sum_{i=1}^{2N}-\Omega_i(\nu_i-\mu_i)\right)+\beta\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^2+i\gamma\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^3\right]\times\exp\left(\sum_{i<j}^{(2N)}\varphi(i,j)(\nu_i\nu_j+\mu_i\mu_j)+\sum_{i=1}^{2N}(\nu_i+\mu_i)\eta_i\right)=0 \tag{2.27}$$

and

$$\gamma\sum_{\mu=0,1}'\sum_{\mu'=0,1}'\left(\sum_{i=1}^{2N}P_i(\mu_i-\mu_i')\right)^2\times\exp\left(\sum_{i<j}^{(2N)}\varphi(i,j)(\mu_i\mu_j+\mu_i'\mu_j')+\sum_{i=1}^{2N}(\mu_i+\mu_i')\eta_i\right)-\alpha\sum_{\nu=0,1}''\sum_{\nu'=0,1}'''\exp\left(\sum_{i<j}^{2N}\varphi(i,j)(\nu_i\nu_j+\nu_i'\nu_j')+\sum_{i=1}^{2N}(\nu_i+\nu_i')\eta_i\right)=0. \tag{2.28}$$

Let the coefficients of the factor

$$\exp\left[\sum_{i=1}^{L}\eta_i+\sum_{i=1}^{L'}\eta_{i+N}+\sum_{i=L+1}^{L+M}2\eta_i+\sum_{i=L'+1}^{L'+M'}2\eta_{i+N}\right] \tag{2.29}$$

in Eqs. (2.27) and (2.28) be D_1 $(1,2,\ldots,L';1^*,2^*,\ldots L^*;\ L+1,\ldots,L+M;(L'+1)^*,\ldots(L'+M')^*)$ and $D_2(1,2,\ldots,L;1^*,2^*,\ldots,L'^*;L+1,\ldots,L+M;(L'+1)^*,\ldots,(L'+M')^*)$, respectively.

We have, from Eq. (2.27)

$$D_1(1,2,\ldots,L;1^*,2^*,\ldots,L'^*:L+1,L+2,\ldots,L+M;(L'+1)^*\ldots,(L'+M')^*)=\sum_{\nu=0,1}''\sum_{\mu=0,1}'\text{condition }(\nu,\mu)\left[i\left(\sum_{i=1}^{2N}-\Omega_i(\nu_i-\mu_i)\right)+\beta\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^2+i\gamma\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^3\right]\times\exp\left(\sum_{i<j}^{2N}\varphi(i,j)(\nu_i\nu_j+\mu_i\mu_j)\right), \tag{2.30}$$

where condition (ν,μ) implies that the summation over μ and ν should be performed under the following conditions:

$$\nu_i+\mu_i=1 \quad \text{for } i=1,2,\ldots,L,$$
$$\nu_i=\mu_i=1 \quad \text{for } i=L+1,L+2,\ldots,L+M,$$
$$\nu_i+\mu_i=0 \quad \text{for } i=L+M+1,L+M+2,\ldots,N,$$
$$\nu_{i+N}+\mu_{i+N}=1 \quad \text{for } i=1,2,\ldots,L',$$
$$\nu_{i+N}=\mu_{i+N}=1 \quad \text{for } i=L'+1,L'+2,\ldots,L'+M',$$

and

$$\nu_{i+N}+\mu_{i+N}=0 \quad \text{for } i=L'+M'+1,L'+M'+2,\ldots,N.$$

Under the above conditions we find that the conditions of μ and ν in Eq. (2.30),

$$\sum_{i=1}^{N}\mu_i=\sum_{i=1}^{N}\mu_{i+N} \quad\text{and}\quad \sum_{i=1}^{N}\nu_i=1+\sum_{i=1}^{N}\nu_{i+N},$$

are compatible and mutually convertible from one to the other if, and only if,

$$L+2M=1+L'+2M'.$$

Let $\sigma_i=1-2\mu_i$, $\sigma_{i+N}=-1+2\mu_i$ for $i=1,2,\ldots,N'$ under the same conditions. We have

$$i\left(\sum_{i=1}^{2N}-\Omega_i(\nu_i-\mu_i)\right)+\beta\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^2+i\gamma\left(\sum_{i=1}^{2N}P_i(\nu_i-\mu_i)\right)^3=i\left(-\sum_{i=1}^{L}\Omega_i\sigma_i-\sum_{i=1}^{L'}(-\Omega_{i+N})\sigma_{i+N}\right)+\beta\left(\sum_{i=1}^{L}P_i\sigma_i+\sum_{i=1}^{L'}(-P_{i+N})\sigma_{i+N}\right)^2+i\gamma\left(\sum_{i=1}^{L}P_i\sigma_i+\sum_{i=1}^{L'}(-P_{i+N})\sigma_{i+N}\right)^3 \tag{2.31}$$

and

$$\begin{aligned}\sum_{i<j}^{(2N)}\varphi(i,j)(\nu_i\nu_j+\mu_i\mu_j)&=\sum_{i<j}^{(L)}\varphi(i,j)\tfrac{1}{2}(1+\sigma_i\sigma_j)+\sum_{i<j}^{(L')}\varphi(i+N,j+N)\tfrac{1}{2}(1+\sigma_{i+N}\sigma_{j+N})\\&+\sum_{i=1}^{L}\sum_{j=1}^{L'}\varphi(i,j+N)\tfrac{1}{2}(1-\sigma_i\sigma_{j+N})+\text{const (independent of }\sigma)\\&=\sum_{i<j}^{(L)}\log[2\gamma(P_i-P_j)^2/\alpha]\cdot\tfrac{1}{2}(1+\sigma_i\sigma_j)\\&+\sum_{i<j}^{(L')}\log[2\gamma(P_{i+N}-P_{j+N})^2/\alpha]\cdot\tfrac{1}{2}(1+\sigma_{i+N}\sigma_{j+N})\\&+\sum_{i=1}^{L}\sum_{j=1}^{L'}\log[2\gamma(P_i-(-P_{j+N}))^2/\alpha]\cdot\tfrac{1}{2}(1+\sigma_i\sigma_{j+N})+\text{const}.\end{aligned} \tag{2.32}$$

The condition of μ in Eq. (2.30),

$$\sum_{i=1}^{N}\mu_i=\sum_{i=1}^{N}\mu_{i+N},$$

is converted to

$$\sum_{i=1}^{L}\sigma_i+\sum_{i=1}^{L'}\sigma_{i+N}=1.$$

Hence, we have, for $L+L'=$ odd,

$$D_1(1,2,\ldots,L;1^*,2^*,\ldots,L'^*:L+1,\ldots,L+M;(L'+1)^*,\ldots,(L'+M')^*)=\text{const}\,\hat{D}_1(\hat{P}_1,\hat{P}_2,\ldots,\hat{P}_{L+L'}), \tag{2.33}$$

with

$$\hat{D}_1(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_{L+L'}) = \sum_{\hat{\sigma}=\pm 1}{}' h_1(\hat{P}_1\hat{\sigma}_1, \hat{P}_2\hat{\sigma}_2, \ldots, \hat{P}_{L+L'}\hat{\sigma}_{L+L'}) \times b(\hat{P}_1, \hat{\sigma}_1, \hat{P}_2, \hat{\sigma}_2, \ldots, \hat{P}_{L+L'}, \sigma_{L+L'}), \tag{2.34}$$

where

$$h_1(\hat{P}_1\hat{\sigma}_1, \hat{P}_2\hat{\sigma}_2, \ldots, \hat{P}_{L+L'}\hat{\sigma}_{L+L'}) = i\left(-\sum_{i=1}^{L+L'} \hat{\Omega}_i\hat{\sigma}_i\right) + \beta\left(\sum_{i=1}^{L+L'} \hat{P}_i\hat{\sigma}_i\right)^2 + i\gamma\left(\sum_{i=1}^{L+L'} \hat{P}_i\hat{\sigma}_i\right)^3 \tag{2.35}$$

and

$$b(\hat{P}_1, \hat{\sigma}_1, \hat{P}_2, \hat{\sigma}_2, \ldots, \hat{P}_{L+L'}, \hat{\sigma}_{L+L'}) = \prod_{i<j}^{(L+L')} [2\gamma(\hat{P}_i - \hat{P}_j)^2/\alpha]^{(1+\hat{\sigma}_i\hat{\sigma}_j)/2} \tag{2.36}$$

with

$$\hat{P}_i = P_i, \quad \hat{\Omega}_i = \Omega_i, \quad \hat{\sigma}_i = \sigma_i \quad \text{for } i = 1, 2, \ldots, L$$

and

$$\hat{P}_{i+L} = -P_i^*, \quad \hat{\Omega}_{i+L} = -\Omega_i^*, \quad \hat{\sigma}_{i+L} = \sigma_{i+N} \quad \text{for } i = 1, 2, \ldots, L',$$

and $\sum'_{\hat{\sigma}=\pm 1}$ implies the summation over all possible combinations of $\hat{\sigma}_1 = \pm 1, \ldots, \hat{\sigma}_{L+L'} = \pm 1$ under the condition $\sum_{i=1}^{L+L'} \hat{\sigma}_i = 1$, and $\prod_{i<j}^{(L+L')}$ indicates the product of all possible combinations of pairs chosen from $L + L'$ elements.

Similar procedures give, for $L + L' =$ even, that

$$D_2(1, 2, \ldots, L; 1^*, 2^*, \ldots, L'^*: L+1, \ldots, L+M; (L'+1)^*, \ldots, (L'+M')^*) = \text{const}\, \hat{D}_2(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_{L+L'}) \tag{2.37}$$

with

$$\hat{D}_2(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_{L+L'}) = \gamma \sum_{\hat{\sigma}=\pm 1}{}'' h_2(\hat{P}_1\hat{\sigma}_1, P_2\hat{\sigma}_2, \ldots, \hat{P}_{L+L'}\hat{\sigma}_{L+L'}) \times b(\hat{P}_1, \hat{\sigma}_1, \hat{P}_2, \hat{\sigma}_2, \ldots, \hat{P}_{L+L'}, \hat{\sigma}_{L+L'}) - \alpha \sum_{\hat{\sigma}=\pm 1}{}''' b(\hat{P}_1, \hat{\sigma}_1, \hat{P}_2, \hat{\sigma}_2, \ldots, \hat{P}_{L+L'}, \hat{\sigma}_{L+L'}), \tag{2.38}$$

where

$$h_2(\hat{P}_1\hat{\sigma}_1, \hat{P}_2\hat{\sigma}_2, \ldots, \hat{P}_{L+L'}\hat{\sigma}_{L+L'}) = \left(\sum_{i=1}^{L+L'} \hat{P}_i\hat{\sigma}_i\right)^2, \tag{2.39}$$

and $\sum''_{\hat{\sigma}=\pm 1}$ and $\sum'''_{\sigma=\pm 1}$ imply that the summations are over all possible combinations of $\hat{\sigma}_1 = \pm 1$, $\hat{\sigma}_2 = \pm 1, \ldots$, $\hat{\sigma}_{L+L'} = \pm 1$ under the condition $\sum_{i=1}^{L+L'} \hat{\sigma}_i = 0$ and $\sum_{i=1}^{L+L'} \hat{\sigma}_i = -2$, respectively. Thus, F and G are solutions of Eqs. (2.23) and (2.24) provided that the following identities hold:

$$\hat{D}_1(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n) = 0, \quad \text{for odd } n \tag{2.40}$$

and

$$\hat{D}_2(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n) = 0, \quad \text{for even } n. \tag{2.41}$$

We shall prove the identities by the same method as used in Papers I, II, and III. $\hat{D}_1$ and $\hat{D}_2$ have the following properties: (i) $\hat{D}_1$ and $\hat{D}_2$ are symmetric polynomials of $\hat{P}_1, \hat{P}_2, \ldots,$ and $\hat{P}_n$; (ii) if $\hat{P}_1 = \hat{P}_2$ then

$$\hat{D}_1(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n) = 2\prod_{j=3}^{n} [2\gamma(\hat{P}_1 - \hat{P}_j)^2/\alpha] \cdot \hat{D}_1(\hat{P}_3, \hat{P}_4, \ldots, \hat{P}_n) \tag{2.42}$$

and

$$\hat{D}_2(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n) = 2\prod_{j=3}^{n} [2\gamma(\hat{P}_1 - \hat{P}_j)^2/\alpha] \cdot \hat{D}_2(\hat{P}_3, \hat{P}_4, \ldots, \hat{P}_n). \tag{2.43}$$

The identities $\hat{D}_1 = 0$ and $\hat{D}_2 = 0$ are easily verified for $n = 1$ and $n = 2$, respectively. Now, assume that the identities hold for $n - 2$. Then, relying on properties (i) and (ii), it is seen that $\hat{D}_1$ and $\hat{D}_2$ can be respectively factorized by a symmetric polynomial,

$$\prod_{k<l}^{(n)} (\hat{P}_k - \hat{P}_l)^2$$

of degree $n(n-1)$. On the other hand, Eqs. (2.35)-(2.39) show that the degree of $\hat{D}_1$ and $\hat{D}_2$ to be $\frac{1}{2}(n-1)^2 + 3$ and $\frac{1}{2}n(n-2) + 2$, respectively. Hence, $\hat{D}_1$ and $\hat{D}_2$ must be zero for n, and the identities have been proved.

APPENDIX A: SOLUTIONS OF THE MODIFIED KORTEWEG-DE VRIES EQUATION

As described in the previous section, in the limit $\beta = \delta = 0$ we have the modified Korteweg-de Vries equation for real ψ $(= G/F)$

$$\frac{\partial \psi}{\partial t} + 3\alpha\psi^2 \frac{\partial \psi}{\partial t} + \gamma \frac{\partial^3 \psi}{\partial x^3} = 0, \tag{A1}$$

and

$$\left[\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right) + \gamma\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^3\right] G(x,t)F(x',t')_{\text{at } x=x', t=t'} = 0, \tag{A2}$$

$$\gamma\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 F(x,t)F(x',t')_{\text{at } x=x', t=t'} - \alpha G(x,t)G(x,t) = 0. \tag{A3}$$

On the other hand, in a previous paper we obtained N-soliton solutions to the equation

$$\frac{\partial v}{\partial t} + 24v^2 \frac{\partial v}{\partial x} + \frac{\partial^3 v}{\partial x^3} = 0 \tag{A4}$$

in the following

$$v = \frac{\partial}{\partial x} \tan^{-1}(g/f) \tag{A5}$$

$$= \frac{gxf - gfx}{f^2 + g^2}, \tag{A6}$$

where f and g satisfy the following equations:

$$\left[\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right) + \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^3\right] g(x,t)f(x',t')_{\text{at } x=x', t=t'} = 0 \tag{A7}$$

and

$$\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial x'}\right)^2[f(x,t)f(x',t') + g(x,t)g(x',t')]_{\text{at } x=x',t=t'}=0. \quad \text{(A8)}$$

A comparison of Eqs. (A1) and (A4) suggests that the functions $F(x,t)$ and $G(x,t)$ can be expressed in terms of $f(x,t)$ and $g(x,t)$ of Eqs. (A7) and (A8). For simplicity, we put $\alpha=2$ and $\gamma=1$, then we have

$$F(x,t)=f^2+g^2, \quad \text{(A9)}$$

$$G(x,t)=2(g_x f-g f_x). \quad \text{(A10)}$$

It is easily shown that F and G defined by the above equations actually satisfy Eqs. (A2) and (A3). Substituting Eqs. (A9) and (A10) into Eqs. (A2) and (A3) gives

$$\begin{aligned}&G_t F-GF_t+G_{xxx}F-3G_{xx}F_x+3G_xF_{xx}-GF_{xxx}\\&=-2[g_tf-gf_t+g_{xxx}f-3g_{xx}f_x+3g_xf_{xx}-gf_{xxx}]\\&\times\frac{\partial}{\partial x}(f^2+g^2)+2(f^2+g^2)\\&\times\frac{\partial}{\partial x}[g_tf-gf_t+g_{xxx}f-3g_{xx}f_x+3g_xf_{xx}-gf_{xxx}]\\&+12(ff_{xx}-f_x^2+gg_{xx}-g_x^2)\frac{\partial}{\partial x}(g_xf-gf_x)\\&-12(g_xf-gf_x)\frac{\partial}{\partial x}(ff_{xx}-f_x^2+gg_{xx}-g_x^2),\end{aligned} \quad \text{(A11)}$$

and

$$\begin{aligned}&FF_{xx}-F_x^2-G^2\\&\quad=2(f^2+g^2)(ff_{xx}-f_x^2+gg_{xx}-g_x^2),\end{aligned} \quad \text{(A12)}$$

respectively, which clearly shows that F and G satisfy Eqs. (A2) and (A3) provided that f and g satisfy Eqs. (A7) and (A8).

APPENDIX B: EXPLICIT FORM OF 2-ENVELOPE-SOLITON SOLUTION

In order to understand the behaviors of two-envelope solitons we rewrite ψ for $N=2$ in a symmetrical form as follows:

$$\psi(x,t)=G(x,t)/F(x,t) \quad \text{(B1)}$$

and

$$F(x,t)=\cosh a+\sinh a\,(\tanh\xi_1\tanh\xi_2 - \operatorname{sech}\xi_1\operatorname{sech}\xi_2\cos(\zeta_1-\zeta_2)), \quad \text{(B2)}$$

$$G(x,t)=A_1\operatorname{sech}\xi_1\exp(i\zeta_1)(\cos\phi_1+i\sin\phi_1\tanh\xi_2) + A_2\operatorname{sech}\xi_2\exp(i\zeta_2)(\cos\phi_2+i\sin\phi_2\tanh\xi_1), \quad \text{(B3)}$$

where

$$\exp a=|P_1-P_2|/|P_1+P_2^*|, \quad \text{(B4)}$$

$$\phi_1=\arg[(P_1-P_2)/(P_1+P_2^*)], \quad \text{(B5)}$$

$$\phi_2=\arg[(P_2-P_1)/(P_2+P_1^*)], \quad \text{(B6)}$$

and, for $i=1,2$,

$$A_i=(\gamma/2\alpha)^{1/2}|P_i+P_i^*|, \quad \text{(B7)}$$

$$\xi_i=\mathrm{Re}(P_ix-\Omega_it)+\text{const}, \quad \text{(B8)}$$

$$\zeta_i=\mathrm{Im}(P_ix-\Omega_it)+\text{const}. \quad \text{(B9)}$$

In the limit $t\to-\infty$, keeping ξ_1 finite, we have for $\mathrm{Re}(\Omega_2-\Omega_1)>0$,

$$F=\cosh a-\sinh a\tanh\xi_1, \quad \text{(B10)}$$

$$G=A_1\operatorname{sech}\xi_1\exp(i(\zeta_1-\phi_1)) \quad \text{(B11)}$$

and

$$\psi(x,t)=\frac{A_1\exp(i(\zeta_1-\phi_1))}{\cosh(\xi_1-a)}. \quad \text{(B12)}$$

In the limit $t\to\infty$, keeping ξ_1 finite, we have for $\mathrm{Re}(\Omega_2-\Omega_1)>0$,

$$F=\cosh a+\sinh a\tanh\xi_1, \quad \text{(B13)}$$

$$G=A_1\operatorname{sech}\xi_1\exp(i(\zeta_1+\phi_1)) \quad \text{(B14)}$$

and

$$\psi(x_1t)=\frac{A_1\exp(i(\zeta_1+\phi_1))}{\cosh(\xi_1+a)}. \quad \text{(B15)}$$

The present form for $\psi(x,t)$ is convenient for calculating detailed behaviors of two envelope-solitons during the overlap time interval.

[1]Ryogo Hirota, Phys. Rev. Lett. **27**, 1192 (1971).
[2]Ryogo Hirota, J. Phys. Soc. Jap. **33**, 1456 (1972).
[3]Ryogo Hirota, J. Phys. Soc. Jap. **33**, 1459 (1972).
[4]N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).
[5]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **60**, 993 (1971) [Sov. Phys.-JETP **33**, 538 (1971)].
[6]Miki Wadati and Morikazu Toda, J. Phys. Soc. Jap. **32**, 1403 (1972).
[7]R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[8]A. C. Scott, Am. J. Phys. **37**, 52 (1969).
[9]G. L. Lamb, Jr., Rev. Mod. Phys. **43**, 99 (1971).
[10]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys.-JETP **34**, 62 (1972)].
[11]Miki Wadati, J. Phys. Soc. Jap. **32**, 1681 (1972).
[12]F. Ursell, Proc. Camb. Philos. Soc. **49**, 685 (1953).
[13]N. J. Zabusky, *A Synergetic Approach to Problems of Nonlinear Dispersive Wave Propagation and Interaction, Nonlinear Partial Differential Equations* (Academic, New York, 1967).
[14]Ryogo Hirota and Kimio Suzuki, J. Phys. Soc. Jap. **28**, 1366 (1970).
[15]M. Toda, J. Phys. Soc. Jap. **23**, 501 (1967).

Exact N-soliton solutions of the wave equation of long waves in shallow-water and in nonlinear lattices

Ryogo Hirota

RCA Research Laboratories, Tokyo International 100-31, Japan
(Received 20 November 1972)

Exact N-soliton solutions have been obtained for the nonlinear wave equation $\partial^2 w/\partial t^2 - \partial^2 w/\partial x^2 - 6(\partial^2 w^2/x^2) - \partial^4 w/\partial x^4 = 0$ which describes motions of long waves in one-dimensional nonlinear lattices and in shallow-water under gravity. The solutions have the same functional form as N-solition solutions of the Korteweg-de Vries equation.

1. INTRODUCTION

This paper is the fifth of a series of papers on exact N-solutions.[1-4] In the present paper we obtain exact N-soliton solutions of the following nonlinear wave equation,

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} - 6\frac{\partial^2 w^2}{\partial x^2} - \frac{\partial^4 w}{\partial x^4} = 0, \tag{1.1}$$

which describes motions of long waves in one-dimensional nonlinear lattices[5] and in shallow water under gravity[6]. The quantities w, x, and t can be rescaled to produce any desired coefficients for the terms of Eq. (1.1). The present choice is convenient for this paper.

It is known that Eq. (1.1) reduces to the Korteweg-de Vries equation in the long wave and weakly nonlinear limit,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \delta\frac{\partial^3 u}{\partial x^3} = 0, \quad \delta = \text{const}, \tag{1.2}$$

for which exact N-soliton solutions have been obtained.[1,7,8]

It is noted that Eq. (1.1) is of the second order in time and Eq. (1.2) is of the first; hence Eq. (1.1) describes solitons[9,10] propagating in opposite directions while Eq. (1.2) does not. In spite of this fact we will find that N-soliton solutions of Eq. (1.1) have the same functional forms as those of Eq. (1.2).

2. FUNDAMENTAL PROPERTIES OF SOLITON SOLUTIONS

In this section we consider fundamental properties of soliton solutions without knowing exact solutions. In a previous paper[10] we explained the fundamental properties of "lattice solitons" in terms of the properties of nonlinear LC networks.

In order to follow a similar procedure here, we transform Eq. (1.1) into the following simultaneous equation by introducing a new function $u(x, t)$

$$\frac{\partial w}{\partial t} = -\frac{\partial u}{\partial x}, \tag{2.1}$$

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(w + 6w^2 + w_{xx}), \tag{2.2}$$

where the subscript indicates the partial differentiation.

For a wave pulse (soliton) solution defined by the condition that $w(x, t), u(x, t)$ and their derivatives vanish at $|t| = \infty$ and/or $|x| = \infty$, we obtain the following conservation laws by integrating Eqs. (2.1) and (2.2):

$$\int_{-\infty}^{\infty} w(x, t)dx = \text{const}, \tag{2.3}$$

$$\int_{-\infty}^{\infty} u(x, t)dt = \text{const}, \tag{2.4}$$

$$\int_{-\infty}^{\infty} u(x, t)dx = \text{const}, \tag{2.5}$$

$$\int_{-\infty}^{\infty} u(x, t)\, w(x, t)dx = \text{const}, \tag{2.6}$$

and

$$\int_{-\infty}^{\infty} [\tfrac{1}{2}(u^2 + w^2) + 4w^3 + ww_{xx} - \tfrac{1}{2}w_x^2]dt = \text{const}, \tag{2.7}$$

where Eqs. (2.6) and (2.7) correspond to the total energy and power conservation laws, respectively. We shall explain the fundamental properties of soliton solutions by using these conservation laws.

(i) Velocity of a soliton: If a wave pulse propagates through a medium without changing its shape and velocity,

$$w(x, t) = w(px - \Omega t) \tag{2.8}$$

and

$$u(x, t) = u(Px - \Omega t), \tag{2.9}$$

we obtain the following relations:

$$\Omega w(x, t) = Pu(x, t), \tag{2.10}$$

$$\Omega u(x, t) = P(w + 6w^2 + w_{xx}). \tag{2.11}$$

From Eqs. (2.10) and (2.11), we have

$$(\Omega/P)^2 = (\int_{-\infty}^{\infty} w dx + 6\int_{-\infty}^{\infty} w^2 dx)/\int_{-\infty}^{\infty} w dx, \tag{2.12}$$

which shows that the velocity of a wave pulse, Ω/P, is greater than unity and increases with increasing pulse height provided that $w(x, t) > 0$.

(ii) Collision of two solitons: As was pointed out in previous papers,[5,10,11] when two solitons collide and overlap with each other, the total energy and power conservation laws play important roles in determining the behavior of the joint amplitude of the two solitons during the overlapping time interval. We shall return to this point after we obtain an explicit form for two-soliton solutions.

(iii) Conservation of the total phase shift: We shall prove that a weighted sum of the phase shifts of individual solitons gained by multiple collisions with other solitons is zero.[8,10]

Consider a sufficiently large time T. For $t < -T < 0$ we assume N independent solitons with phases $\delta_-(i)$ $(i = 1, 2, \ldots, N)$ before collisions,

$$w^{(-)}(x, t) = \sum_{i=1}^{N} w_i(P_i x - \Omega_i t - \delta_-(i)). \tag{2.13}$$

Copyright © 1973 by the American Institute of Physics

For $t > T$, we assume that each soliton changes its phase only, from $\delta_-(i)$ to $\delta_+(i)$, after multiple collisions with other solitons,

$$w^{(+)}(x,t) = \sum_{i=1}^{N} w_i(P_i x - \Omega_i t - \delta_+(i)). \tag{2.14}$$

We rewrite one of the conservation laws, Eq. (2.5),

$$\int_{-\infty}^{\infty} u(x,t)dx = \int_{-\infty}^{\infty} x\left(-\frac{\partial}{\partial x}u(x,t)\right)dx \tag{2.15}$$

$$= \frac{\partial}{\partial t}\int_{-\infty}^{\infty} x\,w(x,t)dx. \tag{2.16}$$

Substituting Eq. (2.13) into Eq. (2.16) gives, for $t < -T$,

$$\int_{-\infty}^{\infty} u(x,t)dx = \sum_{i=1}^{N} \frac{\Omega_i}{P_i^2}\int_{-\infty}^{\infty} w_i(\eta_i)d\eta_i. \tag{2.17}$$

On the other hand, integrating Eq. (2.16) from $T_1 < -T$ to $T_2 > T$ and employing Eqs. (2.13) and (2.14), we have

$$(T_2 - T_1)\int_{-\infty}^{\infty} u(x,t)dx = \sum_{i=1}^{N} [\Omega_i(T_2 - T_1) + \delta_+(i) - \delta_-(i)] \times P_i^{-2}\int_{-\infty}^{\infty} w_i(\eta_i)d\eta_i. \tag{2.18}$$

Substituting Eq. (2.17) into Eq. (2.18) gives

$$\sum_{i=1}^{N} [\delta_+(i) - \delta_-(i)]P_i^{-2}\int_{-\infty}^{\infty} w_i(\eta_i)d\eta_i = 0, \tag{2.19}$$

which shows that the weighted sum of the phase shifts of individual solitons is zero. As is shown later, the explicit form of a single soliton gives

$$P_i^{-2}\int_{-\infty}^{\infty} w_i(\eta_i)d\eta_i = 1 \tag{2.20}$$

and hence

$$\sum_{i=1}^{N} [\delta_+(i) - \delta_-(i)] = 0. \tag{2.21}$$

3. EXACT N-SOLITON SOLUTIONS

We now describe exact N-soliton solutions of the equation

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} - 6\frac{\partial^2 w^2}{\partial x^2} - \frac{\partial^4 w}{\partial x^4} = 0. \tag{3.1}$$

Exact N-soliton solutions can be expressed in the following form:

$$w(x,t) = \frac{\partial^2}{\partial x^2}\log f(x,t), \tag{3.2}$$

$$f(x,t) = \sum_{\mu=0,1} \exp\left(\sum_{i<j}^{(N)} \varphi(i,j)u_i u_j + \sum_{i=1}^{N} u_i\eta_i\right), \tag{3.3}$$

$$\eta_i = P_i x - \epsilon_i\Omega_i t - \eta_i^0 \quad (\epsilon_i = +1 \text{ or } -1), \tag{3.4}$$

$$\Omega_i = P_i(1 + P_i^2)^{1/2}, \tag{3.5}$$

$$\exp[\varphi(i,j)] = -\frac{(\epsilon_i\Omega_i - \epsilon_j\Omega_j)^2 - (P_i - P_j)^2 - (P_i - P_j)^4}{(\epsilon_i\Omega_i + \epsilon_j\Omega_j)^2 - (P_i + P_j)^2 - (P_i + P_j)^4} \tag{3.6}$$

$$= \frac{(\epsilon_i v_i - \epsilon_j v_j)^2 + 3(P_i - P_j)^2}{(\epsilon_i v_i - \epsilon_j v_j)^2 + 3(P_i + P_j)^2}, \tag{3.7}$$

$$v_i = (1 + P_i^2)^{1/2}, \tag{3.8}$$

where P_i and η_i^0 are the real constants relating to the amplitude and phase, respectively, of the ith soliton, $\Sigma_{\mu=0,1}$ implies the summation over all possible combinations of $\mu_1 = 0, 1$, $\mu_2 = 0, 1, \ldots, \mu_N = 0.1$ and $\Sigma_{i<j}^{(N)}$ indicates the summation over all possible pairs chosen from N elements.

We note that the above expression for f is equivalent to the following form:

$$f(x,t) = 1 + \sum_{n=1}^{N}\sum_{{}_NC_n} a(i_1, i_2, \ldots, i_n) \times \exp(\eta_{i_1} + \eta_{i_2} + \cdots + \eta_{i_n}) \tag{3.9}$$

where

$$a(i_1, i_2, \ldots, i_n) = \prod_{k<l}^{(n)} a(i_k, i_l), \tag{3.10}$$

$$a(i_k, i_j) = \exp[\varphi(i_k, i_l)]. \tag{3.11}$$

${}_NC_n$ indicates the summation over all possible combinations of n elements taken from N, and $\prod_{k<l}^{(n)}$ indicates the product of all possible combinations of the n elements. This expression for f has the same functional form as one obtained for N-soliton solutions of the Korteweg–de Vries equation.[1]

As an example we write the form of $f(x,t)$ for $N = 2$,

$$f(x,t) = 1 + e^{\eta_1} + e^{\eta_2} + a(1,2)e^{\eta_1+\eta_2}. \tag{3.12}$$

It is easily seen that $w(x,t)$ defined by Eq. (3.2) is a solution of Eq. (3.1) provided that f satisfies the following equation:

$$\left[\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^2 - \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 - \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^4\right] \times f(x,t)f(x',t')_{\text{at } x=x',\, t=t'} = 0. \tag{3.13}$$

It is noted that the differential operators in Eq. (3.13),

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^2 - \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^2 - \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^4, \tag{3.14}$$

are related to the linear differential operator in Eq. (3.1),

$$\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^4}{\partial x^4}. \tag{3.15}$$

Substituting Eq. (3.3) into Eq. (3.13), we have

$$\sum_{\mu=0,1}\sum_{\nu=0,1}\left[\left(\sum_{i=1}^{N}\epsilon_i\Omega_i(\mu_i - \nu_i)\right)^2 - \left(\sum_{i=1}^{N}P_i(\mu_i - \nu_i)\right)^2 - \left(\sum_{i=1}^{N}P_i(\mu_i - \nu_i)\right)^4\right] \times \exp\left(\sum_{i<j}^{(N)}\varphi(i,j)(\mu_i\mu_j + \nu_i\nu_j) + \sum_{i=1}^{N}(\mu_i + \nu_i)\eta_i\right) = 0. \tag{3.16}$$

812

Let the coefficients of the factor

$$\exp\left(\sum_{i=1}^{n}\eta_i+\sum_{i=n+1}^{m}2\eta_i\right) \tag{3.17}$$

in Eq. (3.16) be $D(1,2,\ldots n;n+1,n+2,\ldots,m)$. We have

$$\begin{aligned}D(1,2,\ldots,n;n+1,n+2,\ldots,m)&=\sum_{\mu=0,1}\sum_{\nu=0,1}\mathrm{cond}(\mu,\nu)\cdot\left[\left(\sum_{i=1}^{N}\epsilon_i\Omega_i(\mu_i-\nu_i)\right)^2\right.\\&\left.-\left(\sum_{i=1}^{N}(\mu_i-\nu_i)P_i\right)^2-\left(\sum_{i=1}^{N}(\mu_i-\nu_i)P_i\right)^4\right]\\&\times\exp\left(\sum_{i<j}^{(N)}\varphi(i,j)(\mu_i\mu_j+\nu_i\nu_j)\right),\end{aligned} \tag{3.18}$$

where $\mathrm{cond}(\mu,\nu)$ implies that the summation over μ and ν should be performed under the conditions

$$\begin{aligned}&\mu_i+\nu_i=1 \quad \text{for } i=1,2,\ldots,n,\\&\mu_i=\nu_i=1 \quad \text{for } i=n+1,n+2,\ldots,m,\\&\mu_i=\nu_i=0 \quad \text{for } i=m+1,m+2,\ldots,N.\end{aligned} \tag{3.19}$$

Let $1-2\mu_i=\sigma_i$, then we find

$$D(1,2,\ldots,n;n+1,n+2,\ldots,m)=\mathrm{const}\cdot\hat{D}(P_1,P_2,\ldots,P_n) \tag{3.20}$$

with

$$\hat{D}(P_1,P_2,\ldots,P_n)=\sum_{\sigma=\pm1}h(P_1\sigma_1,P_2\sigma_2,\ldots,P_n\sigma_n)b(P_1,\sigma_1,P_2,\sigma_2,\ldots,P_n,\sigma_n), \tag{3.21}$$

where

$$h(P_1\sigma_1,P_2\sigma_2,\ldots,P_n\sigma_n)=\left(\sum_{i=1}^{n}\epsilon_iP_iv_i\sigma_i\right)^2-\left(\sum_{i=1}^{n}P_i\sigma_i\right)^2-\left(\sum_{i=1}^{n}P_i\sigma_i\right)^4 \tag{3.22}$$

and

$$b(P_1,\sigma_1,P_2,\sigma_2,\ldots,P_n,\sigma_n)=\prod_{i<j}^{(n)}[(\epsilon_iv_i-\epsilon_jv_j)^2+3(P_i\sigma_i-P_j\sigma_j)^2]. \tag{3.23}$$

Here, we have used the relation

$$\begin{aligned}\varphi(i,j)\tfrac{1}{2}(1+\sigma_i\sigma_j)&=\tfrac{1}{2}(1+\sigma_i\sigma_j)\log\frac{(\epsilon_iv_i-\epsilon_jv_j)^2+3(P_i-P_j)^2}{(\epsilon_iv_i-\epsilon_jv_j)^2+3(P_i+P_j)^2} \qquad (3.24)\\&=\log[(\epsilon_iv_i-\epsilon_jv_j)^2+3(P_i\sigma_i-P_j\sigma_j)^2]\\&\quad-\log[(\epsilon_iv_i-\epsilon_jv_j)^2+3(P_i+P_j)^2]. \qquad (3.25)\end{aligned}$$

Thus, $f(x,t)$ is the solution of Eq. (3.11) provided that the following identity holds:

$$\hat{D}(P_1,P_2,\ldots,P_n)=0 \quad \text{for } n=1,2,\ldots,N. \tag{3.26}$$

The identity can be proved by mathematical induction as used in the previous papers.[1-4] $\hat{D}(P_1,P_2,\ldots,P_n)$ is a symmetric and even function of $P_1,P_2,\ldots,P_n$, and hence if we consider $\hat{D}$ to be a function of $\epsilon_1v_1,\epsilon_2v_2,\ldots,\epsilon_nv_n$ $[v_i=(1+P_i^2)^{1/2}]$, $\hat{D}$ becomes a polynomial of variables $\epsilon_1v_1,\epsilon_2v_2,\ldots$, and ϵ_nv_n.

Let $\epsilon_iv_i=\hat{v}_i$ for $i=1,2,\ldots,n$, and $\hat{D}=\hat{D}(\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_n)$.

We find that the polynomial $\hat{D}(\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_n)$ has the following properties:

(i) $\hat{D}$ is a symmetric polynomial of $\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_n$,

(ii) if $\hat{v}_1=\pm1$,

$$\hat{D}(\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_n)_{\mathrm{at}\,\hat{v}_1=\pm1}=\hat{D}(\hat{v}_2,\hat{v}_3,\ldots,\hat{v}_n)\cdot2\prod_{j=2}^{n}[(\hat{v}_i-\hat{v}_j)^2+3P_j^2] \tag{3.27}$$

and

(iii) if $\hat{v}_1=\hat{v}_2$,

$$\hat{D}(\hat{v}_1,\hat{v}_2,\ldots,\hat{v}_n)_{\mathrm{at}\,\hat{v}_1=\hat{v}_2}=\hat{D}(\hat{v}_3,\hat{v}_4,\ldots,\hat{v}_n)\cdot24P_i^2\times\prod_{j=3}^{n}[(\hat{v}_1-\hat{v}_j)^2+3(P_1-P_j)^2][(\hat{v}_1-\hat{v}_j)^2+3(P_1+P_j)^2] \tag{3.28}$$

The identity is easily verified for $n=1$ and 2. Now, assume that the identity holds for $n-1$ and $n-2$. Then, relying on properties (i), (ii), and (iii), we see that $\hat{D}$ can be factored by a polynomial,

$$\prod_{i<j}^{(n)}(\hat{v}_i-\hat{v}_j)^2\cdot\prod_{i=1}^{n}(\hat{v}_i^2-1) \tag{3.29}$$

of degree $n(n-1)+2n$. On the other hand, Eq. (3.21) shows the degree of $\hat{D}$ to be $n(n-1)+4$. Hence, $\hat{D}$ must be zero for n, and the identity holds.

We shall prove that the solution, Eq. (3.9), splits into N solitons in the limit $|t|\to\infty$, and obtain the phase shift of an arbitrary ith soliton induced by collision with $N-1$ other solitons.

At first we consider the case of $t\to\infty$. In the limit $t\to\infty$, keeping η_i finite, we have

$$\eta_i=P_j[(\Omega_i/P_i)-(\Omega_j/P_j)]t-\eta_j^0 \tag{3.30}$$

and can assume without losing generality,

$$\eta_1,\eta_2,\ldots,\eta_{i-1}=\infty, \tag{3.31}$$

$$\eta_i=\text{finite}, \tag{3.32}$$

and

$$\eta_{i+1},\eta_{i+2},\ldots,\eta_N=-\infty. \tag{3.33}$$

Therefore, we find

$$\lim_{\substack{t\to\infty\\ \eta_i\,\text{finite}}}\exp(-\eta_1-\eta_2-\ldots-\eta_{i-1})f(x,t)=a(1,2,\ldots,i-1)+a(1,2,\ldots,i)\exp(\eta_i) \tag{3.34}$$

$$=a(1,2,\ldots,i-1)\{1+\exp[\eta_i-\delta_+(i)]\}, \tag{3.35}$$

where

$$\delta_+(i)=-\sum_{j=1}^{i-1}\varphi(j,i) \tag{3.36}$$

and obtain an asymptotic form of the ith soliton in the limit $t \to \infty$,

$$w_+(x,t) = (P_i/2)^2 \operatorname{sech}^2\{[\eta_i - \delta_+(i)]/2\}. \tag{3.37}$$

The same procedure as the one used for the case of $t = \infty$ gives an asymptotic form of the ith soliton in the limit $t \to -\infty$. In the limit $t \to -\infty$, keeping η_i finite, we find

$$w_-(x,t) = (P_i/2)^2 \operatorname{sech}^2\{[\eta_i - \delta_-(i)]/2\}, \tag{3.38}$$

where

$$\delta_-(i) = -\sum_{j=i+1}^{N} \varphi(i,j). \tag{3.39}$$

Thus, we have proved that the solution really splits apart into N solitons in the limit of $|t| \to \infty$. This implies that a soliton conserves its identity after colliding with other solitons.[8,12] The effect of collisions appears only in the relative phase shift $\delta(i)$ defined by

$$\delta(i) = \delta_+(i) - \delta_-(i). \tag{3.40}$$

As was proved in the previous section, we find

$$\sum_{i=1}^{N} \delta(i) = 0. \tag{3.41}$$

For a single soliton, we find by using Eq. (3.37) that

$$\int_{-\infty}^{\infty} w_i(\eta_i)d\eta_i = (P_i/2)^2 \int_{-\infty}^{\infty} \operatorname{sech}^2\{[\eta_i - \delta(i)]/2\}d\eta_i = P_i^2, \tag{3.42}$$

which assures Eq. (2.20).

4. TWO-SOLITON SOLUTIONS

The results obtained in the previous section show that the relative phase shift $\delta(i)$ induced by multiple collisions with $N-1$ other solitons is the sum of the relative phase shifts induced by an independent collision with each of $N-1$ other solitons; there is no many-body effect. Therefore, it suffices to study a single collision to obtain an understanding of multiple collisions of N solitons.

We write the two-soliton solutions in the following form:

$$w(x,t) = \frac{q_1^2 \operatorname{sech}^2\xi_1 + q_2^2 \operatorname{sech}^2\xi_2 + A \operatorname{sech}^2\xi_1 \operatorname{sech}^2\xi_2}{[\cosh(\phi/2) + \sinh(\phi/2)\tanh\xi_1 \tanh\xi_2]^2}, \tag{4.1}$$

where

$$\xi_1 = q_1 x - \omega_1 t, \tag{4.2}$$

$$\xi_2 = q_2 x - \omega_2 t, \tag{4.3}$$

$$\omega_1 = \epsilon_1 q_1 v_1, \tag{4.4}$$

$$\omega_2 = \epsilon_2 q_2 v_2, \tag{4.5}$$

$$v_1 = (1 + 4q_1^2)^{1/2}, \tag{4.6}$$

$$v_2 = (1 + 4q_2^2)^{1/2}, \tag{4.7}$$

and

$$\exp(2\phi) = \frac{(\epsilon_1 v_1 - \epsilon_2 v_2)^2 + 12(q_1 - q_2)^2}{(\epsilon_1 v_1 - \epsilon_2 v_2)^2 + 12(q_1 + q_2)^2}, \tag{4.8}$$

$$A = \sinh(\phi/2)[(q_1^2 + q_2^2)\sinh(\phi/2) + 2q_1 q_2 \cosh(\phi/2)], \tag{4.9}$$

where q_1 and q_2 are arbitrary real constants.

The functional form of $w(x,t)$ is the same as that of two-lattice soliton solutions $V_n(t)$,[9,10] which are the solutions of the following equation

$$\frac{\partial^2}{\partial t^2}\log[1 + V_n(t)] = V_{n+1}(t) + V_{n-1}(t) - 2V_n(t). \tag{4.10}$$

We give numerical results of Eq. (4.1) in Figs. 1 and 2 for the case where two solitons at $|t| = \infty$ have disparate amplitudes ($q_1^2 = 25.0$, $q_2^2 = 6.25$) and are moving in the same direction and for the case where

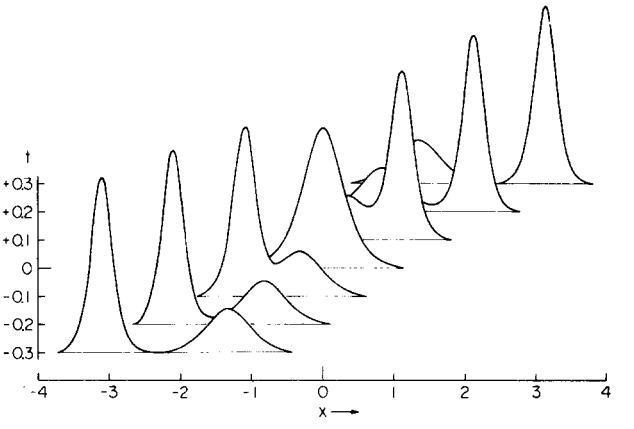

FIG. 1. A collision of two solitons having disparate amplitudes, ($q_1^2 = 25.0, q_2^2 = 6.25$) travelling in the same direction.

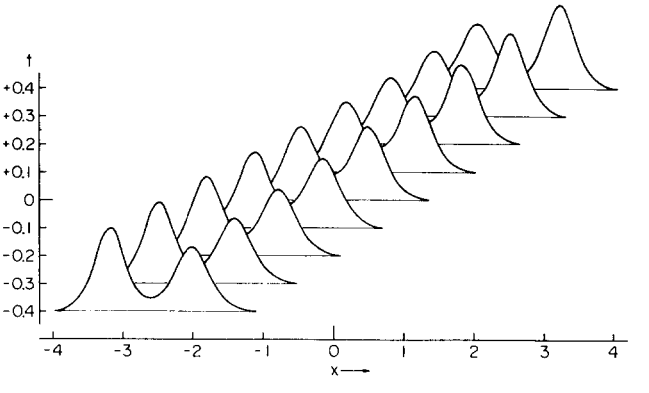

FIG. 2. A collision of two solitons having more nearly equal amplitudes, ($q_1^2 = 12.25, q_2^2 = 9.0$) travelling in the same direction.

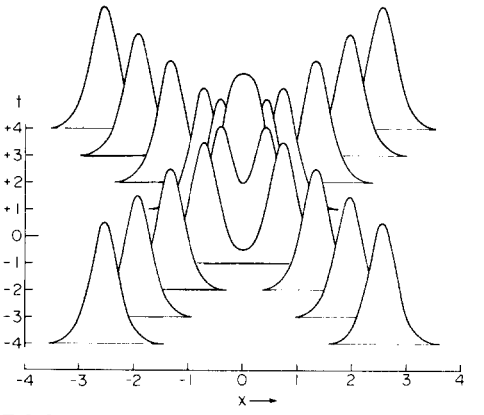

FIG. 3. A head-on-collision of two solitons having the same amplitude ($q_1^2 = q_2^2 = 9.0$) moving in the opposite directions.

the amplitudes of the two solitons are more nearly equal to each other ($q_1^2 = 12.25$, $q_2^2 = 9.0$), respectively. For these two cases, the features of two-soliton interactions are quite similar to those obtained for two-lattice solitons[10] and for solitons of the Korteweg–de Vries equation.[5]

In Fig. 3 we give a numerical result of Eq. (4.1) for the case where two solitons of equal amplitude ($q_1^2 = q_2^2 = 9.0$) are moving in opposite directions. We find a distinction between the present result and that of the two-lattice soliton. When two solitons of equal amplitude collide and overlap, their joint amplitude is *smaller* than twice the amplitude of an individual soliton. We showed in the previous paper[10] that when two lattice solitons of equal amplitude moving in opposite directions collide and overlap, their joint amplitude becomes *greater* than twice of the amplitude of an individual soliton. The difference is of importance if one remembers that the present wave equation (1.1) is a long and weakly nonlinear approximation of Eq. (4.10). This implies that the long and weakly nonlinear approximation breaks down when it describes phenomena such as the head-on collision of two solitons where their joint amplitude becomes large and steep.

[1]Ryogo Hirota, Phys. Rev. Lett. **27**, 1192 (1971).
[2]Ryogo Hirota, J. Phys. Soc. Jap. **33**, 1456 (1972).
[3]Ryogo Hirota, J. Phys. Soc. Jap. **33**, 1459 (1972).
[4]Ryogo Hirota, J. Math. Phys. **14**, 000 (1973).
[5]N. J. Zabusky, "A synergetic approach to problems of nonlinear dispersive wave propagtion and interaction" in *Nonlinear Partial Equations*, edited by W. F. Ames (Academic, New York, 1967).
[6]F. Ursell, Proc. Camb. Philos. Soc. **49**, 685 (1953).
[7]V. E. Zakharov, Zh. Eksp. Teor.-Fiz. **60**, 993 (1971) [Sov. Phys.-JETP **33**, 538 (1971)].
[8]Miki Wadati and Morikazu Toda, J. Phys. Soc. Jap. **32**, 1403 (1972).
[9]M. Toda, Proceedings International Conference, Statistical Mechanics, Kyoto 1968, J. Phys. Soc. Jap. Suppl. **26**, 235 (1969).
[10]Ryogo Hirota and Kimio Suzuki, Proc. IEEE, July (1973) (to be published).
[11]Peter D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).

Dynamics for classical relativistic particles: Circular orbit solutions and the nonrelativistic limit

P. Cordero*

International Centre for Theoretical Physics, Trieste, Italy

G.-C. Ghirardi †

Istituto di Fisica Teorica dell'Università, Trieste, Italy
(Received 24 April 1972; final revised manuscript received 3 October 1972)

This work consists of two parts. First we consider a general classical relativistic direct interaction theory (for two point particles) coming from an action principle such that the equations of motion have, at most, second-order derivatives of the position 4-vectors with respect to proper time. We introduce the general formalism, derive the explicit form of the conserved quantities, and discuss some particular cases which had previously been considered in the literature. We analyze in detail the circular orbit solutions of the general equations of motion, deriving in particular the constraints which the interaction function must satisfy in order that such solutions exist. In the second part we study in detail the nonrelativistic limit of the theory, showing that the limit is perfectly consistent. The obtained Newton equations contain an arbitrary potential that depends on the relative position and velocity.

I. INTRODUCTION

An alternative way to look at the relativistic bound state problem in quantum physics, besides techniques such as the Bethe–Salpeter equation, is to start from the corresponding problem in classical physics and then to quantize. This approach, as far as the general two-body problem is concerned, is relatively new and up to now the only results in this line have been obtained using circular orbit solutions.[1-4]

The classical two-body problem, so simple in principle in nonrelativistic mechanics, has at least two important sources of complications when considered in a relativistic framework.

One of the sources of difficulties—that to a certain extent we can eliminate — is the fact that interactions between particles are usually thought of as a mediated by fields. Fields that carry infinite degrees of freedom, carry away 4-momentum and angular momentum in the form of radiation and which, therefore, transform the problem into a problem of particles *and fields*.

An interesting line of thought developed by various authors[5] allows the problem to be simplified. They assume that there is a direct interaction, i.e., that the fields are not independent objects adding new degrees of freedom, but only agents that carry the interaction from one particle, the emitter, to another, the receiver. It is then assumed that *there is no emission unless there is a receiver*, no matter how far the receiver is.

In the present paper we accept this philosophy as a reasonable generalization of the nonrelativistic interaction where absence of radiation comes naturally. We shall consider a system of just two particles "alone in the universe" and hence no radiation takes place.[6,7]

There is another—this time unavoidable—difficulty. Due to the finite velocity with which a relativistic interaction propagates, the equations of motion are not instantaneous but integro-differential. The simpler electromagnetic interaction case turns these equations into differential-difference equations with the differences (a measure of the retarded effects between the particles) depending on time. No general theory exists for such equations.[8,9] The only situation when the differences do not depend on time, which has been handled with certain success, is the case of circular orbit solutions. Another possibility, less physical but solvable in general, is to transform the equations of motion into pure differential equations. This is achieved by assuming that one particle responds only to retarded "fields" and the other only to advanced "fields" and was treated in Ref. 10 by Rudd and Hill. Even if unphysical, this model has the merit of being the only known example of an exactly solvable classical relativistic two-body problem.

The first papers on these lines[11] tried to build a consistent direct interaction electrodynamics, i.e., a theory that yields the same physical conclusions as classical electrodynamics but avoids the explicit introduction of fields as independent dynamical quantities. Mesodynamics has also been considered.[12] Also along these lines Van Dam and Wigner[13] successively pointed out that it is possible to consider direct interactions which have no adjunct particle-field theory, and have discussed in detail a particular example. Several other authors have been devoted to the analysis of some particular models of direct interaction theories which can be derived from an action principle.

An interesting problem which can be studied with such models is the determination of circular orbit motion, since *in practice* (historically) it is the first step to set up a Bohr-like quantization procedure for a relativistic two-body system and thus we learn something about the corresponding fully quantized system. Also the consistency of the nonrelativistic limit is an interesting problem to consider. Since several attempts have been made along these lines, making similar calculations for different but particular choices of the interaction, we think that it is useful to deal directly with a quite general theory of direct interaction coming from an action principle.

In this paper, then, we deal with a Poincaré invariant direct interaction theory coming from an action integral defined in Sec. II. That section serves to define the basic features of our theory, the equations of motion, and the conserved quantities. We include also the reduction to particular cases considered in the literature.

The general problem of having circular orbit solutions is considered in Sec. III. Since not every interaction will allow circular orbit solutions, certain general constraints on the form of the interaction are found in order that such solutions exist. It is the first time, as far as we know, that such constraints are mentioned and discussed. Moreover, the time-independent equations

Copyright © 1973 by the American Institute of Physics

relating the parameters which characterize the circular orbits and the motion are found. Thus, in this section it is possible to find all the necessary formulas to make, for example, a Bohr-like quantization of the circular motion solutions. This will be done elsewhere.

The problem of proving that our general theory has a consistent nonrelativistic limit is discussed in Secs. IV and V. We show that our formalism is richer than many others considered in the literature also from the point of view of the nonrelativistic limit. In fact, nonrelativistic limits considered explicitly in the literature give the Newton equations of motion for two particles interacting through a potential which depends only on the modulus of the relative position of the two particles. On the contrary, we shall see that it is possible to build an interaction, the limit of which is a potential theory with velocity dependent interaction. There is no obvious physical or mathematical reason to expect that every relativistic theory should have a consistent nonrelativistic limit.

II. RELATIVISTIC DYNAMICS FOR TWO PARTICLES

We shall consider the classical relativistic dynamics of two interacting particles, whose position vectors are $x_a^\mu(\tau_a)$, $(a = 1, 2)$, defined by the action,[12,14]

$$J''_{\tau'} = -m_1c^2\int_{\tau_1'}^{\tau_1''} d\tau_1\sqrt{\zeta_{11}} - m_2c^2\int_{\tau_2'}^{\tau_2''} d\tau_2\sqrt{\zeta_{22}} - c\int_{\tau_1'}^{\tau_1''} d\tau_1\int_{\tau_2'}^{\tau_2''} d\tau_2 F(\rho, \zeta_{11}, \zeta_{22}, \zeta_{12}, \sigma_{12}, \sigma_{21}). \tag{2.1}$$

In this expression the integration variables correspond—at this stage — to arbitrary parametrization of the still undetermined $x_a(\tau_a)$; F is an arbitrary function except that to ensure that the action is independent of the parametrization used, it has to satisfy the following two conditions,

$$F = 2\zeta_{aa}\frac{\partial F}{\partial \zeta_{aa}} + \zeta_{ab}\frac{\partial F}{\partial \zeta_{ab}} + \sigma_{ab}\frac{\partial F}{\partial \sigma_{ab}}, \qquad a, b = 1, 2, \tag{2.2}$$

where the derivatives are with respect to the invariants defined by

$$\rho = [x_1(\tau_1) - x_2(\tau_2)]^2,$$
$$\zeta_{ab} = (1/c^2)\dot{x}_a^\mu(\tau_a)x_{b\mu}(\tau_b),$$
$$\sigma_{ab} = (1/c)\dot{x}_a^\mu(x_a - x_b), \qquad a, b = 1, 2. \tag{2.3}$$

The two conditions (2.2) make F a function of only four of the six invariants; we choose these four arguments to be $\rho, \zeta_{12}, \sigma_{12}, \sigma_{21}$. In the following, whenever we write a derivative of F with respect to ζ_{aa} this derivative must be understood as determined by Eq. (2.2).

After the variational principle is applied we will choose the parameters τ_a to correspond to the proper times of the corresponding world lines, which implies that the components of $\dot{x}_a^\mu$ are not all independent but are subject to the constraint

$$\zeta_{11} = \zeta_{22} = 1. \tag{2.4}$$

We also note that the integrals in (2.1) are taken between arbitrary limits $\tau_a' < \tau_a''$. However, since we want the equations of motion to hold for any time, when deriving them from the action principle we must make $\tau_a' \to -\infty$ and $\tau_a'' \to +\infty$.

Furthermore, if the interaction function F is chosen in such a way that the particles interact for $\rho \geq 0$ (time- or lightlike distances), then an explicit dependence on the sign of $(x_1^0 - x_2^0)$ is allowed since the distinction between past and future light cones is invariant.

The Euler–Lagrange equations of motion are[15]

$$c\int_{-\infty}^{\infty} d\tau_b\frac{\partial F}{\partial x_{a\mu}} = \frac{d}{d\tau_a}\left(m_a\dot{x}_a^\mu + c\int_{-\infty}^{\infty} d\tau_b\frac{\partial F}{\partial \dot{x}_{a\mu}}\right), \tag{2.5}$$

with $\zeta_{11} = \zeta_{22} = 1$ and $a \neq b$.

The explicit expression for the integrands are

$$\frac{\partial F}{\partial x_{a\mu}} = 2\frac{\partial F}{\partial \rho}(x_a - x_b)^\mu + \frac{\partial F}{\partial \sigma_{ab}}\frac{\dot{x}_a^\mu}{c} - \frac{\partial F}{\partial \sigma_{ba}}\frac{\dot{x}_b^\mu}{c}, \tag{2.6a}$$

$$\frac{\partial F}{\partial \dot{x}_{a\mu}} = 2\frac{\partial F}{\partial \zeta_{aa}}\frac{\dot{x}_a^\mu}{c^2} + \frac{\partial F}{\partial \zeta_{ab}}\frac{\dot{x}_b^\mu}{c^2} + \frac{1}{c}\frac{\partial F}{\partial \sigma_{ab}}(x_a - x_b)^\mu. \tag{2.6b}$$

The effect of the operator $d/d\tau_a$ when acting on a pure function of the four invariants $\rho, \zeta_{12}, \sigma_{12}, \sigma_{21}$ is explicitly,

$$\frac{d}{d\tau_a} = 2c\sigma_{ab}\frac{\partial}{\partial \rho} + \frac{\dot{x}_a^\nu\dot{x}_{b\nu}}{c^2}\frac{\partial}{\partial \zeta_{ab}} + \left(\frac{\ddot{x}_a^\nu(x_a - x_b)_\nu}{c} + c\right)\frac{\partial}{\partial \sigma_{ab}} - c\zeta_{ab}\frac{\partial}{\partial \sigma_{ba}}. \tag{2.7}$$

The Poincaré invariance of the action implies that there are ten conserved quantities associated with the ten generators of the group. They are the total energy-momentum four vector

$$P^\mu = \sum_{\substack{a=1,2\\ b\neq a}}\left(m_a\dot{x}_a^\mu + c\int_{-\infty}^{\infty}\frac{\partial F}{\partial \dot{x}_{a\mu}}d\tau_b' - c\int_{\tau_b}^{\infty} d\tau_b'\int_{-\infty}^{\tau_a} d\tau_a'\frac{\partial F}{\partial x_{a\mu}}\right) \tag{2.8}$$

and the skew symmetric angular momentum tensor[16]

$$L^{\mu\nu} = \sum_{\substack{a=1,2\\ b\neq a}}\left[m_a(\dot{x}_a^\mu x_a^\nu - \dot{x}_a^\nu x_a^\mu) + c\int_{-\infty}^{\infty} d\tau_b'\times\left(\frac{\partial F}{\partial \dot{x}_{a\mu}}x_a^\nu - \frac{\partial F}{\partial \dot{x}_{a\nu}}x_a^\mu\right) + c\int_{\tau_a}^{\infty} d\tau_a'\int_{-\infty}^{\tau_b} d\tau_b'\times\left(\frac{\partial F}{\partial \dot{x}_{a\mu}}\dot{x}_a^\nu - \frac{\partial F}{\partial \dot{x}_{a\nu}}\dot{x}_a^\mu + \frac{\partial F}{\partial x_{a\mu}}x_a^\nu - \frac{\partial F}{\partial x_{a\nu}}x_a^\mu\right)\right]. \tag{2.9}$$

For completeness we include their derivation in an appendix. Both quantities satisfy, by definition separate conservation laws with respect to τ_1 and τ_2:

$$\frac{d}{dt}P^\mu(\tau_1\tau_2) = \frac{d}{d\tau_1}P^\mu(\tau_1\tau_2) = \frac{d}{d\tau_2}P^\mu(\tau_1\tau_2), \tag{2.10}$$

and similarly for $L^{\mu\nu}$.

Let us see now how our general case reduces to cases already treated in the literature. The choice of the function F is often made to correspond to that of a field which has been eliminated from the equations of motion of an underlying (or adjunct) particle-field interaction theory. The case of electrodynamics, for example, compels us to choose[1,3,5]

$$F = \lambda\delta(\rho)\zeta_{12}, \tag{2.11}$$

thus obtaining the half-retarded half-advanced interaction-at-a-distance electrodynamics considered by Fokker, Tetrode, Wheeler–Feynman,[5] and more recently by Schild.[1] Also different types of mesodynamics could be considered. But there is no need to have an underlying field theory—as stressed by Van Dam and Wigner[13]—and therefore other choices of F are possible. Of these we mention

$$F = \tfrac{1}{2}V(\rho)\zeta_{12}^{-q}\zeta_{11}^{q/2}\zeta_{22}^{q/2}, \tag{2.12}$$

first given by Katz in the Appendix of Ref. 17. It reduces to the original scalar interaction considered by Katz[18] if the value $q = 1$ is chosen, while for $q = 0$ (2.12) gives rise to the equations of motion considered in Ref. 13. Degasperis[2] also has considered (2.12) to discuss the circular orbit bound state problem.

As in electrodynamics, if the interaction is at lightlike distances only, then the integro-differential equations of motion become differential-difference equations. The very particular choice

$$F = \lambda\theta(x_1^0 - x_2^0)\delta(\rho)\zeta_{12} \tag{2.13}$$

reduces the equations of motion in a two dimensional Minkowsky space to pure differential equations. The general solution to these equations was found in Ref. 10.

To extend the formalism to many interacting particles in a symmetric way, the interaction part of the action has to be of the form:

$$\frac{c}{2}\sum_{\substack{a,b\\a\neq b}}\int_{\tau_a'}^{\tau_a''}d\tau_a\int_{\tau_b'}^{\tau_b''}d\tau_b\, H(\rho_{ab}, \zeta_{aa}, \zeta_{bb}, \zeta_{ab}, \sigma_{ab}, \sigma_{ba}). \tag{2.14}$$

This symmetric type of interaction reduces, in the two-particle case to an action (2.1) with F satisfying the symmetry requirement,

$$F(\rho\zeta_{11}\zeta_{22}\zeta_{12}\sigma_{12}\sigma_{21}) = F(\rho\zeta_{22}\zeta_{11}\zeta_{12}\sigma_{21}\sigma_{12}), \tag{2.15}$$

which is satisfied both by (2.11) and (2.12).

Finally we want to make a strong but inessential assumption, which is that the function F vanishes for timelike distances,

$$F = 0 \quad \text{for} \quad \rho > 0. \tag{2.16}$$

We assume (2.16) for simplicity to avoid weaker, but longer to explain, assumptions at different key points of the paper. It is not difficult to find other weaker assumptions, that do not exclude mesodynamics as condition (2.16) does, but ours is economic and serves our present purpose.

III. CIRCULAR ORBIT SOLUTIONS

We now want to investigate the possibility of having circular orbit solutions to our equations of motion (2.5).

As stressed by several authors[1-4] the derivation of explicit circular orbit solutions for the equations of motion is, in practice, the first step to set up a Bohr-like quantization procedure for a relativistic two-particle system. Quantization with noncircular orbits is a complicated job and the only thing that has been done on this line is in a recent paper by Andersen and von Baeyer.[4] As we shall see, the most general case, i.e., when the function F has a general dependence on all four invariants, yields, for fixed values of the masses, four time-independent equations relating the three parameters: angular velocity ω and the radii R_1, R_2 of the orbits. In this general case, therefore, there cannot always be a solution. A consistency condition that we shall make explicit must be satisfied. In the cases when F is independent of the two invariants σ_{12} and σ_{21}, two of the four equations become identities and we recover a situation similar to that of Ref. 2 even though with a far more general interaction.

As is usual when dealing with circular orbits, we define a unit vector $\hat{\mathbf{u}}(t)$ rotating in a plane with uniform angular velocity ω, so that

$$\frac{d\hat{\mathbf{u}}(t)}{dt} = \omega\hat{\mathbf{v}}(t), \qquad \frac{d\hat{\mathbf{v}}(t)}{dt} = -\omega\hat{\mathbf{u}}(t), \tag{3.1}$$

$\hat{\mathbf{v}}(t)$ also being a uniformly rotating unit vector orthogonal to $\hat{\mathbf{u}}(t)$.

The circular orbits of the two particles are now defined as:

$$\mathbf{x}_a(t) = g_a R_a\hat{\mathbf{u}}(t), \qquad g_1 = -1, \qquad g_2 = 1. \tag{3.2}$$

Hence, the four vectors $x_a^\mu, (x_a - x_b)^\mu, \dot{x}_a^\mu, \ddot{x}_a^\mu$ are

$$\begin{aligned}
-x_a^\mu &= c(t_a, (g_a\beta_a/\omega)\hat{\mathbf{u}}(t_a)),\\
-\dot{x}_a^\mu &= c\gamma_a(1, g_a\beta_a\hat{\mathbf{v}}(t_a)),\\
-\ddot{x}_a^\mu &= c\gamma_a^2(0, -g_a\beta_a\omega\hat{\mathbf{u}}(t_a)),\\
(x_a - x_b)^\mu &= (cg_a/\omega)(2\theta, \beta_a\hat{\mathbf{u}}(t_a) + \beta_b\hat{\mathbf{u}}(t_b)),
\end{aligned} \tag{3.3}$$

where

$$\beta_a = R_a\omega/c \tag{3.4}$$

and

$$2g_a\theta = \omega(t_a - t_b). \tag{3.5}$$

The invariants (2.3) are easily expressed in terms of the constants of the problem and the only relevant variable θ,

$$\begin{aligned}
\rho &= (c^2/\omega^2)[4\theta^2 - 2\beta_a\beta_b\cos 2\theta - (\beta_a^2 + \beta_b^2)],\\
\zeta_{ab} &= \gamma_a\gamma_b(1 + \beta_a\beta_b\cos 2\theta),\\
\sigma_{ab} &= (\gamma_a cg_a/\omega)(2\theta + \beta_a\beta_b\sin 2\theta).
\end{aligned} \tag{3.6}$$

To determine the conditions under which we actually have a solution of the equations of motion, we must replace Eqs. (3.3) in Eq. (2.5). Transforming the integrals over τ_b that appear in (2.5) into integrals over θ and noticing that F is now only a function of θ, we obtain several integrals whose integrands are functions of θ times a vector $\hat{\mathbf{u}}$ or $\hat{\mathbf{v}}$. When the vectors $\hat{\mathbf{u}}(t_b)$ or $\hat{\mathbf{v}}(t_b)$ appear, we express them in terms of $\hat{\mathbf{u}}(t_a)$ and $\hat{\mathbf{v}}(t_a)$, where t_a is the fixed time in (2.5), by using

$$\begin{aligned}
\hat{\mathbf{u}}(t_b) &= \cos 2\theta\hat{\mathbf{u}}(t_a) - g_a\sin 2\theta\hat{\mathbf{v}}(t_a),\\
\hat{\mathbf{v}}(t_b) &= \cos 2\theta\hat{\mathbf{v}}(t_a) + g_a\sin 2\theta\hat{\mathbf{u}}(t_a).
\end{aligned} \tag{3.7}$$

The derivative with respect to τ_a on the right-hand side of (2.5) can easily be performed using Eq. (3.1).

To express the results concisely, it is convenient to define the following integrals:

$$\begin{aligned}
A_\xi &= 2\int_{-\infty}^{\infty}\frac{\partial F}{\partial\xi}d\theta, \qquad C_\xi = 2\int_{-\infty}^{\infty}\frac{\partial F}{\partial\xi}\cos 2\theta d\theta,\\
S_\xi &= 2\int_{-\infty}^{\infty}\frac{\partial F}{\partial\xi}\sin 2\theta d\theta, \qquad T_\xi = 2\int_{-\infty}^{\infty}\frac{\partial F}{\partial\xi}\theta d\theta,
\end{aligned} \tag{3.8}$$

where ξ denotes any one of the six invariants on which F depends. Note that all the above expressions depend only on the parameters ω, R_1, and R_2. Also note that because of (2.16) the integration over θ is on a finite range and the actual limits of integration are, due to the geometry of the problem, equal and of opposite sign.

Some elementary calculations allow us to prove that the equations of motion reduce to the following two equalities ($a = 1, 2$) between 4-vectors:

$$\left(\frac{4c^2g_a}{\omega^2\gamma_b}T_\rho + \frac{c}{\omega}\left(\frac{\gamma_a}{\gamma_b}A_{\sigma_{ab}} - A_{\sigma_{ba}}\right),\right.$$
$$\left(\frac{2c^2g_a}{\omega^2\gamma_b}(\beta_a A_\rho + \beta_b C_\rho) + \frac{c\beta_b}{\omega}S_{\sigma_{ba}}\right)\hat{\mathbf{u}}(t_a)$$
$$+ \left.\left(-\frac{2c^2\beta_b}{\omega^2\gamma_b}S_\rho + \frac{cg_a}{\omega}\beta_b C_{\sigma_{ab}}\right)\hat{\mathbf{v}}(t_a)\right)$$
$$= \left(0, \left(-g_a\gamma_a^2\beta_a m_a c\omega - \frac{2g_a\gamma_a^2\beta_a}{\gamma_b}A_{\zeta_{aa}}\right.\right.$$
$$+ g_a\gamma_a\beta_b C_{\zeta_{ab}} + \frac{\gamma_a c\beta_b}{\omega\gamma_b}S_{\sigma_{ab}}\Big)\hat{\mathbf{u}}(t_a)$$
$$+ \left.\left(-\gamma_a\beta_b S_{\zeta_{ab}} + \frac{cg_a\gamma_a\beta_b}{\omega\gamma_b}C_{\sigma_{ab}}\right)\hat{\mathbf{v}}(t_a)\right). \tag{3.9}$$

Equating the zero components and the coefficients of $\hat{\mathbf{u}}(t_a)$ on the left- and right-hand side of (3.9), respectively, we obtain the equations

$$4cg_aT_\rho = \omega(\gamma_b A_{\sigma_{ba}} - \gamma_a A_{\sigma_{ab}}), \tag{3.10a}$$

$$-\gamma_a^2\beta_a\omega m_a c - (2\gamma_a^2\beta_a/\gamma_b)A_{\zeta_{aa}}$$
$$+ \gamma_a\beta_b C_{\zeta_{ab}} + (g_a\gamma_a c\beta_b/\omega\gamma_b)S_{\sigma_{ab}}$$
$$= (2c^2/\omega^2\gamma_b)(\beta_a A_\rho + \beta_b C_\rho) + (c\beta_b g_a/\omega)S_{\sigma_{ba}}. \tag{3.10b}$$

As regards the equation which arises from equating the coefficients of $\hat{\mathbf{v}}(t_a)$ on both sides of (3.9) we note that it is identically satisfied as a consequence of (3.10a) and of the identity

$$\int_{-\infty}^{\infty}\frac{\partial F}{\partial\theta}d\theta = 0. \tag{3.11}$$

Equation (3.10a) does not involve the masses and it is a constraint on our interaction function F telling us that there cannot always be a circular orbit solution of the equations of motion. As mentioned at the beginning, if there is no σ_{ab}-dependence in F, Eq. (3.10a) is identically satisfied. In fact, in such a case F is an even function of θ [cf. Eq. (3.6)] so that T_ρ vanishes, while the right-hand side of (3.10a) is identically zero. It is immediately verified with the particular choice (2.12) for F, that one obtains exactly Eq. (3.22) of Ref. 2. Obviously when there is no σ_{ab} dependence in F the two Eqs. (3.10b), which are the only equations relating ω, R_1 and R_2, refer to a much more general situation than that considered in Ref. 2. As F given by (2.11) is a particular case of (2.12), our solution also contains those found by Schild[1] and by Anderser and von Baeyer.[3]

IV. PRELIMINARIES TO STUDY THE NONRELATIVISTIC LIMIT

In order to study the nonrelativistic limit of a general theory, such as that defined by the action (2.1), it is necessary to make some specific assumptions concerning the behavior of the function F and its derivatives when $c \to \infty$. Different assumptions will lead, in general, to different nonrelativistic theories. To make our assumptions clearer, it is convenient first to determine the nonrelativistic behavior of the basic quantities appearing in our formalism.

Since we assume that F vanishes for $(x_a - x_b)^2 > 0$, the interaction takes place only between two events, x_a^μ and x_b^μ, which are space- or lightlike separated. The difference in time between two such events will not be zero in every reference frame, but it will tend to zero in the nonrelativistic limit as c^{-1}, so that, putting

$$\eta = g_b c(\tau_b - \tau_a) = c(\tau_2 - \tau_1), \tag{4.1}$$

the nonrelativistic limit of η is, in general, different from zero and finite. In Eq. (4.1), $g_2 = -g_1 = 1$.

When taking the nonrelativistic limit of an integral expression such as those appearing in the equations of motion or in the conserved quantities, we shall make the change of integration variable from proper time τ_b to the variable η. In this way the integration interval remains finite, i.e., $[-r, r]$, where r is the modulus of

$$\mathbf{r} = \mathbf{x}_1(t) - \mathbf{x}_2(t) = g_a[\mathbf{x}_b(t) - \mathbf{x}_a(t)]. \tag{4.2}$$

To deal with double integrals such as those appearing in Eqs. (2.8) and (2.9) with integration variables τ_a' and τ_b', the following changes of variables must be made:

$$\eta_a = g_a c(\tau_a' - \tau_b), \qquad \eta_b = g_b c(\tau_b' - \tau_a). \tag{4.3}$$

It should be made clear that one of the proper times appearing in the expression of η is fixed, and we can expand about this time. Let τ_a be the fixed time in (4.1); then

$$d\eta = cd\tau_b \tag{4.4}$$

(g_b does not appear in this expression since the right ordering of the integration limits takes care of the correct sign). We will often need to expand about τ_a by means of

$$\tau_b = \tau_a - g_a\eta/c\,. \tag{4.5}$$

Thus,

$$\mathbf{x}(\tau_b) \simeq \mathbf{x}_b(\tau_a) - (g_a\eta/c)\dot{x}_b(\tau_a) \tag{4.6}$$

and

$$r^\mu \equiv g_b[x_a(\tau_a) - x_b(\tau_b)]^\mu$$
$$= \{-\eta + [(d/c) + (f/c^2)]g_b, \mathbf{r} - (\eta/c)\dot{\mathbf{x}}_b(\tau_a)\}. \tag{4.7}$$

In the last expression d and f are constants which depend parametrically on τ_a and come from the connection between time and proper time.

From (4.1), (4.2), (4.5), (4.6), and (4.7) the behavior of our four fundamental invariants near the nonrelativistic limit is

$$\rho \simeq \eta^2 - r^2 + (2\eta/c)[\mathbf{r}\cdot\dot{\mathbf{x}}_b + g_a d],\ \zeta_{12} \simeq 1 + v^2/2c^2, \tag{4.8}$$

and

$$\sigma_{ab} \simeq g_a\eta + (1/c)(d - g_b\dot{\mathbf{x}}_a\cdot\mathbf{r})$$
$$+ (1/c^2)[f + g_b\eta(\dot{\mathbf{x}}_a\cdot\dot{\mathbf{x}}_b - \tfrac{1}{2}\dot{\mathbf{x}}_a^2)],$$
$$\sigma_{ba} \simeq -g_a\eta - (1/c)(d - g_b\dot{\mathbf{x}}_b\cdot\mathbf{r})$$
$$- (1/c^2)[f + \eta(\tfrac{1}{2}\dot{\mathbf{x}}_b{}^2 g_b - \ddot{\mathbf{x}}_b\cdot\mathbf{r})]. \tag{4.9}$$

The two remaining invariants are constrained to $\zeta_{aa} = 1$ in the equations of motion. Note that there is an asymmetry between a an b. This is due to the fact that τ_a and τ_b are playing different roles, as is obvious, for example, in (4.5) and (4.6). The relative nonrelativistic velocity $\mathbf{v}$ is defined by

$$\mathbf{v} = \dot{\mathbf{x}}_2 - \dot{\mathbf{x}}_1 = g_a(\dot{\mathbf{x}}_a - \dot{\mathbf{x}}_b). \tag{4.10}$$

For the purpose of having a nonrelativistic limit which yields velocity dependent potentials, it is convenient to redefine some of the invariants so that at least one of the leading terms depends on $\mathbf{v}$. We achieve this by defining

$$\begin{aligned} \sigma_+ &= c(\sigma_{21} + \sigma_{12}) = c(\sigma_{ab} + \sigma_{ba}), \\ \sigma_- &= \sigma_{21} - \sigma_{12} = g_b(\sigma_{ba} - \sigma_{ab}). \end{aligned} \tag{4.11}$$

From (4.9), the behavior near the nonrelativistic limit is

$$\begin{aligned} \sigma_+ &\simeq \mathbf{v}\cdot\mathbf{r} + (\eta/c)[\ddot{\mathbf{x}}_b\cdot\mathbf{r} - \tfrac{1}{2}g_b\mathbf{v}^2], \\ \sigma_- &\simeq 2\eta + (1/c)[(\dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2)\cdot\mathbf{r} - 2g_b d]. \end{aligned} \tag{4.12}$$

The specific assumptions under which we shall study the nonrelativistic limit of our general theory are the following:

The interaction function

$$F(\rho, \zeta_{12}, \sigma_{12}, \sigma_{21}) \tag{4.13}$$

and all its partial derivatives with respect to its four invariant arguments, as specified in (4.13), have a finite nonrelativistic limit.

With this assumption we shall get nonrelativistic equations of motion with a potential U which depends only on r^2 and $\mathbf{r}\cdot\mathbf{v}$. If we want to obtain also a $\mathbf{v}^2$ dependence in the potential, we must make the same assumptions but replace the argument ζ_{12} by $c^2(\zeta_{12}-1)$. Such a change further complicates the analysis of the limiting procedure and teaches us nothing new. Therefore we shall not consider it.

Since our formalism will give, in the limit, a Galilei invariant theory with a potential which depends on the relative position and velocity of the two particles, we summarize briefly, for comparison later, the equations of motion and the conserved quantities for such a nonrelativistic theory. Starting from the Lagrangian

$$L = \tfrac{1}{2}(m_1\dot{\mathbf{x}}_1^2 + m_2\dot{\mathbf{x}}_2^2) - U(r^2, \mathbf{r}\cdot\mathbf{v}), \tag{4.14}$$

where

$$\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2, \quad \mathbf{v} = \dot{\mathbf{x}}_2 - \dot{\mathbf{x}}_1, \quad \mathbf{a} = \mathbf{v}, \tag{4.15}$$

we obtain the equations of motion

$$m_a\ddot{\mathbf{x}}_a = -\frac{\partial U}{\partial \mathbf{x}_a} + \frac{d}{dt}\frac{\partial U}{\partial \dot{\mathbf{x}}_a} \tag{4.16}$$

or equivalently

$$g_a m_a\ddot{\mathbf{x}}_a = \left(2\frac{\partial U}{\partial r^2} - 2\frac{\partial^2 U}{\partial r^2 \partial \mathbf{r}\cdot\mathbf{v}}\mathbf{r}\cdot\mathbf{v} + \frac{\partial^2 U}{\partial(\mathbf{r}\cdot\mathbf{v})^2}(\mathbf{r}\cdot\mathbf{a} - \mathbf{v}^2)\right)\mathbf{r},$$

$$g_2 = -g_1 = 1. \tag{4.17}$$

The conserved total linear momentum and energy are

$$\mathbf{P} = m_1\dot{\mathbf{x}}_1 + m_2\dot{\mathbf{x}}_2 \tag{4.18}$$

and

$$H = \tfrac{1}{2}(m_1\dot{\mathbf{x}}_1^2 + m_2\dot{\mathbf{x}}_2^2) + U - \frac{\partial U}{\partial \mathbf{r}\cdot\mathbf{v}}\mathbf{r}\cdot\mathbf{v}. \tag{4.19}$$

The conserved angular momentum is

$$L^{ij} = \sum_a \Big(m_a(\dot{x}_a^i x_a^j - \dot{x}_a^j x_a^i) + \frac{\partial U}{\partial \mathbf{r}\cdot\mathbf{v}}(x_b^j x_a^i - x_b^i x_a^j) + 2\frac{\partial U}{\partial v^2}[(\dot{x}_a - \dot{x}_b)^j x_a^i - (\dot{x}_a - \dot{x}_b)^i x_a^j]\Big). \tag{4.20}$$

As regards the vector $\mathbf{K}$, which is the generator of infinitesimal special Galilei transformations (boost), we observe that such a transformation does not leave L invariant:

$$L \to L + \frac{d}{dt}(M\mathbf{X}\cdot\delta\mathbf{V}), \tag{4.21}$$

where

$$M = m_1 + m_2, \quad M\mathbf{X} = m_1\mathbf{x}_1 + m_2\mathbf{x}_2,$$

$\delta\mathbf{V}$ being the infinitesimal change of frame velocity. From (4.21) one gets

$$\mathbf{K} = \sum_a \frac{\partial L}{\partial \dot{\mathbf{x}}_a}t - M\mathbf{X} = \mathbf{P}t - M\mathbf{X}. \tag{4.22}$$

V. THE NONRELATIVISTIC LIMIT

A. The equations of motion

The equations of motion whose limit we are interested in considering are those defined by Eq. (2.5), except that, instead of the invariant variables σ_{ab} and σ_{ba}, we use σ_+ and σ_- defined in the previous section. The explicit form of the derivatives of F that replace Eqs. (2.6) are

$$\begin{aligned} \frac{\partial F}{\partial x_{a\mu}} &= 2(x_a - x_b)^\mu\frac{\partial F}{\partial \rho} + (\dot{x}_a^\mu - \dot{x}_b^\mu)\frac{\partial F}{\partial \sigma_+} + g_a\frac{\dot{x}_a^\mu + \dot{x}_b^\mu}{c}\frac{\partial F}{\partial \sigma_-}, \\ \frac{\partial F}{\partial \dot{x}_{a\mu}} &= \frac{\dot{x}_a^\mu}{c^2}F + \frac{\dot{x}_a^\mu}{c^2}(1 - \zeta_{ab})\frac{\partial F}{\partial \zeta_{ab}} \\ &\quad + \left[(x_a - x_b)^\mu - \sigma_{ab}\frac{\dot{x}_a^\mu}{c}\right]\left(\frac{\partial F}{\partial \sigma_+} + \frac{g_a}{c}\frac{\partial F}{\partial \sigma_-}\right). \end{aligned} \tag{5.1}$$

In making the nonrelativistic limit we observe that the equations of motion contain terms which diverge as c, terms that remain finite, and terms which tend to zero. We shall see that the leading divergent terms cancel in the limit, getting therefore, well-defined nonrelativistic equations. It is not necessary to analyze the limit of both the zero component and the space component equations of motion (2.5), (5.1), since the condition $\ddot{x}_a^\mu(\tau_a)\dot{x}_{a\mu}(\tau_a) = 0$ makes the zero component equations logically dependent on the other three equations. After a consistent limit is found for the space component equations of motion, it follows that the time component equations reduce to an identity. In fact this can be checked explicitly.

We consider then, the space component equations. Dropping the terms that tend to zero in the limit, we are left with the equation

$$\int d\eta\left(2g_b\mathbf{r}\frac{\partial F}{\partial \rho} - \dot{\mathbf{x}}_b\frac{\partial F}{\partial \sigma_+}\right) = m_a\ddot{\mathbf{x}}_a + \int d\eta\left[-\dot{\mathbf{x}}_a\frac{\partial F}{\partial \sigma_+} - g_b\mathbf{r}\left(\mathbf{r}\cdot\mathbf{x}_a g_b + \frac{\mathbf{v}^2}{2}\right)\right.$$

$$\times\frac{\partial^2 F}{\partial\sigma_+^2}-2\mathbf{r}\sigma_{ab}\frac{\partial^2 F}{\partial\rho\partial\sigma_-}-2g_a\mathbf{r}\frac{\partial^2 F}{\partial\sigma_-^2}-2\dot{\mathbf{x}}_a\sigma_{ab}^2\frac{\partial^2 F}{\partial\rho\partial\sigma_+}$$
$$-2\dot{\mathbf{x}}_a\sigma_{ab}g_a\frac{\partial^2 F}{\partial\rho\partial\sigma_+}\Big]+2c\int d\eta g_b\mathbf{r}\left(\sigma_{ab}\frac{\partial}{\partial\rho}+g_a\frac{\partial}{\partial\sigma_-}\right)\frac{\partial F}{\partial\sigma_+}. \tag{5.2}$$

All the terms in this equation have already been written at their finite nonrelativistic limit, except for the last integral on the right hand side, which diverges like c. In Appendix B we analyze this term to show that there are cancellations such that the coefficient of the leading term is zero. We evaluate explicitly the finite order contribution of this term. By making use of Eq. (B5) it is possible to show that there is a cancellation between the first term of the first integral on the right-hand side of (5.2) and the last term of the same integral. Taking this into account and Eq. (B7), Eq. (5.2) yields

$$g_a m_a\ddot{\mathbf{x}}_a=-\left(2\int\frac{\partial F}{\partial\rho}\,d\eta-(\mathbf{v}^2-\mathbf{r}\cdot\mathbf{a})\int\frac{\partial^2 F}{\partial\sigma_+^2}\,d\eta\right.$$
$$\left.+2(\mathbf{r}\cdot\mathbf{v})\int\frac{\partial^2 F}{\partial\sigma_+\partial\rho}\,d\eta\right)\mathbf{r}, \tag{5.3}$$

which coincides with the nonrelativistic Eq. (4.17) for particles interacting through a potential depending on the relative position and velocity through $\mathbf{r}^2$ and $\mathbf{r}\cdot\mathbf{v}$, provided we make the identification

$$U(\mathbf{r}^2,\mathbf{r}\cdot\mathbf{v})=\int F(\eta^2-\mathbf{r}^2,\zeta_{12}=1,\mathbf{v}\cdot\mathbf{r},2\eta)\,d\eta, \tag{5.4}$$

where the nonrelativistic values of the invariants have been explicitly written in F. This equation is the natural generalization of the equations derived in Ref. 18 to include $\mathbf{r}\cdot\mathbf{v}$ dependence in the potential.

B. The linear momentum

To obtain the expression of P^μ that we must analyze, Eq. (5.1) has to be replaced in Eq. (2.8). Considering first the limit of the space components, it is not difficult to see that the double integral terms tend to zero like c^{-1} and that in the simple integral terms the only parts that contribute are those corresponding to the derivative with respect to σ_+. There are two of these terms which do not vanish, one for each value of the index a, they are equal and of opposite sign. Therefore they cancel and we obtain

$$\mathbf{P}\xrightarrow[c\to\infty]{}\mathbf{P}^{\mathrm{NR}}=m_1\dot{\mathbf{x}}_1+m_2\dot{\mathbf{x}}_2. \tag{5.5}$$

Concerning the component P^0, the limit to study is that of

$$H\equiv P^0c-m_1c^2-m_2c^2, \tag{5.6}$$

which should be compared, in the nonrelativistic limit, with the function H defined in (4.19), U being given by Eq. (5.4). Although this limit is not independent of the previous one it is easy to obtain it directly. The simple integral terms in (2.8) apparently diverge like c when considering the limit of (5.6). The divergencies, however, cancel and the limit of each one of these two terms yields

$$\int d\eta F+g_a\int d\eta\frac{\partial F}{\partial\mathbf{v}\cdot\mathbf{r}}\,\mathbf{v}_a\cdot\mathbf{r} \tag{5.7}$$

and therefore the sum gives

$$2\int d\eta F-\int d\eta\frac{\partial F}{\partial\mathbf{v}\cdot\mathbf{r}}\,\mathbf{v}\cdot\mathbf{r}. \tag{5.8}$$

The double integral terms also contribute to the limit of (5.6). Defining

$$B\equiv-c^2\int_{\tau_2}^{\infty}d\tau_2'\int_{-\infty}^{\tau_1}d\tau_1'\frac{\partial F}{\partial x_{10}}+c^2\int_{\tau_1}^{\infty}d\tau_1'\int_{-\infty}^{\tau_2}d\tau_2'\frac{\partial F}{\partial x_{10}}, \tag{5.9}$$

and noticing that a derivative with respect to $x_{10}=ct_1$ is, in the limit, equal to a derivative with respect to $c\tau_1$, we have

$$B\approx-c\int_{\tau_2}^{\infty}d\tau_2'F-c\int_{-\infty}^{\tau_2}d\tau_2'F, \tag{5.10}$$

implying

$$B^{\mathrm{NR}}=-\int d\eta F. \tag{5.11}$$

Putting these results together, it is clear that, in fact, H as defined in Eq. (5.6) tends to the Hamiltonian (4.19).

C. Angular momentum and boost

We now take into account the skew-symmetric angular momentum tensor $L^{\mu\nu}$ of Eq. (2.9). Note that when dealing with the L^{0i} components, the relevant quantity which should go under the limit in the vector $\mathbf{K}$ of (4.22) is L^{0i}/c. From (2.19) it can easily be seen that in both cases, i.e., L^{ij} and L^{0i}/c, the term containing a double integral on the right-hand side of that equation tends to zero, and so it can be neglected. As concerns the simple integral appearing in (2.9), it is easily seen that it gives a finite contribution to L^{ij}, while for L^{0i}/c it goes to zero, owing to a cancellation which occurs between the two terms of the integrand. The final results are

$$L^{ij}\to L^{ij}_{\mathrm{NR}}=\sum_a\left(m_a(\dot{x}_a^i x_a^j-\dot{x}_a^j x_a^i)+\int d\eta\frac{\partial F}{\partial\sigma_+}(x_b^i x_a^j-x_b^j x_a^i)\right) \tag{5.12}$$

and

$$\frac{1}{c}L^{0i}\to\mathbf{K}^i=\sum_a m_a(x_a^i-t\dot{x}_a^i)=M\mathbf{X}-\mathbf{P}t. \tag{5.13}$$

Using (5.4), the expression (5.12) coincides with the expression (4.20). Eq. (5.13) already coincides with (4.22). Therefore the theory has a consistent nonrelativistic limit.

ACKNOWLEDGMENTS

One of the authors (P. C.) wishes to thank Professor Abdus Salam and Professor P. Budini, as well as the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX A: LINEAR AND ANGULAR MOMENTUM CONSERVATION

Here we shall briefly sketch the procedure[19] to obtain conservation laws from Poincaré invariance of the action (2.1).

I. Linear momentum

We consider the infinitesimal transformation of the action induced by the following variations:

$$\delta x_1^\mu=\delta x_2^\mu=\epsilon^\mu,\qquad\delta\dot{x}_1^\mu=\delta\dot{x}_2^\mu=0,\qquad\delta\tau_1=\delta\tau_2=0, \tag{A1}$$

which correspond to an infinitesimal translation.

The variation of the action, which must vanish, is

$$\delta J'' = 0 = c \int_{\tau_1'}^{\tau_1''} d\tau_1 \int_{\tau_2'}^{\tau_2''} d\tau_2 \left(\frac{\partial F}{\partial x_{1\mu}} + \frac{\partial F}{\partial x_{2\mu}}\right). \quad \text{(A2)}$$

Using the identity

$$\int_{\tau'}^{\tau''} = \int_{-\infty}^{\infty} - \int_{-\infty}^{\tau'} - \int_{\tau''}^{\infty}, \quad \text{(A3)}$$

and the equations of motion (2.5), (A2) becomes

$$\sum_a \left(m_a \dot{x}_a^\mu + c \int_{-\infty}^{\infty} \frac{\partial F}{\partial \dot{x}_{a\mu}} d\tau_b\right)_{\tau_a'}^{\tau_a''} + \left[\int_{\tau_2'}^{\tau_2''} \left(\int_{-\infty}^{\tau_1'} + \int_{\tau_1''}^{\infty}\right) - \int_{\tau_1'}^{\tau_1''} \left(\int_{-\infty}^{\tau_2'} + \int_{\tau_2''}^{\infty}\right)\right] \frac{\partial F}{\partial x_1} d\tau_1 d\tau_2 = 0, \quad \text{(A4)}$$

where, in the last term, we have used

$$\frac{\partial F}{\partial x_{1\mu}} = -\frac{\partial F}{\partial x_{2\mu}}, \quad \text{(A5)}$$

which follows from translational invariance.

Now we use the identity

$$\int_{\tau_2'}^{\tau_2''} \left(\int_{-\infty}^{\tau_1'} + \int_{\tau_1''}^{\infty}\right) - \int_{\tau_1'}^{\tau_1''} \left(\int_{-\infty}^{\tau_2'} - \int_{\tau_2''}^{\infty}\right) = \left(\int_{-\infty}^{\tau_2} \int_{\tau_1}^{\infty} - \int_{\tau_2}^{\infty} \int_{-\infty}^{\tau_1}\right)''. \quad \text{(A6)}$$

Equation (A4) then becomes

$$P^\mu|_{\tau'}'' = 0, \quad \text{(A7)}$$

where P^μ is given by Eq. (2.8).

II. Angular momentum

The relations corresponding to (A1) in the case of a homogeneous Lorentz transformation are

$$\delta x_a^\mu = \epsilon_\nu^\mu x_a^\nu, \quad \delta \dot{x}_a^\mu = \epsilon_\nu^\mu \dot{x}_a^\nu, \quad \delta \tau_a = 0, \quad \epsilon_{\mu\nu} = -\epsilon_{\nu\mu}, \quad \text{(A8)}$$

$\epsilon_{\mu\nu}$ being an arbitrary infinitesimal; the invariance of F implies

$$\frac{\partial F}{\partial x_{1\mu}} x_1^\nu - \frac{\partial F}{\partial x_{1\nu}} x_1^\mu + \frac{\partial F}{\partial \dot{x}_{1\mu}} \dot{x}_1^\nu - \frac{\partial F}{\partial \dot{x}_{1\nu}} \dot{x}_1^\mu = -\left(\frac{\partial F}{\partial x_{2\mu}} x_2^\nu - \frac{\partial F}{\partial x_{2\nu}} x_2^\mu + \frac{\partial F}{\partial \dot{x}_{2\mu}} \dot{x}_2^\nu - \frac{\partial F}{\partial \dot{x}_{2\nu}} \dot{x}_2^\mu\right). \quad \text{(A9)}$$

The vanishing of the variation of the action gives

$$c \sum_{\substack{a=1,2 \\ b\neq a}} \left[\int_{\tau_a'}^{\tau_a''} \int_{\tau_b'}^{\tau_b''} \left(\frac{\partial F}{\partial x_{a\mu}} x_a^\nu - \frac{\partial F}{\partial x_{a\nu}} x_a^\mu\right) d\tau_a d\tau_b + \int_{\tau_b'}^{\tau_b''} \int_{\tau_b'}^{\tau_b''} \left(\frac{\partial F}{\partial \dot{x}_{a\mu}} \dot{x}_a^\nu - \frac{\partial F}{\partial \dot{x}_{a\nu}} \dot{x}_a^\mu\right) d\tau_a d\tau_b\right] = 0. \quad \text{(A10)}$$

The use of the identity (A3) and of the equation of motion in the first term yields

$$-c \sum_{\substack{a=1,2 \\ b\neq a}} \left[\int_{\tau_a'}^{\tau_a''} \left(\int_{-\infty}^{\tau_b'} + \int_{\tau_b'}^{\infty}\right) \left(\frac{\partial F}{\partial x_{a\mu}} x_a^\nu - \frac{\partial F}{\partial x_{a\nu}} x_a^\mu\right) d\tau_a d\tau_b - \frac{1}{c} \int_{\tau_a'}^{\tau_a''} m_a \frac{d}{d\tau_a} (x_a^\mu x_a^\nu - x_a^\nu x_a^\mu) d\tau_a - \int_{\tau_b'}^{\tau_a''} \int_{\tau_b'}^{\tau_b''} \left(\frac{\partial F}{\partial \dot{x}_{a\mu}} \dot{x}_a^\nu - \frac{\partial F}{\partial \dot{x}_{a\nu}} \dot{x}_a^\mu\right) d\tau_a d\tau_b - \int_{-\infty}^{\infty} d\tau_b \int_{\tau_a'}^{\tau_a''} d\tau_a \left(x_a^\nu \frac{d}{d\tau_a} \frac{\partial F}{\partial \dot{x}_{a\mu}} - x_a^\mu \frac{d}{d\tau_a} \frac{\partial F}{\partial \dot{x}_{a\nu}}\right)\right] = 0. \quad \text{(A11)}$$

Integrating by parts the last term of (A11) gives a double integral whose integrand coincides with that of the last but one term. Adding these two integrals we get a double integral term with integration ranges on τ_a and τ_b that are equal to those of the first term of (A11). At this point we make use of the identity (A9), obtaining

$$\sum_a \left[m_a(\dot{x}_a^\mu x_a^\nu - \dot{x}_a^\nu x_a^\mu) + c \int_{-\infty}^{\infty} \left(x_a^\nu \frac{\partial F}{\partial \dot{x}_{a\mu}} - x_a^\mu \frac{\partial F}{\partial \dot{x}_{a\nu}}\right) d\tau_b\right]'' + c\left[\left(\int_{\tau_1}^{\infty} \int_{-\infty}^{\tau_2} - \int_{-\infty}^{\tau_1} \int_{\tau_2}^{\infty}\right) \times \left(\frac{\partial F}{\partial x_{1\mu}} x_1^\nu - \frac{\partial F}{\partial x_{1\nu}} x_1^\mu + \frac{\partial F}{\partial \dot{x}_{1\mu}} \dot{x}_1^\nu - \frac{\partial F}{\partial \dot{x}_{1\nu}} \dot{x}_1^\mu\right) d\tau_1 d\tau_2\right]'' = 0,$$

which is

$$L^{\mu\nu}|_{\tau'}'' = 0,$$

where $L^{\mu\nu}$ is given by (2.9).

APPENDIX B: THE NONRELATIVISTIC LIMIT OF THE LAST INTEGRAL IN EQ. (5.2)

To determine the nonrelativistic limit of the integral

$$\mathcal{T} = 2c \int d\eta [\mathbf{x}_a(\tau_a) - \mathbf{x}_b(\tau_b)] \left(\sigma_{ab} \frac{\partial^2 F}{\partial \rho \partial \sigma_+} + g_a \frac{\partial^2 F}{\partial \sigma_+ \partial \sigma_-}\right), \quad \text{(B1)}$$

we use Eqs. (2.5), (2.7), and the expansion

$$G(\rho, \zeta_{ab}, \sigma_+, \sigma_-) = G_{NR} + \frac{\partial G}{\partial \rho}|_{NR} \Delta\rho + \frac{\partial G}{\partial \sigma_+}|_{NR} \Delta\sigma_+ + \frac{\partial G}{\partial \sigma_-}|_{NR} \Delta\sigma_- \quad \text{(B2)}$$

of any function of the invariants which will be used for the derivatives of F appearing in (B1). The index NR, which means that the value of the indicated functions must be taken at the limit $c \to \infty$, will be dropped from now on. In (B2) the quantities $\Delta\rho$, $\Delta\sigma_+$, and $\Delta\sigma_-$ are given, according to (4.8) and (4.12), by

$$\begin{aligned} \Delta\rho &= (2\eta/c)(\mathbf{r}\cdot\dot{\mathbf{x}}_b + g_a d), \\ \Delta\sigma_+ &= (\eta/c)(\ddot{\mathbf{x}}_b \cdot \mathbf{r} - \tfrac{1}{2} g_b \mathbf{v}^2), \\ \Delta\sigma_- &= (1/c)[(\dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2)\cdot\mathbf{r} - 2d g_b]. \end{aligned} \quad \text{(B3)}$$

Evaluating explicitly the expression appearing in (B1) and keeping both terms of order c and terms which are finite in the limit, we have

$$\mathcal{T} = -c \int d\eta \frac{\partial}{\partial \sigma_+} \left[\frac{\partial F}{\partial \eta} - \frac{1}{\eta}\left(\frac{\partial F}{\partial \sigma_+} \Delta\sigma_+ + \frac{\partial F}{\partial \rho} \Delta\rho\right)\right] \mathbf{r} + \int \eta d\eta \, \dot{\mathbf{x}}_1 \frac{d}{d\eta} \frac{\partial F}{\partial \sigma_+} + 2 \int d\eta \left[g_b(d - g_b \dot{\mathbf{x}}_a \cdot \mathbf{r}) \frac{\partial^2 F}{\partial \rho \partial \sigma_+}\right] \mathbf{r} - \int d\eta \, \dot{\mathbf{x}}_b \frac{\partial F}{\partial \sigma_+}, \quad \text{(B4)}$$

where use has been made of the expansion

$$\frac{dF}{d\eta} = \left(2\eta \frac{\partial F}{\partial \rho} + 2 \frac{\partial F}{\partial \sigma_-}\right) + \left[\frac{1}{\eta}\left(\frac{\partial F}{\partial \sigma_+} \Delta\sigma_+ + \frac{\partial F}{\partial \rho} \Delta\rho\right) + 2\eta\left(\frac{\partial^2 F}{\partial \rho^2} \Delta\rho + \frac{\partial^2 F}{\partial \rho \partial \sigma_+} \Delta\sigma_+ + \frac{\partial^2 F}{\partial \rho \partial \sigma_-} \Delta\sigma_-\right) + 2\left(\frac{\partial^2 F}{\partial \rho \partial \sigma_-} \Delta\rho + \frac{\partial^2 F}{\partial \sigma_+ \partial \sigma_-} \Delta\sigma_+ + \frac{\partial^2 F}{\partial \sigma_-^2} \Delta\sigma_-\right)\right]. \quad \text{(B5)}$$

The term

$$\int d\eta \frac{\partial}{\partial \sigma_+} \frac{dF}{d\eta} = \int d\eta \frac{d}{d\eta} \frac{\partial F}{\partial \sigma_+}$$

vanishes since F is identically zero for $\rho > 0$. Integration by parts transforms the second integral into

$$\int \eta d\eta \, \dot{\mathbf{x}}_b \frac{d}{d\eta} \frac{\partial F}{\partial \sigma_+} = - \int \frac{\partial F}{\partial \sigma_+} \dot{\mathbf{x}}_b d\eta . \tag{B6}$$

Expressing explicitly the $\Delta\sigma_+$ and $\Delta\rho$ in the first integral of (B4), some cancellations occur with the last integral, so that the final and finite result is

$$\mathcal{T}_{\mathrm{NR}} = \mathbf{r} \left(\ddot{\mathbf{x}}_b \cdot \mathbf{r} - \tfrac{1}{2} g_b \mathbf{v}^2\right) \int d\eta \frac{\partial^2 F}{\partial \sigma_+^2} + 2\mathbf{r} g_b \mathbf{r} \cdot \mathbf{v} \int d\eta \frac{\partial^2 F}{\partial \sigma_+ \partial \rho} - \dot{\mathbf{x}}_b \int d\eta \frac{\partial F}{\partial \sigma_+}, \tag{B7}$$

which is used in Sec. V.

---

*Permanent address: Facultad de Ciencias Físicas y Matemáticas, Departamento de Física, Universidad de Chile, Santiago, Chile.
Supported in part by the Organización de Los Estados Americanos.
†Supported in part by the Istituto Nazionale di Fisica Nucleare, Sottosezione di Trieste, Italy.

[1]A. Schild, Phys. Rev. **131,** 2762 (1963).
[2]A. Degasperis, Phys. Rev. D **3,** 273 (1971).
[3]C. M. Andersen and H. C. von Baeyer, Ann. Phys. (N.Y.) **60,** 67 (1970).
[4]C. M. Andersen and H. C. von Baeyer, Phys. Rev. D **5,** 802 (1972).
[5]A careful discussion of the subject and earlier references on the subject will be found in: J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17,** 157 (1945); Rev. Mod. Phys. **21,** 425 (1949).
[6]F. Rohrlich in his book, Ref. 7, p. 194, states ". . . that a consistent action integral for an electromagnetic n particle system ($n > 1$) requires the introduction of a direct interaction between the charges, not mediated by a field. . .."
[7]F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965).
[8]R. Bellman and K. L. Cooke, *Differential-Difference Equations* (Academic , New York, 1963).
[9]The difficulties with the differential-difference nature of the relativistic equations (when the interaction takes place only for light like distances) are discussed in the following papers: R. D. Driver, Ann. Phys. (N.Y.) **21,** 122 (1953); Phys. Rev. **178,** 2051 (1969).
[10]R. A. Rudd and R. N. Hill, J. Math. Phys. **11,** 2704 (1970).
[11]By "these lines" we mean in the spirit of Refs. 5 and 1 and we explicitly exclude the other approach consisting of a Hamiltonian formulation of dynamics. An up-to-date list of references on this and other theoretical developments is found in Ref. 12.
[12]P. Havas, in *Problems in Foundations of Physics,* edited by M. Bunge, (Springer-Verlag, Berlin, 1971), p. 31. This paper in its generality contains ours as a particular case.
[13]H. Van Dam and E. P. Wigner, Phys. Rev. B **138,** 1576 (1965); Phys. Rev. B **142,** 838 (1966).
[14]The metric tensor is $g_{00} = -g_{ii} = 1$. An event is characterized by the vector $x^\mu = (ct, \mathbf{x})$, so that the 4-velocity is $\dot{x}^\mu \equiv dx^\mu/d\tau = \gamma(c, \mathbf{v})$, where $\mathbf{v} = d\mathbf{x}/dt$ and $\gamma = (1 - v^2/c^2)^{-1/2}$. For a physical world line $x^\mu(\tau)$, the parameter τ is the proper time and the connection with the reference frame time t is $\gamma d\tau = dt$.
[15]A self-consistency check is achieved by multiplying these equations by $\dot{x}_{a\mu}$ (sum over μ) and by means of Eq. (2.2) obtaining an identity.
[16]Having a conserved total linear momentum and angular momentum, it is possible to define the relativistic center of mass world line describing a straight line by $X^\mu = -(L^{\mu\nu}P\nu/P^2) + \beta P^\mu$. Other possible definitions of the center of mass have been discussed in M. Pauri, University of Parma, preprint IFPR-T-019.
[17]A. Katz, J. Math. Phys. **10,** 2215 (1969).
[18]A. Katz, J. Math. Phys. **10,** 1929 (1969).
[19]The treatment given in this appendix follows that given by A. Schild, in *Proceedings of the International School of Physics, Enrico Fermi, Course 20* (Academic , New York, 1963), pp. 69–115.

Perturbation of the $P(\phi)_2$ quantum field Hamiltonian*

Thomas Spencer

Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 15 January 1973)

We consider local energy operators δH_0 as perturbations of the $P(\phi)_2$ Hamiltonian. Our principal result is the estimate $\delta H_0 \leq C[H(g)-E(g)+1]$ where C is a constant independent of the space cutoff g and $E(g)$ is the infimum of the spectrum of $H(g)$.

1. INTRODUCTION

We study perturbations of the Hamiltonian in two dimensional models for quantum field theory. The Hamiltonian H is formally the sum of a free Hamiltonian H_0 and an interaction Hamiltonian H_I and may be expressed as an integral over space of an energy density. In a recent paper, J. Glimm and A. Jaffe[1] proved that a certain class of perturbations δH_1 arising from local changes in the interaction density produces only a finite shift δE in the vacuum energy of the renormalized $P(\phi)_2$ Hamiltonian. Later Guerra, Rosen and Simon[2] used Nelson's symmetry argument to give an elementary proof of this bound on δE with further restrictions on δH_I.

This paper studies perturbations δH_0 of the total Hamiltonian arising from local changes in the free energy density. We show that the resulting shift in the vacuum energy remains finite for a class of such perturbations. Our estimate combined with the methods of Ref. 3 should be useful in establishing higher order estimates in which powers of δH_0 are dominated by powers of H. The proof of our results relies on a simple extention of the results of Ref. 1. Although the methods of Ref. 2 are formally applicable, there appear to be technical difficulties in treating the perturbations considered in this paper.

2. NOTATION AND BASIC RESULTS

Let $\mathfrak{F}$ denote the Fock space over $L^2(R^1)$, $\mathfrak{F} = \sum_{n=0}^{\infty} \otimes \mathfrak{F}_n$, where

$$\mathfrak{F}_0 = \mathbb{C} = \{\mathbb{C}\,\Omega_0\}, \qquad \mathfrak{F}_n = \overset{n}{\underset{1}{\otimes}} L^2(R^1)$$

and $\otimes_1^n$ denotes the n-fold symmetric tensor product. Let

$$\mu(k) = (k^2 + m^2)^{1/2} \tag{2.1}$$

and let

$$\mu = \left(-\frac{dx}{dx^2} + m_0^2\right)^{1/2}, \tag{2.2}$$

so that μ is convolution by the inverse Fourier transform of the function $\mu(\cdot)$. The free Hamiltonian for a scalar boson of mass $m_0 > 0$ is given by the second quantization of μ. Thus,

$$H_0 = d\Gamma(\mu), \tag{2.3}$$

where

$$d\Gamma(\mu) \restriction \mathfrak{F}_0 = 0$$

and

$$d\Gamma(\mu) \restriction \mathfrak{F}_n = \sum_{i=1}^{n} \mu_{x_i},$$

and μ_{x_i} denotes the operator μ acting on the variable x_i.

In configuration space the time zero field is

$$\phi(f) = 2^{-1/2}[a^*(\mu^{-1/2}f) + a(\mu^{-1/2}\bar{f})] \tag{2.4}$$

and we let $\phi(x) = \phi(\delta_x)$, where δ_x is the delta function centered at x. By definition, for $f \in L^2(R^1)$ and $F \in \mathfrak{F}_n$

$$[a(f)F]_{n-1}(x_1, \ldots, x_{n-1}) = n^{1/2}\int f(x)F(x, x_1, \ldots, x_{n-1})dx$$

and $a^*(\bar{f})$ is the formal adjoint of $a(f)$.

Let P be a positive polynomial and let $g(x)$ be a measurable function of compact support such that $0 \leq g \leq 1$. The spatially cutoff Hamiltonian is

$$H(g) = H_0 + H_I(g) = H_0 + \int :P(\phi(x)):g(x)dx. \tag{2.5}$$

More generally let ω be an operator on $L^2(R^1)$ satisfying the conditions

(A. 1) ω is self-adjoint and real in configuration space and for some constant $m > 0$

$$0 < m \leq \omega. \tag{2.6}$$

(A. 2) Let χ_i denote the characteristic function of $[i, i+1]$. There exist constants M_n such that for all $t \geq 0$:

$$\|\chi_i e^{-t\omega}\chi_j\| \leq M_n/(|i-j|^n + 1)(|t|+1)^n. \tag{2.7}$$

Let

$$H_0(\omega) = d\Gamma(\omega) \text{ and } H_1(g) = H_0(\omega) + H_I(g). \tag{2.8}$$

The following results are know to hold for $H_1(g)$. (See Refs. 4–7). Fock space may be realized as $L^2(Q)$ where Q is the maximal ideal space of the weak closure of the algebra of operators generated by

$$\{e^{i\phi(f)} : f \in L^2(R^1)\}.$$

In this representation

(1) $\Omega_0(g) \equiv 1$ and $H_I(g)$ is a multiplication operator. Moreover, H_I and e^{-H_I} belong to $L^p(Q)$ for all $p < \infty$.

(2) $H_I(g)$ is semibounded and essentially self-adjoint on $\mathfrak{D}(H_0) \cap L^p(Q)$ for $2 < p < \infty$.

(3) $e^{-tH_0(\omega)}$ is positively preserving for $t \geq 0$ and for $t > 0$ and $1 < p < \infty$ there is a constant $r < p$ such that $e^{-tH_1(g)}$ is bounded as an operator from $L^r(Q)$ to $L^p(Q)$.

(4) $H_1(g)$ has a unique vacuum $\Omega_1(g)$. As a function on Q space $\Omega_1(g)$ is positive and belongs to $L^p(Q)$ for all $p < \infty$. By definition $\Omega_1(g)$ is the unique eigenvector corresponding to the eigenvalue

$$E_1(g) = \inf \text{ spec } H_1(g). \tag{2.9}$$

Results (1)–(4) follow from (A. 1) alone. We note that (2) implies

Copyright © 1973 by the American Institute of Physics

$$\tfrac{1}{2} H_0(\omega) \leq H_1 + \text{const},$$

so that

$$\|H_0(\omega)^{1/2}(H_1 - E_1 + 1)^{-1/2}\| < \infty.$$

Hence

$$\begin{aligned}\|[H_1(g) - E_1(g) + 1]^{-1/2} a^{\#}(f)\| &\leq \|[H_1(g) - E_1(g) + 1]^{-1/2}[H_0(\omega) + 1]^{1/2}\| \\ &\times \|[H_0(\omega) + 1]^{-1/2} a^{\#}(f)\| \\ &\leq C(g)\|f\|_{L^2}, \end{aligned} \tag{2.10}$$

where $a^{\#} = a$ or a^*.

The choice of Q is not uniquely determined by (1)–(4) and for some purposes the maximum ideal space is unnecessarily large. We may replace Q by $\mathcal{S}'(R^1)$ so that for $f \in \mathcal{S}(R^1)$, the operator $\phi(f) = \int \phi(x) f(x) dx$ becomes the multiplication operator

$$F(q) \to \langle q, f\rangle F(q), \qquad q \in \mathcal{S}'(R^1).$$

The associated measure is Gaussian of covariance $\frac{1}{2}\mu^{-1}$ and (1)–(4) are satisfied. See Ref. 5.

By the Feynman–Kac formula, there is a Gaussian measure dq on $\mathcal{S}'(R^2)$ such that for $\theta_i \in \mathfrak{F} = L^2[\mathcal{S}'(R^1)]$

$$\langle \theta_1, e^{-tH_1(g)} \theta_2\rangle = \int \theta_1 V_0^t T^t \theta_2 dq, \tag{2.11}$$

where

$$V_0^t = \exp\left(-\int_0^t \int : P(q(x,s)) : g(x) dx ds\right)$$

and T^t is time translation. This result is implicit in Nelson[8] (see also Refs. 9–12).

Now we state the principal result of Ref. 1 on which our estimates are based.

Let h be a function of compact support satisfying $-1 \leq h \leq 1$, and suppose

$$\begin{aligned} &0 \leq P(\xi) g(x), \\ &0 \leq P(\xi) g(x) + P_1(\xi) h(x) \end{aligned} \tag{2.12}$$

for all real ξ and x. If $\deg P = \deg P_1$ but $P \neq P_1$, we also require that g is bounded away from zero in a neighborhood of the support of h. Let

$$E(g,h) = \inf \text{spec } H(g) + \int : P_1[\phi(x)] : h(x) dx.$$

Theorem 2.1: There exists a constant M independent of g and h such that

$$|E(g) - E(g,h)| \leq MD, \tag{2.13}$$

where

$$D = (\text{diam supp } h) + 1.$$

We generalize Theorem 2.1 as follows:

Theorem 2.2: Let ω satisfy (A.1)–(A.2). Let P, P_1, g, h be as in (2.12). Then the conclusion of Theorem (2.1) holds, namely,

$$|E_1(g) - E_1(g,h)| \leq MD,$$

We prove this theorem in Sec. 5 by noting the changes required in the proof of Theorem 2.1.

3. STATEMENT OF MAIN THEOREM

Let σ be a real symmetric operator defined on the domain of μ and let σ satisfy the following conditions:

(B.1) $\omega \equiv \mu + \sigma$ is self-adjoint on the domain of μ and for some constant m,

$$0 < m \leq \omega. \tag{3.1}$$

(B.2) There is a constant K such that

$$\|\omega\mu^{-1}\| \leq K, \|\omega^{-1}\mu\| \leq K. \tag{3.2}$$

(B.3) σ has compact support. By this we mean that there is a function $\chi \in C_0^\infty$ such that

$$\chi\sigma\chi = \sigma. \tag{3.3}$$

We define $\delta H_0 = d\Gamma(\sigma)$. Let H_1 and E_1 be given by (2.8) and (2.9).

Theorem 3.1: Let σ satisfy (B.1)–(B.3). Then

$$|E_1(g) - E(g)| \leq C, \tag{3.4}$$

where C depends only on the constants m, K, and χ in (3.1)–(3.3) and on P.

Theorem 3.2: For $\tau \leq 1$, and $\zeta \in C_0^\infty$

$$\sigma = \pm \epsilon \zeta \mu_x^{\tau} \zeta \tag{3.5}$$

satisfies (B.1)–(B.3) provided $\zeta \in C_0^\infty$ and ϵ is sufficiently small.

The proof of this theorem is given in Sec. 4. We set

$$N_{\tau\,\text{loc}} = d\Gamma(\zeta \mu_x^{\tau} \zeta).$$

Theorems 3.1 and 3.2 imply that $\langle \Omega(g), N_{\tau\,\text{loc}} \Omega(g)\rangle$ is bounded uniformly in g, $0 \leq g \leq 1$. By a result of Glimm and Jaffe,[13] this implies that the generalized sequence of vacuum states

$$\nu_g = \langle \Omega(g), \cdot \Omega(g)\rangle$$

is norm compact on restriction to each local algebra

$$\mathfrak{A}(O) = \{\exp i[\phi(f_1) + \pi(f_2)] : \text{supp} f_i \subset O\}''.$$

Here O is a bounded open set in R^1. Hence any limit point of the generalized sequence ν_g is locally normal. The previous norm compactness of Ref. 13 applied only to the space averaged vacuums.

For any limit stated ν of the generalized sequence $\{\nu_g\}$ the GNS construction gives rise to a Hilbert space $\mathfrak{F}_\nu$ and a representation of the algebra $\mathfrak{A}(O)$ on $\mathfrak{F}_\nu$. By another result of Ref. 13, this representation of $\mathfrak{A}(O)$ on $\mathfrak{F}_\nu$ and the free representation of $\mathfrak{A}(O)$ on $\mathfrak{F}$ are unitarily equivalent. However, the corresponding global representations of the norm closure of the union

$$\mathfrak{A} = \{\bigcup_{O \subset R^1} \mathfrak{A}(O)\}$$

are believed to be inequivalent.

4. ONE PARTICLE ESTIMATES

In this section we study operators $\omega = \mu + \sigma$. We show that ω satisfies (A.1) and (A.2). This will enable us to use Theorem 2.2 in the proof of Theorem 3.1. To derive our estimates we use properties (B.1)–(B.3) for

σ, the Duhamel formula for the different of two semigroups, and the pseudolocal property of μ.

The Duhamel formula is

$$e^{-t\omega} = e^{-t\mu} - \int_0^t e^{-s\omega}\sigma e^{-(t-s)\mu}\,ds \tag{4.1}$$

and may be obtained by first differentiating and then integrating the function

$$f(s) = e^{-s\omega}e^{-(t-s)\mu}.$$

An operator D: $\mathcal{S}' \to \mathcal{S}'$ is said to be pseudolocal if for $f \in \mathcal{S}'$, $(Df)(\cdot)$ is C^∞ at x whenever f is C^∞ at x. Thus D does not enlarge the singular support of a distribution f. We note that if η and ξ belong to C_0^∞ and if $\eta\xi = 0$ then $\eta D\xi$ is infinitely smoothing.

To study the properties of a class of pseudodifferential operators, let $f(k)$ be a smooth function which for each $n \geq 0$ satisfies a bound of the form

$$|D_k^n f(k)| \leq C_n(|k| + 1)^{-n+a}. \tag{4.2}$$

Here C_n and a are constants and $D_k = d/dk$. Let η_i be C_0^∞ functions supported in $[i-1, i+1]$ whose derivatives $D_k^N\eta_i$ are bounded uniformly in i for fixed N. We define C_f to be convolution by $\check{f}$, the inverse Fourier transform of f.

Lemma 4.1: Let f satisfy (4.2). Given n and $p \geq 0$ and η_i as above, there exists a constant $C_{n,p}$ such that

$$\|\mu^p\eta_i C_f \eta_j\mu^p\| \leq C_{n,p}(|i-j|+1)^{-n}.$$

provided $\eta_i\eta_j = 0$.

Proof: From (4.2) we see that for each p and for n sufficiently large,

$$[(ik)^p D_k^n f]^\vee \in L^2.$$

Since

$$[(ik)^p D_k^n f]^\vee = D_x^p[(ik)^n\check{f}] \in L^2,$$

we have for x bounded away from zero and for all n

$$|D_x^p\check{f}(x)| \leq C'_{n,p}/|x|^n.$$

Thus the distribution

$$G(x,y) = \eta_i(x)\check{f}(x-y)\eta_j(y)$$

belongs to C_0^∞ and

$$\|\mu^p\eta_i C_f\eta_j\mu^p\| \leq \max_{r,s\leq p}\text{const}\|D_x^r D_y^s G(x,y)\|_{L^2}$$
$$\leq C_{n\,p}(|i-j|+1)^{-n}.$$

See Ref. 14 for a systematic treatment of pseudodifferential operators.

Lemma 4.2: There are constants $a(\tau,k)$ such that for real x, y, τ, k and $y > 0, k \geq 0$,

$$\sup_{x\geq k} x^\tau e^{-yx} \leq a(\tau,k)e^{-ky}(1+y^{-\tau}).$$

Remark: The function $f(k) = e^{-t[\mu(k)-m/2]}$ satisfies (4.2) uniformly in $t \geq 0$; consequently, μ satisfies (A.2) [while (A.1) for μ is obvious].

Lemma 4.3: Let A and B be self-adjoint operators such that

$$0 \leq A \leq B.$$

Then for $0 \leq \tau \leq 1$

$$A^\tau \leq B^\tau.$$

(See Ref. 15.)

Theorem 4.1: Let σ satisfy (B.1)–(B.3). Then $\omega = \mu + \sigma$ satisfies (A.1) and (A.2).

Proof: (A.1) follows from (B.1) together with the assumption that σ is real. We now verify (A.2). We apply (4.1) to $e^{-t\omega}$. The first term resulting from (4.1) is

$$e^{-t\mu} = e^{-t(\mu-m/2)}e^{-t\,m/2}$$

and the required bound (A.2) follows from Lemma 4.1 with

$$f(k) = e^{-t[\mu(k)-m/2]}.$$

If both $\eta_i\chi \neq 0$ and $\eta_j\chi \neq 0$, we have $|i-j| = 0(1)$ and the required bound follows from 3.1. Thus without loss of generality we assume that $0(1)|j| \geq |i|$ and $\eta_j\chi = 0$. The norm of the second term resulting from (4.1) is dominated by

$$\int_0^t ds\|\eta_i e^{-s\omega}\sigma\chi e^{-(t-s)\mu}\eta_j\|$$
$$\leq \int_0^t ds\|\eta_i e^{-s\omega}\|\,\|\sigma\mu^{-1}\|\,\|\mu\chi e^{-(t-s)\mu}\eta_j\|.$$

By Lemma 4.1 the last factor is bounded by

$$C_{n_1}e^{-(t-s)\,m/2}(|j|+1)^{-n}.$$

The first two factors are bounded by $e^{-s\,m}$ and $K+1$, respectively. These bounds yield the theorem.

The following estimate will be useful in Sec. 6.

Proposition 4.1: Suppose that $\eta_i\chi = 0$. For each real p, there is a constant C'_p such that for all $t_1, t_2 \geq 0$.

$$\|\mu^{1/2}\chi e^{-t_1\mu}\,e^{-t_2\,\omega}\mu^\tau\eta_i\mu^p\| \leq (C'_p e^{-t_2\,m/2})/(|i|+1)^3.$$

Proof: We apply (4.1) to $e^{-t\omega}$. As in the proof of Theorem 4.1, the required bound holds for the first term resulting from (4.1). The second term is bounded by

$$\int_0^{t_2} ds\|\mu^{1/2}\chi e^{-t_1\mu}e^{-s\omega}\|\,\|\sigma\mu^{-1}\|\,\|\mu\chi e^{-(t_2-s)\mu}\mu^\tau\eta_i\mu^p\|.$$

Applying (3.2) and Lemma (4.1), we bound the last two factors by

$$(K+1)C_{3\,p}(|i|+1)^{-3}e^{-m(t_2-s)/2}.$$

For the first factor we write

$$\mu^{1/2}\chi = \chi\mu^{1/2} + [\mu^{1/2},\chi].$$

Since $[\mu^{1/2},\chi]$ is a bounded operator,

$$\|\mu^{1/2}\chi e^{-t_1\mu}\,e^{-s\omega}\|$$
$$\leq \text{const}\{\|\mu^{1/2}e^{-s\omega}\| + \|e^{-s\omega}\|\}.$$

Lemmas 4.2 and 4.3 combined with (3.2) and the spectral theorem show that

$$\|\mu^{1/2}e^{-s\omega}\| \leq \text{const}\,(1+s^{-1/2})e^{-sm}.$$

After performing the s integration, the proof follows.

We conclude this section with some elementary estimates needed to establish properties (B. 1)–(B. 3) for $\sigma = \zeta\mu_x\zeta$.

Proof of Theorem 3.2: The operator σ is easily seen to be real and symmetric on $\mathfrak{D}(\mu)$. Since

$$\|\zeta\mu\zeta\mu^{-1}\| \leq \|\zeta^2\| + \|\zeta[\mu, \zeta]\mu^{-1}\| \tag{4.3}$$

is finite for smooth ζ, $\epsilon\zeta\mu\zeta$ is a small perturbation of μ in the sense of Kato; consequently, ω is self-adjoint and (3. 1) holds. Condition (B. 3) is clearly satisfied, hence it remains to establish (B. 2). The bound $\|\omega\mu^{-1}\| \leq K$ is immediate from (4. 3). The bound $\|\mu\omega^{-1}\| \leq K$ follows from

$$\mu^2 \leq 2(\mu \pm \epsilon\zeta\mu\zeta)^2$$

or

$$\pm 2\epsilon(\mu\zeta\mu\zeta + \zeta\mu\zeta\mu) \leq \mu^2 + 2(\epsilon\zeta\mu\zeta)^2.$$

Thus we need only show that for ϵ sufficiently small

$$\pm 2\epsilon(\mu\zeta\mu\zeta + \zeta\mu\zeta\mu) \leq \mu^2,$$

but this is immediate from (4. 3).

5. THE PULL THROUGH FORMULA

To prove Theorem 2. 2. we reformulate and justify the pull through formula which generates the graph expansion of Ref. 1. After making a few comments on condition (A. 2), the rest of the proof of Theorem 2. 2 follows exactly as in Ref. 1.

We note that by using (A. 2) in place of Lemma 3. 1. 2 of Ref. 1, the exponential bound $\gamma e^{-(m-\epsilon)d}$ in (3. 2. 3) may be replaced by $\gamma'_N(d+1)^{-N}$ for large N. Thus l_ν in (2. 3. 3) of Ref. 1 now equals $\gamma'_N(d+1)^{-N/4}$. This is sufficient to control the combinatoric factors of Ref. 1. Hence Proposition 3. 3. 1 of Ref. 1 remains valid under the conditions of Theorem 2. 2.

Let $f \in L^2(R^2)$. Then for $t > 0$

$$\begin{aligned} a(f)e^{-tH_1} &= e^{-tH_1}a(f_t) - \int_0^t e^{-sH_1}[a(f_s), H_I(g)]e^{-(t-s)H_1}ds \\ &= e^{-tH_1}a(f_t) - \int_0^t e^{-sH_1}P'_s e^{-(t-s)H_1}ds, \end{aligned} \tag{5.1}$$

where

$$f_s = e^{-s\omega}f \tag{5.2}$$

and

$$P'_s = \int : P'(\phi(x)) : g(x)(\mu^{-1/2}f_s)(x)dx. \tag{5.3}$$

In (5. 3) P' denotes the derivative P.

Equation (5. 1) may be derived by first differentiating and then integrating from $s = 0$ to $s = t$, the function

$$B(s) = e^{-sH_1}a(f_s)e^{-(t-s)H_1}.$$

To justify this argument we first note that $B(s)$ is norm differentiable for $0 < s < t$ and its derivative is easily seen to be the integrand in (5. 1). The following lemma shows that $B'(s)$ is bounded in norm for $0 < s < t$.

Lemma 5. 1: Let $F_i(\phi)$ be functions of ϕ which belong to $L^p(Q)$ for all $p < \infty$. Then for $t > 0$

$$\|e^{-s_1H_1}F_1(\phi)e^{-(s_2-s_1)H_1}F_2(\phi)\cdots e^{-(t-s_m)H_1}\| \tag{5.4}$$

is bounded uniformly in s_i, for s_i in the sector $0 \leq s_i \leq \cdots \leq s_i \leq s_{i+1} \cdots \leq t$. Here the norm is the operator norm on $\mathfrak{F}$.

Proof: By the Feynman–Kac formula (2. 11) and the Holder inequality, we can bound (5. 4) by

$$\sup_{\|\theta_i\|_{p_2} \leq 1} \|\theta_1(T^t\theta_2)\|_{p_1} \|V_0^t \Pi_i^m T^{s_i}F_i\|_{p_2}, \tag{5.5}$$

where $p_1^{-1} + p_2^{-1} = 1$. Another application of the Holder inequality bounds the second factor by a constant independent of s_i. [Note: for $r < \infty$, by (2. 11), $\|V_0^t\|_r = \langle \Omega_0, e^{-tH_0 + rH_1}\Omega_0\rangle^{1/r} < \infty$]. We bound the first factor of (5. 5)

$$\langle {\theta_1}^{p_1}, e^{-tH_0(\omega)}{\theta_2}^{p_1}\rangle^{1/p_1}_{L^2(Q)} \leq \text{const}\|\theta_1\|_2\|\theta_2\|_2$$

by property (3) of Sec. 2.

By (2. 10) and Lemma 5. 1, (5. 1) holds as an identity of bounded operators defined on the domain $\mathfrak{D}[a(f_t)]$.

More generally, let $R(\phi)$ and $R_i(\phi)$ be polynomials in ϕ with L^2 kernels. Let $t > 0$ and $f \in L^2$. We assert that

$$\text{range} R(\phi)e^{-tH_1} \subset \mathfrak{D}(a(f)), \tag{5.6}$$

$$\text{range} a(f)e^{-tH_1} \subset \mathfrak{D}(R(\phi)), \tag{5.7}$$

and

$$\begin{aligned} a(f)R(\phi)e^{-tH_1} &= R(\phi)e^{-tH_1}a(f_t) \\ &+ [a(f), R(\phi)]e^{-tH_1} - R(\phi)\int_0^t e^{-sH_1}P'_s e^{-(t-s)H_1}ds \end{aligned} \tag{5.8}$$

on the domain of $a(f_t)$. By Lemma 5. 1 and the inclusions (5. 6) and (5. 7) each term in (5. 8) is a bounded operator defined on $\mathfrak{D}(a(f_t))$. Thus we may apply (5. 8) iteratively to the expression $a(f)\Pi_i(R_i\phi e^{-t_iH_1})\Omega_0$ which justifies the graph expansion of Ref. 1.

To establish (5. 6), we note that by (5. 1), (5. 8) holds as a bilinear form on $\mathfrak{D} \times \mathfrak{D}$ where $\mathfrak{D}$ is the space of vectors of finite particle number. Hence using (2. 10), Lemma 5. 1 and the fact that $[a(f), R(\phi)]$ is again a polynomial in ϕ, we have

$$\langle a(f)^* d_1, R(\phi)e^{-tH_1}d_2\rangle \leq \text{const}\|d_1\|\ \|d_2\|$$

for $d_1, d_2 \in \mathfrak{D}$. It follows that the range of $R(\phi)e^{-tH_1}$ is contained in the domain of $a(f)^{**} = a(f)$. Here we have used the fact that $\mathfrak{D}$ is a core for $a(f)^*$, the adjoint of $a(f)$. This fact holds because the domain of $N(f) = a(f)^*a(f)$ is a core for $a(f)^*$ by Ref. 16, because $\mathfrak{D}$ is a core for $N(f)$ and because $\|a(f)^*\psi\| = \|N(f)^{1/2}\psi\|$ for $\psi \in \mathfrak{D}$. Hence (5. 6) follows. The inclusion (5. 7) follows from (5. 1) and Lemma 5. 1. Equation (5. 8) is a consequence of (5. 1), (5. 6) and (5. 7).

6. PROOF OF THEOREM 3.1

Before beginning the proof of Theorem 3. 1 we make two simplifying assumptions. By Theorems 2. 2 and 4. 1, we can assume that $g(x)$ vanishes in a neighborhood of the support of σ. This will enable us to control the singularities arising from σ. We may also assume that

$$0 \leq E(g) - E_1(g). \tag{6.1}$$

If the reverse inequality $0 \leq E_1(g) - E(g)$ holds, we interchange the roles of H and H_1 in the following proof.

The proof of Theorem 3. 1 relies on Theorem 2. 2, the pull through formula (5. 1), and the Duhamel formula

$$e^{-H_1} = e^{-H} - \int_0^1 e^{-sH}\delta H_0 e^{-(1-s)H_1}ds. \quad (6.2)$$

Let T be a large positive number to be specified later. We apply (6.2) to the factor e^{-H_1} below:

$$\begin{aligned} e^{-(2T+1)H_1} &= e^{-TH_1}e^{-H_1}e^{-TH_1} \\ &= e^{-TH_1}e^{-H}e^{-TH_1} \\ &\quad - \int_0^1 e^{-TH_1}e^{-sH}\delta H_0 e^{-(1-s)H_1}e^{-TH_1}ds. \end{aligned} \quad (6.3)$$

Next we want to pull through the annihilation and creation operators of

$$\delta H_0 = d\Gamma(\sigma) = \int a^*(x)a(y)\sigma(x,y)dxdy$$

in (6.3) using (5.1) and its adjoint. Suppose σ is Hilbert–Schmidt. After expressing σ in the form

$$\sigma(x,y) = \Sigma_{ij}a_{ij}g_i(x)g_j(y),$$

$\|g_i\|_{L^2} \le 1$, we apply (5.1) to obtain a new expression for the integrated term of (6.3). In general, we approximate σ by the Hilbert–Schmidt operator

$$\sigma_\epsilon = \chi\, e^{-\epsilon\mu}\sigma e^{-\epsilon\mu}\chi$$

and set $\delta H_\epsilon = d\Gamma(\sigma_\epsilon)$. After applying (5.1) to δH_ϵ in (6.3), it may be shown that the limit as $\epsilon \to 0$ is the expression obtained by formal application of (5.1). In order to identify this expression with the last term in (6.3), we show that

$$\int_0^1 e^{-TH_1}e^{-sH}(\delta H_0 - \delta H_\epsilon)e^{-(1-s)H_1}e^{-TH_1}ds$$

tends weakly to 0 as $\epsilon \to 0$. Since

$$\delta H_\epsilon \le \text{const}\, H_0 \le \text{const}(H_1 - E_1 + 1),$$

if suffices to show

$$(H_0 + 1)^{-1/2}(\delta H_0 - \delta H_\epsilon)(H_0 + 1)^{-1/2}$$

tends weakly to zero on a dense set of vectors of finite particle number. This follows from the strong continuity of $e^{-\epsilon\mu}$.

Application of (5.1) and its adjoint to the integrand in (6.3) produces four terms. It is convenient to express each of these terms as a sum of local terms. Let η_i and ξ_i belong to $C_0^\infty(i-1, i+1)$ and satisfy

$$\xi_i\eta_i = \eta_i \quad \text{and} \quad \Sigma_i\eta_i = 1. \quad (6.4)$$

As in Sec. 4 we require the derivatives D_x^N of η_i and ξ_i to be bounded uniformly in i for fixed N.

We write the term with no contractions in the form

$$\int \Sigma_{i,j} a^*(f_{ix})e^{-TH_1}e^{-sH}e^{-(T+1-s)H_1}a(f_{jy}) \times K_0(i,j,x,y)\,dxdyds, \quad (6.5)$$

where

$$K_0(i,j) = \eta_i e^{-(T-1)\omega}e^{-s\mu}\sigma\, e^{-(T-s)\omega}\eta_j \quad (6.6)$$

and

$$f_{ix}(\cdot) = \xi_i(x)[e^{-\omega}\delta_x](\cdot). \quad (6.7)$$

Two terms have just one contraction. The term in which only the operator a contracts may be expressed as

$$\int \Sigma_{i,j} a^*(f_{ix})e^{-TH_1}e^{-sH}e^{-s_2H_1}P_j'(y) \times e^{-(T+1-s-s_2)H_1}K_1(i,j,x,y)dx\,dy\,ds\,ds_2, \quad (6.8)$$

where

$$K_1(i,j) = \eta_i e^{-(T-1)\omega}e^{-s\mu}\sigma\, e^{-s_2\omega}\mu^{-1/2}\zeta\eta_j\mu^{5/4} \quad (6.9)$$

and

$$P_j'(x) = \int : P'(\phi(z)) : g(z)\xi_j(z)\mu^{-5/4}(x-z)dz. \quad (6.10)$$

Here we define $\mu^\tau(x)$ to be the Fourier transform of $\mu^\tau(k)$ and we choose ζ to be a smooth function such that $\zeta\chi = 0$ and $\zeta g = g$. The term in which only the operator a^* contracts is similar. We write it as a sum of two terms (contraction to e^{-sH} or to e^{-TH_1}) with kernels $K_2(i,j)$ and $K_2'(i,j)$ replacing K_1 above.

We also write the term in which both a and a^* contract as a sum of two terms. The term in which a^* contracts to e^{-sH} is

$$\int \Sigma_{i,j} e^{-TH_1}e^{-(s-s_1)H}P_i'(x)e^{-s_1H}e^{-s_2H_1}P_j'(y) \times e^{-(T+1-s-s_2)H_1}K_3(i,j,x,y)\,dx\,dy\,ds_1\,ds_2\,ds, \quad (6.11)$$

where

$$K_3(i,j) = \mu^{5/4}\zeta\eta_i\mu^{-1/2}e^{-s_1\mu}\sigma\, e^{-s_2\omega}\mu^{-1/2}\zeta\eta_j\mu^{5/4}. \quad (6.12)$$

The term in which a^* contracts to e^{-TH_1} is

$$\int \Sigma_{ij} e^{-(T-s_1)H_1}P_i'(x)e^{-s_1H_1}e^{-sH}e^{-s_2H_1}P_j'(y)e^{-(T+1-s-s_2)H_1} \times K_3'(i,j,x,y)\,dx\,dy\,ds_1\,ds_2 ds,$$

where

$$K_3'(i,j) = \mu^{5/4}\zeta\eta_i\,\mu^{-1/2}e^{-s_1\omega}e^{-s\mu}\sigma\, e^{-s_2\omega}\mu^{-1/2}\zeta\eta_j\mu^{5/4}.$$

We estimate each of these terms in norm. For the factors $a(f_{jy})$ and $a^*(f_{iy})$ we use the g dependent bound (2.10). We note that $\|f_{ix}(\cdot)\|$ is bounded uniformly in x and i and vanishes for $|x-i| \ge 2$.

The following proposition gives a g independent bound on $P_i'(x)$.

Proposition 6.1: Let $P_i'(x)$ be defined by (6.10). There exists a constant independent of g, x, and i such that

$$\|[H_1(g) - E_1(g) + 1]^{-1/2}P_i'(x)[H_1(g) - E_1(g) + 1]^{-1/2}\| \le \text{const}(1 + |i-x|)^{-1} \quad (6.13)$$

Remark: The L^2 norm of the left side of (6.13) as a function of x is bounded uniformly in i. Note that the same bound holds with H replacing H_1.

Proposition 6.2: Let K be one of K_α, K_β', $\alpha = 1, 2, 3$, $\beta = 2, 3$. Then there exists a constant independent of g such that

$$\|K(i,j)\| \le \text{const}(|i| + |j| + 1)^{-3}e^{-mS}, \quad (6.14)$$

where $S = s_1 + s_2$ if α or $\beta = 3$, and $S = T$ otherwise. Here the norm denotes the operator norm on $L^2(R^1)$.

We postpone the proof of the two propositions and complete the proof of Theorem 3.1 The spectral theorem combined with Lemma 4.2 implies

$$\|[H_1 - E_1(g) + 1]^{-1/2}e^{-tH_1}\| \le 0(1 + t^{-1/2})e^{-tE_1(g)}. \quad (6.15)$$

To apply Lemma 4.2 in (6.15), we write

$$e^{-tH_1} = e^{-t[H_1 - E_1(g)+1]} e^{-t[E_1(g)-1]}.$$

Combining (2.10) and (6.15), we bound (6.5) by

$$C(g)^2 e^{-2TE_1} \int_0^1 e^{-sE} e^{-(1-s)E_1} ds \text{ const} \sum_{i,j} \|K_0(i,j)\|. \quad (6.16)$$

Here we have used the fact that the L^2 norm of $\|f_{ix}(\cdot)\|$ as a function of x is uniformly bounded in i and so

$$\int \|f_{ix}(\cdot)\| K_0(i,j,x,y) \|f_{jy}(\cdot)\| dxdy \leq \text{const} \|K_0(i,j)\|.$$

From (6.1) we have

$$\frac{e^{-E_1} - e^{-E}}{E - E_1} = \int_0^1 e^{-sE} e^{-(1-s)E_1} ds \leq e^{-E_1}. \quad (6.17)$$

Hence by Proposition 6.2, we can bound (6.5) by

$$\text{const} C(g)^2 e^{-(2T+1)E_1} e^{-mT}. \quad (6.18)$$

Similarly, after applying Propositions 6.1 and 6.2 and (6.15), we bound (6.11) by

$$C_1 e^{2TE_1} \frac{e^{-E_1} - e^{-E}}{E - E_1},$$

where C_1 is independent of g. We have used

$$\int_0^t (t-s)^{-1/2} s^{-1/2} ds$$

as uniformly bounded for $t \geq 0$. Bounds on the other terms follow similarly. Moreover, using (6.3) these bounds may be combined to give us an estimate of the form

$$e^{-2TE_1} e^{-E_1} \leq e^{-2TE_1} e^{-E} + \text{const} C(g)^2 e^{-Tm} e^{-2TE_1} e^{-E_1} + C_1 e^{-2TE_1} \frac{e^{-E_1} - e^{-E}}{E - E_1}.$$

We cancel the common factor of e^{-2TE_1}, and let $T \to \infty$. Suppose that $|E_1 - E|$ is greater than $2C_1$. (If not then we have the desired bound.) Hence we have

$$e^{-E_1} \leq e^{-E} + 1/2 e^{-E_1}.$$

or

$$1/2 e^{-E_1} \leq e^{-E}.$$

This bound together with (6.1) completes the proof of Theorem 3.1.

Proof of Proposition 6.1: We observe the function

$$h_i(z) = \mu^{-5/4}(x - z)\xi_i(z)$$

has compact support and is bounded in the sup norm by const$(|x - i| + 1)^{-1}$. (See the proof of Lemma 4.1). Let $P_1 = P'$. Since deg$P_1 <$ deg$P$, there is a constant such for real ξ, z, and i

$$0 \leq P(\xi)g(z) \pm P_1(\xi)h_i(z)g(z) + \text{const}.$$

Hence by Theorem 2.2 and a simple scaling argument we have

$$\pm P_i'(x) \leq \text{const}(|x - i| + 1)^{-1}[H_1(g) - E_1(g) + 1]$$

or

$$\|(H_1(g) - E_1(g) + 1)^{-1/2} P_i'(x)(H_1(g) - E_1(g) + 1)^{-1/2}\| \leq \frac{\text{const}}{|x - i| + 1}.$$

Proof of Proposition 6.2: We dominate $K_0(i,j)$ by

$$\|\eta_i e^{-(T-1)\omega} e^{-s\mu} \chi \mu^{1/2}\| \; \|\mu^{-1/2} \sigma \mu^{-1/2}\| \; \|\mu^{1/2} \chi e^{-(T-s)\omega} \eta_j\|.$$

The middle factor is bounded by (3.2) and Lemma 4.3. When $\eta\chi = 0$, we apply Proposition 4.1 to the other two factors. In applying Proposition 4.1 to the last factor, we set $\tau = 0$, $t_1 = 0$, and $t_2 = (T - s)$. For the first factor we set $\tau = 0$, $t_1 = s$, and $t_2 = T - 1$. If $\eta\chi \neq 0$ we bound these factors using (3.2) and Lemmas 4.2 and 4.3 by a constant.

Similarly, we bound the remaining $K(i,j)$ by a product of three factors and apply Proposition 4.1.

ACKNOWLEDGMENTS

I am happy to thank Professor James Glimm for proposing the problem and providing encouragement and guidance throughout the course of this work. I also thank Jon Dimock for helpful conversations.

---

*Supported in part by the National Science Foundation Grant NSF-GP-24003

[1] J. Glimm and A. Jaffe, J. Math. Phys. **10**, 1568 (1972).
[2] F. Guerra, L. Rosen, and B. Simon, Commun. Math. Phys. **27**, 19 (1972).
[3] L. Rosen, Commun. Pure Appl. Math. **24**, 415 (1971).
[4] J. Glimm and A. Jaffe, "Quantum field models" in *Statistical mechanics and quantum field theory* edited by C. de Witt and R. Stora (Gordon and Breach, New York, 1971).
[5] J. Glimm and A. Jaffe, "Boson quantum field models" in *Mathematics of Contemporary Physics*, edited by R. Streater (Academic, New York, to appear).
[6] I. Segal, I. Ann. Math. **92**, 462 (1970).
[7] B. Simon and R. Hoegh-Krohn, J. Funct. Anal. **9**, 121 (1972).
[8] E. Nelson, "Quantum fields and Markov fields," Proc. A.M.S. summer conference, 1971.
[9] E. Nelson, "Construction of quantum fields from Markoff fields," preprint.
[10] E. Nelson, "The free Markoff field," preprint.
[11] T. Spencer, "Perturbation of the $P(\phi)_2$ Hamiltonian," thesis, New York Univeristy, 1972.
[12] J. Feldman, "A relativisitic Feynman–Kac formula," preprint.
[13] J. Glimm and A. Jaffe, Acta Math. **125**, 203 (1970).
[14] L. Nirenberg, "Pseudo differential operators" in *Proceedings of symposia in pure mathematics* (American Mathematical Society, Providence, Rhode Island, 1970), Vol. XVI.
[15] L. Nirenberg, "Functional analysis," lecture notes, New York University, 1961.
[16] T. Kato, *Perturbation theory for linear operators* (Springer-Verlag, New York, 1966).

A note on Kato's perturbation theory

F. E. Schroeck Jr.

Department of Mathematics, Florida Atlantic University, Boca Raton, Florida 33432
(Received 5 June 1972; final revised manuscript received 3 October 1972)

The results of T. Kato are expanded by generalizing the relative bound condition on the perturbation to determine the domain of powers of the perturbed operator and by exhibiting some useful relative bounds between the unperturbed operator, the perturbed operator, and the perturbation.

In applying the perturbation theory of T. Kato to physical models,[1] two natural generalizations of Kato's result arise:

(i) a set of relative bounds between the unperturbed operator, the perturbed operator, and the perturbation;
(ii) further results on the domain of the perturbed operator.

In this paper we present these results in a general form.

Kato has proven[2] that under the conditions

(i) A is self-adjoint, B symmetric,
(ii) $D(B) \supseteq D(A)$,
(iii) there exist constants α, β, $0 \leq \alpha < 1$, $0 \leq \beta < \infty$, such that for all χ in $D(A)$,

$$\|B\chi\| \leq \alpha\|A\chi\| + \beta\|\chi\|,$$

then $A + B$ is self-adjoint and $D(A + B) = D(A)$.

Furthermore, if A is semibounded, then so is $A + B$.

We shall refer to (i), (ii), (iii) collectively as (K) or alternatively we shall say that B is a Kato perturbation of A. One of the useful aspects of the above is that the inequality (iii) is required on a presumably known domain $D(A)$.

We next observe that if A, B are any operators satisfying $B: D(A^n) \to D(A^{n-1})$ for $n = 1, 2, \cdots N$, then, by induction, $D((A + B)^m) \supseteq D(A^m)$ for $m = 1, 2, \cdots, N$.[3] This domain mapping condition will be referred to as (D).

Kato's theorem is easily extended as follows: We notice that any operator of the form λI, λ real, is a Kato perturbation of any self-adjoint operator. Let A, B satisfy (K), with A bounded below. Then there exist constants δ, ρ such that $A + \rho I$, $A + B + \delta I$ are self-adjoint, strictly positive, and have domains equal to $D(A)$. Without loss of generality, we may then assume A, $A + B$ are positive.

Corollary 1: Let A, B satisfy condition (K) for some α, β. Let A be positive, and choose δ, ρ positive with $\rho > \beta(1 - \alpha)^{-1}$. Then there exist positive constants $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ such that, for all χ in $D(A)$,

(i) $\|B\chi\| \leq \gamma_1\|(A + \rho I)\chi\|$,
(ii) $\|(B + \delta I)\chi\| \leq \gamma_2\|(A + \rho I)\chi\|$,
(iii) $\|(A + \rho I)\chi\| \leq \gamma_3\|(A + B + \rho I)\chi\|$,
(iv) $\|(A + B + \rho I)\chi\| \leq \gamma_4\|(A + \rho I)\chi\|$.

Proof: Let $\chi \in D(A)$. Then by the positivity of A, we have $\|(A + \rho I)\chi\| \geq \rho\|\chi\|$, $\|(A + \rho I)\chi\| \geq \|A\chi\|$. Therefore, $\|B\chi\| \leq \alpha\|A\chi\| + \beta\|\chi\| \leq (\alpha + \beta/\rho)\|(A + \rho I)\chi\|$, proving (i). Proving (ii), (iv) is similar. For (iii) we have

$$\begin{aligned}\|(A + \rho I)\chi\| &= \|(A + \rho I + B - B)\chi\| \\ &\leq \|(A + \rho I + B)\chi\| + \|B\chi\| \\ &\leq \|(A + \rho I + B)\chi\| + (\alpha + \beta/\rho)\|(A + \rho I)\chi\|.\end{aligned}$$

Since $\rho > \beta(1 - \alpha)^{-1}$, it follows that $\alpha + \beta/\rho < 1$ so that result (iii) holds by choosing $\gamma_3 = [1 - \alpha - \beta/\rho]^{-1}$.

We notice that "$A \geq 0$ implies $A^m \geq 0$ for all $m > 0$" admits the following generalization:

Lemma 1: Let A be a positive, self-adjoint operator. Let $0 \leq n < m$ and $\epsilon > 0$. Then there exists b finite, such that on $D(A^m)$, $A^n \leq \epsilon A^m + bI$.

Proof: Letting $\{E_\lambda\}$ denote the spectral family for A we have $A^n = \int_0^\infty \lambda^n dE_\lambda$. Furthermore, for any $\lambda_0 > 0$,

$$\begin{aligned}\int_0^{\lambda_0} \lambda^n dE_\lambda &\leq \lambda_0^n \int_0^{\lambda_0} dE_\lambda \leq \lambda_0^n \int_0^\infty dE_\lambda = \lambda_0^n I, \\ \int_{\lambda_0}^\infty \lambda^n dE_\lambda &= \int_{\lambda_0}^\infty \lambda^m \lambda^{n-m} dE_\lambda \leq \lambda_0^{n-m} \int_{\lambda_0}^\infty \lambda^m dE_\lambda \\ &\leq \lambda_0^{n-m} \int_0^\infty \lambda^m dE_\lambda = \lambda_0^{n-m} A^m.\end{aligned}$$

Choosing λ_0 by $\epsilon = \lambda_0^{n-m}$ and choosing $b = \lambda_0^n = \epsilon^{n/(n-m)}$, we obtain

$$A^n = \int_0^{\lambda_0} \lambda^n dE_\lambda + \int_{\lambda_0}^\infty \lambda^n dE_\lambda \leq \epsilon A^m + bI.$$

We will use multiple commutator notation defined by $(\mathrm{ad}A)^0(B) = B$, $(\mathrm{ad}A)^n(B) = [A, (\mathrm{ad}A)^{n-1}(B)]^\sim$ defined on the domain of the right-hand side, where $^\sim$ denotes the smallest closed extension.

Lemma 2: Let A, B be linear operators such that domain mapping condition (D) holds for some N. Then for all χ in $D(A^{m+1})$, $1 \leq m \leq N - 1$, we have

$$A^m B\chi = \sum_{p=0}^m \binom{m}{p} (\mathrm{ad}A)^p(B) A^{m-p}\chi.$$

Proof: By condition (D), all operators are well defined. The combinatorial factors are then obtained by induction using the known result

$$\binom{a}{b} + \binom{a}{b-1} = \binom{a+1}{b}.$$

We may now state

Corollary 2: Let A be self-adjoint. Let B be symmetric and satisfy

(i) $B: D(A^n) \to D(A^{n-1})$ for $n = 1, 2, \cdots, N$,
(ii) for all $p = 0, 1, 2, \cdots, N-1$, there exist positive constants c_p, d_p such that for all χ in $D(A^{p+1})$,

$$\|(\mathrm{ad}A)^p(B)\chi\| \leq c_p\|A^{p+1}\chi\| + d_p\|\chi\|$$

with

$$\prod_{j=0}^{N-1} F_j < 2 \quad \text{where } F_j = 1 + \sum_{p=0}^j \binom{j}{p} c_p.$$

Then $A + B$ is self-adjoint and $D(A^m) = D((A + B)^m)$ for $m = 1, 2, \cdots, N$.

Copyright © 1973 by the American Institute of Physics

Proof: $F_j > F_{j-1}$ since $\binom{j}{p} > \binom{j-1}{p}$. Hence $F_0 = 1 + c_0 < 2$ or $c_0 < 1$. From (i) for the case $n = 1$ we have $D(B) \supseteq D(A)$ so that $A + B$ is self-adjoint with $D(A) = D(A + B)$ by Kato's theorem. Thus A^n, $(A + B)^n$ are self-adjoint for all n and, by (i), $D((A + B)^n) \supseteq D(A^n)$ for $n = 1, 2, \cdots, N$. We shall show that the remainders $R_n = (A + B)^n - A^n$ are Kato perturbations of A^n for $n = 2, 3, \cdots, N$ as well. Since $D(R_n) = D((A + B)^n - A^n) = D(A^n)$, R_n is symmetric. It remains to prove that for each $n = 2, \cdots, N$, there exist constants α_n, β_n with $0 \le \alpha_n < 1$, $0 \le \beta_n < \infty$ such that, for all χ in $D(A^n)$, $\|R_n\chi\| \le \alpha_n\|A^n\chi\| + \beta_n\|\chi\|$.

We show the inequality to hold for all $n = 1, 2, \cdots, N$ with the choice $\alpha_n = (\prod_{j=0}^{n-1} F_j) - 1 + \epsilon_n$, where $\epsilon_n > 0$ may be chosen arbitrarily small. The inequality is true for $n = 1$ with $\alpha_1 = c_0 = F_0 - 1$. Assume that the result is true for $n = m < N$. Then, for χ in $D(A^{m+1})$,

$$\begin{aligned}\|R_{m+1}\chi\| &= \|[(A + B)^{m+1} - A^{m+1}]\chi\| \\ &= \|[A^mB + R_m(A + B)]\chi\| \\ &\le \|A^mB\chi\| + \alpha_m\|A^m(A + B)\chi\| + \beta_m\|(A + B)\chi\| \\ &\le (1 + \alpha_m)\|A^mB\chi\| + \alpha_m\|A^{m+1}\chi\| \\ &\quad + \beta_m\|A\chi\| + \beta_m\|B\chi\|.\end{aligned}$$

Consider

$$\begin{aligned}\|A^mB\chi\| &\le \sum_{p=0}^{m}\binom{m}{p}\|(\mathrm{ad}A)^p(B)A^{m-p}\chi\| \\ &\le \sum_{p=0}^{m}\binom{m}{p}c_p\|A^{m+1}\chi\| + \sum_{p=0}^{m}\binom{m}{p}d_p\|A^{m-p}\chi\|.\end{aligned}$$

Then

$$\begin{aligned}\|R_{m+1}\chi\| \le &\left[(1 + \alpha_m)\sum_{p=0}^{m}\binom{m}{p}c_p + \alpha_m\right]\|A^{m+1}\chi\| \\ &+ (1 + \alpha_m)\sum_{p=0}^{m}\binom{m}{p}d_p\|A^{m-p}\chi\| \\ &+ \beta_m\|A\chi\| + \beta_m c_0\|A\chi\| + \beta_m d_0\|\chi\|.\end{aligned}$$

By Lemma 1 and the positivity of A^2, all terms after the first may be bounded by

$$\epsilon\|A^{m+1}\chi\| + b_{m+1}\|\chi\| \quad \text{where } \epsilon > 0 \text{ is arbitrary.}$$

Therefore,

$$\|R_{m+1}\chi\| \le \left[\epsilon + (1 + \alpha_m)\sum_{p=0}^{m}\binom{m}{p}c_p + \alpha_m\right]\|A^{m+1}\chi\| + b_{m+1}\|\chi\|,$$

and

$$\begin{aligned}&\epsilon + (1 + \alpha_m)\sum_{p=0}^{m}\binom{m}{p}c_p + \alpha_m \\ &\quad = \epsilon - 1 + F_m + F_m\alpha_m \\ &\quad = \epsilon - 1 + F_m + F_m\left(\prod_{j=0}^{m-1}F_j - 1 + \epsilon_m\right) \\ &\quad = \prod_{j=0}^{m}F_j - 1 + (\epsilon + \epsilon_m F_m)\end{aligned}$$

proving the result.

This corollary is augmented by the following:

(i) Since R_m is a Kato perturbation of A^m, the estimates of Corollary 1 apply with A replaced by A^m and B by R_m whenever A^m is semibounded.

(ii) If the c_p may be chosen arbitrarily small for $p = 0, 1, 2, \cdots$, then the result holds for arbitrarily large N, and hence the C_∞ domains coincide:

$$C_\infty(A) = C_\infty(A + B).$$

(iii) By Lemma 1, the condition

$$\|(\mathrm{ad}A)^p(B)\chi\| \le e_p\|A^{p+1-\delta}\chi\| + g_p\|\chi\|,$$

for $\delta > 0$, e_p, g_p finite, allows one to choose c_p arbitrarily small.

(iv) The following lemmas indicate that in some cases it suffices to know the assumptions of corollary 2 on a smaller domain:

Lemma 3: Let A, B be closed, A^{-1} bounded. Then A^nB is closed for $n = 0, 1, 2, \cdots$.

Proof: Pick a sequence $\{u_n\} \subset D(A^nB)$ such that $u_n \to u$, $A^nBu_n \to \chi$ in $\mathcal{H}$. Then since A^{-n} is bounded, $Bu_n = A^{-n}A^nBu_n \to A^{-n}\chi$. Since B is closed, $u \in D(B)$ and $A^{-n}\chi = Bu$; i.e., $\chi = A^nBu$.

Lemma 4: Let A, B be closed, A^{-1} closed and bounded, and let $(\mathrm{ad}A)^n(B)$ be A^{n+1}-bounded for $n = 0, 1, \cdots, N$. Then there exists a core for A^{n+1}, D_0, with $BD_0 \subseteq D(A^n)$ if and only if $BD(A^{n+1}) \subseteq D(A^n)$.

Proof: The indirect proof is trivial. Assume such a D_0 exists. Let $u \in D(A^{n+1})$. Then there exists $\{u_p\} \subset D_0$ such that $u_p \to u$, $A^{n+1}u_p \to A^{n+1}u$. Since A^{-1} is bounded, it follows that $A^mu_p \to A^mu$ for all $m = 0, 1, \ldots, n + 1$. On D_0 we have

$$A^nBu_p = \sum_{m=0}^{n}\binom{n}{m}(\mathrm{ad}A)^m(B)A^{n-m}u_p.$$

Using the A^{m+1} bounds, the right-hand side is seen to converge. Since A^nB is closed it follows that $u \in D(A^nB)$, i.e. $Bu \in D(A^n)$.

Lemma 5: Let A be normal. Let A^nB be closed and $(\mathrm{ad}A)^n(B)$ be A^{n+1}-bounded for $n = 0, 1, \ldots, N$. Then there exists a core for A^{n+1}, D_0, with $BD_0 \subseteq D(A^n)$ if and only if $BD(A^{n+1}) \subseteq D(A^n)$.

The proof is similar to the previous ones, but we use Lemma 1 to obtain convergence of A^mu_p from that of $A^{n+1}u_p$, $m = 0, 1, \ldots, n + 1$.

ACKNOWLEDGMENT

The author is indebted to Professor G. G. Emch for his guidance and suggestions in the development of these results. He also wishes to thank Professor T. P. Schonbek for pointing out a significant extension of the original results.

---

[1]The spirit of these generalizations may be found in "Constructive Quantum Field Theory", Lectures given by A. M. Jaffe at the Eidgenossishe Technische Hochschule, Zurich, Switzerland, 1968. A purely mathematical treatment of the forerunner to the present results is contained in F. E. Schroeck, Jr., "On the Field Theoretic Approach to the Yukawa Interaction of Scalar Nucleons in the Absence of Antinucleons", thesis (University of Rochester, 1970). Some of the results of the above may also be found in the appendixes to the paper J. T. Cannon, J. Funct. Anal. 8, 101 (1971).

[2]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1969), pp. 287–91.

[3]Reference 2, p. 163.

C^∞-perturbations of a cosmological model*

Paul C. White

Department of Physics, St. Edward's University, Austin, Texas 78704
(Received 20 September 1971)

It is shown that the so-called "moment condition" imposed by R. K. Sachs and A. M. Wolfe in their general treatment of the pertrubation problem for $k=0$ Robertson–Walker universes both exclude physically reasonable solutions and are unnecessary. This restriction is removed by establishing that the Sachs–Wolfe solution gives the most general C^∞ solution provided the potentials are C^∞.

I. INTRODUCTION

In 1966, R. K. Sachs and A. M. Wolfe presented an elegant treatment of the perturbations of $k = 0$, Robertson-Walker universes filled with either pure radiation or incoherent matter.[1] Notably, the components of the perturbed metric were expressed in terms of a 'small' number of potentials, and a physical interpretation of each was given. The Fourier transformation method used by Sachs and Wolfe required the class of functions eligible for representing the components of the perturbed metric to be restricted by certain moment conditions. We wish to show that, in fact, the Sachs–Wolfe solution contains all C^∞-perturbations, provided the potentials are C^∞. A simple, physically reasonable example shows that the moment conditions form an undesirable and, as our main theorem shows, unnecessary restriction. Also, a shortening and simplification of the proof is achieved by working in spacetime throughout. We employ units in which $c = 8\pi G = 1$. Latin indices will run from 1 to 4, while Greek indices run from 1 to 3. A comma denotes partial differentiation, while a semicolon denotes covariant differentiation.

II. THE SACHS-WOLFE PROCEDURE

To formulate the perturbation problem, we consider a one-parameter family

$$g_{ab}(x^c, \epsilon)$$

of solutions to the Einstein equations with a perfect fluid source,

$$G^a_b = -(\rho + p)u^a u_b + p\delta^a_b \tag{1}$$

which coincides at $\epsilon = 0$ with the $k = 0$ Robertson–Walker models. Then,

$$g_{ab}(x^c, 0) = a^2(\eta)\eta_{ab} \tag{2}$$

where $\eta_{ab} = \text{diag}(-1, -1, -1, 1)$ and η is a time coordinate with range $0 < \eta < \infty$, scaled to have present value unity. In the cases of incoherent matter and pure radiation sources, we have, respectively,

$$\begin{aligned} &p_0 := p(\epsilon = 0) = 0: \\ &\quad a = 2\eta^2/H, \quad \rho_0 := \rho(\epsilon = 0) = 3H^2/\eta^6, \\ &\quad t = 2\eta^3/3H, \quad p_0 = \rho_0/3: \\ &\quad a = \eta/H, \quad \rho_0 = 3H^2/\eta^4, \quad t = \eta^2/2H, \end{aligned} \tag{3}$$

where

$$H := (a'/a)\eta = 1$$

is the Hubble parameter now, t is the cosmological proper time, and the prime denotes differentiation with respect to η.

The first order perturbations of this "background" are defined as those elements of the one parameter family "near" $\epsilon = 0$, in the sense

$$g_{ab}(x^c, \epsilon) \cong g_{ab}(x^c, 0) + \delta g_{ab}(x^c, \epsilon), \tag{4}$$

where

$$\delta g_{ab}(x^c, \epsilon) := \epsilon\left(\frac{\partial g_{ab}(x^c, \epsilon)}{\partial \epsilon}\right)_{\epsilon=0}.$$

Because of (2), we write δg_{ab} in the form

$$\delta g_{ab} = a^2(\eta)h_{ab}, \tag{5}$$

and employ comoving coordinates

$$u^a = \delta^a_4/a(\eta) \leftrightarrow G^\alpha{}_4 = 0, \quad h_{44} = 0. \tag{6}$$

The coefficients h_{ab} then have indices raised and lowered by η_{ab}.

The coordinates in the background metric are fixed up to the Euclidean transformations

$$\hat{x}^4 = x^4, \quad \hat{\mathbf{x}} = 0\mathbf{x} + \mathbf{d}, \tag{7a}$$

where

$$00^T = 1, \quad 0 = \text{const}, \quad \mathbf{d} = \text{const}. \tag{7b}$$

The remaining coordinate transformations

$$\hat{x}^4 = f^a(x^b, \epsilon)$$

in the perturbed metric

$$\hat{x}^a = x^a + \epsilon\left(\frac{\partial f^a(x^b, \epsilon)}{\partial \epsilon}\right)_{\epsilon=0},$$

when restricted by the comoving coordinate conditions, lead to changes in h_{ab} of the form

$$\hat{h}_{\alpha\beta} = h_{\alpha\beta} + d_{(\alpha,\beta)} + 2(a'/a^2)b\eta_{\alpha\beta}, \tag{8a}$$

$$\hat{h}_{\alpha 4} = h_{\alpha 4} + a^{-1}b,\alpha, \tag{8b}$$

where b and d_α are arbitrary spatial functions of first order in ϵ.

The perturbed field equations are

$$\delta G^4{}_4 = -\delta\rho, \quad \delta G^\alpha{}_4 = 0, \quad \delta G^\alpha{}_\beta = \delta^\alpha{}_\beta \delta P, \tag{9}$$

where

$$\delta G^a{}_b = \epsilon\left(\frac{\partial G^a{}_b(x^c, \epsilon)}{\partial \epsilon}\right)_{\epsilon=0},$$

and $G^a{}_b(x^c, \epsilon)$ is the Einstein tensor associated with $g_{ab}(x^c, \epsilon)$. The Sachs–Wolfe procedure for solving these

Copyright © 1973 by the American Institute of Physics

equations utilized the flatness of the preferred timelike three surfaces in the background to Fourier-transform the unknown h_{ab}:

$$h_{\alpha\beta}(\mathbf{x},\eta) = \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{h}_{\alpha\beta}(k,\eta),$$

$$h_{\alpha 4}(\mathbf{x},\eta) = \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{h}_{\alpha 4}(k,\eta).$$

The field equations then become a set of coupled, ordinary differential equations for the coefficients $\tilde{h}_{\alpha\beta}$ and $\tilde{h}_{\alpha 4}$. If further regularity conditions (moment conditions) are imposed on these coefficients, then the field equations can be uniquely separated into parts longitudinal or transverse with respect to $\mathbf{k}$. Their solutions can be expressed in terms of powers of η and spherical Bessel functions of low order. Transforming these solutions back to position space, and reducing the number of arbitrary functions by means of the gauge freedom of Eqs. (8) yields the Sachs–Wolfe solution:

$$p_0 = 0, \quad \delta p = 0 : h_{\alpha\beta} = \frac{1}{\eta}\frac{\partial}{\partial\eta}\left(\frac{D_{\alpha\beta}}{\eta}\right) - 2\left(\frac{8}{\eta^3} - \frac{\nabla^2}{\eta}\right) \times (C_{\alpha,\beta} + C_{\beta,\alpha}) + \left(\eta_{\alpha\beta} - \frac{\eta^2}{10}\frac{\partial^2}{\partial x^\alpha \partial x^\beta}\right)B,$$

$$h_{\alpha 4} = -(2/\eta^2)\nabla^2 C_\alpha, \quad \delta\rho = \frac{H^2}{4}\nabla^2\left(\frac{6A}{\eta^9} - \frac{3B}{5\eta^4}\right), \tag{10a}$$

$$p_0 = \rho_0/3, \quad \delta p = \delta\rho/3 : h_{\alpha\beta} = \frac{D_{\alpha\beta}}{\eta} - (\eta\nabla^2 + \frac{8}{\eta}) \times (C_{\alpha,\beta} + C_{\beta,\alpha}) + \frac{\eta^2}{2}\frac{\partial}{\partial\eta}\left(\frac{E_{,\alpha\beta}}{\eta^2}\right) - \frac{\eta_{\alpha\beta}}{\eta^2}\frac{\partial E}{\partial\eta},$$

$$h_{\alpha 4} = -\nabla^2 C_\alpha + \frac{\eta^2}{4}\frac{\partial}{\partial\eta}\left(\frac{1}{\eta^2}\frac{\partial E_{,\alpha}}{\partial\eta}\right),$$

$$\partial\rho = \frac{3H^2}{\eta^4}\frac{\partial}{\partial\eta}\left[\eta^2\frac{\partial}{\partial\eta}\left(\frac{\partial E/\partial\eta}{\eta^2}\right)\right]. \tag{10b}$$

Here, the potentials A, B, and C_α are functions of $\mathbf{x}$, while the potentials $D_{\alpha\beta}$ and E are functions of $\mathbf{x}$ and η. They are subject to the constraints

$$D^{\alpha\beta}{}_{,\beta} = 0, \quad \left(\frac{\partial^2}{\partial\eta^2} - \nabla^2\right)D_{\alpha\beta} = 0, \quad C^\alpha{}_{,\alpha} = 0, \quad \left(3\frac{\partial^2}{\partial\eta^2} - \nabla^2\right)E = 0, \tag{10c}$$

where ∇^2 is the Laplacian of Euclidean 3-space E^3.

The 'moment conditions' which have been imposed on the coefficients $\tilde{h}_{\alpha\beta}$ and $\tilde{h}_{\alpha 4}$ in order to guarantee the uniqueness of the splitting with respect to $\mathbf{k}$ require representations of the form

$$\tilde{h}_{\alpha\beta}(\mathbf{k},\eta)k^\alpha k^\beta = k^4 f(\mathbf{k},\eta), \tag{11a}$$

$$\tilde{h}_{\alpha\beta}(\mathbf{k},\eta)k^\beta = k^2 g_\alpha(\mathbf{k},\eta), \tag{11b}$$

$$\tilde{h}^\alpha{}_\alpha(\mathbf{k},\eta) = k^4 j(\mathbf{k},\eta), \tag{11c}$$

$$\tilde{h}_{\alpha 4}(\mathbf{k},\eta)k^\alpha = -ik^2 m(\mathbf{k},\eta), \tag{11d}$$

where f, g_α, j and m are generalized functions, coinciding with continuous ordinary functions in the neighborhood of $\mathbf{k} = 0$. All physically distinct, first order perturbations of the given background model, Eqs. (3), which obey the "moment conditions" are described by the Sachs–Wolfe solution. On the other hand, a straightforward calculation shows that (10) always give solutions of the perturbation Eqs. (9) irrespective of moment conditions, provided $D_{\alpha\beta}$ is C^3, C_α is C^5, A and B are C^4, and E is C^5. Among them are simple, physically acceptable solutions which fail to satisfy the moment condition. As an example, we choose, in Eqs. (10a),

$$D_{\alpha\beta} = 0, \quad C_\alpha = 0, \quad A = 0, \tag{12a}$$

and

$$B(r) = B_0\int_0^r \frac{dr'}{r'^2}\int_0^{r'} r''^2 dr''\, \exp(-r''/r_0{}^2), \quad r^2 = x^2{}_1 + x_2{}^2 + x_3{}^2, \tag{12b}$$

and B_0 is a constant. The perturbed matter density then has the form of a Gaussian

$$\delta\rho(r,\eta) = -(3H^2B_0/20\eta^2)\exp(-r^2/r_0{}^2).$$

On the other hand, we readily find that

$$\tilde{h}_{\alpha\beta}k^\alpha k^\beta = \frac{\sqrt{\pi}r_0{}^3\exp(-k^2r_0{}^2/4)}{2(2\pi)^2}(1 - \eta^2k^2/10),$$

so that from Eq. (11a)

$$f(\mathbf{k},\eta) = \frac{\sqrt{\pi}r_0{}^3\exp(-k^2r_0{}^2/4)}{2(2\pi)^2}\left(\frac{1}{k^4} - \frac{\eta^2}{10k^2}\right).$$

Clearly $f(\mathbf{k},\eta)$ is not continuous at $k = 0$, and so this simple solution violates the moment conditions.

We shall show, in the following section, that in fact the Sachs–Wolfe solution (10) contains all C^∞-solutions of the perturbed field equations.

III. AN ALTERNATE TREATMENT

Our procedure here will be first to integrate the field equations directly, and then vigorously to employ the gauge freedom in order to cast the results into the form of the Sachs–Wolfe solution. It is well known (see, for example, Synge[2]) that the field equations

$$G^a{}_b + T^a{}_b = 0$$

in space–time (M^4) are locally and equivalent to

$$G^\alpha{}_\beta + T^\alpha{}_\beta = 0, \tag{13a}$$

$$T^a{}_{b;a} = 0 \tag{13b}$$

in M^4, together with

$$G^4{}_a + T^4{}_a = 0 \tag{13c}$$

on S^3, where S^3 is a three-dimensional hypersurface defined locally by

$$x^4 = \text{const}$$

in coordinates chosen so that

$$g^{44} \neq 0.$$

With the metric written in the form of Eq. (4), the perturbed field Eqs. (9) can be rewritten as

$$\delta G^\alpha{}_\beta + \delta T^\alpha{}_\beta = 0 \tag{14a}$$

and

$$\delta(T^{ab};b) := (\delta T^{ab})_{,b} + \overset{0}{\Gamma}{}^{a}{}_{bc}\delta T^{bc} + \overset{0}{\Gamma}{}^{b}{}_{cb}\delta T^{ac} + \delta\Gamma^{a}{}_{bc}\overset{0}{T}{}^{bc} + \delta\Gamma^{b}{}_{cb}\overset{0}{T}{}^{ac} = 0, \quad (14b)$$

$$\delta G^{4}{}_{a} + \delta T^{4}{}_{a} = 0$$

on S^3, where we shall use the perturbed perfect fluid matter tensor

$$\delta T^{ab} = (\delta\rho + \delta p)\overset{0}{u}{}^{a}\overset{0}{u}{}^{b} - \delta p\overset{0}{g}{}^{ab} - \overset{0}{p}\delta g^{ab} + 2(\overset{0}{\rho} + \overset{0}{p})\delta\mu^{(a}\overset{0}{u}{}^{b)}.$$

Let us call any solution to the field equations which can be represented by C^∞ functions a C^∞ solution. Then, in order to show that the Sachs–Wolfe solution is the general C^∞ solution to Eqs. (14), we repeatedly use the following important fact:

Lemma: If g is any C^∞ function on E^3, then there exists a C^∞ function f on E^3 such that

$$\nabla^2 f := -\eta^{\alpha\beta}f_{,\alpha\beta} = g.$$

This lemma is a special case of a theorem for more general mth order differential operators in E^n. For a proof, see Friedman.[3]

We now state our main result in the form of a theorem:

Theorem: All C^∞ solutions to the perturbed field equations (14) can, modulo C^∞ gauge transformations of the type (8), be expressed in the form of the Sachs–Wolfe solution, Eqs. (10), with C^∞ potentials $A, B, C^\alpha, D^{\alpha\beta}$ and E.

Proof: If we solve Eqs. (14a) for the second time derivatives of the $h_{\alpha\beta}$, and then take the trace with respect to $\eta_{\alpha\beta}$, we find that in M^4

$$h'' = -2(a'/a)h' + 2h^{4\mu}{}_{,\mu}{}' + 4(a'/a)h^{4\mu}{}_{,\mu} + \tfrac{1}{2}S^{\mu\nu}{}_{,\mu\nu} + \tfrac{1}{3}\nabla^2 h - 3a^2\delta p, \quad (15a)$$

where $h := \eta^{ab}h_{ab} = \eta^{\alpha\beta}h_{\alpha\beta}$. Also, we find that the trace free-part of the metric, $S^{\alpha}{}_{\beta} := h^{\alpha}{}_{\beta} - \tfrac{1}{3}\delta^{\alpha}{}_{\beta}h$, satisfies

$$S^{\alpha}{}_{\beta}{}'' = -2(a'/a)S^{\alpha}{}_{\beta}{}' + \nabla^2 S^{\alpha}{}_{\beta} + S^{\alpha\mu}{}_{,\beta\mu} + S_{\beta\mu}{}^{,\alpha\mu} - \tfrac{2}{3}\delta^{\alpha}{}_{\beta}S^{\mu\nu}{}_{,\mu\nu} + h^{\alpha 4}{}_{,\beta}{}' + h_{\beta 4}{}^{,\alpha\prime} - \tfrac{1}{3}h_{,\beta}{}^{\alpha} + 2(a'/a)(h^{\alpha 4}{}_{,\beta} + h_{\beta 4}{}^{,\alpha}) - \tfrac{1}{3}\delta^{\alpha}{}_{\beta}[\tfrac{1}{3}\nabla^2 h + 4(a'/a)h^{4\mu}{}_{,\mu} + 2h^{4\mu}{}_{,\mu\prime}]. \quad (15b)$$

Similarly, we solve Eqs. (14b) for the first time derivatives of $h^{4\alpha}$,

$$h^{4\alpha\prime} = (\rho_0 + p_0) - 1(\delta p^{,\alpha} - \overset{0}{p}{}'h^{4\alpha}) - (a'/a)h^{4\alpha}, \quad (15c)$$

and $\delta\rho$,

$$\delta\rho' = -\tfrac{1}{2}(\rho_0 + p_0)h' - 3(a'/a)(\delta\rho + \delta p). \quad (15d)$$

The constraint equations on S^3 are then found by expanding Eqs. (14c):

$$S^{\mu\nu}{}_{,\mu\nu} + \tfrac{2}{3}\nabla^2 h + 2(a'/a)(2h^{4\mu}{}_{,\mu} - h') + 2a^2\delta p = 0, \quad (16a)$$

and

$$\nabla^2 h^{4\alpha} - \tfrac{2}{3}h_{,}{}^{\alpha\prime} + S^{\alpha\mu}{}_{,\mu}{}' + h^{4\mu}{}_{,\mu}{}^{\alpha} - [8(a'^2/a^2) - 4(a''/a)]h^{4\alpha} = 0. \quad (16b)$$

It can readily be checked that Eqs. (10) give a particular solution to the above field equations. It remains to show that in both cases (i.e., radiation and dust) Eqs. (10) form the most general C^∞ solution to (15) and (16) modulo gauge transformations. We shall integrate the equations directly, eliminating some functions of position alone by means of gauge transformations. The details of the procedure are tedious; however, enough of the solution is presented here to make the procedure clear.

A. The case of incoherent matter

With $\overset{0}{p} = \delta p = 0$, the field equations reduce to

$$S^{\alpha}{}_{\beta}{}'' = -(4/\eta)S^{\alpha}{}_{\beta}{}' + \nabla^2 S^{\alpha}{}_{\beta} + S_{\beta\mu}{}^{,\alpha\mu} + S^{\alpha\mu}{}_{,\beta\mu} - \tfrac{2}{3}\delta^{\alpha}{}_{\beta}S^{\mu\nu}{}_{,\mu\nu} + h^{\alpha 4\prime}{}_{,\beta} + h_{\beta 4}{}^{,\alpha\prime} - \tfrac{1}{3}h_{,}{}^{\alpha}{}_{\beta} + (4/\eta)(h^{\alpha 4}{}_{,\beta} + h_{\beta 4}{}^{,\alpha}) - \tfrac{1}{3}\delta^{\alpha}{}_{\beta}(\tfrac{1}{3}\nabla^2 h + (8/\eta)h^{4\mu}{}_{,\mu} + 2h^{4\mu\prime}{}_{,\mu}), \quad (17a)$$

$$h'' = -(4/\eta)h' + 2h^{4\mu}{}_{,\mu}{}' + (8/\eta)h^{4\mu}{}_{,\mu} + \tfrac{1}{2}S^{\mu\nu}{}_{,\mu\nu} + \tfrac{1}{3}\nabla^2 h, \quad (17b)$$

$$h^{4\alpha\prime} = -(2/\eta)h^{4\alpha}, \quad (17c)$$

$$\delta\rho' = -\tfrac{3}{2}(H/\eta^6)h' - (6/\eta)\delta\rho, \quad (17d)$$

in M^4, and

$$S^{\mu\nu}{}_{,\mu\nu} + \tfrac{2}{3}\nabla^2 h + (4/\eta)(2h^{4\mu}{}_{,\mu} - h') + (4\eta^4/H^2)\delta\rho = 0, \quad (18a)$$

$$\nabla^2 h^{4\alpha} + \tfrac{2}{3}h_{,}{}^{\alpha\prime} + S^{\alpha\mu}{}_{,\mu}{}' + h^{4\mu}{}_{,\mu}{}^{\alpha} - (24/\eta^2)h^{4\alpha} = 0, \quad (18b)$$

on S^3. The allowed gauge transformations are

$$\hat{S}_{\alpha\beta} = S_{\alpha\beta} + d_{(\alpha,\beta)} - \tfrac{1}{3}\eta_{\alpha\beta}d^{\mu}{}_{,\mu}, \quad (19a)$$

$$\hat{h} = h + d^{\alpha}{}_{,\alpha} + (6H/\eta^3)b, \quad (19b)$$

$$\hat{h}_{4\alpha} = h_{4\alpha} + (H/2\eta^2)b_{,\alpha}. \quad (19c)$$

Integration of Eq. (17c) immediately yields

$$h^{4\alpha}(\mathbf{x},\eta) = -(2/\eta^2)F^{\alpha}(\mathbf{x}),$$

where the sign and numerical coefficients are chosen for convenience. We can, however, always find a gauge in which $h^{4\alpha}$ has the representation

$$h^{4\alpha}(\mathbf{x},\eta) = -(2/\eta^2)\nabla^2 C^{\alpha}(\mathbf{x}) \quad (20a)$$

with

$$C^{\alpha}{}_{,\alpha} = 0. \quad (20b)$$

For, under the gauge transformations of Eqs. (19c),

$$\hat{h}^{4\alpha}(\mathbf{x},\eta) = h^{4\alpha}(\mathbf{x},\eta) + (H/2\eta^2)b(\mathbf{x})^{,\alpha} = (2/\eta^2)[F^{\alpha}(\mathbf{x}) - (H/4)b(\mathbf{x})^{,\alpha}],$$

and given $F^{\alpha}(\mathbf{x})$, there exists a $b(\mathbf{x})$ and $C^{\alpha}(\mathbf{x})$ such that

$$\nabla^2 C^{\alpha} = F^{\alpha} - (H/4)b^{,\alpha}, \qquad C^{\alpha}{}_{,\alpha} = 0.$$

To see this, choose $q^{\alpha}(\mathbf{x})$ so that $\nabla^2 q^{\alpha} = F^{\alpha}$. This is possible by the above lemma. Then define $e(\mathbf{x})$ by $\nabla^2 e = q^{\alpha}{}_{,\alpha}$. It follows that

$$\nabla^2 e^{,\alpha} + F^{\alpha} = \nabla^2(e^{,\alpha} + q^{\alpha}).$$

Then, taking $C^{\alpha} = q^{\alpha} + e^{,\alpha}$ and $b = -(4/H)\nabla^2 e$ we accomplish the objective. Clearly, in this gauge,

$$h^{4\alpha}{}_{,\alpha} = 0.$$

If we now take the divergence of Eq. (18b), we find

$$(S^{\mu\nu}{}_{,\mu\nu} + \tfrac{2}{3}\nabla^2 h)' = 0.$$

So, defining $B(\mathbf{x})$ by

$$\nabla^2 B = \tfrac{1}{2} S^{\mu\nu}{}_{,\mu\nu} + \tfrac{1}{3}\nabla^2 h,$$

it follows from Eq. (17b) that

$$(\eta^4 h' - \tfrac{1}{5}\eta^5 \nabla^2 B)' = 0.$$

Thus, there exists an $A(\mathbf{x})$ such that

$$\nabla^2 A = \tfrac{1}{3}(\eta^4 h' - \tfrac{1}{5}\eta^5 \nabla^2 B).$$

Integration then gives

$$h(\mathbf{x}, \eta) = \frac{\eta^2}{10}\nabla^2 B(\mathbf{x}) - (1/\eta^3)\nabla^2 A(\mathbf{x}) + G(\mathbf{x}).$$

The gauge transformation which would bring this result into accord with Sachs–Wolfe also alters $S^{\mu\nu}{}_{,\nu}$, so we first integrate Eq. (18b), obtaining

$$S^{\alpha\nu}{}_{,\nu} = (16/\eta^3)\nabla^2 C^\alpha - (2/\eta)\nabla^2\nabla^2 C^\alpha + (\eta^2/15)\nabla^2 B^{,\alpha} - \tfrac{2}{3}(\nabla^2 A^{,\alpha}/\eta^2) + J^\alpha(\mathbf{x}).$$

Now, the remaining gauge freedom allows us to alter $h(\mathbf{x}, \eta)$ by

$$\hat{h}(\mathbf{x}, \eta) = h(\mathbf{x}, \eta) + 2d^\mu{}_{,\mu}(x),$$

while at the same time changing $S^{\mu\nu}{}_{,\nu}$ by

$$\hat{S}^{\mu\nu}{}_{,\nu} = S^{\mu\nu}{}_{,\nu} - \nabla^2 d^\mu + \tfrac{1}{3} d^\nu{}_{,\nu}{}^{,\mu}.$$

It is possible to find a d^μ such that

$$h(\mathbf{x}, \eta) = (\eta^2/10)\nabla^2 B(\mathbf{x}) - (1/\eta^3)\nabla^2 A(\mathbf{x}) + 3B(\mathbf{x}) \tag{21}$$

and at the same time

$$S^{\alpha\nu}{}_{,\nu} = (16/\eta^3)\nabla^2 C^\alpha - (2/\eta)\nabla^2\nabla^2 C^\alpha + (\eta^2/15)\nabla^2 B^{,\alpha} - (2/3)(\nabla^2 A^{,\alpha}/\eta^3) \tag{22}$$

as follows. First write $d^\mu = u^\mu + v^\mu$ where $u^\mu{}_{,\mu} = 0$. Then set

$$v^\mu{}_{,\mu} = 3B - G$$

which always has a solution for v^μ. The lemma assures the existence of a u^μ satisfying

$$\nabla^2 u^\mu = J^\mu + \nabla^2 v^\mu + \tfrac{1}{3}(3B - G)^{,\mu},$$

and this u^μ leads us to the required d^μ. Incidentally, this gauge permits us to solve Eq. (18a) for $\delta\rho$, giving

$$\delta\rho = \frac{H^2}{4}\nabla^2\left(\frac{6A}{\eta^9} - \frac{3}{5}\frac{B}{\eta^4}\right),$$

and a check with Eq. (17d) shows this result to be consistant.

Finally, we must find the trace-free part of the metric as a solution to

$$S^\alpha{}_\beta{}'' + \frac{4}{\eta}S^\alpha{}_\beta{}' - \nabla^2 S^\alpha{}_\beta = 2\left(\frac{6}{\eta}\nabla^2 - \frac{1}{\eta}\nabla^2\nabla^2\right)(C^\alpha{}_{,\beta} - C_\beta{}^{,\alpha}) - \frac{1}{\eta^3}\nabla^2 A_{,\beta}{}^\alpha - \frac{1}{\eta^3}\nabla^2\nabla^2 A\delta^\alpha{}_\beta + \frac{\eta^2}{10}\nabla^2 B^{,\alpha}{}_\beta + \frac{\eta^2}{30}\nabla^2\nabla^2 B\delta^\alpha{}_\beta - B^{,\alpha}{}_\beta - \tfrac{1}{3}\delta^\alpha{}_\beta\nabla^2 B.$$

We may easily construct, from Sachs–Wolfe, a particular solution $S^\alpha{}_\beta$, compatible with Eq. (22) and the general solution is then given by

$$S^\alpha{}_\beta = \overset{S}{0}{}^\alpha{}_\beta + \overset{S}{1}{}^\alpha{}_\beta,$$

where $\overset{S}{0}{}^\alpha{}_\beta$ satisfies

$$\overset{S}{0}{}^\alpha{}_\beta{}'' + \frac{4}{\eta}\overset{S}{0}{}^\alpha{}_\beta{}' - \nabla^2\overset{S}{0}{}^\alpha{}_\beta = 0,$$

with

$$\overset{S}{0}{}^\alpha{}_{\beta,\alpha} = 0.$$

The general solution to this homogeneous equation can be found by first writing

$$\overset{S}{0}{}^{\alpha\beta} = \frac{1}{\eta}\frac{\partial}{\partial\eta}\left(\frac{D^{\alpha\beta}}{\eta}\right),$$

so that $D^{\alpha\beta}$ satisfies

$$\frac{1}{\eta}\frac{\partial}{\partial\eta}\left(\frac{\Box D^{\alpha\beta}}{\eta}\right) = 0,$$

where

$$\Box := \frac{\partial^2}{\partial\eta^2} - \nabla^2.$$

We must therefore have

$$\Box D^{\alpha\beta}(\mathbf{x}, \eta) = \eta Q^{\alpha\beta}(\mathbf{x})$$

for arbitrary spatial function $Q^{\alpha\beta}$. However, the defining relation for $D^{\alpha\beta}$ permits changes of the form

$$\tilde{D}^{\alpha\beta}(\mathbf{x}, \eta) = D^{\alpha\beta}(\mathbf{x}, \eta) + \eta K^{\alpha\beta}(\mathbf{x}),$$

and under such a change

$$\Box\tilde{D}^{\alpha\beta} = \Box D^{\alpha\beta} + \eta\nabla^2 K^{\alpha\beta}.$$

Thus, by choosing $K^{\alpha\beta}$ to satisfy

$$\nabla^2 K^{\alpha\beta} = Q^{\alpha\beta},$$

we may always demand that $D^{\alpha\beta}(\mathbf{x}, \eta)$ be a solution of

$$\Box D^{\alpha\beta} = 0, \tag{23a}$$

with

$$D^{\alpha\beta}{}_{,\alpha} = 0. \tag{23b}$$

Assembling the parts to the solution for $S^{\alpha\beta}$ we have

$$S^{\alpha\beta} = \frac{1}{\eta}\frac{\partial}{\partial\eta}\left(\frac{D^{\alpha\beta}}{\eta}\right) - 2\left(\frac{8}{\eta^3} - \frac{\nabla^2}{\eta}\right)(C^{\alpha,\beta} + C^{\beta,\alpha}) + \frac{A}{\eta^3}^{,\alpha\beta} + \left(\eta^{\alpha\beta}B - \frac{\eta^2}{10}B^{,\alpha\beta}\right) - \tfrac{1}{3}\eta^{\alpha\beta}\left(\frac{\eta^2}{10}\nabla^2 B + 3B - \frac{\nabla^2 A}{\eta^3}\right) \tag{24}$$

which brings the entire pressure-free solution into accord with Sachs–Wolfe.

B. The case of radiation

If $\overset{0}{p} = \frac{1}{3}\rho_0$ and $\delta p = \frac{1}{3}\delta\rho$ then the perturbed field equations become

$$S^\alpha{}_\beta{}'' = -(2/\eta)S^\alpha{}_\beta{}' + \nabla^2 S^\alpha{}_\beta + S_{\beta\mu}{}^{,\alpha\mu} + S^{\alpha\mu}{}_{,\beta\mu} - \tfrac{2}{3}\delta^\alpha{}_\beta S^{\mu\nu}{}_{,\mu\nu} + h^{\alpha 4}{}_{,\beta}{}' + h_{\beta 4}{}^{,\alpha'} - \tfrac{1}{3}h^{,\alpha}{}_\beta + (2/\eta)(h^{\alpha 4}{}_{,\beta} + h_{\beta 4}{}^{,\alpha}) - \tfrac{1}{3}\delta^\alpha{}_\beta[\tfrac{1}{3}\nabla^2 h + (4/\eta)h^{4\mu}{}_{,\mu} + 2h^{4\mu}{}_{,\mu}{}'], \tag{25a}$$

$$h'' = -(2/\eta)h' + 2h^{4\mu}{}_{,\mu}{}' + (4/\eta)h^{4\mu}{}_{,\mu} - (\eta^2/H^2)\delta\rho + \tfrac{1}{2}S^{\mu\nu}{}_{,\mu\nu} + \tfrac{1}{3}\nabla^2 h, \tag{25b}$$

$$h^{4\alpha'} = (\eta^4/12H^2)\delta\rho{}^{,\alpha}, \tag{25c}$$

$$\delta\rho' = -(2H^2/\eta^4)h' - (4/\eta)\delta\rho \tag{25d}$$

in M^4, together with

$$S^{\mu\nu}{}_{,\mu\nu} + \tfrac{2}{3}\nabla^2 h + (2/\eta)(2h^{4\mu}{}_{,\mu} - h') + (2\eta^2/H^2)\delta\rho = 0, \tag{26a}$$

$$\nabla^2 h^{4\alpha} - \tfrac{2}{3}h,{}^{\alpha'} + S^{\alpha\mu}{}_{,\mu}{}' + h^{4\mu}{}_{,\mu}{}^{\alpha} - (8/\eta^2)h^{4\alpha} = 0, \tag{26b}$$

on S^3. Furthermore, the permissible gauge transformations are of the form

$$\hat{S}^{\alpha\beta} = S^{\alpha\beta} + d^{(\alpha,\beta)} - \tfrac{1}{3}\eta^{\alpha\beta}d^\mu{}_{,\mu}, \tag{27a}$$

$$\hat{h} = h + d^\mu{}_{,\mu} + (6H/\eta^2)b, \tag{27b}$$

$$\hat{h}^{\alpha 4} = h^{\alpha 4} + (H/\eta)b{}^{,\alpha}. \tag{27c}$$

Integration of Eq. (25d) implies that

$$\eta^4\delta\rho + 2H^2 h = L(\mathbf{x}).$$

However, by integrating

$$d^\alpha{}_{,\alpha} = L/2H, \tag{28}$$

we may always find a gauge in which

$$\delta\rho(\mathbf{x},\eta) = -(2H^2/\eta^4)h(\mathbf{x},\eta). \tag{29}$$

Now, guided solely by Sachs–Wolfe, we represent $h(\mathbf{x},\eta)$ by

$$h(\mathbf{x},\eta) = -\tfrac{3}{2}\frac{\partial}{\partial\eta}\left[\eta^2\frac{\partial}{\partial\eta}\left(\frac{1}{\eta^2}\frac{\partial E(\mathbf{x},\eta)}{\partial\eta}\right)\right].$$

Clearly $E(\mathbf{x},\eta)$ is not unique, and $h(\mathbf{x},\eta)$ is not affected by changes of the form

$$\tilde{E}(\mathbf{x},\eta) = E(\mathbf{x},\eta) + \eta^3\alpha(\mathbf{x}) + \eta^2\beta(\mathbf{x}) + \gamma(\mathbf{x}).$$

This nonuniqueness will be used, like the remaining gauge freedom, to eliminate some of the arbitrary spatial functions arising from integration of the other field equations.

From Eq. (25c) we find that

$$h^{4\alpha}(\mathbf{x},\eta) = \frac{\eta^2}{4}\frac{\partial}{\partial\eta}\left(\frac{1}{\eta^2}\frac{\partial E{}^{,\alpha}(x,\eta)}{\partial\eta}\right) + M^\alpha(\mathbf{x}). \tag{30}$$

However, the nonuniqueness of E leaves the function M^α ambiguous up to a gradient

$$\tilde{h}^{4\alpha}(\mathbf{x},\eta) = \frac{\eta^2}{4}\frac{\partial}{\partial\eta}\left(\frac{1}{\eta^2}\frac{\partial E{}^{,\alpha}(\mathbf{x},\eta)}{\partial\eta}\right) + \tfrac{1}{2}\beta{}^{,\alpha}(\mathbf{x}) + M^\alpha(\mathbf{x}).$$

Then, as in the case of incoherent matter, we may find a $C^\alpha(\mathbf{x})$ and a $\beta(\mathbf{x})$, such that

$$\nabla^2 C^\alpha = -M^\alpha - \tfrac{1}{2}\beta{}^{,\alpha}, \qquad C^\alpha{}_{,\alpha} = 0.$$

Thus,

$$h^{4\alpha}(\mathbf{x},\eta) = \frac{\eta^2}{4}\frac{\partial}{\partial\eta}\left(\frac{1}{\eta^2}\frac{\partial E{}^{,\alpha}(\mathbf{x},\eta)}{\partial\eta}\right) - \nabla^2 C^\alpha(\mathbf{x}). \tag{31}$$

Then, rewriting Eq. (25b) in the form

$$\eta\frac{\partial}{\partial\eta}\left(\frac{1}{\eta}\frac{\partial^2}{\partial\eta^2}\right)(\nabla^2 E - 3E'') = 0,$$

we see that

$$\nabla^2 E - 3E'' = \eta^3 N(\mathbf{x}) + \eta P(\mathbf{x}) + R(\mathbf{x}).$$

Moreover, the remaining gauge freedom, and nonuniqueness of E still permit changes of the form

$$\hat{\tilde{E}} = E + \eta^3\alpha(\mathbf{x}) - 2H\eta b(\mathbf{x}) + \gamma(\mathbf{x}).$$

Thus, by choosing α, γ, and b to satisfy

$$\nabla^2\alpha = -N, \qquad \nabla^2\gamma = -R,$$

and

$$\nabla^2 b = (1/2H)(P + 6\alpha)$$

we can always require $E(\mathbf{x},\eta)$ to be a solution of

$$\nabla^2 E - 3E'' = 0. \tag{32}$$

From Eq. (26b) we determine that

$$S^{\mu\nu}{}_{,\nu}(\mathbf{x},\eta) = \tfrac{2}{3}h{}^{,\mu}(\mathbf{x},\eta) + \eta\nabla^2\nabla^2 C^\mu(\mathbf{x}) + \frac{8}{\eta}\nabla^2 C^\mu(\mathbf{x}) + \frac{2}{\eta^2}\frac{\partial E{}^{,\mu}(\mathbf{x},\eta)}{\partial\eta} + I^\mu(\mathbf{x}), \tag{33}$$

where, by Eq. (26a), $I^\mu(\mathbf{x})$ is divergence free. Since the choice of d^μ in Eq. (28) is fixed only up to a divergence free vector $e^\mu(\mathbf{x})$, we may eliminate I^μ by choosing e^μ to satisfy

$$\nabla^2 e^\alpha = I^\alpha.$$

In this gauge, the trace-free part of the metric satisfies

$$S^\alpha{}_\beta{}'' + (2/\eta)S^\alpha{}_\beta{}' - \nabla^2 S^\alpha{}_\beta = E{}_{,}{}^\alpha{}_\beta{}''' + (3/\eta)E{}_{,}{}^\alpha{}_\beta{}'' + (\eta\nabla^2\nabla^2 + (6/\eta)\nabla^2)(C^\alpha{}_{,\beta} + C_{\beta,}{}^\alpha) + \tfrac{1}{3}\delta^\alpha{}_\beta[(3/\eta)\nabla^2 E'' - \nabla^2 E''{}'].$$

Again we may use Sachs–Wolfe to construct a particular solution $\overset{1}{S}{}^\alpha{}_\beta$ compatable with Eq. (33), so that the general solution is given by

$$S^\alpha{}_\beta = \overset{0}{S}{}^\alpha{}_\beta + \overset{1}{S}{}^\alpha{}_\beta,$$

where

$$\overset{0}{S}{}^\alpha{}_\beta{}'' + (2/\eta)\overset{0}{S}{}^\alpha{}_\beta{}' - \nabla^2\overset{0}{S}{}^\alpha{}_\beta = 0,$$

with

$$\overset{0}{S}{}^\alpha{}_{\beta,\alpha} = 0.$$

Writing

$$\overset{0}{S}{}^\alpha{}_\beta = (1/\eta)D^{\alpha\beta},$$

we see that $D^{\alpha\beta}$ must be a solution of D'Alembert's equation

$$\Box D^{\alpha\beta} = 0. \tag{34}$$

Reconstructing the trace-free part of the metric, we have

$$S^\alpha{}_\beta = (1/\eta) D^\alpha{}_\beta - [\eta\nabla^2 + (8/\eta)](C^\alpha{}_{,\beta} + C_{\beta,}{}^\alpha) + \tfrac{1}{2} E_{,}{}^\alpha{}_\beta{}' - (1/\eta) E^\alpha{}_{,\beta} - \tfrac{1}{3}\delta^\alpha{}_\beta[(1/\eta)\nabla^2 E - \tfrac{1}{2}\nabla^2 E'], \tag{35}$$

in agreement with Sachs–Wolfe.

IV. SUMMARY

The general C^∞ solution which we have constructed for the perturbations of a Robertson–Walker, $k = 0$, universe containing either dust or radiation is precisely that first obtained by Sachs and Wolfe. However, we have shown that the admissible, metric determining functions $h_{\alpha\beta}$ need not be restricted by moment conditions. Rather, these functions, first reported by Sachs–Wolfe, form the most general C^∞ solution if the potentials are C^∞. Note, in this connection, that the illustrative solution given in Eq. (12) is indeed a C^∞ solution. We should stress that our success with this simplified approach was aided immensely by an *a priori* knowledge of the Sachs–Wolfe results. We refer the interested reader to their paper for an interesting discussion of the physical interpretation of the potentials appearing in the solution. Finally, attempts to construct the general solutions the the perturbation problems for Robertson–Walker universes with $k \neq 0$ have been plagued by difficulties in constructing complete sets of orthonormal functions on the 3-surfaces of constant density. It may be hoped that the simplified approach presented here will facilitate solutions to these problems by working entirely in configuration space.

ACKNOWLEDGMENT

This work would not have been possible without many helpful suggestions from Professor Jurgen Ehlers of the Center for Relativity Theory at the University of Texas at Austin.

*Supported in part by the National Science Foundation (Grant Nos. GU-1598 and GZ974.

[1]R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
[2]J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1964).
[3]A. Friedmam, *Generalized Functions and Partial Differential Equations* (Prentice-Hall, Englewood Cliffs, N. J., 1963), p. 320.

Correlation inequalities and phase transition in the generalized X-Y model*

Masuo Suzuki

The Institute for Solid State Physics, The University of Tokyo, Roppongi, Minato-ku, Tokyo, Japan
(Received 18 October 1971)

Correlation inequalities $<\sigma_A^z> \geq 0$, $<\sigma_A^z\sigma_B^z> \geq <\sigma_A^z><\sigma_B^z>$, $\partial<\sigma_A^z>/\partial J_B^z \geq 0$, and $\partial<\sigma_A^z>/\partial J_B^x \leq 0$ are proved for the generalized X-Y model with the Hamiltonian of the form $\mathcal{H} = -\Sigma(J_A^z\sigma_A^z + J_A^x\sigma_A^x)$, where $\sigma_A^z = \Pi_{j\in A}\sigma_j^z$, $\sigma_A^x = \Pi_{j\in A}\sigma_j^x$, $J_A^z \geq 0$, $J_A^x \geq 0$, A denotes an arbitrary subset of the N lattice points, and σ_j^x, σ_j^z are the Pauli matrices. This yields a simple extension of the Griffiths–Kelly–Sherman inequalities to the above quantal system. Applications to phase transitions are also discussed briefly.

1. INTRODUCTION

We consider the following Hamiltonian

$$\mathcal{H} = -\sum_A (J_A^z\sigma_A^z + J_A^x\sigma_A^x), \tag{1.1}$$

where $J_A^z \geq 0, J_A^x \geq 0$, and

$$\sigma_A^z = \prod_{j\in A}\sigma_j^z \quad \text{and} \quad \sigma_A^x = \prod_{j\in A}\sigma_j^x. \tag{1.2}$$

This includes as an example the ordinary X-Y model, the Hamiltonian of which is given by

$$\mathcal{H}_{XY} = -\sum_{i<j}(J_{ij}^z\sigma_i^z\sigma_j^z + J_{ij}^x\sigma_i^x\sigma_j^x) - \sum_j (H_j^z\sigma_j^z + H_j^x\sigma_j^x) \tag{1.3}$$

with $J_{ij}^z \geq 0$, $J_{ij}^x \geq 0$, $H_j^z \geq 0$, and $H_j^x \geq 0$.

In this paper we prove that

$$\langle\sigma_A^z\rangle \geq 0, \quad \langle\sigma_A^z\sigma_B^z\rangle \geq \langle\sigma_A^z\rangle\langle\sigma_B^z\rangle, \tag{1.4}$$

$$\frac{\partial\langle\sigma_A^z\rangle}{\partial J_B^z} \geq 0, \quad \text{and} \quad \frac{\partial\langle\sigma_A^z\rangle}{\partial J_B^x} \leq 0, \tag{1.5}$$

for the Hamiltonian (1.1) under the conditions that $J_A^z \geq 0$ and $J_A^x \geq 0$. Recently Gallavotti[1] has obtained the above inequalities (1.4) and (1.5) for the ordinary X-Y model (1.3) in the absence of magnetic fields (i.e., for $H_j^z = 0$ and $H_j^x = 0$), by using the reduction formula of the quantal partition function to a classical one derived by Suzuki and Fisher,[2] and consequently by applying the Griffiths–Kelly–Sherman inequalities[3] to a classical system thus reduced.

2. PROOF OF CORRELATION INEQUALITIES

Here, a slightly different approach is made to give a simple proof to a more general case (1.1). Following Ginibre,[4] we reduce the partition function of the Hamiltonian (1.1) to that of the Ising model with many-spin interaction and with *higher dimensions*, by using Trotter's formula:

$$\begin{aligned}\exp(-\beta\mathcal{H}) &= \lim_{n\to\infty}\{\exp(-n^{-1}\beta\mathcal{H}_0)\exp(-n^{-1}\beta\mathcal{H}_1)\}^n \\ &= \lim_{n\to\infty}\{\exp(-n^{-1}\beta\mathcal{H}_0)(1-n^{-1}\beta\mathcal{H}_1)\}^n,\end{aligned} \tag{2.1}$$

where $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$. This is easily extended to the following form

$$\exp(-\beta\mathcal{H}) = \lim_{n\to\infty}\{\exp(-n^{-1}\beta\mathcal{H}_0)\prod_{v=1}^{V}(1-n^{-1}\beta\mathcal{H}_v)\}^n \tag{2.2}$$

for a bounded operator $\mathcal{H}$ of the form $\mathcal{H} = \sum_{v=0}^{V}\mathcal{H}_v$ (for detail, see Ref. 2.)

From (1.1) and (2.2) we then obtain

$$Z_N = \lim_{n\to\infty}\Phi_{N,n}, \tag{2.3}$$

where

$$\begin{aligned}\Phi_{N,n} = \sum_{\sigma_1=\pm1}\cdots\sum_{\sigma_N=\pm1}&\langle\sigma_1,\sigma_2,\cdots,\sigma_N| \\ &\times\{\exp(-n^{-1}\beta\mathcal{H}_0)\prod_A(1+\epsilon_A\sigma_A^x)\}^n|\sigma_1,\sigma_2,\cdots,\sigma_N\rangle,\end{aligned} \tag{2.4}$$

with $\epsilon_A = n^{-1}\beta J_A^x$ and $\mathcal{H}_0 = -\sum J_A^z\sigma_A^z$ is diagonal in the above representation $|\sigma_1,\sigma_2,\cdots,\sigma_N\rangle$.

Consider now the matrix elements of $(1+\epsilon_A\sigma_A^x)$:

$$\exp f(\sigma_A,\sigma_A') \equiv \langle\{\sigma_A\}|(1+\epsilon_A\sigma_A^x)|\{\sigma_A'\}\rangle, \tag{2.5}$$

where $\{\sigma_A\}$ denotes a set $(\{\sigma_j\}; j\in A)$. It is easily seen that if the matrix elements (2.5) represent "*ferromagnetic* partial Boltzmann factors" for all A, then the "partition function" $\Phi_{N,n}$ (and consequently Z_N) is equivalent to that of the Ising model with "*ferromagnetic*" many-spin interaction in the sense that

$$\mathcal{H}_{\text{Is}}^{\text{ferro}} = -\sum J_A\sigma_A, \quad J_A \geq 0. \tag{2.6}$$

For more details, see Ref. 2. Then, the key point of the present argument is to find out the explicit expression of the function $f(\sigma_A,\sigma_A')$ in (2.5) and to show that $-f(\sigma_A,\sigma_A')$ has the form (2.6). Without loss of generality, we assume that σ_A^x takes the form $\sigma_A^x = \sigma_1^x\sigma_2^x\cdots\sigma_m^x$ (m = arbitrary positive integer). This is expressed as a direct product

$$\sigma_A^x = \begin{pmatrix}0&1\\1&0\end{pmatrix}_1 \dot\times \begin{pmatrix}0&1\\1&0\end{pmatrix}_2 \dot\times\cdots\dot\times \begin{pmatrix}0&1\\1&0\end{pmatrix}_m. \tag{2.7}$$

The matrix elements of the σ_A^x are given by

$$\langle\{\sigma_A\}|\sigma_A^x|\{\sigma_A'\}\rangle = \prod_{j=1}^{m}\tfrac{1}{4}(\sigma_j-\sigma_j')^2 = \prod_{j=1}^{m}\tfrac{1}{2}(1-\sigma_j\sigma_j'). \tag{2.8}$$

From the property (2.8) it is easily shown that

$$\begin{aligned}&\langle\{\sigma_A\}|1+\epsilon_A\sigma_A^x|\{\sigma_A'\}\rangle \\ &\quad= \lim_{\lambda\to\infty}\exp\left[-\lambda\sum_{j=1}^{m}(\sigma_j-\sigma_j')^2\sum_{k=1}^{m}\left(1-\frac{(\sigma_k-\sigma_k')^2}{4}\right)\right] \\ &\qquad\times\exp\left(\frac{1}{4m}(\log\epsilon_A^{-1})\sum_{j=1}^{m}(\sigma_j-\sigma_j')^2\right) \\ &\quad= \exp\left[\frac{1}{2m}\log\left(\frac{nk_BT}{J_A^x}\right)\sum_{j=1}^{m}(\sigma_j\sigma_j'-1)\right. \\ &\qquad\left.+\infty\sum_{j,k=1}^{m}(\sigma_j\sigma_k\sigma_j'\sigma_k'-1)\right].\end{aligned} \tag{2.9}$$

Copyright © 1973 by the American Institute of Physics

For a sufficiently large n, we have $\log(nk_BT/J_A^x) > 0$. Thus, it is proved that the partition function $\Phi_{N,n}$ is equivalent to that of the Ising model *"ferromagnetic"* two-spin and four-spin interactions. Furthermore, the effective ferromagnetic interaction $(2m)^{-1}\log(nk_BT/J_A^x)$ in (2.9) is a *decreasing* function of J_A^x. Therefore, by applying the Griffiths–Kelly–Sherman inequalities for the Ising model with interaction of the form (2.6) to the above reduced effective Ising model with *"ferromagnetic"* many-spin interaction, we arrive immediately at the correlation inequalities (1.4) and (1.5) for all nonzero temperatures.

In particular, for the ordinary X–Y model (1.3), we have the inequalities

$$\frac{\partial}{\partial H_j^z}\langle\sigma_A^z\rangle \geq 0 \quad \text{and} \quad \frac{\partial}{\partial H_j^x}\langle\sigma_A^z\rangle \leq 0, \tag{2.10}$$

in addition to the results obtained by Gallavotti.[1]

3. APPLICATIONS TO PHASE TRANSITIONS

In this section we consider as an example the one-dimensional X–Y model, the Hamiltonian of which is described by

$$\mathcal{H} = -\sum_j \sum_{n\geq 1} [J^z(n)\sigma_j^z\sigma_{j+n}^z + J^x(n)\sigma_j^x\sigma_{j+n}^x], \tag{3.1}$$

where

$$J^z(n) \geq 0, \quad J^x(n) \geq 0, \quad \sum_{n\geq 1} J^z(n) < \infty, \quad \text{and} \quad \sum_{n\geq 1} J^x(n) < \infty. \tag{3.2}$$

Hereafter without loss of generality we assume that $J^z(n) \geq J^x(n)$.

(a) If the interaction $J^z(n)$ satisfies the condition

$$\lim_{N\to\infty} \{\log(\log N)\}^{-1} \sum_{n=1}^{N} nJ^z(n) = 0, \tag{3.3}$$

then there is no long-range order with respect to σ^z — σ^z and σ^x — σ^x correlations (i.e., $M_s^z \equiv 0$ and $M_s^x \equiv 0$) for all nonzero temperatures. This is easily derived from Dyson's results[5] on the Ising model [i.e., (3.1) with $J^x(n) \equiv 0$], by utilizing the correlation inequalities

$$\frac{\partial}{\partial J^x(n)}\langle\sigma_j^z\sigma_k^z\rangle \leq 0 \quad \text{etc.}, \tag{3.4}$$

proved in the previous section.

(b) When the interaction $J^z(n)$ takes the form $J^z(n) = n^{-s}$ $(1 < s < 2)$, Dyson[5] has proved the existence of the phase transition for the Hamiltonian (3.1) with $J^x(n) \equiv 0$. Now, we define the critical temperatures $T_c^{(Is)}$ and $T_c^{(XY)}$ by singular points of the susceptibilities χ_{Is}^{zz} $[\propto \sum_{j,k}\langle\sigma_j^z\sigma_k^z\rangle$ for $J^x(n) \equiv 0]$ and χ_{XY}^{zz} [which is defined by the canonical correlation $\sum_{j,k}(\sigma_j^z, \sigma_k^z)$ for $J^x(n) \neq 0$], respectively. Then we obtain the inequality

$$T_c^{(XY)} \leq T_c^{(Is)}, \tag{3.5}$$

by applying the correlation inequalities

$$\langle\sigma_j^z\sigma_k^z\rangle_{XY} \leq \langle\sigma_j^z\sigma_k^z\rangle_{Is} \tag{3.6}$$

to the susceptibility χ_{Is}^{zz} and to the asymptotic relation[6]

$$\chi_{XY}^{zz} \sim \sum_{j,k}\langle\sigma_j^z\sigma_k^z\rangle_{XY} \tag{3.7}$$

near the critical point. The above relation (3.7) is confirmed from the following inequalities concerning the canonical and direct correlations[6,7]:

$$(1 - e^{-\beta\bar{\omega}})/\beta\bar{\omega} \leq (\delta Q, \delta Q)/\langle(\delta Q)^2\rangle \leq 1 \tag{3.8}$$

for an Hermitian quantity Q, where $\bar{\omega}$ is the first moment defined by

$$\bar{\omega} = \tfrac{1}{4}\langle[\delta Q, [\mathcal{H}, \delta Q]]\rangle/\langle(\delta Q)^2\rangle, \tag{3.9}$$

being finite at T_c in usual situations.

It should be remarked that the above inequality (3.5) is, in general, valid for the ordinary X–Y model (1.3) with $H_j^z = H_j^x = 0$. In more general, we have

$$\partial T_c^{(XY)}/\partial J_{ij}^z \geq 0 \quad \text{and} \quad \partial T_c^{(XY)}/\partial J_{ij}^x \leq 0, \tag{3.10}$$

for $J_{ij}^z > J_{ij}^x \geq 0$. The numerical results on the critical temperature $T_c^{(XY)}$ obtained by Betts *et al.*[8] are consistent with the above general relation (3.5).

These results on the correlation inequalities (1.4)–(1.5), and the relations (3.10) on the critical temperature may also be applied to discussions on the critical behavior of the quantum lattice gas model originally proposed by Matsubara and Matsuda.[9]

4. DISCUSSIONS

The physical reason why the Griffiths type inequalities (1.4) and (1.5) are valid for the Hamiltonian (1.1) is that the off-diagonal interaction $\mathcal{H}_1 = -\sum_A J_A^x\sigma_A^x$ $(J_A^x \geq 0)$ plays a role to decrease the ferromagnetic correlation among σ_j^z-spins, but that it has not enough of a cooperative effect to cause an "antiferromagnetic" correlation. Namely, one can say that $\mathcal{H}_1$ is a dynamical random force acting on z–z correlations.

It seems impossible to extend our arguments to higher spin problems.

The present results on spin correlation functions may be applied to the discussions on critical exponents in a separate paper.[10]

ACKNOWLEDGMENTS

The author should like to express his sincere thanks to Professor M. E. Fisher for supplying the initial stimulus. The support of the National Science Foundation (U.S.A.) is gratefully acknowledged.

---

*This was mainly performed on leave at Cornell University, Ithaca, New York.

[1] G. Gallavotti, Stud. Appl. Math. **50**, 89 (1971).
[2] M. Suzuki and M. E. Fisher, J. Math. Phys. **12**, 235 (1971).
[3] R. B. Griffiths, J. Math. Phys. **8**, 478 (1967); J. Math. Phys. **8**, 484 (1967). D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968). See also R. B. Griffiths, in *Phase Transitions and Critical Points*, edited by C. Domb and M. S. Green (Academic, London, 1972).
[4] J. Ginibre, Lectures at the Cargese Summer School, 1969.
[5] F. J. Dyson, Commun. Math. Phys. **12**, 91 (1969); Commun. Math. Phys. **12**, 212 (1969). See also D. Ruelle, Commun. Math. Phys. **9**, 267 (1968).
[6] H. Falk and L. W. Bruch, Phys. Rev. **180**, 442 (1969). See also M. Suzuki, Prog. Theor. Phys. **46** 1054, (1971).
[7] J. M. Luttinger, Prog. Theor. Phys. Suppl. **37**, 35 (1966). B. D. Josephson, Proc. Phys. Soc. Lond. **92**, 269 (1967). A. B. Harris, J. Math. Phys. **8**, 1044 (1967).
[8] D. D. Betts, C. J. Elliott, and M. H. Lee, Phys. Lett. A **29**, 150 (1969); Can J. Phys. **48**, 1566 (1970).
[9] T. Matsubara and H. Matsuda, Prog. Theor. Phys. **16**, 569 (1956). R. Whitlock and P. Zilsel, Phys. Rev. **131**, 2409 (1963). R. H. Parmenter, Phys. Rev. **170**, 194 (1968).
[10] M. Suzuki, to be submitted to Prog. Theor. Phys.

The transformation from a harmonic single-particle basis to the self-consistent harmonic approximation*

Laurent G. Caron

Département de physique, Université de Sherbrooke, Sherbrooke, Québec, Canada
(Received 28 November 1972)

The transformation rules between a harmonic single-particle basis and the self-consistent harmonic approximation in quantum crystal theory are deduced. In the process, the thermal renormalization of the phonon spectrum is calculated to all orders in the phonon–phonon interaction. The generalization of the transformation technique to a Hartree single-particle basis is speculated upon.

I. INTRODUCTION

In a recent paper,[1] a low temperature formalism for quantum crystals in the single-particle picture was developed. This formalism required mapping the single-particle space into a quasiparticle, boson space. This was done, however, under the implicit guideline of the random phase approximation (RPA) in that the excitation spectrum was the same as Werthamer's.[2] For a Bravais crystal, in addition to the three acoustic phonon branches the RPA spectrum exhibits some higher energy ones even in the limit of a harmonic interatomic interaction. These single-particle type excitations prove to be the unsatisfactory outcome of the method.

It is our feeling that any decent treatment of atomic motion originating from the single-particle picture should be able to correctly reproduce the harmonic limit. Consequently, we shall endeavour in this paper to formulate a modified mapping procedure for a harmonic single-particle basis which does culminate in the equivalent phonon basis. For anharmonic systems, this leads to the self-consistent harmonic approximation.[3] Although Brenig[4] has shown how the harmonic phonon spectrum can be unraveled from a small subset of harmonic single-particle wave-functions the proper treatment of the complete set has not yet been achieved.

The one-dimensional harmonic oscillator is first examined. The insight gained from this simple analysis paves the way for a more complicated problem namely the harmonic crystal. Subsequently, the anharmonic crystal is studied and the full transformation procedure is developed. It is seen that thermal renormalization due to phonon–phonon interaction can be accounted for quite naturally. We end by speculating on the approach to a Hartree self-consistent basis using the knowledge gained from the treatment of the harmonic basis and from the RPA.

For the sake of simplicity the analysis will be concentrated on a Bravais crystal of orthorhombic or higher symmetry. The interatomic interaction will be of the two-body type.

II. ONE-DIMENSIONAL HARMONIC OSCILLATOR

Any transformation procedure used to obtain the excitation spectrum of atoms starting from the single-particle picture should be able to reproduce the results of the one-dimensional harmonic oscillator. This aspect has not been dwelled upon in Ref. 1. Had this been done it would have been easily observed that the proposed transformation rules did lead to spurious excitations. It then seems more logical to start from this simple case and proceed from there.

The one-dimensional harmonic oscillator Hamiltonian is

$$\mathcal{H} = -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial X^2} + \tfrac{1}{2}M\omega^2 X^2, \tag{1}$$

where $\hbar$ = Planck's constant $/2\pi$, M = oscillator mass, ω = oscillator frequency $\times 2\pi$.

This Hamiltonian is normally quantized[5] using the quasiparticle, boson operators $b^\dagger, b$ which create or annihilate an excitation. The end result is

$$H = \hbar\omega(b^\dagger b + \tfrac{1}{2}). \tag{2}$$

The immediate aim is then to be able to reproduce Eq. (2) starting from the single-particle representation.

The configuration space eigenfunctions of the Hamiltonian, Eq. (1), are

$$\phi_n(X) = \left(\frac{\Gamma}{\pi}\right)^{1/4}\left(\frac{1}{2^n n!}\right)^{1/2} H_n(\Gamma^{1/2}X)\, e^{-\Gamma X^2/2}, \qquad n = 0, 1, 2, \ldots, \tag{3}$$

where

$$\Gamma / = M\omega/\hbar \tag{4}$$

and H_n is the Hermite polynomial of order n. The eigenvalues are

$$E_n = \hbar\omega(n + \tfrac{1}{2}). \tag{5}$$

The Hamiltonian can then be second quantized in the single-particle picture using the complete set of eigenfunctions. The occupation space Hamiltonian is

$$H = \sum_m \sum_n \int dX \phi_n^*(X)\mathcal{H}\phi_m(X) a_n^\dagger a_m = \sum_n \hbar\omega(n + \tfrac{1}{2}) a_n^\dagger a_n, \tag{6}$$

where the $a_n^\dagger, a_n$ create or annihilate the oscillator in the state with quantum number n. The oscillator spin index is unnecessary since the spin is conserved in Eq. (1). These operators commute or anticommute for integer or half-integer spin values respectively.

In the same spirit as Ref. 1 but under the constraint that there be a single type of quasiparticle operator as in Eq. (2), let us postulate a correspondence between the set of oscillator eigenvectors

$$|\phi_n\rangle = \text{oscillator in eigenfunction } \phi_n(X), \tag{7}$$

and the complete set of boson eigenvectors

$$|n\rangle = \text{state with } n \text{ bosons}. \tag{8}$$

Associated with this boson vector space are the opera-

Copyright © 1973 by the American Institute of Physics

tors $b^\dagger, b$ which create or annihilate a boson. They obey the following rules

$$\begin{aligned} &[b, b] = [b^\dagger, b^\dagger] = 0, \\ &[b, b^\dagger] = 1, \\ &b\,|n\rangle = n^{1/2}\,|n-1\rangle, \\ &b^\dagger\,|n\rangle = (n+1)^{1/2}\,|n\rangle. \end{aligned} \tag{9}$$

The single occupancy of the oscillator states suggests the following equivalence relationship:

$$a_n^\dagger a_m \longleftrightarrow \frac{(b^\dagger)^n X(0)(b)^m}{(n!m!)^{1/2}}, \tag{10}$$

where

$$X(0) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} (b^\dagger)^r (b)^r \tag{11}$$

projects out the boson vacuum state $|0\rangle$.

From the operator rules, Eq. (9), and because of the single occupancy of the oscillator, it can be verified that the following operation

$$a_n^\dagger a_m\,|\phi_q\rangle = |\phi_n\rangle\,\delta_{mq} \tag{12}$$

is equivalent to

$$\frac{(b^\dagger)^n X(0)(b)^m}{(n!m!)^{1/2}}\,|q\rangle = |n\rangle\,\delta_{mq} \tag{13}$$

thus asserting the validity of Eq. (10).

The Hamiltonian in the single-particle representation, Eq. (6), is transformed with the use of the mapping rule, Eq. (10), into

$$H = \sum_n \frac{\hbar\omega(n+\frac{1}{2})(b^\dagger)^n X(0)(b)^n}{n!}. \tag{14}$$

Using the expression for the projection operator $X(0)$ in Eq. (11), this becomes

$$H = \sum_{s=0}^{\infty} \hbar\omega(F_s + \tfrac{1}{2}G_s)(b^\dagger)^s(b)^s, \tag{15}$$

where

$$F_s = \sum_{n=0}^{s} \frac{(-1)^{s-n} n}{n!(s-n)!} = \begin{cases} 0, & s = 0 \text{ or } s > 1 \\ 1, & s = 1 \end{cases} \tag{16}$$

and

$$G_s = \sum_{n=0}^{s} \frac{(-1)^{s-n}}{n!(s-n)!} = \begin{cases} 1, & s = 0 \\ 0, & s > 0 \end{cases}. \tag{17}$$

Consequently, the Hamiltonian in the single-particle picture has been mapped into

$$H = \hbar\omega(b^\dagger b + \tfrac{1}{2}) \tag{18}$$

using Eqs. (15) to (17), which is precisely the form of the quasiparticle Hamiltonian, Eq. (2).

This result assures us that the proposed equivalence relationship, Eq. (10), is indeed the correct transformation from the single-particle picture to the quasiparticle representation for a harmonic oscillator.

III. HARMONIC CRYSTAL

The harmonic crystal is a multidimensional generalization of the harmonic oscillator. The transformation from the single-particle representation can be achieved by an extrapolation of the analysis of the previous section. The harmonic crystal Hamiltonian in configuration space is

$$H = -\sum_i \frac{\hbar^2\nabla_i^2}{2M} + \tfrac{1}{2}\sum_i\sum_j (\mathbf{r}_i - \mathbf{R}_i)\cdot\overleftrightarrow{v}_{ij}\cdot(\mathbf{r}_j - \mathbf{R}_j), \tag{19}$$

where $\overleftrightarrow{v}_{ij}$ is the interaction diadic, $\mathbf{R}_i$ represents the equilibrium position of the ith atom and is chosen to correspond to a lattice vector, and $\mathbf{r}_i$ is the instantaneous position of this atom.

Second quantization of this Hamiltonian in the single-particle picture can be achieved through the set of three-dimensional harmonic oscillator wavefunctions

$$\phi_{\mathbf{n}}(\mathbf{r} - \mathbf{R}_i) = \phi_{n(1)}(x - X_i)\phi_{n(2)}(y - Y_i)\phi_{n(3)}(z - Z_i), \tag{20}$$

where

$$\phi_{n(\alpha)}(\xi_\alpha) = \left(\frac{\Gamma_\alpha}{\pi}\right)^{1/4}\left(\frac{1}{2^{n(\alpha)}n(\alpha)!}\right)^{1/2} H_{n(\alpha)}(\Gamma_\alpha^{1/2}\xi_\alpha) \times e^{-\Gamma_\alpha\xi_\alpha/2}\ \alpha \tag{21}$$

and

$$\Gamma_\alpha = M\omega_\alpha/\hbar. \tag{22}$$

The quantized field operator is then

$$\psi_{(\mathbf{r})} = \sum_i\sum_{\mathbf{n}}\sum_\sigma \phi_{\mathbf{n}}(\mathbf{r} - \mathbf{R}_i)a_{i\mathbf{n}\sigma}, \tag{23}$$

where $a_{i\mathbf{n}\sigma}$ destroys an atom of nuclear spin σ with quantum number $\mathbf{n}$ at site i. In the event that wavefunction overlap effects can be neglected and single occupancy at each site is preserved (see Ref. 1 for a discussion of this approximation), the Hamiltonian takes the form of

$$H = H_0 + H_1, \tag{24}$$

where

$$H_0 = \sum_i \sum_{\mathbf{n}\mathbf{n}'} \sum_\sigma \left(i\mathbf{n}'\left|-\frac{\hbar^2\nabla^2}{2M} + \tfrac{1}{2}(\mathbf{r} - \mathbf{R}_i)\cdot\overleftrightarrow{v}_{ii}\cdot(\mathbf{r} - \mathbf{R}_i)\right|i\mathbf{n}\right) a_{i\mathbf{n}'\sigma}^\dagger a_{i\mathbf{n}\sigma} \tag{25}$$

and

$$H_1 = \tfrac{1}{2}\sum_{i\neq j}\sum_{\mathbf{n}\mathbf{n}'}\sum_{\mathbf{m}\mathbf{m}'}\sum_{\sigma\sigma'} (i\mathbf{n}'j\mathbf{m}'\,|(\mathbf{r}_i - \mathbf{R}_i)\cdot\overleftrightarrow{v}_{ij}\cdot(\mathbf{r}_j - \mathbf{R}_j)\,|\,i\mathbf{n}j\mathbf{m})\, a_{i\mathbf{n}\sigma}^\dagger a_{i\mathbf{n}\sigma} a_{j\mathbf{m}'\sigma'}^\dagger a_{j\mathbf{m}\sigma'}. \tag{26}$$

The definitions of the matrix elements in Eqs. (25) and (26) are

$$(i\mathbf{n}'\,|f(\mathbf{r})\,|i\mathbf{n}) = \int d\mathbf{r}\,\phi_{\mathbf{n}'}^*(\mathbf{r} - \mathbf{R}_i)f(\mathbf{r})\phi_{\mathbf{n}}(\mathbf{r} - \mathbf{R}_i) \tag{27}$$

and

$$(i\mathbf{n}'j\mathbf{m}'\,|f(\mathbf{r}_i, \mathbf{r}_j)\,|i\mathbf{n}j\mathbf{m}) = \int d\mathbf{r}_i\,d\mathbf{r}_j\,\phi_{\mathbf{n}'}^*(\mathbf{r}_i - \mathbf{R}_i) \times \phi_{\mathbf{n}}(\mathbf{r}_i - \mathbf{R}_i)f(\mathbf{r}_i, \mathbf{r}_j)\phi_{\mathbf{m}'}^*(\mathbf{r}_j - \mathbf{R}_j)\phi_{\mathbf{m}}(\mathbf{r}_j - \mathbf{R}_j). \tag{28}$$

Under the restrictive conditions just imposed, the Hamiltonian conserves spin at each lattice site. It is then possible to consider only a single spin configuration of the system and relinquish the degeneracy to a $-NkTln(2s+1)$ term in the free energy, s being the

spin value of the nuclei. The effect of overlap and spin coupling can be treated perturbatively. The spin index on the annihilation and creation operators can then be dropped while working on this spin subspace.

The proposed transformation rules are a simple extension of those for the harmonic oscillator

$$a^{\dagger}_{i\mathbf{n}'}\, a_{i\mathbf{n}} \longleftrightarrow \prod_{\alpha=1}^{3} \frac{(b^{\dagger}_{i\alpha})^{n'(\alpha)} X_{i\alpha}(0) b^{n(\alpha)}_{i\alpha}}{[n'(\alpha)!\, n(\alpha)!\,]^{1/2}}, \tag{29}$$

where

$$X_{i\alpha}(0) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} (b^{\dagger}_{i\alpha})^r (b_{i\alpha})^r \tag{30}$$

and where $b^{\dagger}_{i\alpha}, b_{i\alpha}$ create or annihilate a pseudo excitation of Cartesian component α at site i. Each set of one-dimensional wavefunctions of Eq. (20) then corresponds to a boson space

$$|\phi_{n(\alpha)}(r_\alpha - R_{i\alpha})\rangle \longleftrightarrow |n(\alpha)\rangle_i\,, \tag{31}$$

as in Eq. (7) and Eq. (8), each with its own quasiparticle operators. If one chooses the axes such that the interaction diadic is diagonal and the oscillator wavefunction frequencies ω_α such that

$$\omega_\alpha = [M^{-1}(\overleftrightarrow{v}_{ii})_{\alpha\alpha}]^{1/2}, \tag{32}$$

in other words,

$$(\overleftrightarrow{v}_{ii})_{\alpha\beta} = M\omega^2_\alpha \delta_{\alpha\beta}\,, \tag{33}$$

one can verify that on using Eqs. (29) and (30) one obtains

$$\begin{aligned} H_0 &= \sum_i \sum_{\mathbf{nn}'} \left\langle i\mathbf{n}\left| -\frac{\hbar^2\nabla^2}{2M} + \tfrac{1}{2}\sum_\alpha M\omega^2_\alpha (r_\alpha - R_{i\alpha})^2 \right| i\mathbf{n}\right\rangle \\ &\quad \times \prod_\alpha \frac{(b^{\dagger}_{i\alpha})^{n'(\alpha)} X_{i\alpha}(0) (b_{i\alpha})^{n(\alpha)}}{[n'(\alpha)!\, n(\alpha)!\,]^{1/2}} \\ &= \sum_i \sum_\alpha \Big\{ \sum_{s(\alpha)=0}^{\infty} \hbar\omega_\alpha [F_{s(\alpha)} + \tfrac{1}{2} G_{s(\alpha)}] \\ &\quad \times (b^{\dagger}_{i\alpha})^{s(\alpha)} (b_{i\alpha})^{s(\alpha)} \Big\} P_\alpha, \end{aligned} \tag{34}$$

where

$$P_\alpha = \prod_{\beta\neq\alpha} \left[\sum_{s(\beta)=0}^{\infty} G_{s(\beta)} (b^{\dagger}_{i\beta})^{s(\beta)} (b_{i\beta})^{s(\beta)} \right] = 1, \tag{35}$$

where F_s is given by Eq. (16) and G_s by Eq. (17).

Consequently, this same term is simply

$$H_0 = \sum_i \sum_\alpha \hbar\omega_\alpha (b^{\dagger}_{i\alpha} b_{i\alpha} + \tfrac{1}{2}). \tag{36}$$

On the other hand, the recursion relations for Hermite polynomials lead to

$$\begin{aligned} &(i\mathbf{n}'j\mathbf{m}' \,|\, (\mathbf{r}_i - \mathbf{R}_j) \cdot \overleftrightarrow{v}_{ij} \cdot (\mathbf{r}_j - \mathbf{R}_j) \,|\, i\mathbf{n}j\mathbf{m}) \\ &= \sum_\alpha \sum_\beta \frac{(\overleftrightarrow{v}_{ij})_{\alpha\beta}}{(4\Gamma_\alpha\Gamma_\beta)^{1/2}} \{[n(\alpha)+1][m(\beta)+1]\}^{1/2} \\ &\quad \times \delta(\mathbf{n}' - \mathbf{n} - \hat{\alpha})\delta(\mathbf{m}' - \mathbf{m} - \hat{\beta}) + \{[n(\alpha)+1]m(\beta)\}^{1/2} \\ &\quad \times \delta(\mathbf{n}' - \mathbf{n} - \hat{\alpha})\delta(\mathbf{m}' - \mathbf{m} + \hat{\beta}) + \{n(\alpha)[m(\beta)+1]\}^{1/2} \\ &\quad \times \delta(\mathbf{n}' - \mathbf{n} + \hat{\alpha})\delta(\mathbf{m}' - \mathbf{m} - \hat{\beta}) + [n(\alpha)m(\beta)]^{1/2} \\ &\quad \times \delta(\mathbf{n}' - \mathbf{n} + \hat{\alpha})\delta(\mathbf{m}' - \mathbf{m} + \hat{\beta})). \end{aligned} \tag{37}$$

Consequently, with the use of Eqs. (29) and (30), the H_1 contribution to the Hamiltonian can be written as

$$\begin{aligned} H_1 &= \frac{1}{2} \sum_{i\neq j} \sum_{\mathbf{nn}'} \sum_{\mathbf{mm}'} (i\mathbf{n}'j\mathbf{m}' \,|\, (\mathbf{r}_i - \mathbf{R}_i) \cdot \overleftrightarrow{v}_{ij} \cdot \\ &\quad \times (\mathbf{r}_j - \mathbf{R}_j) \,|\, i\mathbf{n}j\mathbf{m}) \prod_\alpha \frac{(b^{\dagger}_{i\alpha})^{n'(\alpha)} X_{i\alpha}(0) (b_{i\alpha})^{n(\alpha)}}{[n'(\alpha)!\, n(\alpha)!\,]^{1/2}} \\ &\quad \times \prod_\beta \frac{(b^{\dagger}_{j\beta})^{m'(\beta)} X_{j\beta}(0) (b_{j\beta})^{m(\beta)}}{[m'(\beta)!\, m'(\beta)!\,]^{1/2}} \\ &= \frac{1}{2} \sum_{i\neq j} \sum_\alpha \sum_\beta \frac{(\overleftrightarrow{v}_{ij})_{\alpha\beta}}{(4\Gamma_\alpha\Gamma_\beta)^{1/2}} \Big\{ \sum_{s(\alpha)=0}^{\infty} G_{s(\alpha)} \\ &\quad \times [(b^{\dagger}_{i\alpha})^{s(\alpha)+1} (b_{i\alpha})^{s(\alpha)} + (b^{\dagger}_{i\alpha})^{s(\alpha)} (b_{i\alpha})^{s(\alpha)+1}] \Big\} \\ &\quad \times \Big\{ \sum_{s(\beta)=0}^{\infty} G_{s(\beta)} [(b^{\dagger}_{j\beta})^{s(\beta)+1} (b_{j\beta})^{s(\beta)} \\ &\quad + (b^{\dagger}_{j\beta})^{s(\beta)} (b_{j\beta})^{s(\beta)+1}] \Big\} P_\alpha P_\beta\,, \end{aligned} \tag{38}$$

or finally as

$$H_1 = \frac{\hbar}{4M} \sum_{i\neq j} \sum_\alpha \sum_\beta \frac{(\overleftrightarrow{v}_{ij})_{\alpha\beta}}{(\omega_\alpha\omega_\beta)^{1/2}} (b^{\dagger}_{i\alpha} + b_{i\alpha}) \times (b^{\dagger}_{j\beta} + b_{j\beta}). \tag{39}$$

Using the results from Eqs. (36) and (39) one can synthesize the full Hamiltonian

$$\begin{aligned} H &= \sum_i \sum_\alpha \hbar\omega_\alpha (b^{\dagger}_{i\alpha} b_{i\alpha} + \tfrac{1}{2}) \\ &\quad + \frac{\hbar}{4M} \sum_{i\neq j} \sum_\alpha \sum_\beta \frac{(\overleftrightarrow{v}_{ij})_{\alpha\beta}}{(\omega_\alpha\omega_\beta)^{1/2}} (b^{\dagger}_{i\alpha} + b_{i\alpha}) \\ &\quad \times (b^{\dagger}_{j\beta} + b_{j\beta}). \end{aligned} \tag{40}$$

It is in the form of Eq. (16) of Ref. 1 except that here the indices run over the three Cartesian axes. It can be diagonalized by the following canonical transformation

$$b_{i\alpha} = N^{-1/2} \sum_{\mathbf{k}} \sum_{\nu=1}^{3} \exp(i\mathbf{k}\cdot\mathbf{R}_i)[A_{\alpha\nu}(\mathbf{k}) d_{\mathbf{k}\nu} + B_{\alpha\nu}(\mathbf{k}) d^{\dagger}_{-\mathbf{k}\nu}], \tag{41}$$

where

$$A(\mathbf{k}) = A^*(-\mathbf{k}), \qquad B(\mathbf{k}) = B^*(-\mathbf{k}),$$

N is the number of atoms in the crystal, $\mathbf{k}$ is a wavevector in the first Brillouin zone, and

$$A(\mathbf{k})\, A^{\dagger}(\mathbf{k}) - B(\mathbf{k})\, B^{\dagger}(\mathbf{k}) = I, \tag{42}$$

$$A(\mathbf{k})\, B^{\dagger}(\mathbf{k}) - B(\mathbf{k})\, A^{\dagger}(\mathbf{k}) = 0. \tag{43}$$

The eigenvalue equation is

$$U_{\alpha\nu}(\mathbf{k})\, \hbar^2\omega^2_{\mathbf{k}\nu} = \sum_\beta M_{\alpha\beta}(\mathbf{k})\, U_{\beta\nu}(\mathbf{k}), \tag{44}$$

where

$$M_{\alpha\beta}(\mathbf{k}) = (\hbar\omega_\alpha)^{1/2} \left[\hbar\omega_\alpha \delta_{\alpha\beta} + \frac{\hbar v_{\alpha\beta}(\mathbf{k})}{M(\omega_\alpha\omega_\beta)^{1/2}} \right] (\hbar\omega_\beta)^{1/2}, \tag{45}$$

$$v_{\alpha\beta}(\mathbf{k}) = N^{-1} \sum_{i\neq j} (\overleftrightarrow{v}_{ij})_{\alpha\beta} \exp[i\mathbf{k}\cdot(\mathbf{R}_i - \mathbf{R}_j)], \tag{46}$$

$$A_{\alpha\nu}(\mathbf{k}) = \tfrac{1}{2}[(\omega_\alpha/\omega_{\mathbf{k}\nu})^{1/2} U_{\alpha\nu}(\mathbf{k}) + (\omega_{\mathbf{k}\nu}/\omega_\alpha)^{1/2} U_{\alpha\nu}(\mathbf{k})], \tag{47}$$

$$B_{\alpha\nu}(\mathbf{k}) = \tfrac{1}{2}[(\omega_\alpha/\omega_{\mathbf{k}\nu})^{1/2} U_{\alpha\nu}(\mathbf{k}) - (\omega_{\mathbf{k}\nu}/\omega_\alpha)^{1/2} U_{\alpha\nu}(\mathbf{k})]. \tag{48}$$

The necessary invariance of the Hamiltonian with respect to a translation of the crystal requires

$$\sum_i \sum_j \overleftrightarrow{v}_{ij} = 0, \tag{49}$$

which in addition to Eqs. (33) and (46) leads to

$$(\overleftrightarrow{v}_{ii})_{\alpha\beta} = M\omega_\alpha^2 \delta_{\alpha\beta} = -\sum_{i\neq j} (\overleftrightarrow{v}_{ij})_{\alpha\beta} = -v_{\alpha\beta}(0). \tag{50}$$

Consequently, the eigenvalue equation can be written down as

$$U_{\alpha\nu}(\mathbf{k})\omega_{\mathbf{k}\nu}^2 = \sum_\beta M^{-1}[v_{\alpha\beta}(\mathbf{k}) - v_{\alpha\beta}(0)]U_{\beta\nu}(\mathbf{k}). \tag{51}$$

This equation is recognized as the one encountered in conventional harmonic phonon theory[6] where $\hbar\omega_{\mathbf{k}\nu}$ is the νth acoustic branch phonon energy and $U_{\alpha\nu}(\mathbf{k})$ ($\alpha = 1, 2, 3$), its polarization vector components. Furthermore, under this canonical transformation, the Hamiltonian becomes

$$H = \sum_{\mathbf{k}} \sum_\nu \hbar\omega_{\mathbf{k}\nu}(d_{\mathbf{k}\nu}^\dagger d_{\mathbf{k}\nu} + \tfrac{1}{2}) \tag{52}$$

which is again identified with the usual phonon Hamiltonian in the collective representation.[6]

IV. ANHARMONIC CRYSTAL

The anharmonic crystal case is the logical extension of the analysis. The Hamiltonian for such systems is

$$H = \sum_i \left(-\frac{\hbar^2\nabla_i^2}{2M} + \tfrac{1}{2}\sum_{i\neq j} V(\mathbf{r}_i - \mathbf{r}_j)\right), \tag{53}$$

where $V(\mathbf{r}_i - \mathbf{r}_j)$ is the two-body interatomic interaction. Second quantization is again achieved through the quantized field operator of Eq. (23) where the spin index is dropped for the reasons mentioned in Sec. III. The coordinate axes for the single-particle wavefunctions in Eq. (20) are chosen along the principal axes of the crystal. This is the reason for considering only Bravais crystals with orthogonal translation vectors.

The kinetic energy contribution to the second quantized Hamiltonian is

$$H_0 = \sum_i \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \left(i\mathbf{n}'\left| -\frac{\hbar^2\nabla^2}{2M}\right| i\mathbf{n}\right) a_{i\mathbf{n}'}^\dagger a_{i\mathbf{n}}. \tag{54}$$

Since

$$\left(i\mathbf{n}'\left| -\frac{\hbar^2\nabla^2}{2M}\right| i\mathbf{n}\right) = \sum_\alpha \tfrac{1}{2}\hbar\omega_\alpha[n(\alpha) + \tfrac{1}{2}]\delta(\mathbf{n}' - \mathbf{n}) - \tfrac{1}{4}\hbar\omega_\alpha\{(n(\alpha)+2)(n(\alpha)+1)\}^{1/2}\delta(\mathbf{n}' - \mathbf{n} - 2\hat{\alpha}) - \tfrac{1}{4}\hbar\omega_\alpha\{[n'(\alpha)+2][n'(\alpha)+1]\}^{1/2}\delta(\mathbf{n} - \mathbf{n}' - 2\hat{\alpha}), \tag{55}$$

and with the use of the equivalence relationships, Eqs. (29) and (30), H_0 can be transformed into

$$H_0 = \sum_i \sum_\alpha \left(\sum_{s(\alpha)=0}^\infty \tfrac{1}{2}\hbar\omega_\alpha(F_{s(\alpha)} + \tfrac{1}{2}G_{s(\alpha)})(b_{i\alpha}^\dagger)^{s(\alpha)} \times (b_{i\alpha})^{s(\alpha)}\right)P_\alpha - \sum_i \sum_\alpha \left(\sum_{s(\alpha)=0}^\infty \tfrac{1}{4}\hbar\omega_\alpha G_{s(\alpha)} \times [(b_{i\alpha}^\dagger)^{s(\alpha)+2}(b_{i\alpha})^{s(\alpha)} + (b_{i\alpha}^\dagger)^{s(\alpha)}(b_{i\alpha})^{s(\alpha)+2}]\right)P_\alpha = \sum_i \sum_\alpha \tfrac{1}{4}\hbar\omega_\alpha[1 + 2b_{i\alpha}^\dagger b_{i\alpha} - (b_{i\alpha}^\dagger)^2 - (b_{i\alpha})^2], \tag{56}$$

where the F_s, G_s and P_α functions are given by Eqs. (16), (17) and (35), respectively.

On the other hand, the potential energy contribution to the Hamiltonian is

$$H_1 = \tfrac{1}{2}\sum_{i\neq j}\sum_{\mathbf{nn}'}\sum_{\mathbf{mm}'} (i\mathbf{n}'j\mathbf{m}'|V(\mathbf{r}_i - \mathbf{r}_j)|i\mathbf{n}j\mathbf{m}) \times a_{i\mathbf{n}'}^\dagger a_{i\mathbf{n}} a_{j\mathbf{m}'}^\dagger a_{j\mathbf{m}} \tag{57}$$

which, using the result of the Appendix, can be written as

$$H_1 = \frac{1}{2}\sum_{i\neq j}\sum_{\mathbf{n}'\geq\mathbf{n}}\sum_{\mathbf{m}'\geq\mathbf{m}}\sum_{\mathbf{s}\leq\mathbf{n}}\sum_{\boldsymbol{\mu}\leq\mathbf{m}} \times (iojo|\prod_{\alpha\beta}\nabla_{i\alpha}^{[n'(\alpha)-n(\alpha)+2s(\alpha)]}\nabla_{j\beta}^{[m'(\beta)-m(\beta)+2\mu(\beta)]} \times V(\mathbf{r}_i - \mathbf{r}_j)|iojo) \times \prod_\alpha\{[n(\alpha)-s(\alpha)]!\,s(\alpha)!\,(2\Gamma_\alpha)^{[n'(\alpha)-n(\alpha)+2s(\alpha)]/2} \times [n'(\alpha)-n(\alpha)+s(\alpha)]!\}^{-1} \times \{(b_{i\alpha}^\dagger)^{n'(\alpha)}X_{i\alpha}(0)(b_{i\alpha})^{n(\alpha)} + [1-\delta(n'(\alpha)-n(\alpha))] \times (b_{i\alpha}^\dagger)^{n(\alpha)}X_{i\alpha}(0)(b_{i\alpha})^{n'(\alpha)}\} \times \prod_\beta\{[m(\beta)-\mu(\beta)]!\,\mu(\beta)!\,(2\Gamma_\beta)^{[m'(\beta)-m(\beta)+2\mu(\beta)]/2} \times [m'(\beta)-m(\beta)+\mu(\beta)]!\}^{-1} \times \{(b_{j\beta}^\dagger)^{m'(\beta)}X_{j\beta}(0)(b_{j\beta})^{m(\beta)} + [1-\delta(m'(\alpha)-m(\alpha))] \times (b_{j\beta}^\dagger)^{m(\beta)}X_{j\beta}(0)(b_{j\beta})^{m'(\beta)}\}. \tag{58}$$

This last equation can be rewritten as

$$H_1 = \frac{1}{2}\sum_{i\neq j}\sum_{\mathbf{pq}}(iojo|\prod_\alpha\prod_\beta(\nabla_{i\alpha})^{p(\alpha)}(\nabla_{j\beta})^{q(\beta)}V(\mathbf{r}_i - \mathbf{r}_j)|iojo) \times\left[\sum_{\mathbf{s}\leq\mathbf{p}/2}\prod_\alpha\left(\frac{(b_{i\alpha}^\dagger)^{p(\alpha)-s(\alpha)}(b_{i\alpha})^{s(\alpha)} + [1-\delta(p(\alpha)-2s(\alpha))](b_{i\alpha}^\dagger)^{s(\alpha)}(b_{i\alpha})^{p(\alpha)-s(\alpha)}}{s(\alpha)!\,(2\Gamma_\alpha)^{p(\alpha)/2}\,[p(\alpha)-s(\alpha)]!}\right)\right] \times\left[\sum_{\boldsymbol{\mu}\leq\mathbf{q}/2}\prod_\beta\left(\frac{(b_{j\beta}^\dagger)^{q(\beta)-\mu(\beta)}(b_{j\beta})^{\mu(\beta)} + [1-\delta(q(\beta)-2\mu(\beta))](b_{j\beta}^\dagger)^{\mu(\beta)}(b_{j\beta})^{q(\beta)-\mu(\beta)}}{\mu(\beta)!\,(2\Gamma_\beta)^{q(\beta)/2}[q(\beta)-\mu(\beta)]!}\right)\right]. \tag{59}$$

Let us now define an ordering operator $\underline{O}$ which puts the creation operators first when acting on a product of boson operators with the same indices. For instance, one has

$$\underline{O}b_{i\alpha}b_{i\alpha}^\dagger b_{j\gamma}b_{i\beta}^\dagger = b_{i\alpha}^\dagger b_{i\alpha}b_{j\gamma}b_{i\beta}^\dagger. \tag{60}$$

With the help of this operator, Eq. (59) is simplified into

$$H_1 = \frac{1}{2}\sum_{i\neq j}\sum_{\mathbf{pq}}(iojo|\prod_\alpha\prod_\beta(\nabla_{i\alpha})^{p(\alpha)}(\nabla_{j\beta})^{q(\beta)} \times V(\mathbf{r}_i - \mathbf{r}_j)|iojo) \times\prod_\alpha\left(\frac{\underline{O}(b_{i\alpha}^\dagger + b_{i\alpha})^{p(\alpha)}}{p(\alpha)!\,(2\Gamma_\alpha)^{p(\alpha)/2}}\right)\prod_\beta\left(\frac{\underline{O}(b_{j\beta}^\dagger + b_{j\beta})^{q(\beta)}}{q(\beta)!\,(2\Gamma_\beta)^{q(\beta)/2}}\right). \tag{61}$$

The Fourier transform of the interatomic potential is now introduced as

$$V(\mathbf{k}) = \frac{1}{(2\pi)^3}\int d\mathbf{r}V(\mathbf{r})\exp(-i\mathbf{k}\cdot\mathbf{r}). \tag{62}$$

The inverse Fourier transform is

$$V(\mathbf{r}) = \int d\mathbf{k}V(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{63}$$

Using Eq. (63), the potential energy contribution is

$$H_1 = \tfrac{1}{2}\sum_{i\neq j}\int d\mathbf{k}V(\mathbf{k})U_{ij}(\mathbf{k})\langle iojo\,|\exp[i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)]\,|iojo\rangle, \tag{64}$$

where

$$U_{ij}(\mathbf{k}) = \sum_{\mathbf{pq}}\prod_{\alpha}\left(\frac{\underline{O}(b_{i\alpha}^{\dagger}+b_{i\alpha})^{p(\alpha)}(ik_{\alpha})^{p(\alpha)}}{p(\alpha)!\,(2\Gamma_{\alpha})^{p(\alpha)/2}}\right) \times\prod_{\beta}\left(\frac{\underline{O}(b_{j\beta}^{\dagger}+b_{j\beta})^{q(\beta)}(-ik_{\beta})^{q(\beta)}}{q(\beta)!\,(2\Gamma_{\beta})^{q(\beta)/2}}\right). \tag{65}$$

Because of the anharmonicity of the interatomic potential, H_1 exhibits nonlinear terms in the boson operators. These contain the information on the phonon-phonon interaction. Linearization of H_1 can be achieved by a Hartree–Foch type factorization whereby each term decomposes into a sum of all possible products of thermodynamic averages of separate pairs of operators except for one or two unpaired operators depending on whether their total number in the given term is odd or even. For instance, one would have under there rules

$$\begin{aligned} b_{i\alpha}^{\dagger}b_{i\beta}b_{j\gamma}b_{j\delta} \rightarrow\ & b_{i\alpha}^{\dagger}b_{i\beta}\langle b_{j\gamma}b_{j\delta}\rangle \\ & + b_{i\alpha}^{\dagger}b_{j\gamma}\langle b_{i\beta}b_{j\delta}\rangle + b_{i\alpha}^{\dagger}b_{j\delta}\langle b_{i\beta}b_{j\gamma}\rangle \\ & + b_{i\beta}b_{j\gamma}\langle b_{i\alpha}^{\dagger}b_{j\delta}\rangle + b_{i\beta}b_{j\delta}\langle b_{i\alpha}^{\dagger}b_{j\delta}\rangle \\ & + b_{j\gamma}b_{j\delta}\langle b_{i\alpha}^{\dagger}b_{i\beta}\rangle, \end{aligned} \tag{66}$$

where $\langle\cdots\rangle$ represents a thermodynamic average. All possible pairings are considered since the boson operators do not correspond to those of the true excitations, e.g., Eq. (41).

Out of all the terms of H_1 there is one and only one which does not contain any boson operators. It is the one for which $\mathbf{p}=\mathbf{q}=0$. It is equal to

$$\tfrac{1}{2}\sum_{i\neq j}\langle iojo\,|\,V(\mathbf{r}_i-\mathbf{r}_j)\,|\,iojo\rangle. \tag{67}$$

Under the factorization rules there will be no term with a single unpaired operator. These can only arise from terms containing an odd number of operators, that is when

$$\sum_{\alpha}[p(\alpha)+q(\alpha)] = \text{odd}. \tag{68}$$

Such terms vanish because of the inversion symmetry of the crystal.

Let us now consider the coefficient of the factorized terms of $U_{ij}(\mathbf{k})$ having the unpaired operators

$$\underline{O}(b_{i\mu}^{\dagger}+b_{i\mu})^2. \tag{69}$$

Only the terms of Eq. (65) for which

$$\sum_{\alpha}[p(\alpha)+q(\alpha)] = 2(n+1), \qquad n=0,1,2,\cdots \tag{70}$$

can contribute a factorized term like Eq. (69). The products of thermodynamic averages of separate pairs of operators will yield contributions to $U_{ij}(\mathbf{k})$ of the form

$$\begin{aligned} & -\prod_{\alpha}\prod_{\beta}[p(\alpha)!\,q(\beta)!\,]^{-1}\left(\frac{\langle(b_{i\alpha}^{\dagger}+b_{i\alpha})(b_{j\beta}^{\dagger}+b_{j\beta})\rangle k_{\alpha}k_{\beta}}{(4\Gamma_{\alpha}\Gamma_{\beta})^{1/2}}\right)^{s(\alpha,\beta)} \\ & \times\prod_{\gamma}\left(\frac{-\langle\underline{O}(b_{i\gamma}^{\dagger}+b_{i\gamma})^2\rangle k_{\gamma}^2}{2\Gamma_{\gamma}}\right)^{s(\gamma)}\frac{\underline{O}(b_{i\mu}^{\dagger}+b_{i\mu})^2}{2\Gamma_{\mu}}k_{\mu}{}^2 \end{aligned} \tag{71}$$

where

$$\begin{aligned} & 2s(\alpha)+\sum_{\beta}[s(\alpha,\beta)+s(\beta,\alpha)]+2\delta_{\alpha\mu} = p(\alpha)+q(\alpha), \\ & p(\alpha) = 2m(\alpha)+\sum_{\beta}s(\alpha,\beta)+2\delta_{\alpha\mu}, \quad 0\leq m(\alpha)\leq s(\alpha). \end{aligned} \tag{72}$$

Owing to the translational invariance of the $\langle\underline{O}(b_{i\gamma}^{\dagger}+b_{i\gamma})^2\rangle$ average, all values of $p(\alpha)$ and $q(\alpha)$ which are consistent with Eq. (72) will contribute a term the likes of Eq. (71). Moreover, for a given $\mathbf{p}$, $\mathbf{q}$, $\mathbf{s}$, and $\overleftrightarrow{s}$ there are

$$\prod_{\alpha}\frac{p(\alpha)!\,q(\alpha)!}{[p(\alpha)-\sum_{\beta}s(\alpha,\beta)]![q(\alpha)-\sum_{\beta}s(\beta,\alpha)]!}\prod_{\gamma}\prod_{\delta}\frac{1}{s(\gamma,\delta)!} \tag{73}$$

possible ways of obtaining distinct intersite pair averages and

$$\begin{aligned} & \prod_{\alpha}\left[1+\delta_{\alpha\mu}\left(\frac{1}{2}a(\alpha)-1\right)\right]\frac{a(\alpha)!}{2^{a(\alpha)/2}[\frac{1}{2}a(\alpha)]!} \\ & \qquad\times\prod_{\beta}\frac{b(\beta)!}{2^{b(\beta)/2}[\frac{1}{2}b(\beta)]!}, \\ & a(\alpha) = p(\alpha)-\sum_{\beta}s(\alpha,\beta), \\ & b(\beta) = q(\beta)-\sum_{\alpha}s(\alpha,\beta), \end{aligned} \tag{74}$$

different ways of pairing the remaining operators to obtain the intrasite averages. The first bracketed term of Eq. (74) differentiates the unpaired operators from the paired intrasite ones. Note that $a(\alpha)$ and $b(\beta)$ are even numbers as per Eq. (72).

As a consequence of Eqs. (72) to (74) the total number of different ways of obtaining a term like the one in Eq. (71) from the linearization procedure multiplied by $[p(\alpha)!\,q(\beta)!\,]^{-1}$ is

$$\begin{aligned} & (-1)^{\Sigma_{\alpha}s(\alpha)}\left(\prod_{\alpha}\prod_{\beta}s(\alpha,\beta)!\right)^{-1} \\ & \quad\times\prod_{\gamma}\sum_{m(\gamma)=0}^{s(\gamma)}\{2^{m(\gamma)+\delta_{\gamma\mu}}m(\gamma)!\,2^{s(\gamma)-m(\gamma)}[s(\gamma)-m(\gamma)]!\}^{-1} \\ & \quad = \tfrac{1}{2}(-1)^{\Sigma_{\alpha}s(\alpha)}\left(\prod_{\alpha}\prod_{\beta}s(\alpha,\beta)!\prod_{\gamma}s(\gamma)!\right)^{-1}. \end{aligned} \tag{75}$$

If we now sum over all values of $\mathbf{s}$ and $\overleftrightarrow{s}$ consistent with Eqs. (70) and (72), that is

$$\sum_{\alpha}\left[2s(\alpha)+\sum_{\beta}\left(s(\alpha,\beta)+s(\beta,\alpha)\right)+2\delta_{\gamma\mu}\right] = 2(n+1), \tag{76}$$

and then sum over all values of n, we find that the total coefficient of the $\underline{O}(b_{i\mu}^{\dagger}+b_{i\mu})^2$ operators in $U_{ij}(\mathbf{k})$ is

$$
\begin{aligned}
&-(4\Gamma_\mu)^{-1}k_\mu^2\sum_n\sum_{\mathbf{s}}\sum_{\mathbf{s}}(-1)^{\Sigma_\alpha s(\alpha)}\prod_\alpha\prod_\beta\\
&\times\left(\frac{\langle(b_{i\alpha}^\dagger+b_{i\alpha})(b_{j\beta}^\dagger+b_{j\beta})\rangle k_\alpha k_\beta}{(4\Gamma_\alpha\Gamma_\beta)^{1/2}}\right)^{s(\alpha,\beta)}\frac{1}{s(\alpha,\beta)!}\\
&\times\prod_\gamma\left(\frac{\langle\underline{O}(b_{i\gamma}^\dagger+b_{i\gamma})^2\rangle k_\gamma^2}{2\Gamma_\gamma}\right)^{s(\gamma)}\frac{1}{s(\gamma)!}\\
&=-(4\Gamma_\mu)^{-1}k_\mu^2\sum_n\sum_{r=0}^{n}[r!(n-r)!]^{-1}\\
&\times\left(\sum_\alpha\sum_\beta\frac{\langle(b_{i\alpha}^\dagger+b_{i\alpha})(b_{j\beta}^\dagger+b_{j\beta})\rangle k_\alpha k_\beta}{(4\Gamma_\alpha\Gamma_\beta)^{1/2}}\right)^r\\
&\times\left(-\sum_\gamma\frac{\langle(\underline{O}(b_{i\gamma}^\dagger+b_{i\gamma})^2\rangle k_\gamma^2}{2\Gamma_\gamma}\right)^{n-r}\\
&=-(4\Gamma_\mu)^{-1}k_\mu^2\sum_n(n!)^{-1}\left(-\sum_\alpha\frac{\langle(\underline{O}(b_{i\alpha}^\dagger+b_{i\alpha})^2\rangle k_\alpha^2}{2\Gamma_\alpha}\right.\\
&\left.+\sum_\alpha\sum_\beta\frac{\langle(b_{i\alpha}^\dagger+b_{i\alpha})(b_{j\beta}^\dagger+b_{j\beta})\rangle}{(4\Gamma_\alpha\Gamma_\beta)^{1/2}}k_\alpha k_\beta\right)^n\\
&=-(4\Gamma_\mu)^{-1}k_\mu^2\exp[-\tfrac{1}{2}\tilde{K}D_{ii}K+\tfrac{1}{2}\tilde{K}G_{ij}^{-1}K],
\end{aligned}
\tag{77}
$$

where the matrices K, D_{ii} and G_{ij}^{-1} are defined as

$$K_\alpha=k_\alpha, \tag{78}$$

$$(D_{ii})_{\alpha\beta}=\langle\underline{O}(b_{i\alpha}^\dagger+b_{i\alpha})^2\rangle\Gamma_\alpha^{-1}\delta_{\alpha\beta}, \tag{79}$$

$$(G_{ij}^{-1})_{\alpha\beta}=\langle(b_{i\alpha}^\dagger+b_{i\alpha})(b_{j\beta}^\dagger+b_{j\beta})\rangle(\Gamma_\alpha\Gamma_\beta)^{-1/2}. \tag{80}$$

One should note at this point that in addition to a $\underline{O}(b_{i\mu}^\dagger+b_{i\mu})^2$ contribution to the linearized $U_{ij}(\mathbf{k})$ there is also a symmetrical $\underline{O}(b_{j\mu}^\dagger+b_{j\mu})^2$ contribution. Consequently, their contribution to H_1 is

$$
\begin{aligned}
&\tfrac{1}{2}\sum_{i\neq j}\sum_\mu\int d\mathbf{k}V(\mathbf{k})(iojo\,|\exp[i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)]\,|iojo)(ik_\mu)^2\\
&\times(2\Gamma_\mu)^{-1}\exp[-\tfrac{1}{2}\tilde{K}D_{ii}K+\tfrac{1}{2}\tilde{K}G_{ij}^{-1}K]\underline{O}(b_{i\mu}^\dagger+b_{i\mu})^2.
\end{aligned}
\tag{81}
$$

Since

$$
\begin{aligned}
(iojo\,|\exp[i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)]\,|iojo)&=\exp\left[-\sum_\alpha(2\Gamma_\alpha)^{-1}k_\alpha^2\right]\\
&\times\exp[i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)]
\end{aligned}
\tag{82}
$$

Eq. (81) becomes

$$
\begin{aligned}
&\tfrac{1}{2}\sum_{i\neq j}\sum_\mu\int d\mathbf{k}V(\mathbf{k})(ik_\mu)^2\exp[-\tfrac{1}{2}\tilde{K}G_{ii}^{-1}K+\tfrac{1}{2}\tilde{K}G_{ij}^{-1}K]\\
&\times\exp[i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)](2\Gamma_\mu)^{-1}\underline{O}(b_{i\mu}^\dagger+b_{i\mu})^2,
\end{aligned}
\tag{83}
$$

where

$$(G_{ii}^{-1})_{\alpha\beta}=(D_{ii})_{\alpha\beta}+\Gamma_\alpha^{-1}\delta_{\alpha\beta}. \tag{84}$$

Defining a matrix

$$G_{xx}^{-1}=\begin{bmatrix}G_{ii}^{-1} & G_{ij}^{-1}\\ G_{ji}^{-1} & G_{jj}^{-1}\end{bmatrix}, \tag{85}$$

it can be shown that

$$
\begin{aligned}
\exp[-\tfrac{1}{2}\tilde{K}G_{ii}^{-1}K+\tfrac{1}{2}\tilde{K}G_{ij}^{-1}K]&=\pi^{-3}\,|G_{xx}^{-1}\,|^{-1/2}\\
&\times\int d^6u\exp[i\mathbf{k}\cdot(\mathbf{u}_i-\mathbf{u}_j)+\tilde{U}_{xx}(G_{xx}^{-1})^{-1}U_{xx}],
\end{aligned}
\tag{86}
$$

where

$$\tilde{U}_{xx}=(u_{i1},u_{i2},u_{i3},u_{j1},u_{j2},u_{j3}). \tag{87}$$

Using the result of the analysis of Koehler, Eq. (31) of Ref. 7, Eq. (86) becomes

$$
\begin{aligned}
\exp(-\tfrac{1}{2}\tilde{K}G_{ii}^{-1}K+\tfrac{1}{2}\tilde{K}G_{ij}^{-1}K)&=\pi^{-3N/2}\,|G\,|^{1/2}\\
&\times\int d^{3N}u\exp(-\tilde{U}GU)\exp[i\mathbf{k}\cdot(\mathbf{u}_i-\mathbf{u}_j)],
\end{aligned}
\tag{88}
$$

where N is the number of atoms in the crystal, U is a column vector containing the $3N$ coordinate variables $u_{i\alpha}$ and the matrix G is a $3N\times3N$ matrix whose inverse is

$$(G^{-1})_{ij,\alpha\beta}=(G_{ij}^{-1})_{\alpha\beta}. \tag{89}$$

After a change of variables from $u_{i\alpha}$ to $(\mathbf{r}_i-\mathbf{R}_i)_\alpha$ in Eq. (88), the contribution of the intrasite term of Eq. (69) to the linearized H_1 is finally

$$\tfrac{1}{2}\sum_{i\neq j}\sum_\mu\int d^{3N}r\Psi^*[\nabla_{i\mu}^2V(\mathbf{r}_i-\mathbf{r}_j)]\Psi(2\Gamma_\mu)^{-1}\underline{O}(b_{i\mu}^\dagger+b_{i\mu})^2, \tag{90}$$

where

$$\Psi=\pi^{-3N/4}\,|G\,|^{1/4}\exp[-\tfrac{1}{2}\tilde{R}GR] \tag{91}$$

and R is a column vector containing the $3N$ coordinate variables $(\mathbf{r}_l-\mathbf{R}_l)_\alpha$.

As far as the $(b_{i\alpha}^\dagger+b_{i\alpha})(b_{i\beta}^\dagger+b_{i\beta})$ for $\alpha\neq\beta$ linearized term of H_1 goes its coefficient vanishes because of the imposed crystal symmetry.

There only remain factorized terms containing intersite unpaired operators like $(b_{i\mu}^\dagger+b_{i\mu})(b_{j\nu}^\dagger+b_{j\nu})$ in the linearization of H_1. Following a procedure similar to the one described for the intrasite terms one would arrive at an additional contribution to H_1 of

$$
\begin{aligned}
&\tfrac{1}{2}\sum_{i\neq j}\sum_\mu\sum_\nu\int d^{3N}r\Psi^*[\nabla_{i\mu}\nabla_{j\nu}V(\mathbf{r}_i-\mathbf{r}_j)]\Psi\\
&\times(4\Gamma_\mu\Gamma_\nu)^{1/2}(b_{i\mu}^\dagger+b_{i\mu})(b_{j\nu}^\dagger+b_{j\nu}),
\end{aligned}
\tag{92}
$$

where again Ψ is the wavefunction of Eq. (91).

Consequently, synthesizing the full linearized Hamiltonian from Eqs. (56), (67), (90) and (92) one has

$$
\begin{aligned}
H_L&=H_0+(H_1)_{\text{linearized}}+\langle H_0+H_1\rangle_{\text{HF}}\\
&\quad-\langle H_0+(H_1)_{\text{linearized}}\rangle\\
&=\sum_i\sum_\alpha\tfrac{1}{4}\hbar\omega_\alpha[1+2b_{i\alpha}^\dagger b_{i\alpha}-(b_{i\alpha}^\dagger)^2-(b_{i\alpha})^2]\\
&\quad+\tfrac{1}{2}\sum_{i\neq j}(iojo\,|V(\mathbf{r}_i-\mathbf{r}_j)\,|iojo)\\
&\quad+\tfrac{1}{2}\sum_{i\neq j}\sum_\alpha\int d^{3N}r\Psi^*[\nabla_{i\alpha}^2V(\mathbf{r}_i-\mathbf{r}_j)]\Psi\\
&\quad\times(2\Gamma_\alpha)^{-1}[2b_{i\alpha}^\dagger b_{i\alpha}+(b_{i\alpha}^\dagger)^2+(b_{i\alpha})^2]\\
&\quad+\tfrac{1}{2}\sum_{i\neq j}\sum_\alpha\sum_\beta\int d^{3N}r\Psi^*[\nabla_{i\alpha}\nabla_{j\beta}V(\mathbf{r}_i-\mathbf{r}_j)]\Psi\\
&\quad\times(4\Gamma_\alpha\Gamma_\beta)^{-1/2}(b_{i\alpha}^\dagger+b_{i\alpha})(b_{j\beta}^\dagger+b_{j\beta})\\
&\quad+\langle H_0+H_1\rangle_{\text{HF}}-\langle H_0+(H_1)_{\text{linearized}}\rangle.
\end{aligned}
\tag{93}
$$

Here $\langle H_0+H_1\rangle_{\text{HF}}$ represents the Hartree–Fock average of the full Hamiltonian. The reason for the impromtu appearance of the last two thermodynamic averages of Eq. (93) stems from the well-known fact that a Hartree-Fock procedure overcounts the average

interatomic energy but not the excitation energies. These additional terms insure that

$$\langle H_L \rangle = \langle H_0 + H_1 \rangle_{\mathrm{HF}}, \tag{94}$$

i.e., that the thermodynamic average of the linearized Hamiltonian is equal to the thermodynamic average of the full Hamiltonian in the Hartree–Fock approximation. This can be further evidenced from the following. After going through the procedure which led to Eq. (90), it can be shown that

$$\langle H_0 + H_1 \rangle_{\mathrm{HF}} = \sum_i \sum_\alpha \tfrac{1}{4} \hbar \omega_\alpha [1 - \langle \underline{Q}(b^\dagger_{i\alpha} - b_{i\alpha})^2 \rangle] + \tfrac{1}{2} \sum_{i \neq j} \int d^{3N} r \, \Psi^* V(\mathbf{r}_i - \mathbf{r}_j) \Psi, \tag{95}$$

which is obviously different from $\langle H_1 + (H_1)_{\text{linearized}} \rangle$ as calculated from Eq. (93). The difference originates in the interatomic interaction part H_1, i.e., $\langle H_1 \rangle_{\mathrm{HF}} \neq \langle (H_1)_{\text{linearized}} \rangle$.

The first step towards diagonalization of H_L is to set

$$\omega_\alpha = \{ M^{-1} N^{-1} \sum_{i \neq j} \int d^{3N} r \, \Psi^* [\nabla^2_{i\alpha} V(\mathbf{r}_i - \mathbf{r}_j)] \Psi \}^{1/2} \tag{96}$$

which is the generalization of Eq. (32). One then has

$$\begin{aligned} H_L = & \sum_i \sum_\alpha \hbar \omega_\alpha b^\dagger_{i\alpha} b_{i\alpha} \\ & + \tfrac{1}{2} \sum_{i \neq j} \sum_\alpha \sum_\beta \int d^{3N} r \, \Psi^* [\nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j)] \Psi \\ & \times (4 \Gamma_\alpha \Gamma_\beta)^{-1/2} (b^\dagger_{i\alpha} + b_{i\alpha})(b^\dagger_{j\beta} + b_{j\beta}) \\ & + \sum_i \sum_\alpha \tfrac{1}{4} \hbar \omega_\alpha + \tfrac{1}{2} \sum_{i \neq j} (iojo \mid V(\mathbf{r}_i - \mathbf{r}_j) \mid iojo) \\ & + \langle H_0 + H_1 \rangle_{\mathrm{HF}} - \langle H_0 + (H_1)_{\text{linearized}} \rangle. \end{aligned} \tag{97}$$

This Hamiltonian is of the same form as the harmonic crystal Hamiltonian, Eq. (40), and is diagonalized in the same way. The canonical transformation is again

$$b_{i\alpha} = N^{-1/2} \sum_{\mathbf{k}} \sum_\nu \exp(i \mathbf{k} \cdot \mathbf{R}_i) [A_{\alpha\nu}(\mathbf{k}) d_{\mathbf{k}\nu} + B_{\alpha\nu}(\mathbf{k}) d^\dagger_{-\mathbf{k}\nu}], \tag{98}$$

while the eigenvalue equation is also

$$U_{\alpha\nu}(\mathbf{k}) \hbar^2 \omega^2_{\mathbf{k}\nu} = \sum_\beta M_{\alpha\beta}(\mathbf{k}) U_{\beta\nu}(\mathbf{k}), \tag{99}$$

where

$$M_{\alpha\beta}(\mathbf{k}) = (\hbar \omega_\alpha)^{1/2} \left(\hbar \omega_\alpha \delta_{\alpha\beta} + \frac{\hbar v_{\alpha\beta}(\mathbf{k})}{M(\omega_\alpha \omega_\beta)^{1/2}} \right) (\hbar \omega_\beta)^{1/2}, \tag{100}$$

$$v_{\alpha\beta}(\mathbf{k}) = N^{-1} \sum_{i \neq j} \int d^{3N} r \, \Psi^* [\nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j)] \times \exp[i \mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)] \Psi, \tag{101}$$

$$A_{\alpha\nu}(\mathbf{k}) = \tfrac{1}{2} [(\omega_\alpha / \omega_{\mathbf{k}\nu})^{1/2} U_{\alpha\nu}(\mathbf{k}) + (\omega_{\mathbf{k}\nu} / \omega_\alpha)^{1/2} U_{\alpha\nu}(\mathbf{k})], \tag{102}$$

$$B_{\alpha\nu}(\mathbf{k}) = \tfrac{1}{2} [(\omega_\alpha / \omega_{\mathbf{k}\nu})^{1/2} U_{\alpha\nu}(\mathbf{k}) - (\omega_{\mathbf{k}\nu} / \omega_\alpha)^{1/2} U_{\alpha\nu}(\mathbf{k})]. \tag{103}$$

Furthermore, the imposed crystal symmetry and the rules of partial differentiation lead to

$$\begin{aligned} & N^{-1} \sum_{i \neq j} \int d^{3N} r \, \Psi^* [\nabla_{i\alpha} \nabla_{i\beta} V(\mathbf{r}_i - \mathbf{r}_j)] \Psi \\ & = \delta_{\alpha\beta} N^{-1} \sum_{i \neq j} \int d^{3N} r \, \Psi^* [\nabla^2_{i\alpha} V(\mathbf{r}_i - \mathbf{r}_j)] \Psi \\ & = - N^{-1} \sum_{i \neq j} \int d^{3N} r \, \Psi^* [\nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j)] \Psi. \end{aligned} \tag{104}$$

Consequently, the eigenvalue equation becomes

$$U_{\alpha\nu}(\mathbf{k}) \omega^2_{k\nu} = \sum_\beta M^{-1} [v_{\alpha\beta}(\mathbf{k}) - v_{\alpha\beta}(0)] U_{\beta\nu}(\mathbf{k}). \tag{105}$$

This is equivalent to the eigenvalue equation of a harmonic crystal with the interaction diadic

$$\overleftrightarrow{v}_{ij} = \int d^{3N} r \, \Psi^* [\nabla_i \nabla_j V(\mathbf{r}_i - \mathbf{r}_j)] \Psi. \tag{106}$$

Because of Eqs. (98), (102), and (103) the different thermodynamic averages encountered in the formalism can be evaluated. One obtains

$$\begin{aligned} \langle \underline{Q}(b^\dagger_{i\alpha} \pm b_{i\alpha})(b^\dagger_{j\beta} \pm b_{j\beta}) \rangle &= \pm N^{-1} \sum_{\mathbf{k}} \exp[i \mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)] \\ &\times \Big\{ \sum_\nu [A_{\alpha\nu}(\mathbf{k}) \pm B_{\alpha\nu}(\mathbf{k})](1 + 2 n_{\mathbf{k}\nu})[A_{\beta\nu}(\mathbf{k}) \pm B_{\beta\nu}(\mathbf{k})] \\ &- \delta_{ij} \delta_{\alpha\beta} \Big\}, \\ &= \pm (\omega_\alpha \omega_\beta)^{\pm 1/2} N^{-1} \sum_{\mathbf{k}} \exp[i \mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)] \\ &\times \Big\{ \sum_\nu \omega^{\pm 1}_{\mathbf{k}\nu} (1 + 2 n_{\mathbf{k}\nu}) U_{\alpha\nu}(\mathbf{k}) U_{\beta\nu}(\mathbf{k}) - \delta_{ij} \delta_{\alpha\beta} \Big\}, \end{aligned} \tag{107}$$

where

$$n_{\mathbf{k}\nu} = \langle d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu} \rangle. \tag{108}$$

Consequently, from Eqs. (79), (80), (84), and (88) it is deduced that

$$(G^{-1}_{ij})_{\alpha\beta} = \hbar M^{-1} N^{-1} \sum_{\mathbf{k}} \exp[i \mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)] \times \sum_\nu \omega^{-1}_{\mathbf{k}\nu} (1 + 2 n_{\mathbf{k}\nu}) U_{\alpha\nu}(\mathbf{k}) U_{\beta\nu}(\mathbf{k}) \tag{109}$$

for all i, j. Note that crystalline symmetry leads to

$$(G^{-1}_{ii})_{\alpha\beta} = (G^{-1}_{ii})_{\alpha\alpha} \delta_{\alpha\beta}. \tag{110}$$

This expression for the matrix G^{-1}, Eq. (109), is seen to be of the same form as the one quoted in Koehler's work.[7] As a matter of fact it is identical at the absolute zero of temperature when $n_{\mathbf{k}\nu} = 0$. Therefore, at 0° K the normalized wavefunction Ψ defined in Eq. 91 is the ground state wavefunction of a harmonic crystal whose interaction diadic is the one in Eq. (106). This is the self-consistent harmonic approximation of Koehler. At finite temperature, the wavefunction Ψ and the self-consistent spring constants are renormalized by the factor $n_{\mathbf{k}\nu}$ in Eq. (109). As a consequence of this thermal renormalization, the wavefunction becomes more spread out, thus reflecting the increased atomic motion, and the self-consistent spring constants become larger. The self-consistency condition is thus more physical at nonzero temperature than the zero temperature one. The linearized Hamiltonian is also reduced to its self-consistent harmonic form

$$H_L = \sum_{\mathbf{k}} \sum_\nu \tfrac{1}{4} \hbar \omega_{\mathbf{k}\nu} + \tfrac{1}{2} \sum_{i \neq j} \int d^{3N} r \, \Psi^* V(\mathbf{r}_i - \mathbf{r}_j) \Psi + \sum_{\mathbf{k}} \sum_\nu \hbar \omega_{\mathbf{k}\nu} (d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu} - \tfrac{1}{2} n_{\mathbf{k}\nu}). \tag{111}$$

Note again that the $n_{\mathbf{k}\nu}$ term in Eq. (111) assures that Eq. (94) is satisfied.

Let us define at this point a new operator $\mathcal{L}$ which excludes all lower order linearized terms when acting on a product of phonon operators. Thus, one would have

$$\mathcal{L} d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu} = d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu} - \langle d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu} \rangle, \tag{112}$$

$$\mathcal{L}\, d^\dagger_{\mathbf{k}_1\nu_1} d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_3\nu_3} d_{\mathbf{k}_4\nu_4} = d^\dagger_{\mathbf{k}_1\nu_1} d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_3\nu_3} d_{\mathbf{k}_4\nu_4}$$
$$- \langle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_3\nu_3}\rangle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_4\nu_4} - \langle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_4\nu_4}\rangle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_3\nu_3}$$
$$- \langle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_4\nu_4}\rangle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_4\nu_4} - \langle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_4\nu_4}\rangle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_3\nu_3}$$
$$+ \langle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_3\nu_3}\rangle \langle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_4\nu_4}\rangle + \langle d^\dagger_{\mathbf{k}_2\nu_2} d_{\mathbf{k}_4\nu_4}\rangle \langle d^\dagger_{\mathbf{k}_1\nu_1} d_{\mathbf{k}_3\nu_3}\rangle. \tag{113}$$

In other words, when $\mathcal{L}$ acts on a product of phonon operators, all lower order terms arising from a Hartree–Fock decoupling scheme are removed in a way that insures the proper thermodynamic limit. Using this operator, the linearized Hamiltonian becomes

$$H_L = E_0 + \sum_{\mathbf{k}}\sum_{\nu} \hbar\omega_{\mathbf{k}\nu} \mathcal{L}\, d^\dagger_{\mathbf{k}\nu} d_{\mathbf{k}\nu}, \tag{114}$$

where

$$E_0 = \sum_{\mathbf{k}}\sum_{\nu} \tfrac{1}{2}\hbar\omega_{\mathbf{k}\nu}(n_{\mathbf{k}\nu} + \tfrac{1}{2}) + \tfrac{1}{2}\sum_{i\neq j}\int d^{3N}r\, \Psi^* V(\mathbf{r}_i - \mathbf{r}_j)\Psi \tag{115}$$

is the internal energy of the crystal in the Hartree–Fock approximation.

The nonlinear terms of the full Hamiltonian can also be obtained using the procedure outlined previously. For instance, the Hartree–Fock coefficient of the term of H_1 containing the operator products

$$\underline{O}\prod_{\alpha} (b^\dagger_{i\alpha} + b_{i\alpha})^{u(\alpha)} (b^\dagger_{j\alpha} + b_{j\alpha})^{v(\alpha)} \tag{116}$$

is found to be

$$\tfrac{1}{2}\prod_{\alpha}\prod_{\beta} \int d^{3N}r\, \Psi^* [\nabla^{u(\alpha)}_{i\alpha} \nabla^{v(\alpha)}_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j)]\Psi$$
$$\times [u(\alpha)!v(\beta)!]^{-1} (2\Gamma_\alpha)^{-u(\alpha)} (2\Gamma_\beta)^{-v(\beta)}. \tag{117}$$

Upon substitution of the canonical transformation Eq. (98), into Eq. (117) followed by a summation over all

$$\sum_{\alpha}[u(\alpha) + v(\alpha)] = n, \tag{118}$$

the total nonlinear contribution of H_1 is calculated to be

$$\delta H = \tfrac{1}{2}\sum_{i\neq j}\sum_{n=3}^{\infty} (n!)^{-1} \sum_{\{\mathbf{k}\}}\sum_{\{\nu\}} \Delta(\mathbf{k}_1 + \mathbf{k}_2 + \cdots \mathbf{k}_n)$$
$$\times \hbar^n (2^n M^n \omega_{\mathbf{k}_1\nu_1}\omega_{\mathbf{k}_2\nu_2}\cdots\omega_{\mathbf{k}_n\nu_n})^{-1/2}$$
$$\times \int d^{3N}r\, \Psi^* [\nabla_{\mathbf{k}_1\nu_1}\nabla_{\mathbf{k}_2\nu_2}\cdots\nabla_{\mathbf{k}_n\nu_n} V(\mathbf{r}_i - \mathbf{r}_j)]\Psi$$
$$\times \mathcal{L}(d^\dagger_{-\mathbf{k}_1\nu_1} + d_{\mathbf{k}_1\nu_1})(d^\dagger_{-\mathbf{k}_2\nu_2} + d_{\mathbf{k}_2\nu_2})\cdots$$
$$\times (d^\dagger_{-\mathbf{k}_n\nu_n} + d_{\mathbf{k}_n\nu_n}), \tag{119}$$

where

$$\Delta(\mathbf{k}) = \begin{cases} 1 \text{ if } \mathbf{k} = 0 \text{ or a reciprocal lattice vector} \\ 0 \text{ otherwise,} \end{cases}$$

and

$$\nabla_{\mathbf{k}\nu} = N^{-1}\sum_i\sum_\alpha U_{\alpha\nu}(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{R}_i)\nabla_{i\alpha}. \tag{120}$$

At the absolute zero of temperature, this form of δH is identical to Koehler's one. It should be noted that the operator $\mathcal{L}$ prevents interference between the terms of Eq. (119) with different n values. Since any Hartree–Fock reduction of an nth order term is excluded, it cannot contribute to any lesser order term. This explicit exclusion is required since the effect of Hartree–Fock reduction is already included in the wavefunction Ψ. Although the operator $\mathcal{L}$ might seem tedious to carry along in an equation of the motion approach it is simplicity itself in a diagram expansion in perturbation theory.[8] One can finally write the full Hamiltonian as

$$H = H_L + \delta H, \tag{121}$$

where H_L and δH are defined in Eqs. (114) and (119), respectively.

V. QUANTUM CRYSTAL

The quantum crystal problem requires a departure from the harmonic single-particle basis in order to handle the inherent nonlinearities. A normal procedure would be to use a basis of self-consistent wavefunctions $\phi_{\mathbf{n}}(\mathbf{r}_i - \mathbf{R}_i)$ obeying the following eigenvalue equation

$$\left(-\frac{\hbar^2\nabla_i^2}{2M} + \sum_{j\neq i}(j\,0|\,V(\mathbf{r}_i - \mathbf{r}_j)\,|j\,0)\right)\phi_{\mathbf{n}}(\mathbf{r}_i - \mathbf{R}_i)$$
$$= \epsilon_{\mathbf{n}}\phi_{\mathbf{n}}(\mathbf{r}_i - \mathbf{R}_i). \tag{122}$$

The matrix elements are still defined as in Eqs. (27) and (28) with the Hartree wavefunctions replacing the harmonic ones. The $\mathbf{n}$ value here is a convenience to represent the three quantum numbers associated with each wavefunction. Hence, one can no longer use these $\mathbf{n}$ vectors as a basis for the transformation in Eq. (29). This is borne out by the experience gained from the RPA whereby in the Gillis approximation, Eq. (39) of Ref. 9, the single-phonon modes sample all excited Hartree states. Yet, whatever transformation procedure is used it must be reducible to the one of Sec. III in the harmonic limit. This suggests the more general transformations rules

$$a^\dagger_{i\mathbf{n}'} a_{i\mathbf{n}} = \sum_{\mathbf{m}\mathbf{m}'} U^*_i(\mathbf{n}', \mathbf{m}') U_i(\mathbf{n}, \mathbf{m})$$
$$\times \frac{\prod_\alpha (b^\dagger_{i\alpha})^{m'(\alpha)} X_{i\alpha}(0)(b_{i\alpha})^{m(\alpha)}}{[m'(\alpha)!m(\alpha)!]^{1/2}}, \tag{123}$$

where the U_i matrix is unitary, i.e.,

$$\sum_{\mathbf{n}} U^*_i(\mathbf{n}, \mathbf{m}')U_i(\mathbf{n}, \mathbf{m}) = \delta(\mathbf{m} - \mathbf{m}') \tag{124}$$

and

$$U_i(\mathbf{n}, 0) = \delta(\mathbf{n}). \tag{125}$$

This last equation means that the ground state at lattice site i is to be the boson vacuum as in the previous sections. The problem is then of finding this transformation matrix for a Hartree basis.

Keeping in mind the asymptotic limit of the transformation rules for a harmonic potential as well as the Gillis approximation[9] to the RPA spectrum, the following single-boson eigenvectors are postulated.

$$U_i(\mathbf{n}, \hat{\alpha}) = A_{\hat{\alpha}}(\epsilon_{\mathbf{n}} - \epsilon_0)(i\mathbf{n}\,|X_\alpha\,|io), \tag{126}$$

where

$$A_{\hat{\alpha}} = (M/\hbar^2)[-(io\,|\nabla^2_\alpha\,|io)]^{-1/2}. \tag{127}$$

The orthonormality of these three vectors can be verified from Eq. (124) by making use of Eq. (41) of Ref. 9. And indeed one has

$$U_i(\mathbf{n}, \hat{\alpha}) = \delta(\mathbf{n} - \hat{\alpha}) \tag{128}$$

in the limit of a harmonic interaction which is precisely the correct value of Sec. III. It is interesting at this point to linearize the Hamiltonian at the two-boson level as in spin wave theory, that is keeping only those terms of the Hamiltonian which involve at most a product of two $b_{i\alpha}$ or $b^\dagger_{i\alpha}$ operators.

Only the eigenvectors of Eq. (126) are required for such a procedure. It follows that the Hamiltonian

$$\begin{aligned} H = {} & N\epsilon_0 + \sum_i \sum_{n\neq 0} (\epsilon_{\mathbf{n}} - \epsilon_0) a^\dagger_{i\mathbf{n}} a_{i\mathbf{n}} \\ & + \tfrac{1}{2} \sum_{i\neq j} \sum_{\mathbf{nn'}} \sum_{\mathbf{mm'}} (i\mathbf{n'}j\mathbf{m'} \,|\, V(\mathbf{r}_i - \mathbf{r}_j) \,|\, i\mathbf{n}j\mathbf{m}) \\ & \times a^\dagger_{i\mathbf{n'}} a^\dagger_{j\mathbf{m'}} a_{j\mathbf{n}} a_{i\mathbf{n}} \\ & - \sum_{i\neq j} \sum_{\mathbf{nn'}} (i\mathbf{n'}jo \,|\, V(\mathbf{r}_i - \mathbf{r}_j) \,|\, i\mathbf{n}jo) a^\dagger_{i\mathbf{n'}} a_{i\mathbf{n}} \end{aligned} \tag{129}$$

when linearized this way becomes

$$\begin{aligned} H_L = {} & N\epsilon_0 - \tfrac{1}{2} \sum_{i\neq j} (iojo \,|\, V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo) \\ & + \tfrac{1}{2} \sum_{i\neq j} \sum_{\alpha} [-(io \,|\, \nabla^2_\alpha \,|\, io)]^{-1} (iojo \,|\, \nabla^2_{i\alpha} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo, \\ & \times b^\dagger_{i\alpha} b_{i\beta} + \tfrac{1}{8} \sum_{i\neq j} \sum_{\alpha\beta} [(iojo \,|\, \nabla^2_{i\alpha} \nabla^2_{j\alpha} \,|\, iojo)]^{-1/2} \\ & \times (iojo \,|\, \nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo) (b^\dagger_{i\alpha} b_{j\beta} \\ & + b^\dagger_{i\alpha} b^\dagger_{j\beta} + \text{h.c.}). \end{aligned} \tag{130}$$

Extensive use has been made of Eqs. (41) and (42) of Ref. 9. in arriving at this last equation. The eigenvalue equation resulting from the diagonalization of this Hamiltonian is

$$\epsilon_{\alpha\nu}(\mathbf{k}) \hbar^2 \omega^2_{\mathbf{k}\nu} = \sum_\beta M_{\alpha\beta}(\mathbf{k}) \epsilon_{\beta\nu}(\mathbf{k}), \tag{131}$$

where

$$\begin{aligned} M_{\alpha\beta}(\mathbf{k}) = {} & \frac{1}{4} \left\{ N^{-1} \sum_{i\neq j} \frac{(iojo \,|\, \nabla^2_{i\alpha} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo)}{[-(io \,|\, \nabla^2_{i\alpha} \,|\, io)]} \right\}^{1/2} \\ & \times \left\{ N^{-1} \sum_{i\neq j} \frac{(iojo \,|\, \nabla^2_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo)}{[-(jo \,|\, \nabla^2_{j\beta} \,|\, jo)]} \right\}^{1/2} \\ & \times \left\{ N^{-1} \sum_{i\neq j} \frac{(iojo \,|\, \nabla^2_i V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo)}{[-(io \,|\, \nabla^2_{i\alpha} \,|\, jo)]} \delta\alpha\beta + N^{-1} \sum_{i\neq j} \right. \\ & \left. \frac{(iojo \,|\, \nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo) \exp[i\mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)]}{(iojo \,|\, \nabla^2_{i\alpha} \nabla^2_{j\beta} \,|\, iojo)^{1/2}} \right\}, \end{aligned} \tag{132}$$

and where $\epsilon_\nu(\mathbf{k})$ is the polarization vector of the νth mode of the phonon with wavevector $\mathbf{k}$. Except for a renormalization factor, this solution is similar to the one of Gillis,[9] Eq. (43), in which the spring constants are just

$$(iojo \,|\, \nabla_{i\alpha} \nabla_{j\beta} V(\mathbf{r}_i - \mathbf{r}_j) \,|\, iojo). \tag{133}$$

Our postulated single-phonon eigenvectors of Eq. (126) are then compatible with the RPA result and do asymptotically go the proper harmonic limit. This does not imply, however, that Eq. (126) is the only possibility for the single-boson eigenvectors. It is only a speculative one which does lead to physically acceptable results.

As for the many-boson eigenvectors, their construction is even more speculative and does not seem to be unique. Extrapolating on the idea behind Eq. (126) that the boson operators $b_{i\alpha}$ and $b^\dagger_{i\alpha}$ be associated with a linear displacement operator one may postulate that

$$U_i(\mathbf{n}, \mathbf{m}) = (\epsilon_{\mathbf{n}} - \epsilon_0) \left(i\mathbf{n} \left| \sum_{\mathbf{p}}{}' P(\mathbf{m}, \mathbf{p}) \prod_\gamma X^{p(\gamma)}_\gamma \right| io \right), \tag{134}$$

where the primed summation indicates that

$$\sum_\alpha p(\alpha) \leq \sum_\alpha m(\alpha). \tag{135}$$

In other words, the polynomial in Eq. (134) is of order $\sum_\alpha m(\alpha)$ and the corresponding eigenvector is to be associated with a product of boson operators of the same order, $\prod_\alpha (b_{i\alpha})^{m(\alpha)}$. The $P(\mathbf{m}, \mathbf{p})$ coefficients are determined by the orthonormality equation (124) between all $U_i(\mathbf{n}, \mathbf{m}')$ for which

$$\sum_\alpha m'(\alpha) \leq \sum_\alpha m(\alpha), \tag{136}$$

that is

$$\begin{aligned} -(\hbar^2/2M)^2 \sum_{\mathbf{p}}{}' \sum_{\mathbf{q}}{}' P(\mathbf{m}', \mathbf{p}) P(\mathbf{m}, \mathbf{q}) (io \,|\, [H, \prod_\alpha X^{p(\alpha)}_\alpha] \\ \times [H, \prod_\beta X^{q(\beta)}_\beta] \,|\, io) = \delta(\mathbf{m}' - \mathbf{m}). \end{aligned} \tag{137}$$

There is, unfortunately, an undeterminacy in choosing an orthogonal basis from a set of nonorthogonal vectors. One can only impose that the proper harmonic limit

$$\begin{aligned} \sum_{\mathbf{p}}{}' P(\mathbf{n}, \mathbf{p}) \prod_\alpha X^{p(\alpha)}_\alpha \to \prod_\alpha (\Gamma_\alpha/\pi)^{1/4} [2^{m(\alpha)} m(\alpha)!]^{1/2} \\ \times H_{m(\alpha)}(\Gamma^{1/2}_\alpha X_\alpha) \end{aligned} \tag{138}$$

be respected.

The problem with a Hartree basis is then twofold. First, a systematic procedure must be found to generate the transformation eigenvectors of Eq. (123). Second, the hierarchy of boson operators as encountered in Sec. IV should as much as possible be summed in order to generate a closed form solution to the thermal renormalization problem. Failure to achieve a closed form solution restricts the method to temperatures lower than the Debye temperature in which range a series expansion in powers of $\langle b^\dagger_{i\alpha} b_{i\alpha} \rangle$ is convergent.

VI. CONCLUDING REMARKS

The transformation procedure as applied to a harmonic single particle basis does not offer a new approach to quantum crystals. As a matter of fact it seems an unnecessarily tedious procedure compared to the more natural collective approach. But it does, however, pave the way for a possible approach to a Hartree single-particle basis. It must be borne in mind that any generalized approach to the single-particle picture must be asymptotic to our transformation procedure in the harmonic limit. One should also be aware that the effect of overlap, that is of particle indistinguishability, can be treated systematically via the effective exchange operator described in Ref. 1.

APPENDIX A

As indicated by Eq. (30) of Ref. 3 the matrix element of a function $f(x)$ between the harmonic oscillator wavefunctions of Eq. (3) is

$$\begin{aligned} (m \,|\, f(x) \,|\, n) = {} & \int dx \phi_m(x) f(x) \phi_n(x), \\ = {} & (2m\Gamma)^{-1/2} (m - 1 \,|\, \nabla_x f(x) \,|\, n) \\ & + (n/m)^{1/2} (m - 1 \,|\, f(x) \,|\, n - 1). \end{aligned} \tag{A1}$$

With repeated applications of this recursion relation it can be deduced that

$$(m\,|f(x)\,|n) = \sum_{s=0}^{m} (m!n!)^{1/2}[(n-m+s)!\,(m-s)!\,s! \times (2\Gamma)^{(n-m+2s)/2}]^{-1}(0\,|\nabla_x^{n-m+2s} f(x)\,|0) \quad \text{(A2)}$$

for $m \leq n$. Consequently, the interatomic potential energy matrix element of Eq. (57) becomes

$$\begin{aligned}(i\,\mathbf{n}'j\,\mathbf{m}'\,|V(\mathbf{r}_i-\mathbf{r}_j)\,|\,i\,\mathbf{n}j\,\mathbf{m}) = \sum_{\mathbf{s}\leq\mathbf{n}} \sum_{\mathbf{u}\leq\mathbf{m}} & (iojo\,|\prod_\alpha \prod_\beta \nabla_i^{n'(\alpha)-n(\alpha)+2s(\alpha)} \nabla_j^{m'(\beta)-m(\beta)+2u(\beta)} \\ & \times V(\mathbf{r}_i-\mathbf{r}_j)\,|\,iojo) \prod_\alpha [n'(\alpha)!\,n(\alpha)!\,]^{1/2} \\ & \times \{[n'(\alpha)-n(\alpha)+s(\alpha)]! \\ & \times [n(\alpha)-s(\alpha)]!\,s(\alpha)!\,(2\Gamma_\alpha)^{[n'(\alpha)-n(\alpha)+2s(\alpha)]/2}\}^{-1} \\ & \times \prod_\beta [m'(\beta)!\,m(\beta)!\,]^{1/2}\{[m'(\beta)-m(\beta)+u(\beta)]! \\ & \times [m(\beta)-u(\beta)]!\,u(\beta)!\,(2\Gamma_\beta)^{[m'(\beta)-n(\beta)+2u(\beta)]/2}\}^{-1} \end{aligned} \quad \text{(A3)}$$

for the case $\mathbf{n}' \geq \mathbf{n}$ and $\mathbf{m}' \geq \mathbf{m}$. The other cases $\mathbf{n}' < \mathbf{n}$ and/or $\mathbf{m}' < \mathbf{n}$ are easily generated from Eq. (A3) by permuting $\mathbf{n}'$, $\mathbf{n}$ and/or $\mathbf{m}'$, $\mathbf{m}$ on the right hand side.

---

*Work supported by the Canadian National Research Council.

[1]L. G. Caron, Phys. Rev. B **6**, 1081 (1972).
[2]N. R. Werthamer, Am. J. Phys. **37**, 763 (1969).
[3]T. R. Koehler, Phys. Rev. **165**, 942 (1968).
[4]W. Brenig, Z. Phys. **171**, 60 (1963).
[5]See any text dealing with the quantization of the harmonic oscillator, the lattice vibrations or the electromagnetic field. For instance, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), 2nd ed., Sec. 46.
[6]*Phonons and Phonon Interactions,* edited by T. A. Bak (Benjamin, New York, 1964).
[7]T. R. Koehler, Phys. Rev. **144**, 789 (1966).
[8]A. A. Abrikosov, L. O. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[9]N. S. Gillis and N. R. Werthamer, Phys. Rev. **167**, 607 (1968).

An algebra of the Yang–Mills field

A. M. Buoncristiani

Department of Mathematics, The Ohio State University, Columbus, Ohio 43210
(Received 15 February 1972)

It is shown how the classical Yang–Mills (and the Maxwell) field equations can be written in a simple form by introducing differential operators based on the split-Cayley algebra. As a result, a field algebra accommodating both space–time symmetry (Lorentz invariance) and internal symmetry evolves. The connection between the view of the Yang–Mills field as a split-Cayley algebra and the exceptional Jordan angebra M_8^3, discovered by Jordon, von Neumann, and Wigner, is discussed.

INTRODUCTION

The purpose of this paper is to introduce a mathematical formalism involving the equations of the classical Yang–Mills[1] field. This formalism utilizes the split-Cayley algebra, which seems well-suited to a treatment of local gauge fields. Both the gauge fields and the partial derivative operators, ∂_μ, will be expressed as elements of this algebra and combined to form a new operator which we call the covariant Cayley derivative. In terms of this new operator, the equations of the Yang–Mills field assume a simple form. While this result is purely formal, it does exhibit a remarkable interplay between the algebraic and differential-geometric aspects of the gauge fields. In particular, as we shall show, the space–time symmetry (Lorentz invariance) and the internal symmetry of the fields are synthesized into a single algebraic object, which, incidentally, need not be a Lie algebra. Although the discussion in this paper is limited to classical fields, we would conjecture that this synthesis is also a feature of the corresponding quantum fields. At any rate, the Cayley derivative is an interesting object and should prove useful in other contexts. As an illustration of its utility we briefly describe its application to the Maxwell equations.

Cayley algebras have made infrequent appearances in the physics literature. There seem to be three relevant references. The first was in 1934 when Jordan, von Neumann, and Wigner[2] initiated a program to classify the algebras appropriate to quantum mechanics.[3] The results of their classification turned up an exceptional case, the Jordan algebra M_3^8, which, at that time, they considered "too narrow for the generalized need." This exceptional Jordan algebra is the set of all three rowed Hermitian matrices whose elements are the Cayley numbers [essentially the Lie algebra $U(3)$ over a Cayley algebra]. In 1961, Pais[4] pointed out a striking similarity between the algebra of interactions and the split-Cayley algebra. Using this algebra he constructed meson and baryon multiplets which possess many strong interaction symmetries. He also indicated how this scheme might be extended to the theory of weak interactions. Later, in 1965, Gamba[5] demonstrated that a theory of quantum mechanics which includes $SU(3)$ symmetry among its particles enjoys many of the mathematical properties of the exceptional Jordan algebra M_3^8, discarded earlier by Jordan, von Neumann, and Wigner.

Although the primary aim of this paper is to describe some formal properties of the classical Yang–Mills field, it is interesting to examine how these properties are related to the work of Pais and Gamba. Both hint at the possibility of a connection between the algebra M_3^8 or the Cayley algebra and the unitary symmetry of elementary particles. Both, however, are algebraic in content and do not discuss how this algebra could be fitted into a dynamical theory. In the concluding section of this paper, we will show how local gauge fields, which are defined by dynamical equations, can be elements of M_3^8.

Finally, it is worth mentioning, we introduce the split-Cayley algebra via a vector-matrix representation (due to Zorn). This representation is a useful tool and simplifies computation. The multiplets introduced by Pais, for example, are much easier to deal with when considered as vector-matrices.

In the next section the Yang–Mills field equations are rewritten in a form which allows easy comparison with Maxwell's equations, which are needed for our presentation. In the second section the split-Cayley algebra is described. In the third and fourth sections this information is combined to present a new representation of Maxwell's and the Yang–Mills equations. A discussion of these results follows.

1. THE YANG-MILLS FIELD EQUATIONS (A)

Our starting point is the equations for the classical Yang–Mills field, Eqs. (I) and (II) below. These can be developed from several points of view, but for our purpose the following brief description is sufficient. We consider a Minkowski space–time manifold with points coordinated by x_μ, $\mu = 1,2,3,4$. Associated with each point is a finite dimensional linear space, $\mathcal{U}(x)$, spanned by the basis $\{e_A(x)\}$, $A = 1,2,3,\ldots,N$. The gauge fields $\Gamma_\mu(x)$ are four linear operators acting on $\mathcal{U}(x)$. We could consider these operators to be the "lift" of the tangent vectors ∂_μ to $\mathcal{U}(x)$,[6]

$$\partial_\mu e_A(x) = \Gamma^B_{\mu A}(x) e_B(x).$$

The dimension of $\mathcal{U}(x)$ depends upon the fields to which the $\Gamma_\mu(x)$ may couple. For example, following Yang and Mills, the gauge fields may be decomposed into a linear combination of generators of the gauge group,

$$\tau_s, \quad s = 1,2,\ldots,R,$$

as

$$\Gamma_\mu(x) = \sum_{s=1}^{R} \phi^s_\mu(x)\tau_s.$$

Here the ϕ^s_μ are the universal gauge fields, independent of their coupling to other fields, i.e. they interact in the same way with all fields.

The Yang–Mills tensor, $R_{\mu\nu}$, is defined in terms of the gauge fields by

$$\text{(I)}: R_{\mu\nu} = \partial_\mu\Gamma_\nu - \partial_\nu\Gamma_\mu + \Gamma_\mu\Gamma_\nu - \Gamma_\nu\Gamma_\mu.$$

Terms like $\Gamma_\mu\Gamma_\nu$ above represent the matrix multiplication

$$(\Gamma_\mu\Gamma_\nu)^A_B = \Gamma^A_{\mu c}\Gamma^c_{\nu B}.$$

Copyright © 1973 by the American Institute of Physics

Their appearance in $R_{\mu\nu}$ as a commutator reflects their relation to the generators of a (non-Abelian) gauge group. The field tensor, $R_{\mu\nu}$, also admits a similar matrix realization, $R^{A}_{\mu\nu B}$. As first pointed out by Utiyama,[7] in such a realization the field tensor is similar to the Riemann-Christoffel curvature tensor if the fields are identified as generalized coefficients of linear connection. In fact, Eq. (I) could be written

$$(\mathrm{I}'): R_{\mu\nu} = \nabla_\mu \Gamma_\nu - \nabla_\nu \Gamma_\mu,$$

where ∇_μ is a derivative covariant with respect to the connection Γ_μ.

Next we examine the inhomogeneous equation

$$(\mathrm{II}): \partial_\mu R_{\mu\nu} = -J_\nu.$$

The current J_ν, is conserved and has contributions due to interactions of the gauge fields with other fields; it also has contributions from the gauge fields alone. The latter contribution can be found by considering the covariant divergence of $R_{\mu\nu}$,

$$(\mathrm{II}'): \nabla_\mu R_{\mu\nu} = -j_\nu,$$

The covariant divergence of a second rank tensor is

$$\nabla_\mu R_{\mu\nu} = \partial_\mu R_{\mu\nu} + \Gamma_\mu R_{\mu\nu} - R_{\mu\nu}\Gamma_\mu,$$

so that if we denote

$$-K_\nu = \Gamma_\mu R_{\mu\nu} - R_{\mu\nu}\Gamma_\mu,$$

then

$$(\mathrm{III}): j_\nu = J_\nu + K_\nu.$$

The four vector, K_ν, is that identified by Yang and Mills as the contribution to the current due to the gauge fields alone.

Our purpose now is to rewrite the field equations in a form suitable to the introduction of an algebraic formalism. We first destroy the manifest covariance. The gauge fields, Γ_μ, can be split into spatial and temporal parts and written, as is usual for 4-vectors, $(\boldsymbol{\Gamma}, \Gamma_4)$. Here, however, each component is a matrix, the realization of a linear operator. Some formulas pertaining to the calculus of "vectors" with matrix components are collected in the Appendix. Similarly, splitting the field tensor into its spatial and temporal parts we define the two vectors V and U by

$$R_{4k} = V_k, \tag{1}$$

$$R_{kl} = \epsilon_{klm} U_m, \tag{2}$$

where ϵ_{klm} is the permutation symbol in three dimensions ($\epsilon_{123} = +1$). In the vector notation defined in the Appendix these equations become

$$\boldsymbol{V} = \partial_4 \boldsymbol{\Gamma} - \nabla\Gamma_4 + \boldsymbol{\lambda}, \tag{3}$$

$$\boldsymbol{U} = \nabla \times \boldsymbol{\Gamma} + \boldsymbol{\omega}, \tag{4}$$

with

$$\boldsymbol{\lambda} \equiv [\Gamma_4, \boldsymbol{\Gamma}] \tag{5}$$

and

$$\boldsymbol{\omega} \equiv \boldsymbol{\Gamma} \times \boldsymbol{\Gamma}. \tag{6}$$

The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ arise from the spatial and temporal part of the commutator in $R_{\mu\nu}$. The cross-product, $\boldsymbol{\Gamma} \times \boldsymbol{\Gamma}$ (which is not zero since the components of Γ do not commute) should not be confused with a similar product in the paper of Yang and Mills. The latter is the Lie product defined on the generators of the given symmetry group. The cross-product above is defined as,

$$(\boldsymbol{\Gamma} \times \boldsymbol{\Gamma})_k = \epsilon_{klm} \Gamma_l \Gamma_m.$$

Following directly from the definition of V and U and from the vector calculus of the Appendix, we have

$$\nabla\cdot(\mathbf{U} - \boldsymbol{\omega}) = 0 \tag{7}$$

$$\nabla\times(\mathbf{V} - \boldsymbol{\lambda}) - \partial_4(\mathbf{U} - \boldsymbol{\omega}) = 0. \tag{8}$$

Splitting Eq. (II) into spatial and temporal parts yields the two equations

$$\nabla\cdot\mathbf{V} = J_4 \tag{9}$$

and

$$\nabla \times \mathbf{U} + \partial_4 \mathbf{V} = \mathbf{J}. \tag{10}$$

Thus Eqs. (I) and (II) reduce to Eqs. (3) through (10). It is interesting to compare these latter equations directly with Maxwell's equations for stationary media; these are, in appropriate units[8]

$$\nabla\cdot\mathbf{B} = 0,$$
$$\nabla \times \mathbf{E} + \partial_4 \mathbf{B} = 0,$$
$$\nabla\cdot(\mathbf{E} + \mathbf{P}) = \rho,$$
$$\nabla \times (\mathbf{B} - \mathbf{M}) - \partial_4(\mathbf{E} + \mathbf{P}) = \mathbf{j}.$$

Comparison with Eqs. (7) through (10) prompts the identifications

$$\mathbf{V} = \mathbf{E} + \mathbf{P}, \quad \mathbf{U} = -\mathbf{B} + \mathbf{M},$$

so that

$$\mathbf{E} = \partial_4 \boldsymbol{\Gamma} - \nabla\Gamma_4, \quad \mathbf{B} = -\nabla \times \boldsymbol{\Gamma},$$

$$\mathbf{P} = \boldsymbol{\lambda}, \quad \mathbf{M} = \boldsymbol{\omega}.$$

It is convenient to make one further transformation. If we denote,

$$\begin{aligned} \mathbf{V}_\pm &= \mathbf{V} \pm \mathbf{U}, \\ J'_4 &= \nabla\cdot\boldsymbol{\omega}, \\ \mathbf{J}' &= \nabla \times \boldsymbol{\lambda} - \partial_4 \boldsymbol{\omega}, \end{aligned} \tag{11}$$

then Eqs. (7) through (10) become

$$\begin{aligned} &\nabla\cdot\mathbf{V}_\pm = J_4 \pm J'_4, \\ &\nabla \times \mathbf{V}_\pm \pm \partial_4 \mathbf{V}_\mp = \mathbf{J} \pm \mathbf{J}'. \end{aligned} \tag{12}$$

With these preliminaries completed we now discuss the split-Cayley algebra.

2. THE SPLIT-CAYLEY ALGEBRA

The following brief description of the split-Cayley algebra is based on an explicit representation, sometimes called the Zorn vector-matrix algebra. It is a form of the algebra well suited to calculation.

Consider a (real or complex) vector space of dimen-

sion eight. We label each element, $\mathfrak{A}$, of this space by the pair of four vectors $q = (\mathbf{q}, q_4)$ and $p = (\mathbf{p}, p_4)$ and arrange them in the matrix-like array

$$\mathfrak{A} = \begin{pmatrix} q_4 & \mathbf{p} \\ \mathbf{q} & p_4 \end{pmatrix}. \tag{13}$$

Addition of any two such elements and multiplication of an element by a scalar are defined in the same manner as for 2-by-2 matrices. For convenience, we will often denote this array by $(q;p)$ or by $(\mathbf{q}, q_4; \mathbf{p}, p_4)$. The bilinear multiplication of any two such elements $\mathfrak{A} = (q;p)$ and $\mathfrak{B} = (r;s)$ is defined by

$$\begin{aligned} \mathfrak{A}\mathfrak{B} &= \begin{pmatrix} q_4 & \mathbf{p} \\ \mathbf{q} & p_4 \end{pmatrix} \begin{pmatrix} r_4 & \mathbf{s} \\ \mathbf{r} & s_4 \end{pmatrix} \\ &= \begin{pmatrix} q_4 r_4 + \mathbf{p}\cdot\mathbf{r} & q_4\mathbf{s} + p s_4 - \mathbf{q}\times\mathbf{r} \\ \mathbf{q} r_4 + p_4\mathbf{r} + \mathbf{p}\times\mathbf{s} & \mathbf{q}\cdot\mathbf{s} + p_4 s_4 \end{pmatrix}. \end{aligned} \tag{14}$$

This is similar to ordinary matrix multiplication with the addition of cross-product terms insuring the non-associativity of the multiplication.

The element $I = (0,1;0,1)$ is evidently the unit element. The operation taking $\mathfrak{A} = (\mathbf{q}, q_4; \mathbf{p}, p_4)$ into $\bar{\mathfrak{A}} = (-\mathbf{q}, p_4; -\mathbf{p}, q_4)$ is an involution; it is formally the same operation as taking the inverse of a unimodular 2-by-2 matrix. From Eq. (14) we can verify that

$$\mathfrak{A}\bar{\mathfrak{A}} = \bar{\mathfrak{A}}\mathfrak{A} = (q_4 p_4 - \mathbf{q}\cdot\mathbf{p})I, \tag{15}$$

and if we introduce the quadratic form

$$Q(\mathfrak{A}) = q_4 p_4 - \mathbf{q}\cdot\mathbf{p} \tag{16}$$

it follows that this form is multiplicative, that is

$$Q(\mathfrak{A}\mathfrak{B}) = Q(\mathfrak{A})Q(\mathfrak{B}). \tag{17}$$

The quadratic form can be used to define an inner product,

$$\langle \mathfrak{A}, \mathfrak{B} \rangle = \tfrac{1}{2}[Q(\mathfrak{A} + \mathfrak{B}) - Q(\mathfrak{A}) - Q(\mathfrak{B})].$$

The "real" part of a product is

$$\mathfrak{A}\bar{\mathfrak{B}} + \mathfrak{B}\bar{\mathfrak{A}} = 2\langle \mathfrak{A}, \mathfrak{B} \rangle I. \tag{18}$$

If $t(\mathfrak{A})$ denotes the trace of $\mathfrak{A}$,

$$\mathfrak{A} + \bar{\mathfrak{A}} = t(\mathfrak{A})I,$$

it can be checked that each element of the algebra satisfies the quadratic equation

$$\mathfrak{A}^2 - t(\mathfrak{A})\mathfrak{A} + Q(\mathfrak{A})I = 0.$$

Notice that the quadratic form of the element $\mathfrak{A}$ is exactly the negative of the Minkowski inner product of the two four vectors which label it. This is a property of the split-Cayley algebra and is quite independent of any geometrical considerations.

Next consider that each entry in the array (13) is a function of the space-time point, x. We introduce the linear differential operator,

$$D = \begin{pmatrix} \partial_4 & \nabla \\ \nabla & \partial_4 \end{pmatrix}. \tag{19}$$

Here ∂_4 and ∇ are the spatial and temporal partial derivative operators. D acts from the left on an element of the algebra according to the multiplication specified in Eq. (14). Forming the involution of D we calculate that

$$D\bar{D} = \bar{D}D = -(\nabla^2 - \partial_{44})I, \tag{20}$$

so that the split-Cayley algebra provides another solution to the problem of factoring the d'Alembertian.

This introduction is sufficient for our present needs. However, in order to give some perspective, we describe a theorem (cf. Jacobson[9]) to clarify the relation between Cayley algebras and the more familiar real, complex, and quaternion algebras.

Let $\mathcal{U}$ be a real linear space with elements $\{x, y, \cdots\}$ and equipped with a multiplication of elements, xy. Although it is not necessary[10] we assume a unit element, I. Suppose also that $\mathcal{U}$ has a quadratic form, $Q(x)$; that is, $Q(x)$ satisfies,

$$Q(\alpha x) = \alpha^2 Q(x),$$

with α a real number, and where

$$\langle x, y \rangle = \tfrac{1}{2}[Q(x + y) - Q(x) - Q(y)]$$

is a (nondegenerate) bilinear form. Further assume that the quadratic form is multiplicative:

$$Q(xy) = Q(x)Q(y).$$

From these hypotheses we conclude

(A) The dimension, n, of this space can be only 1, 2, 4 or 8.

(B) For each n ($n \neq 1$, an obvious exception) there are two and only two possible quadratic forms. Either the signature is n (diagonal form has all $+1$'s or it is zero (diagonal form has half $+1$'s and half -1's). If the signature is n then the algebra is a division algebra, i.e., every nonzero element has an inverse. The division algebras are respectively the real, complex, quaternion and octonion numbers. These algebras have positive definite quadratic form. If the signature of the form is zero, the algebra is called split; there are, for example, nonzero elements, x, with $Q(x) \neq 0$. The Pauli matrices with the 2-by-2 identity adjoined form a basis of the split-quaternion algebra.

Thus the hypotheses of this theorem are quite restrictive, allowing only seven solutions. The Cayley algebra, split or not, is the most structured of a very restrictive class of algebras. All the other algebras of this class are familar to physicists.

3. THE MAXWELL EQUATIONS

By way of example, we use the split-Cayley algebra introduced in the previous section to discuss several of the equations of classical electromagnetic theory.[11] It is convenient to introduce the notation

$$\mathbf{E}_\pm = \mathbf{E} \pm \mathbf{B},$$

in terms of which the Maxwell equations become

$$\nabla\cdot\mathbf{E}_\pm = \rho,$$

$$\nabla\times\mathbf{E}_\pm \mp \partial_4\mathbf{E}_\mp = \pm\,\mathbf{j}.$$

The role of the four-potential, A_μ, is assumed by the split-Cayley element $\mathfrak{A} = (A;A)$. Application of the derivative operator $\bar{D}$ to $\mathfrak{A}$ gives immediately

$$\bar{D}\mathfrak{A} = (\mathbf{E}_+, Z; \mathbf{E}_-, Z),$$

where

$$Z = \partial_4\phi - \nabla\cdot\mathbf{A}.$$

Thus if the Lorentz gauge condition $Z = 0$ is assumed, the role of the electromagnetic field tensor can be assigned to the Cayley element $m = (\mathbf{E}_+, 0; \mathbf{E}_-, 0)$ and this is "derived" from the potential $\mathfrak{A}$ by

$$\bar{D}\mathfrak{A} = m. \tag{21}$$

Application of D to the above equation gives

$$Dm = \bar{J}, \tag{22}$$

where $J = (\mathbf{j}, \rho; \mathbf{j}, \rho)$, and which, in view of Eq. (19) implies that

$$D\bar{D}\mathfrak{A} = -(\nabla^2 - \partial_{44})\mathfrak{A} = \bar{J}. \tag{23}$$

The following relations containing the elements of the conventional stress-energy tensor can be easily verified

$$\tfrac{1}{2}(mm^T + m^Tm) = (\mathbf{E}^2 + \mathbf{B}^2)I$$

and

$$\tfrac{1}{4}(mm^T - m^Tm) = (\mathbf{E}\times\mathbf{B}, \mathbf{E}\cdot\mathbf{B}; -\mathbf{E}\times\mathbf{B}, -\mathbf{E}\cdot\mathbf{B}).$$

The symbol T denotes the transpose of a Cayley element.

Thus, from a strictly formal point of view, these fundamental relations of electromagnetic theory are realized on the split-Cayley algebra. In the next section, we extend these results to the Yang–Mills field, where there is additional algebraic structure.

4. THE YANG-MILLS FIELD EQUATIONS (B)

In order to discuss the Yang–Mills equations we proceed, as in the previous section, and in place of the gauge potentials, Γ_μ, introduce the element

$$\gamma = \begin{pmatrix} \Gamma_4 & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \Gamma_4 \end{pmatrix}. \tag{24}$$

Notice, however, that γ is not an element of the split-Cayley algebra since its "matrix elements", Γ_μ, are themselves matrices. This additional algebraic structure distinguishes a non-Abelian gauge field from an Abelian one. It is characteristic of the Yang–Mills field.

We adopt the formal multiplication defined by Eq. (14); the "dot" and "cross" products involved are those suitable to matrices as described in the Appendix. We also adopt the definition of $\bar{\mathfrak{A}}$, although this need not be an involution. With these conventions established, a straightforward calculation yields

$$(\bar{D} + \bar{\gamma})\gamma = \begin{pmatrix} Z & \mathbf{V}+\mathbf{U} \\ \mathbf{V}-\mathbf{U} & Z \end{pmatrix},$$

where V and U are defined in Eqs. (3) and (4), and where

$$Z = \nabla_\mu\Gamma_\mu = g^{\mu\lambda}(\partial_\mu + \Gamma_\mu)\Gamma_\lambda. \tag{25}$$

If we adopt the gauge condition, $Z \equiv 0$, the Yang–Mills field tensor is replaced by the element $R = (\mathbf{V}-\mathbf{U}, 0; \mathbf{V}+\mathbf{U}, 0)$. By identifying $D + \gamma$, acting on the gauge potential, γ, as a covariant Cayley derivative, δ, the field tensor is "derived" from the potential via

$$\bar{\delta}\gamma = R. \tag{26}$$

Continuing, we differentiate R by calculating

$$DR = \begin{pmatrix} J_4 - J_4' & \mathbf{J}-\mathbf{J}' \\ \mathbf{J}+\mathbf{J}' & J_4 + J_4' \end{pmatrix}, \tag{27}$$

where J_μ and J_μ' are defined by Eqs. (II) and (11), respectively. This formula can be expressed in a recognizable form by first using Eq. (III) to replace J_μ by $j_\mu - K_\mu$, so that

$$DR = \mathcal{J} - \begin{pmatrix} K_4 + J_4' & \mathbf{K}+\mathbf{J}' \\ \mathbf{K}-\mathbf{J}' & K_4 - J_4 \end{pmatrix} \tag{28}$$

with $\mathcal{J} = (j;j)$. The last term in Eq. (28) can be explicitly evaluated using the definition of K_μ, and the identity (A11) from the Appendix. Specifically, we obtain the relation,

$$\gamma R - (R^T\gamma)^T = \begin{pmatrix} K_4 + J_4' & \mathbf{K}+\mathbf{J}' \\ \mathbf{K}-\mathbf{J}' & K_4 - J_4' \end{pmatrix}.$$

Finally, we can write,

$$DR + \gamma R - (R^T\gamma)^T = \mathcal{J}. \tag{29}$$

Recalling the expression for the covariant divergence of the field tensor,

$$\nabla_\mu R_{\mu\nu} = \partial_\mu R_{\mu\nu} + \Gamma_\mu R_{\mu\nu} - R_{\mu\nu}\Gamma_\mu,$$

we identify the left hand side of Eq. (29) with the covariant Cayley derivative of R, and write

$$\delta R = \mathcal{J}. \tag{30}$$

Thus from the gauge potential, γ, the first covariant Cayley derivative gives the field tensor

$$\bar{\delta}\gamma = R, \tag{31}$$

and this formula summarizes the information in Eq. (I), the definition of R. The second derivative relates the field to the current

$$\delta R = \mathcal{J} \tag{32}$$

and summarizes the information in Eq. (II′). The last two formulas then are the dynamical equations of the Yang–Mills field.

5. SUMMARY AND DISCUSSION

The result of the previous sections has been to rewrite the Yang–Mills field equations. This has been a purely formal procedure, but one emphasizing both the algebraic and differential-geometric aspects of the gauge fields. These fields (referred to a basis in the internal symmetry space) are multicomponent functions of space-time, $\Gamma^A_{\mu B}(x)$. They carry indices which refer to both the space-time geometry and to an internal symmetry space. The internal symmetry indices are absorbed by considering the collection of matrices Γ_μ,

$\mu = 1,2,3,4$. The space–time index, indicating that these Γ_μ are defined on the space tangent to the Minkowski manifold, is absorbed by the split-Cayley algebra. In this way, the entire collection $\{\Gamma^A_{\mu B}\}$ is compacted into a single algebraic object γ, the gauge field. By introducing the differential operators δ (covariant Cayley derivative), which act upon these gauge fields, the field equations are reproduced. Thus, an algebra into which both space–time and internal symmetry have been synthesized results. Accompanying this algebra are differential operators which allow the dynamical equations of the field to be stated succintly.

In general then, the gauge fields are drawn from the tensor product $\mathcal{C} \otimes M_n$, of the split-Cayley algebra and the ring of n-by-n matrices, M_n. The full matrix ring is, of course, too general for physical gauge fields and suitable restrictions are imposed upon it. Conventionally these restrictions require that the gauge fields are in the Lie algebra of the gauge group. However, it is possible to limit the gauge fields by means of a different sort of restriction, namely the requirement that $\mathcal{C} \otimes M_n$ have a given algebraic structure. To illustrate this possibility we shall examine the requirement that $\mathcal{C} \otimes M_n$ be the Jordan algebra M_3^8. The choice of this example is motivated by the central role it plays in the work of Gamba[5] and the fact that it is an algebra of unitary 3-by-3 matrices over the split-Cayley algebra.

With this in mind, we shall now examine the condition that the gauge field algebra is in fact M_3^8. First, consider objects in the form of Zorn vector-matrices, but with each entry a 3-by-3 matrix, that is,

$$\gamma = \begin{pmatrix} \rho & \mathbf{r} \\ \mathbf{r} & \rho \end{pmatrix}$$

with

$$\rho = \{p^A_B\}$$

$$\mathbf{r} = \{\mathrm{r}^A_B\} = \{\gamma_{1B}^A, \gamma_{2B}^A, \gamma_{3B}^A\}, \quad A,B = 1,2,3.$$

This array (of 36 entries) can be rearranged into a 3-by-3 matrix with split-Cayley matrix elements. Specifically,

$$\gamma = \{\gamma^A_B\}, \quad A,B = 1,2,3,$$

where

$$\gamma^A_B = \begin{pmatrix} \rho^A_B & \mathbf{r}^A_B \\ \mathbf{r}^A_B & \rho^A_B \end{pmatrix}.$$

Now an element of M_3^8 must have the form

$$\begin{pmatrix} W_1 & X & Y \\ \overline{X} & W_2 & Z \\ \overline{Y} & \overline{Z} & W_3 \end{pmatrix},$$

where W_1, W_2, and W_3 are real numbers; X, Y, and Z are elements of the split-Cayley algebra with $\overline{X}, \overline{Y}$, and $\overline{Z}$ their involutes. Requiring that γ have this form implies

$$\Gamma_4^T = \Gamma_4, \quad \mathbf{\Gamma}^T = -\mathbf{\Gamma}.$$

These symmetry conditions on the gauge fields operators suggest that Γ_4 is a unitary scalar and tensor field, while $\mathbf{\Gamma}$ is a unitary vector-field. Apparently, with respect to the unitary symmetry space Γ_4 and $\mathbf{\Gamma}$ are orthogonal, i.e., transverse to each other. We shall not pursue this topic further here.

APPENDIX

Here we collect the definitions and some properties of the Γ_μ matrices used in the text. $\Gamma_\mu = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4\}$ are four matrices of unspecified dimension; each element, $\Gamma^A_{\mu B}$ is a function of the space–time point, x. These functions are assumed sufficiently differentiable to allow the commutation of mixed second partial derivatives. From these we define the six matrices.

$$\lambda_k = [\Gamma_4, \Gamma_k], \quad k = 1,2,3 \tag{A1}$$

and

$$\omega_k = \epsilon_{klm}\, \Gamma_l \Gamma_m. \tag{A2}$$

So, for example,

$$\omega_1 = [\Gamma_2, \Gamma_3]. \tag{A3}$$

Those vectors with $k = 1,2,3$ are indicated by $\mathbf{\Gamma}, \boldsymbol{\lambda}$ and $\boldsymbol{\omega}$. Among any two such vectors $\mathbf{A}$ and $\mathbf{B}$ we define scalar and vector multiplication

$$\mathbf{A}\cdot\mathbf{B} = A_k B_k,$$
$$(\mathbf{A} \times \mathbf{B})_k = \epsilon_{klm}\, A_l B_m.$$

The result in both cases is again a matrix. The vector algebra of such objects is straightforward, care being taken to account for the fact that the components of such vectors do not commute.

The following identities are easily checked.

$$\mathbf{\Gamma}\cdot\boldsymbol{\omega} = \boldsymbol{\omega}\cdot\mathbf{\Gamma}, \tag{A4}$$

$$\mathbf{\Gamma} \times \boldsymbol{\lambda} + \boldsymbol{\lambda} \times \mathbf{\Gamma} = \Gamma_4\boldsymbol{\omega} - \boldsymbol{\omega}\Gamma_4. \tag{A5}$$

The first of the above corresponds to the usual identity arising from the "exchange of the dot and the cross".

In a similar manner we can introduce differential operators ∂_4 and $\boldsymbol{\nabla}$ which act on each element of the matrices. For any such vector $\mathbf{A}$ or scalar ϕ

$$\begin{aligned} \boldsymbol{\nabla}\cdot\mathbf{A} &= \partial_k A_k, \\ (\boldsymbol{\nabla} \times \mathbf{A})_k &= \epsilon_{klm}\, \partial_l A_m, \\ (\boldsymbol{\nabla}\phi)_k &= \partial_k \phi. \end{aligned} \tag{A6}$$

The familiar relations

$$\begin{aligned} &\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla} \times \mathbf{A}) = 0, \quad \boldsymbol{\nabla} \times (\boldsymbol{\nabla}\phi) = 0, \\ &\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A} \end{aligned} \tag{A7}$$

follow immediately. These relations are used in the derivation of $\mathbf{V}$ and $\mathbf{U}$ from the gauge potentials.

The following relations evidence the noncommutivity of the Γ_μ:

$$\boldsymbol{\nabla}\cdot\boldsymbol{\omega} = (\boldsymbol{\nabla} \times \mathbf{\Gamma})\cdot\mathbf{\Gamma} - \mathbf{\Gamma}\cdot(\boldsymbol{\nabla} \times \mathbf{\Gamma}), \tag{A8}$$

$$\boldsymbol{\nabla} \times \boldsymbol{\lambda} = (\boldsymbol{\nabla}\Gamma_4) \times \mathbf{\Gamma} + \mathbf{\Gamma} \times (\boldsymbol{\nabla}\Gamma_4) + [\Gamma_4, \boldsymbol{\nabla} \times \mathbf{\Gamma}], \tag{A9}$$

$$\partial_4\boldsymbol{\omega} = (\partial_4\mathbf{\Gamma}) \times \mathbf{\Gamma} + \mathbf{\Gamma} \times \partial_4\mathbf{\Gamma}. \tag{A10}$$

From the above equations, together with Eqs. (A2) through (A4), we derive the identities

$$\begin{aligned} &\boldsymbol{\nabla}\cdot\boldsymbol{\omega} = \mathbf{U}\cdot\mathbf{\Gamma} - \mathbf{\Gamma}\cdot\mathbf{U}, \\ &\boldsymbol{\nabla} \times \boldsymbol{\lambda} - \partial_4\boldsymbol{\omega} = [\Gamma_4, \mathbf{U}] - \mathbf{V} \times \mathbf{\Gamma} - \mathbf{\Gamma} \times \mathbf{V}. \end{aligned} \tag{A11}$$

[1]C. N. Yang and R. L. Mills, Phys. Rev. **96,** 191 (1954).
[2]P. Jordan, J. von Neumann, and E. Wigner, Ann. of Math **35,** 29 (1934).
[3]P. Jordan, Z. Phys. **80,** 285 (1933).
[4]A. Pais, Phys. Rev. Lett. **7,** 291 (1961).
[5]A. Gamba, "Proceedings of the Seminar on High Energy Physics and Elementary Particles," Trieste, published by the International Atomic Energy Agency, 1965.
[6]In all that follows, Greek indices run over (1,2,3,4), lower case Latin indices run over (1,2,3), and upper case Latin indices refer to the internal symmetry space which is usually left unspecified. We take the Minkowski metric $g_{11}=g_{22}=g_{33}=-g_{44}=+1$. The usual summation convention is assumed. The B_μ and $F_{\mu\nu}$ of Ref. 1 are related to quantities defined here by $\Gamma_\mu=i\epsilon B_\mu$ and $R_{\mu\nu}=i\epsilon F_{\mu\nu}$.
[7]R. Utiyama, Phys. Rev. **101,** 1597 (1956).
[8]See W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley, Reading, Mass., 1955)
[9]N. Jacobson, Rend. Circ. Mat. Palermo **7,** 55 (1958). I am grateful to Professor H. Brown and Professor R. Brown of The Ohio State University for valuable discussions of this paper.
[10]As long as there is one element x such that $Q(x)\neq 0$, then this element can be "normalized" to give a unit.
[11]The vector potential used here differs by exactly a sign from that defined in many texts, e.g., Ref. 8. This sign difference has no effect on the equations for the electric and magnetic field intensities; our choice allows the electric field intensity to be expressed as the spatial part (F_{4k}) of the 4-curl of A.

Linear operators and transfer of radiation with spherical symmetry

Alan P. Wang

Department of Mathematics, Arizona State University, Tempe, Arizona 85281
(Received 15 January 1973)

This paper is concerned the construction of linear-operator equations for transfer of radiation taking place on a spherical shell. A complete set of equations is obtained for inhomogeneous, anisotropically scattering media with internal or external illumination and with an arbitrary reflecting core. In application, linear-operator equations are reduced to a class of familiar functional equations. Our general result provides answers to a set of well-known problems in astrophysics.

I. INTRODUCTION

In recent years the radiative transfer problem has been enlarged in scope in two directions. In the field of planetary and stellar atmosphere physics and in the field of neutron transport, extension of the slab geometry is needed to take curved surfaces into consideration. This has led to the study of radiative transfer in spherical media, by Bellman and Ueno and others.[1-25] The other direction is in the extension of mathematical construction to a class of linear operators. The resulting theory gives a simple and unified derivation of many integral-differential equations for transport problems. Redheffer and Wang[26-29] have successfully developed the theory governing the transmission and reflection operators for microwave and transport problems in slab geometry.

The problem of radiative transfer in a slab was attacked by Ambarzumian[30] by constructing an auxiliary equation of the Fredholm type. His method has been developed and extended by many others.[31-34] Later, Chandrasekhar[35] presented a complete set of integral-differential equations for the scattering and transmission functions in a slab, based on the principle of invariance. His powerful method has been applied and extended to the problems in plantary and stellar atmospheres.[36] The principle of invariance has been applied to many other transport problems, in rather diversified fields.[37-41] The earliest examples of principles of invariance are in works of McClelland[42] and Schmidt[43] dating from early in this century.

Bellman's method of invariant imbedding, originally developed for a slab, is now being used with spherical geometry.[1,2,44,45] Redheffer's operator equations, based on star-products, and used to construct a class of general transport equations in the slab case, can also be adopted to spherical geometry. This is done by a proper interpretation of the meaning of intensities and linear operators.

We shall construct a rather general linear-operator equation which governs the specular (directly reduced) and diffused radiation field. Two classes of problems are considered, one with external illumination and the other with internal illumination. When our linear-operator equations are reduced to functional equations, we have the results of Bellman and Ueno,[1,2,3] Allen, Shampine, and Wing,[22] Leong and Sen,[24,45] Aronson and Yarmush,[46] and Ueno and Wang.[47] From the astrophysics point of view we have succeeded in solving the Chandrasekhar problem in a spherical shell, Schuster's problem in the theory of line formation and the Milne problem of the diffusion of light from a center star as special cases.

II. INTENSITY AND OPERATIONS

The medium in which the transfer of radiation taking place has spherical symmetry, and the medium properties are functions of the distance from the center of symmetry only. Incident radiation is spherically uniform. The medium may be an inhomogeneous spherical shell of atmosphere which scatters anisotropically, with radii x and y, $0 < x \leq y$. The intensity in the total radiation field at radius z, $x \leq z \leq y$, at inclination $\cos^{-1}u$, $0 < u \leq 1$, to the radius vector directed in the inward direction is denoted by $I(z,-u)$ and the intensity in the outward direction with the inclination $\cos^{-1}u$, $0<u<1$, to the radius vector is $I(z,u)$.

Let the total radiation field in the inward direction at the outer surface of the spherical shell be $I(y,-u)$, $0 < u \leq 1$. Then the total radiation fields transmitted and reflected are denoted by $I(x,-v)$ and $I(y,v)$, see Fig. 1. The transmitted radiation is composed of two parts,

$$I(x,-v) = II(x,-v) + I^*(x,-v), \tag{1}$$

the specular part (the directly reduced part) and the diffuse part. However, the reflected radiation field has the diffuse part only

$$I(y,v) = I^*(y,v). \tag{2}$$

Viewing $II(x,-v)$, $I^*(x,-v)$, and $I^*(y,v)$ as outputs of a linear system due to an input $I(y,-u)$, we may write

$$\begin{aligned} &II(x,-v) = \mathbf{Q}\cdot I(y,-u), \qquad I^*(x,-v) = \tau\cdot I(y,-u), \\ &I^*(y,v) = \rho\cdot I(y,-u), \end{aligned} \tag{3}$$

where $\mathbf{Q}$, τ, and ρ are called the *specular*, *transmission*, and *reflection operators*. The operators have integral representations with kernels $Q(x,y,-v,-u)$, $\tau(x,y,-v,-u)$, and $\rho(x,y,v,-u)$, respectively. For example,

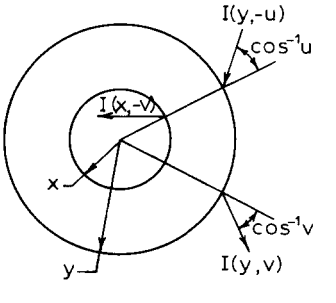

FIG. 1. Spherical symmetrical shell with incident radiation field $I_i = I(y,-u)$ at the outer surface.

Copyright © 1973 by the American Institute of Physics

$$[\tau \cdot I(y,-u)](x,-v) = \int_0^1 \tau(x,y,-v,-u)\, I(y,-u)\, du. \quad (4)$$

Similarly, if there is an incident radiation field $I(x,u)$ at x in the outward direction then the corresponding operators are denoted by $\mathbf{P}, \mathbf{t}, \mathbf{r}$ and their kernels are $P(x,y,v,u)$, $t(x,y,v,u)$, $r(x,y,-v,u)$, respectively.

It should be noted that the intensities we defined here have taken all multiplication of scattering into account. With the exception that the core is a perfect absorber, there is multiple scattering taking place between the core and the spherical shell. This included the case when the core is a vacuum. For example, if the core radiates an amount I_i of intensity at the surface in the outward direction the intensity $I(x,v) = (\mathbf{E} - \mathbf{K} \cdot \mathbf{r})^{-1} I_i$, where $\mathbf{K}$ is the reflection operator for the core, (see Sec. IV). In the slab geometry case, when the slab is imbedded in a vacuum or a nonreflective space, $I(x,v) = I_i$, because there is no multiple scattering taking place between radiative transfer in a slab and in a spherical shell. Another distinction is that while an intensity travels in a straight line in a spherical shell its parameter u changes with radius, see Fig. 2. In the slab geometry it remains the same.

III. TRANSFER OF RADIATION

Let us consider a thin spherical shell with radii $z - \Delta z$ and z for $\Delta z > 0$ small. Then the change of the radiation fields with respect to z due to the existence of this thin spherical shell involves three parts. The first and second parts are the volume attenuation and the diffusion parts,

$$\frac{\alpha(z)}{v} I(z,v) - \frac{\sigma(z)}{2v} \int_{-1}^1 p(z,v,u)\, I(z,u)\, du, \qquad 0 < |v| \leq 1, \quad (5)$$

as in the slab case, where $\alpha(z)$, $\sigma(z)$, and $p(z,v,u)$ are, respectively, the volume attenuation, scattering coefficient, and azimuth independent phase function. Intensity is function of z and v. The change in cosine of the angle due to the change of radius constitutes the third part,

$$-\frac{1-v^2}{zv} \frac{\partial}{\partial v} I(z,v). \quad (6)$$

(See Fig. 2.) The total effect is given by superposition.

For the moment, let us consider the specular part of the radiation only, i.e., the diffuse part of the contribution is ignored. The inward specular equation is given by

$$\frac{\partial}{\partial z} II(z,-v) = \left(\frac{\alpha(z)}{v} - \frac{1-v^2}{zv} \frac{\partial}{\partial v}\right) II(z,-v), \qquad x \leq z \leq y,\ 0 < v \leq 1. \quad (7)$$

This partial differential equation has a general solution of the form

$$\Psi[z(1-v^2)^{1/2}],$$

where Ψ is a differentiable function in z and v. In particular, if the boundary condition at y in the inward direction is given by $\delta(v-u)$, then a particular solution for (7) is

$$II(z,-v) = \delta(v_* - u) e^{-h(z,y,u)}, \quad (8)$$

where the optical thickness

$$h(z,y,u) = \int_z^y \frac{\alpha(\xi)}{u^*(\xi)}\, d\xi$$

and the superscript * and subscript $_*$ are used to denote the following relationship:

$$\left.\begin{aligned} f_* &= f_*(z,x) = \left[1 - \left(\frac{z}{x}\right)^2 (1-f^2)\right]^{1/2} \\ f^* &= f^*(z,x) = \left[1 - \left(\frac{x}{z}\right)^2 (1-f^2)\right]^{1/2} \end{aligned}\right\}, \qquad z < x.$$

We see that, $(f_*)^* = (f^*)_* = f$.

Upon substituting the Q defined in (3) into Eq. (7), we have

$$\frac{\partial}{\partial z}\mathbf{Q} = \left(\frac{\alpha(z)}{v} - \frac{1-v^2}{zv}\frac{\partial}{\partial v}\right)\mathbf{Q}, \quad (9)$$

where the operator $\mathbf{Q}$ has kernel $Q(z,y,-v,-u)$. It is clear that

$$Q(z,y,-v,-u) = \delta(v_* - u) e^{-h(z,y,u)} \quad (10)$$

is a solution of Eq. (9) with initial value $\delta(v-u)$ as $z = y$. Under the integral convention, see Eq. (4), $\mathbf{Q}$ acts as an identity when $z = y$ which agrees with the physics of the system.

We have chosen a simple case to show the derivation of the intensity and the operator equations. The solution for the intensity equation depends on the value of the incident radiation while the operator equation is independent of the incident radiation field. Also it should be noted that the Dirac delta δ used in Eq. (10) should be considered as an operator in the sense of distribution theory. More precisely, in the integral representation in Eq. (4), all operators and the Dirac delta are considered as in the sense of regular distribution, while the Dirac delta used in (8) is not a well-defined function.

Let us compute the total flux at z, $x \leq z \leq y$ with $\alpha(z) \equiv 0$ in Eq. (9) and $u_c \leq u \leq 1$, with $u_c^2 = 1 - (x/y)^2$. Then

$$\begin{aligned} &4\pi z^2 \int_0^1 \int_0^1 v_* Q(z,y,-v,-u)\, II(y,-u)\, dv du \\ &= 4\pi z^2 \int_{u_c}^1 \int_0^1 v_* \delta(v_* - u)\, II(y,-u) \frac{dv}{dv_*}\, dv_* du \\ &= 2\pi x^2 \int_0^1 u\, II(y,-u)\, du. \end{aligned}$$

That is, in the case of $\alpha(z) \equiv 0$ and $u_c \leq u \leq 1$, the total specular flux is independent of z.

For $o < u < u_c$, the specular part of the intensity will not reach the area in the spherical shell with radius less than $y_c = y(1-u^2)^{1/2}$. However, this part of the intensity is passing through the spherical shell and it becomes a reflected part (see Fig. 3). In fact,

$$II(y,v) = \delta(v-u) e^{-2h(y_c,y,u)}, \qquad o < u < u_c.$$

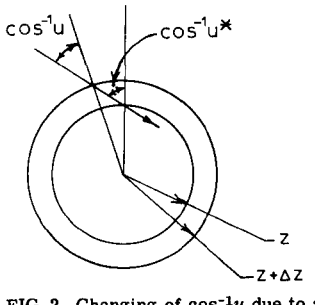

FIG. 2. Changing of $\cos^{-1}u$ due to a small change of radius, where $u^* = \{1 - [(z + \Delta z)/z]^2(1-u)^2\}^{1/2}$.

This situation does not appear in the slab case, since $u_c = 0$. And this situation also holds for a very thin spherical shell for a somewhat different reason. For incident intensity with inclination $\cos^{-1}u$, $o < u < 1$, we can always choose Δz so small that

$$u \geq u_c = \{1 - [(z - \Delta z)/z^2]\}^{1/2}.$$

In a similar manner, the specular operator equation for $\mathbf{P}$ is

$$\frac{\partial}{\partial z}\mathbf{P} = -\left(\frac{\alpha(z)}{v} + \frac{1-v^2}{zv}\frac{\partial}{\partial v}\right)\mathbf{P} \tag{11}$$

with initial valve

$$\mathbf{P} = \mathbf{E} = \text{identity for } z = y.$$

Here the operator $\mathbf{P}$ has kernel $P(y,z,v,u)$, and

$$P(y,z,v^*,u) = \delta(v^* - u)e^{-h(y,z,u)} \tag{12}$$

is a solution. We shall not discuss the properties of P, since they are analogs of the properties of Q.

IV. MEDIUM COEFFICIENTS

The differential operator-equations for the radiative transfer may be obtained when the added spherical shell is very thin. The medium coefficients at y are specified by behavior of such a thin shell extended from y to $y + \Delta y$ where Δy is small. The diffusion coefficients $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$, and $\mathbf{d}^*$ are defined as the limits of diffusion operators $\boldsymbol{\rho}, \mathbf{t}, \mathbf{r}$ and $\boldsymbol{\tau}$ associated with the thin shell. Kernels of $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*, \mathbf{d}^*$, can be represented by phase functions, which we have already used in the previous section. Phase functions are physical parameters and are also called differential cross section. In the operator notation we write

$$\begin{pmatrix} \mathbf{b}^* & \mathbf{a}^* \\ \mathbf{c}^* & \mathbf{d}^* \end{pmatrix} = \frac{\sigma}{2v}\begin{pmatrix} \mathbf{p}^{++} & \mathbf{p}^{+-} \\ \mathbf{p}^{-+} & \mathbf{p}^{++} \end{pmatrix}, \tag{13}$$

where the operator $p^{\pm\pm}$ has kernel $p(y, \pm v, \pm u)$.

To obtain the specular coefficients, we shall consider the case operators $\mathbf{Q}$ and $\mathbf{P}$ to be associated with the same thin spherical shell. Since $\mathbf{Q}$ and $\mathbf{P}$ involve $\delta(v_* - u)$ and $\delta(v^* - u)$, the limits of a left-hand and right-hand specular operators under the composite operation are somewhat different. For this reason let us compute two typical cases.

In the first case, we consider the limit of $\mathbf{P}$ on the left-hand side of a composite operation $\mathbf{P}$. That is,

$$\lim_{\Delta y\to 0}\frac{1}{\Delta y}[\mathbf{P}\cdot\boldsymbol{\rho} - \boldsymbol{\rho}](y,v) = \lim_{\Delta y\to 0}\frac{1}{\Delta y}\int_0^1 \times [\delta(v^* - w)e^{-h(y,y+\Delta y,w)} - \delta(v - w)]\rho(x,y,w,-u)\,dw$$
$$= -\left(\frac{\alpha(y)}{v} + \frac{1-v^2}{yv}\frac{\partial}{\partial v}\right)\rho(x,y,v,-u), \tag{14}$$

where the last equality follows from Eq. (11) and $P(y, y + \Delta y, -v, -w) = \delta(v - w)$ when $\Delta y \to 0$.

In the second case we consider the limit of $\mathbf{Q}$ on the right-hand side of a composite operation $\boldsymbol{\rho}\cdot\mathbf{Q}$. Let us compute

$$[\boldsymbol{\rho}\cdot\mathbf{Q}](y,v) = \int_0^1 \rho(x,y,v,-w)\delta(w_* - u)e^{-h(y,y+\Delta y,u)}dw$$
$$= \left(\frac{y+\Delta y}{y}\right)^2\left(\frac{u}{u^*}\right)\int_0^{u_c}\rho(x,y,v,-w)\delta(w - u^*)\,dw,$$

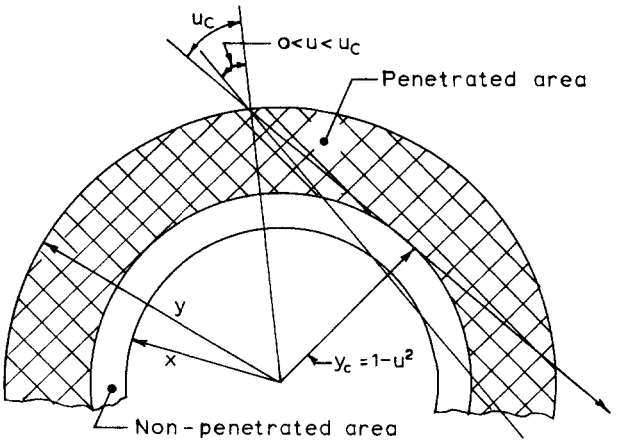

FIG. 3. Section of spherical shell with specular part of radiation only, showing nonpenetrated area.

where $u^* = u^*(y, y + \Delta y)$ and $u_c = \{1 - [(y + \Delta y)/y]\}^{1/2}$. From this it follows that by taking the limit,

$$\lim_{\Delta y\to 0}\frac{1}{\Delta y}[\boldsymbol{\rho}\cdot\mathbf{Q} - \boldsymbol{\rho}](y,v)$$
$$= -\int_0^{u_c}\rho(x,y,v,-w) \times \left(\frac{\alpha(y)}{w} - \frac{2}{y} - \frac{1-w^2}{yw^2} + \frac{1-w^2}{yw}\frac{\partial}{\partial w}\right)\delta(w - u^*)\,dw|_{u^*=u}$$
$$= -\left(\frac{\alpha(y)}{u} - \frac{2}{y} - \frac{1-u^2}{u^2y} - \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)\rho(x,y,v,-u), \tag{15}$$

where we used Eq. (9) and the fact that for $o < u \leq 1$,

$$o = \frac{\partial}{\partial w}\int_0^{u_c}\rho(x,y,v,-w)\delta(w - u^*)\,dw$$
$$= \rho(x,y,v,-w)\delta(w - u^*)|_{w=0}^{w=u_c}$$

since

$$u^* = u^*(y, y + \Delta y) = \left[1 - \left(\frac{y+\Delta y}{y}\right)^2(1-u^2)\right]^{\frac{1}{2}} \neq u_c$$

or 0.

In view of (14) and (15) where ρ is arbitrary, we have that the specular coefficients at y corresponding to $\mathbf{P}$ and $\mathbf{Q}$ are coefficients $\mathbf{B}$ and $\tilde{\mathbf{D}}$, and their kernels have the form

$$B(y,v,u) = -\left(\frac{\alpha(y)}{v} + \frac{1-v^2}{yv}\frac{\partial}{\partial u}\right)\delta(v - u),$$
$$\tilde{D}(y,-v,-u) = -\left(\frac{\alpha(y)}{v} - \frac{2}{y} - \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)\delta(v - u), \tag{16}$$

with the understanding that $B(y,v,u)$ is used in the left-hand of a composite operation and $\tilde{D}(y,-v,-u)$ in the right-hand, signified by superscript $\sim$. In a similar manner, the operator $\tilde{B}$ and D have kernels

$$\tilde{B}(y,v,u) = -\left(\frac{\alpha(y)}{u} + \frac{2}{y} + \frac{1-u^2}{yu^2} - \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)\delta(v - u),$$
$$D(y,-v,-u) = -\left(\frac{\alpha(y)}{u} + \frac{1-v^2}{yu}\frac{\partial}{\partial v}\right)\delta(v - u). \tag{17}$$

By (16) and (17), Eqs. (9) and (12) can be expressed as

$$\frac{\partial}{\partial y}\mathbf{Q} = \mathbf{Q}\cdot\tilde{\mathbf{D}} \quad \text{and} \quad \frac{\partial}{\partial y}\mathbf{P} = \mathbf{B}\cdot\mathbf{P}. \tag{18}$$

V. STATE AND LOCAL FORM

With coefficients as constructed in Eqs. (13), (16), and (17), one can obtain a complete system of equations which govern the radiative transfer under consideration. This is done by Redheffer's method of adding a thin layer and using the technique of star-products. The method was developed originally for the slab. Mathematically, the theory is constructed on the class of linear operators on Hilbert space. Therefore, it can apply equally well to the spherical shell provided the meaning of the intensity is understood in the way defined here, see remarks at the end of the second section. We take the star-product of the scattering matrix associated with radii x and y, $0 < x \leq y$, defined by

$$S = \begin{pmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{Q} \end{pmatrix} + \begin{pmatrix} \mathbf{t} & \rho \\ \mathbf{r} & \tau \end{pmatrix}, \tag{19}$$

with another scattering matrix associated with radii y and $y + \Delta y$, then take the limit as $\Delta y \to 0$. With the aide of Eqs. (13), (16), (17), and (18), we obtain a set of operator equations (for more detail, see Ref. 27),

$$t_y = (\mathbf{b} + \rho \cdot \mathbf{c})(\mathbf{t} + \mathbf{P}) - \mathbf{B} \cdot \mathbf{P}, \tag{20a}$$

$$\rho_y = \mathbf{a} + \mathbf{b} \cdot \rho + \rho \cdot \mathbf{d} + \rho \cdot \mathbf{c} \cdot \rho, \tag{20b}$$

$$r_y = (\tau + \mathbf{Q}) \cdot \mathbf{c} \cdot (\mathbf{t} + \mathbf{P}), \tag{20c}$$

$$\tau_y = (\tau + \mathbf{Q}) \cdot (\mathbf{d} + \mathbf{c} \cdot \rho) - \mathbf{Q} \cdot \tilde{\mathbf{D}}, \tag{20d}$$

where the subscript y denotes partial differentiation, and where the coefficients

$$\begin{pmatrix} \mathbf{b} & \mathbf{a} \\ \mathbf{c} & \mathbf{d} \end{pmatrix} = \begin{pmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{D} \end{pmatrix} + \begin{pmatrix} \mathbf{b}^* & \mathbf{a}^* \\ \mathbf{c}^* & \mathbf{d}^* \end{pmatrix} \tag{21}$$

are evaluated at y.

In the slab case, one will obtain another set of operator-equations by adding a thin layer at x. There is a remarkable symmetry between the partial differentiation with respect to y and that with respect to x. However, for the spherical shell one should not expect complete symmetry, since from the physical point of view the shell being added at y is different from that added at x. This also can be explained from the mathematical point of view: under the composite operation the left-hand coefficient is different from the right-hand coefficient, as discussed in the previous section. However, the theory of an additional layer and star-products is still valid. Details are not presented here, we merely state the result:

$$-\mathbf{t}_x = \mathbf{t} \cdot (\bar{\mathbf{b}} + \mathbf{a} \cdot \mathbf{r}) - \mathbf{P} \cdot \tilde{\mathbf{B}}, \tag{22a}$$

$$-\rho_x = (\mathbf{t} + \mathbf{P}) \cdot \mathbf{a} \cdot (\tau + \mathbf{Q}), \tag{22b}$$

$$-\mathbf{r}_x = \mathbf{c} + \bar{\mathbf{d}} \cdot \mathbf{r} + \mathbf{r} \cdot \bar{\mathbf{b}} + \mathbf{r} \cdot \mathbf{a} \cdot \mathbf{r}, \tag{22c}$$

$$-\tau_x = (\bar{\mathbf{d}} + \mathbf{r} \cdot \mathbf{a}) \cdot \tau - \mathbf{D} \cdot \mathbf{Q}, \tag{22d}$$

with $\bar{\mathbf{b}} = \tilde{\mathbf{B}} + \mathbf{b}^*$, $\bar{\mathbf{d}} = \mathbf{D} + \mathbf{d}^*$ and the coefficients evaluated at x. Equations (20) and (22) are called the *state forms for a symmetric spherical shell*. They correspond to the state form for a slab if $u^* = u_* = u$ and $v^* = v_* = v$. In this case x and y are slab depth.

Let us consider the intensities on both sides of a thin spherical shell with radii z and $z + \Delta z$ where Δz is small. We have

$$I(z + \Delta z, v) = (\mathbf{P} + \mathbf{t}) \cdot I(z, u) + \rho \cdot I(z + \Delta z, -u),$$
$$I(z, -v) = \mathbf{r} \cdot I(z, u) + (\mathbf{Q} + \tau) \cdot I(z + \Delta z, -u),$$

where operators are associated with the thin spherical shell. Upon taking the limits as $\Delta z \to 0$ and using coefficients as stated in Eqs. (13), (16), and (17), the following linear system is obtained:

$$\frac{\partial}{\partial z} \begin{pmatrix} I(z, v) \\ I(z, -v) \end{pmatrix} = \left[\begin{pmatrix} \mathbf{B} & 0 \\ 0 & -\mathbf{D} \end{pmatrix} + \begin{pmatrix} \mathbf{b}^* & \mathbf{a}^* \\ -\mathbf{c}^* & -\mathbf{d}^* \end{pmatrix} \right] \cdot \begin{pmatrix} I(z, v) \\ I(z, -v) \end{pmatrix}. \tag{23}$$

This intensity-equation is called the *local form* for a symmetric spherical shell and all coefficients are evaluated at z, $x \leq z \leq y$. It is presented in the decomposed form, the first term on the right-hand side of Eq. (23) being the specular part and the second term being the diffuse part. Using Eq. (21), the above local form appears identical to that for a slab, with somewhat different meanings for the coefficients.

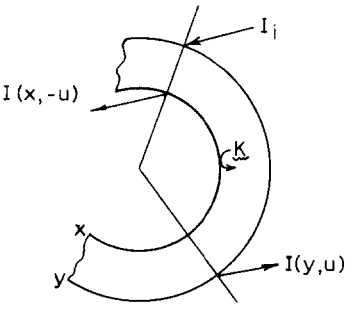

FIG. 4. A section of spherical shell with external illumination.

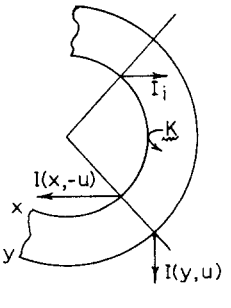

FIG. 5. A section of spherical shell with internal illumination.

VI. THE REFLECTING CORE PROBLEMS

Let us consider a spherical symmetric shell of atmosphere surrounding a reflecting core with a reflection governed by an operator **K**. We consider two types of problems. In problem type a the shell is externally illuminated; while in problem type b the shell is internally illuminated. See Figs. 4 and 5, where I_i is the illuminating radiation. These are the fundamental problems in the theories of the illumination of the sky and of the planetary illumination.

The basic tasks for problems of type a and b are the determination of the intensity field at the outside of the spherical shell in the outward direction, $I(y, u)$, and the specification of the radiation field, $I(x, -u)$, as seen by an observer at the surface of the core looking at the atmosphere.

Before we proceed with the physical meanings of such problems, we shall give a discussion of the operator **K**, see Eqs. (4) and (5). The precise mathematical meaning of **K** is that, when there is no spherical shell, it relates $I(x,u)$ and $I(x,-v)$ by

$$I(x,u) = \mathbf{K}\cdot I(x,-v). \tag{24}$$

When the core is a vacuum, then $\mathbf{K} = \mathbf{V}$ where **V** has kernel $\delta(v+u)$ since there is no absorbtion or diffusion taking place, and all incident intensities $I(x,u)$ go through the core and constitute an output $I(x,-u)$ where the total flux is preserved (see Sec. 2), but the direction is reversed. Whereas when a slab is imbedded in a vacuum, since there is no intensity returning from the vacuum, we have $\mathbf{K} = \mathbf{0}$ = zero operator. Another case in a spherical shell $\mathbf{K} = \mathbf{V}$ occurs when the core is made of a perfect reflecting material. On the other hand, if the core is made of a perfect absorbing material, such as a perfect black body, then $\mathbf{K} = \mathbf{0}$. Of course similar situations hold if a slab is imbedded in a perfectly reflecting or absorbing material.

Problem type a, when $\mathbf{K} = \mathbf{0}$ corresponds to the "standard radiative transfer" with spherical symmetry. When **K** has kernel $K = K(v,u) = Avu$, where A is a constant, we have the "Chandrasekhar's problem" with spherical symmetry. When $\mathbf{K} = \mathbf{V}$, we have a problem of a spherical shell imbedded in a vacuum. We may view problem type a as a sort of generalized Chandrasekhar's problem in a spherically symmetric shell.

In problem type b, when $\mathbf{K} = 0$, we have the "Schuster's problem" in the theory of line formation by a perfectly absorbing core. When $\mathbf{K} = Avu$ we have the problem of a Lambert reflecting core, which radiates intensities into the surrounding spherical shell of atmosphere. When $\mathbf{K} = \mathbf{V}$, we have the "Milne's problem" of the diffusion of light from a center star.

We shall attack those two types of problems by constructing a single ideal model and take advantage of the star-product.

Instead of considering a core with given reflection operator K, we assume that there is another spherical shell inside of the given one with radii x_1 and x, $0 < x_1 \leq x$. This inner shell has scattering matrix

$$S_1 = \begin{pmatrix} \mathbf{t}_1 & \mathbf{K} \\ \mathbf{r}_1 & \tau_1 \end{pmatrix},$$

where t_1 is chosen so that $t_1 \cdot I(x_1,v) = I_i$ and τ_1 and r_1 are unrestricted. The condition imposed on t_1 is required only for problem type b. The replacing of a core by such an inner spherical shell does not alter the results of our problems.

Using the star-product of S_1 and S as given by Redheffer[27] and $(E - r\cdot K)$ nonsingular, we have

$$S_1 * S = \begin{pmatrix} (\mathbf{P}+\mathbf{t})\cdot(\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}t_1 & \rho + (\mathbf{P}+\mathbf{t})\cdot\mathbf{K}\cdot(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot(\mathbf{Q}+\tau). \\ \mathbf{r}_1 + \tau_1\cdot\mathbf{r}\cdot(\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}\cdot t_1 & \tau_1\cdot(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot(\mathbf{Q}+\tau) \end{pmatrix}. \tag{25}$$

Equation (25) denotes the overall transmission and reflection by taking account of multiple scattering. The overall reflection at radius y is

$$\rho(K) = \rho + (\mathbf{P}+\mathbf{t})\cdot\mathbf{K}\cdot(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot(\mathbf{Q}+\tau), \tag{26}$$

where t, τ, ρ, and r are solutions of (20) and (22) subjected to the initial conditions

$$S = \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{E} \end{pmatrix}, \quad \text{when } x = y.$$

The reflected radiation at y in the outward direction due to the incident radiation I_i at y is

$$I(y,v) = I^*(y,v) = \rho(K)\cdot I_i. \tag{27}$$

This result is the desired first part of the answer for problem of type a. To obtain the remaining part of the result, we note that the overall transmission at x_1 due to the incident radiation I_i at y is

$$\tau_1\cdot(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot(\mathbf{Q}+\tau)\cdot I_i.$$

Assuming τ_1 is nonsingular (this restruction can be removed, see Appendix), the total transmissitted intensity at the surface of the core in the inward direction is given by

$$I(x,-v) = (\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot(\mathbf{Q}+\tau)\cdot I_i. \tag{28}$$

Equation (28) should be separated into two parts, the specular part and the diffuse part, by

$$\begin{aligned} II(x,-v) &= \mathbf{Q}\cdot I_i, \\ I^*(x,-v) &= \tau(K)\cdot I_i, \end{aligned} \tag{29}$$

where

$$\tau(K) = (\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1}\cdot[\tau + \mathbf{r}\cdot\mathbf{K}\cdot\mathbf{Q}]. \tag{30}$$

Thus, the problem of type a is solved for arbitrary **K** provided $(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})$ is nonsingular.

For the problem of type b, we use the fact that the intensities at x can be expressed as

$$I(x,-v) = \mathbf{r}\cdot I(x,u)$$

and

$$I(x,v) = \mathbf{t}_1\cdot I(x_1,u) + \mathbf{K}\cdot I(y,-u) = I_i + \mathbf{K}\cdot I(y,-u).$$

The last equality follows from the construction of $\mathbf{S}_1$. By eliminating $I(x,-v)$, we have

$$I(x,v) = (\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}\cdot I_i. \tag{31}$$

It follows immediately that

$$I(y,v) = (\mathbf{P}+\mathbf{t})\cdot I(x,u) = (\mathbf{P}+\mathbf{t})\cdot(\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}\cdot I_i$$

since there is no incident radiation at y in the inward direction. Therefore, the specular and diffuse parts at y in the outward direction are

$$\begin{aligned} II(y,v) &= \mathbf{P}\cdot I_i, \\ I^*(y,v) &= \mathbf{t}(K)\cdot I_i, \end{aligned} \tag{32}$$

where

$$\mathbf{t}(K) = [\mathbf{t} + \mathbf{P}\cdot\mathbf{K}\cdot\mathbf{r}]\cdot(\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}. \tag{33}$$

Equations (32) are the transmitted intensities at y due to

an incident radiation I_i at the surface of the core in the outward direction. The reflected intensity at x is

$$I(x,-v) = I^*(x,-v) = \mathbf{r}\cdot I(x,u) = \mathbf{r}(K)\cdot I_i, \tag{34}$$

where

$$\mathbf{r}(K) = \mathbf{r}\cdot(\mathbf{E}-\mathbf{K}\cdot\mathbf{r})^{-1}. \tag{35}$$

The initial valves are $\rho(K) = \mathbf{K}$ and zero for $\mathbf{t}(K)$, $\tau(K)$ and $\mathbf{r}(K)$ at $x = y$. This statement can be easily checked by taking the limit as $x \to y$. It also agrees with the physics of our system.

Equations (32)–(35) are the desired results for the problem of type b.

VII. SPECIAL CASES AND APPLICATIONS

The complete system of operator equations introduced in previous sections may seem to be unfamiliar and somewhat abstract. To present them in a more familiar form and also to give us a partial check of our results we shall reduce some of the above operator equations to a set of functional equations. Also a brief discussion on systems with various reflecting operators $\mathbf{K}$ is presented.

To discuss in more detail the generalized Chandrasekhar's radiative problem in spherical geometry, the problem of type a, we assume that the core has an arbitrary reflecting kernel $K(v,u)$. The conical flux of radiation of unit intensity per unit area per unit solid angle is assumed to be spherically uniformly incident on the outer surface with radius y at inclination $\cos^{-1}u$, $0 < u \leq 1$, to the inward-directed radius vector, i.e.,

$$I_i = \delta(v-u).$$

To obtain the reflection functional equation at y, we let operators on both sides of Eq. (20c) operate on I_i under our integral convention, see Eq. (4). With the aide of Eqs. (13), (16), and (17), we have

$$\begin{aligned}
&\left[\frac{\partial}{\partial y} + \alpha(y)\left(\frac{1}{u}+\frac{1}{v}\right) + \frac{1-v^2}{yv}\frac{\partial}{\partial v} + \frac{1-u^2}{yu}\frac{\partial}{\partial u} - \frac{1+u^2}{yu^2}\right]\\
&\quad\times \rho(x,y,v,-u)\\
&= \frac{\sigma}{2}\left(p(y,v,-u) + \int_0^1 p(y,v,w)\rho(x,y,w,-u)\,dw + v\right.\\
&\quad\left.\times\int_0^1 \rho(x,y,v,-w)p(y,-w,-u)\frac{dw}{w}\right)\\
&\quad + v\int_0^1\int_0^1 \rho(x,y,-v,-w)p(y,-w,w')\\
&\quad\times\rho(x,y,w',u)\,dw'\frac{dw}{w},
\end{aligned} \tag{36}$$

with $\rho = 0$ when $x = y$. Likewise using Eqs. (10) and (12), Eq. (26) is reduced to

$$\begin{aligned}
\rho(K;y,v,-u) &= \rho(y,v,-u) + e^{-2h(x,y,u)}(y/x)^2\\
&\quad\times(u/u^*)k(K,v^*,-u_*) + e^{-h(x,y,u)}\int_0^1 k(K,v^*,-w)\\
&\quad\times\tau(x,y,-w,-u)\frac{dw}{w} + e^{-h(x,y,u)}(y/x)^2(u/u^*)\\
&\quad\times\int_0^1 t(x,y,v,w)\,k(K,w,-u^*)\,dw\\
&\quad + \int_0^1\int_0^1 t(x,y,v,w)k(K,w,-w')\\
&\quad\times\tau(x,y,-w',-u)\,dw'\frac{dw}{w},
\end{aligned} \tag{37}$$

where $\rho(K,x,y,v,-u)$ is the kernel for $\rho(K)$, $u^* = u^*(x,y)$, $u_* = u_*(x,y)$ and the resolvent kernel satisfies

$$\begin{aligned}
k(K,v,-u) &= K(v,-u)\\
&\quad + \int_0^1\int_0^1 k(K,v,-w)r(x,y,-w,w')\\
&\quad\times K(x,y,w,-u)\,dw'\frac{dw}{w}
\end{aligned} \tag{38}$$

with $K(v,-u)$ as the kernel for $\mathbf{K}$.

The desired intensity at the surface of the spherical shell of atmosphere in the outward direction is given by

$$I^*(y,u) = \rho(K,x,y,v,-u). \tag{39}$$

The corresponding functional equations for Eqs. (20d) and (30) can be obtained by a similar method. They are, respectively,

$$\begin{aligned}
&\left(\frac{\partial}{\partial y} + \frac{\alpha(y)}{u} + \frac{1-u^2}{yu}\frac{\partial}{\partial u} - \frac{1+u^2}{yu^2}\right)\tau(x,y,-v,-u)\\
&\quad = \frac{\sigma}{2}\left[\frac{1}{v_*}e^{-h(x,y,u)}\left(p(y,-v_*,-u)\right.\right.\\
&\qquad\left.\left. + \int_0^1 p(y,-v_*,w)\rho(x,y,w,-u)\,dw\right)\right.\\
&\qquad + \int_0^1 \tau(x,y,-v,-w)p(y,-w,-u)\frac{dw}{w}\\
&\qquad + \int_0^1\int_0^1 \tau(x,y,-v,-w)p(y,-w,w')\\
&\qquad\left.\times\rho(x,y,w',-u)\,dw'\frac{dw}{w}\right],
\end{aligned} \tag{40}$$

with initial value $\tau = 0$ when $x = y$, and

$$\begin{aligned}
\tau(K,x,y,-v,-u) &= e^{-h(x,y,u)}(y/x)^2(u/u_*)\,k(K,-v,-u^*)\\
&\quad + \int_0^1 k(K,-v,w)\tau(x,y,-w,-u)\,dw,
\end{aligned} \tag{41}$$

where $\tau(K,x,y,-v,-u)$ is the kernel of the operator $\tau(K)$. The radiation fields as seen by an observer at the surface of the core looking at the atmosphere are given by the following specular and diffuse parts,

$$II(x,-v) = \delta(v_*-u)\,e^{-h(x,y,u)} \tag{42}$$

and

$$I^*(x,-v) = \tau(K,x,y,-v,-u). \tag{43}$$

For the "standard problem" in a spherically symmetric shell, the core is a perfect absorber $\mathbf{K} = 0$. Then

$$\rho(K,x,y,v,-u) = \rho(x,y,v,-u)$$

and

$$\tau(K,x,y,-v,-u) = \tau(x,y,-v,-u).$$

For the case $\mathbf{K} = \mathbf{V}$, Eq. (38) reduces to

$$\begin{aligned}
k(K,v,-u) &= \delta(v-u) + r(x,y,-v,u)\\
&\quad + \int_0^1 r(x,y,-v,w)r(x,y,-w,u)\,dw + \cdots.
\end{aligned} \tag{44}$$

One may use Eqs. (37), (41), and (44) to obtain $\rho(V,x,y,v,-u)$ and $\tau(V,x,y,-v,-u)$.

For the "Chandrasekhar's problem" in a spherically symmetric shell, i.e., with $\mathbf{K}$ according to Lambert's law with a constant albeds A,

$$[\mathbf{K}\cdot I(x,-u)](v) = Av\int_0^1 uI(x,-u)\,dn \tag{45}$$

for any $I(x,-u)$, for all v, $0 < v \leq 1$. The reflected radiation field is isotropic. By using Neumann's series

$$\mathbf{K}\cdot(\mathbf{E}-\mathbf{r}\cdot\mathbf{K})^{-1} = \mathbf{K}+\mathbf{K}\cdot\mathbf{r}\cdot\mathbf{K}+\mathbf{K}\cdot\mathbf{r}\cdot\mathbf{K}\cdot\mathbf{r}\cdot\mathbf{K}+\cdots$$

and repeating the operation (45) by taking the advantage of the isotropic property, we obtain

$$\begin{aligned} I^*(y,v) &= \rho(K)\cdot I_i = \rho(x,y,v,-u) + (1/vu)\bar{A}\,\underline{t}(v)\bar{\tau}(u), \\ I^*(x,-v) &= \tau(K)I_i = (1/u)\bar{A}\,\bar{\tau}(u) + (\bar{A}-1)(y/x)^2 \\ &\quad\times (u/u_*)e^{-h(x,y,u)}, \end{aligned} \tag{46}$$

where

$$\bar{A} = A(1-A\bar{r})^{-1},$$

and

$$\begin{aligned} \underline{f}(v) &= \int_0^1 vf(x,y,v,u)\,du, \\ \bar{f}(u) &= \int_0^1 vf(x,y,-v,-u)\,dv. \end{aligned} \tag{47}$$

$II^*(x,-v)$ has the same value as in Eq. (42), since it is independent of $\mathbf{K}$. In summary, the results presented in Eqs. (39), (42), and (43) are the solutions for the general spherical shell of atmosphere with an arbitrary reflecting core and with external illumination. In the case $\mathbf{K}$ given by Lambert's law, we have the Chandrasekhar's radiation problem in spherical geometry. The results are presented in Eqs. (46) and (47). Equations (46) are identical to these given by Ueno and others,[2,21,22] upon replacing ρ, τ, x, and y by $(1/v)S$, $(1/v)T, y$, and x.

As for the problem of type b, with internal illumination, $I_i = \delta(v-u)$ at x in the outward direction, we shall write down the functional equations from operator-equations (20a) and (20b), with the aide of Eqs. (13), (16), (17), and (18). The result is

$$\begin{aligned} &\left(\frac{\partial}{\partial y}+\frac{\alpha(y)}{v}+\frac{1-v^2}{yv}\frac{\partial}{\partial v}\right)t(x,y,v,u) \\ &\quad= \frac{\sigma(y)}{2}\left[\left(\frac{x}{y}\right)^2 e^{-h(x,y,u)}\left(p(y,v,u_*)\right.\right. \\ &\qquad\left.+\int_0^1 \rho(x,y,v,-w)p(y,-w,u)\frac{dw}{w}\right) \\ &\qquad+\int_0^1 p(y,v,w)t(x,y,w,u) \\ &\qquad\left.+\int_0^1\int_0^1 \rho(x,y,v,-w)p(y,-w,w')t(x,y,w',u)\,dw'\,\frac{dw}{w}\right] \end{aligned} \tag{48}$$

and

$$\begin{aligned} \frac{\partial}{\partial y}r(x,y,-v,u) &= \left(\frac{x}{y}\right)^2\left(\frac{u}{u_*}\right)\left(e^{-2h(x,y,u)}p(y,-v_*,u^*)\right. \\ &\quad+e^{-h(x,y,u)}\int_0^1 \tau(x,y,-v,-w)p(y,-w,u_*)\frac{dw}{w}\Big) \\ &\quad+e^{-h(x,y,u)}\int_0^1 p(y,-v_*,w)t(x,y,w,u)\,dw \\ &\quad+\int_0^1\int_0^1 \tau(x,y,-v,-w)p(y,-w,w')t(x,y,w',u) \\ &\quad\times dw'\,\frac{dw}{w}. \end{aligned} \tag{49}$$

with initial conditions $r = t = 0$ at $x = y$. By using Eqs. (32) and (33) the specular and transmitted intensities at y in the outward direction are

$$\begin{aligned} II(y,v) &= \delta(v^*-u)\,e^{-h(x,y,u)}, \\ I^*(y,v) &= -\left[\delta(v^*-u)-\tilde{k}(K,v^*,-u)\right]e^{-h(x,y,u)} \\ &\quad+\int_0^1 t(x,y,v,w)\tilde{k}(K,w,-u)\,dw \end{aligned} \tag{50}$$

where $\tilde{k}(K,-v,u)$ is the modified resolvent which satisfies the equation

$$\begin{aligned} \tilde{k}(K,v,-u) &= \delta(v-w) \\ &\quad+\int_0^1\int_0^1 K(v,-w)r(x,y,-w,w')k(V,w,-u)\,dw'\,\frac{dw}{w}, \end{aligned}$$

where $k(V,w,-u)$ satisfies (38) with $\mathbf{K} = \mathbf{V}$ or $K(v,-u) = \delta(v+u)$. The reflected intensity seen by the observer on the core surface looking into the atmosphere is, by Eqs. (34) and (35),

$$I^*(x,-v) = \int_0^1 r(x,y,-v,w)\tilde{k}(K,w,-u)\,dw. \tag{51}$$

For the "Schuster's problem", the core is an emitter and a perfect absorber. In this case $r(K,x,y,-v,u) = r(x,y,-v,u)$ and $t(K,x,y,v,u) = t(x,y,v,u)$. For $I_i = \delta(v-u)$, answers for this problem are given by

$$\begin{aligned} II(y,v) &= \delta(v^*-u)\,e^{-h(x,y,u)}, \\ I^*(y,v) &= t(x,y,v,u), \end{aligned}$$

and

$$I^*(x,-v) = r(x,y,-v,u),$$

where t and r are given by Eqs. (48) and (49).

For Milne's problem, $\tilde{k}(K,v,-u) = k(K,v,-u)$ has the expression as given in Eq. (44). If $I_i = \delta(v-1)$, i.e., the core radiates intensities in the normal direction. This is equivalent to the physical problem of the transfer of light from a central star surrounded by a spherical planetary nebula. Since incident radiation is normal to the inner surface of the spherical shell, $u = u^* = u_*$ and $v = v^* = v_*$.

For $\mathbf{K}$ given by Lambert's law, as discussed in the problems of type a, the output under $\mathbf{K}$ is isotropic and K has kernel Au. All integrations can be easily computed, for $I_i = \delta(v-u)$. The results are

$$\begin{aligned} II(y,v) &= \delta(v^*-u)\,e^{h(x,y,u)}, \\ I^*(y,v) &= \mathbf{r}(K)\cdot I_i = (1/u)\bar{A}\,\bar{r}(u), \end{aligned}$$

and

$$I^*(y,v) = \mathbf{t}(K)\cdot I_i = (1/v)\bar{A}\underline{t}(v) + (\bar{A}-1)\delta(v^*-u)e^{-h(x,y,u)},$$

where functions $\bar{A}, \bar{r}, \underline{t}$ are given in Eq. (47).

APPENDIX

For problem type a with external illumination I_i, the intensities at both sides of the spherical shell with radii x and y are related by

$$I(y,v) = (\mathbf{P}+\mathbf{t})\cdot I(x,u) + \boldsymbol{\rho}\cdot I_i, \tag{A1}$$

$$I(x,-v) = \mathbf{r}\cdot I(x,u) + (\mathbf{Q}+\boldsymbol{\tau})\cdot I_i. \tag{A2}$$

From Eq. (A1),

$$(\mathbf{P} + \mathbf{t})\, I(x,u) = I(y,v) - \rho \cdot I_i = [\rho(K) - \rho] \cdot I_i .$$

By the results of Eq. (26) and the fact that $(\mathbf{P} + \mathbf{t})$ is not a zero operator, we have

$$I(x,u) = \mathbf{K} \cdot (\mathbf{E} - \mathbf{r} \cdot \mathbf{K})^{-1} (\mathbf{Q} + \tau) \cdot I_i . \tag{A3}$$

Upon substitution of Eq. (A3) into (A2), using Neumann's series, we obtain

$$I(x,-v) = (\mathbf{E} - r \cdot K)^{-1} \cdot (\mathbf{Q} + \tau) \cdot I_i ,$$

which agrees with Eq. (28).

[1] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and S. Ueno, J. Math. Phys. **9**, 909 (1968).
[2] S. Ueno, H. Kagiwada, and R. Kalaba, J. Math. Phys. **12**, 1279 (1971).
[3] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and S. Ueno, Icarus **7**, 365 (1969); Icarus **11**, 417 (1969).
[4] D. Barbier and G. W. Curtis, Ann. Astrophys. **19**, 129 (1956).
[5] V. I. Barkov, Opt. Spektrosk. **14**, 537 (1963) [Opt. Spectrosc. **14**, 285 (1963)].
[6] I. N. Minin and V. V. Sobolev, Astron. Zh. **40**, 496 (1963) [Sov. Astron.-AJ **7**, 379 (1963)].
[7] I. N. Minin and V. V. Sobolev, Kosm. Issled. **1**, 227 (1963) [Cosm. Res. **1**, 190 (1963)].
[8] I. N. Minin and V. V. Sobolev, Kosm. Issled. **2**, 610 (1964) [Cosm. Res. **2**, 529 (1964)].
[9] V. V. Sobolev and I. N. Minin, Iskusstv. Sputniki Zemli **14**, 7 (1962) [Planet. Space Sci. **11**, 657 (1963)].
[10] V. V. Sobolev, Astron. Zh. **37**, 3 (1960) [Sov. Astron.-AJ **4**, 1 (1960);)Astron. Zh. **39**, 632 (1962) [Sov. Astron.-AJ **6**, 497 (1963)].
[11] H. K. Sen, Astrophys. J. **110**, 276 (1949).
[12] R. D. Chapman, Astrophys. J. **143**, 61 (1966).
[13] J. Lenoble and Z. Sekera, Proc. Natl. Acad. Sci. USA **47**, 372 (1961).
[14] I. N. Minin, Astron. Zh. **41**, 662 (1964) [Sov. Astron.-AJ **8**, 528 (1965)].
[15] S. J. Wilson and K. K. Sen, Can. J. Phys. **43**, 432 (1965).
[16] S. J. Wilson and K. K. Sen, Ann. Astrophys. **28**, 348 (1965).
[17] S. J. Wilson and K. K. Sen, Ann. Astrophys. **28**, 855 (1965).
[18] P. B. Bailey, J. Math. Anal. Appl. **8**, 144 (1964).
[19] P. B. Bailey and G. M. Wing, J. Math. Anal. Appl. **8**, 170 (1964);
[20] M. A. Heaslet and R. F. Warming, J. Quant. Spectrosc. Radiat. Transfer **5**, 669 (1965).
[21] A. Uesugi and J. Tsujita, Publ. Astron. Soc. Jap. **21**, 370 (1966).
[22] R. C. Allen, Jr., L. F. Shampine, and G. Milton Wing, DASA-2421, University of New Mexico, Albuquerque, New Mexico (1970).
[23] G. B. Rybicki, J. Comput. Phys. **6**, 131 (1970).
[24] T. K. Leong and K. K. Sen, Publ. Astron. Soc. Jap. **21**, 167 (1969).
[25] J. Gruschinske and S. Ueno, Publ. Astron. Soc. Jap. **22**, 365 (1970); J. Quant. Spectrosc. Radiat. Transfer (to be published).
[26] E. F. Beckenback, *Modern Mathematics for the Engineer* (McGraw-Hill, New York, 1961), 2nd Ser., Chap. 12.
[27] R. Redheffer, J. Math. & Phys. **41**, 1 (1962)..
[28] A. P. Wang, J. Franklin Inst. **285**, 409 (1969).
[29] R. Redheffer and A. P. Wang, J. Math. Mech. **19**, 765 (1970).
[30] V. A. Ambarzumian, Russ. Astr. J. **19**, 30 (1942).
[31] V. Kourganoff, *Basic Methods in Transfer Problems* with the collaboration of I. W. Busbridge (Oxford U. P., Oxford, 1952).
[32] V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, New Jersey, 1962).
[33] I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge U.P., Cambridge, 1960).
[34] Milton G. Wing, *An Introduction to Transport Theory* (Wiley, New York, 1962).
[35] S. Chandrasekhar *Radiative Transfer* (Dover, New York, 1960).
[36] H. C. Van de Hulst, J. Quant. Spectrosc. Radiat. Transfer **11**, 785 (1971).
[37] R. Bellman and R. Kalaba, Proc. Natl. Acad. Sci. USA **44**, 629 (1965).
[38] J. Mycielski and S. Paszkowski, Stud. Math. **15**, 188 (1965).
[39] G. H. Peebles and M. S. Plesset, Phys. Rev. **81**, 430 (1951).
[40] G. Raisbeck, J. Appl. Phys. **25**, 1510 (1954).
[41] R. Preisendorfer, Proc. Natl. Acad. Sci. USA **44**, 323 (1958).
[42] J. A. McClelland, Royal Dublin Soc. Sci. Trans. **9**, 9 (1906).
[43] H. W. Schmidt, Ann. Phys. (Leipz.) **23**, 671 (1907).
[44] R. E. Bellman and R. E. Kalaba, Proc. Natl. Acad. Sci. USA **43**, 514 (1957).
[45] T. K. Leong and K. K. Sen, Ann. Astrophys. **31**, 467 (1968).
[46] R. Aronson and D. L. Yarmush, J. Math. Phys. **7**, 221 (1966).
[47] S. Ueno and A. P. Wang, University of Southern California Technical Report, USC-113P19-29 (to appear).

Asymptotic solutions of integral equations with a convolution kernel. I

C. van Trigt

Phillips Research Laboratories, Eindhoven, The Netherlands
(Received 11 August 1972)

Homogeneous eigenvalue problems for integral equations with a kernel of the convolution type, defined on a finite volume in N-dimensional space, are discussed. It is shown that they can be reduced asymptotically to eigenvalue problems for simpler integral equations. The integral equations to be derived also yield the asymptotic solution of the inhomogeneous problem for the original integral equations.

I. INTRODUCTION

In many problems arising in radiative transfer, neutron transport theory, etc., it is required to solve the following homogeneous or inhomogeneous integral equations for x in L':

$$f(x) - \int_{L'} K(x - x')f(x')dx' = \mu f(x), \tag{1a}$$

$$f(x) - \int_{L'} K(x - x')f(x')dx' = g(x). \tag{1b}$$

In these equations the integration extends over a finite, half-infinite, or infinite interval L'. In Eq. (1b) g is a known function defined for $x \in L'$. If L' is the interval $(-\infty, +\infty)$ or $[0, \infty)$, then, under certain conditions to be imposed on the function K, Eqs. (1a) and (1b) can be solved by standard Fourier techniques.[1] Let L' be a finite interval. Suppose that K in Eqs. (1a) and (1b) is an even and real function of x, belonging to $L_1(-\infty, +\infty)$. Suppose furthermore that its Fourier transform $\hat{K}$ defined by

$$\hat{K}(k) \equiv \int_{-\infty}^{+\infty} e^{ikx}K(x)dx, \tag{2}$$

assumes its maximum $M > 0$ at $k = 0$ and only at $k = 0$. For the sake of simplicity, we shall assume that K has been normalized such that M can be taken to be equal to unity. Often $K(x - x')$ is a probability kernel, giving the chance that, for instance, a photon or neutron emitted at x' will be absorbed at x. In that case, the chance of being absorbed anywhere is 1 and $\hat{K}(k = 0) = \int K(x)dx = 1$. Finally, let the behaviour of $\hat{K}$ (which is a real and even function of k) near $k = 0$ be given by

$$\hat{K}(k) \sim 1 - c|k|^{\alpha}, \quad k \to 0, \tag{3}$$

for certain values of the real parameters c and α, $0 < c, \alpha < \infty$.

On the basis of these conditions, Widom[2] has obtained the following results, taking L' as a finite but large segment $[-L/2, +L/2]$.

(i) Consider the solutions of the following eigenvalue problem for $|\xi| \leq 1$:

$$f_L(\xi) - \frac{L}{2}\int_{-1}^{+1} K\left[\frac{L}{2}(\xi - \xi')\right]f_L(\xi')d\xi' = \mu_L f_L(\xi),$$

obtained from Eq. (1a) by introducing $\xi = 2x/L$, $\xi' = 2x'/L$, $-1 \leq \xi, \xi' \leq +1$, $\mu_L = \mu$, $f_L(\xi) = f(\frac{1}{2}L\xi)$. This eigenvalue problem has a discrete set of eigenvalues and eigenfunctions for any value of L, labelled by j. For j fixed, $L \to \infty$, its solutions approach in mean square the solutions for $|\xi| < 1$ of the eigenvalue problem,

$$f^{(\alpha)}(\xi) = c^{-1}(L/2)^{\alpha}\mu_L\int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')f^{(\alpha)}(\xi')d\xi', \tag{4}$$

the kernel $K^{(\alpha)}(\xi, \xi')$ being defined for $0 < \alpha < 2$ by

$$K^{(\alpha)}(\xi, \xi') = \frac{\cos(\alpha\pi/2)}{\Gamma(\alpha)}|\xi - \xi'|^{\alpha-2}\max(0, \xi - \xi') - \frac{\sin(\alpha\pi/2)}{\pi\Gamma(\alpha)}(1 - \xi'^2)^{\alpha/2}\int_{-1}^{\xi}\frac{(\xi - \zeta)^{\alpha-1}d\zeta}{(1 - \zeta^2)^{\alpha/2}(\zeta - \xi')}. \tag{5}$$

The integral in Eq. (5) has to be interpreted as a principal value if $\xi > \xi'$. Widom[2] has also derived $K^{(\alpha)}(\xi, \xi')$ for $\alpha \geq 2$ by analytic continuation of Eq. (5). In other words, for j fixed, $L \to \infty$,

$$\mu_{j,L} \sim c\left(\frac{2}{L}\right)^{\alpha}\lambda_j^{-1}(\alpha), \qquad f_{j,L} \sim f_j^{(\alpha)} \quad \text{(in mean square)}, \tag{6}$$

the $\lambda_j(\alpha)$ and $f_j^{(\alpha)}$ being the solutions of

$$\lambda(\alpha)f^{(\alpha)}(\xi) = \int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')f^{(\alpha)}(\xi')d\xi'. \tag{7}$$

(ii) (implicit in the analysis) Consider the following inhomogeneous problem for $|\xi| \leq 1$:

$$f_L(\xi) - \frac{L}{2}\int_{-1}^{+1} K\left[\frac{L}{2}(\xi - \xi')\right]f_L(\xi')d\xi' = g(\xi).$$

If g is sufficiently well behaved, its solution f_L approaches asymptotically for $L \to \infty$ the function f defined by the following integral:

$$f(\xi) = c^{-1}\left(\frac{L}{2}\right)^{\alpha}\int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')g(\xi')d\xi'. \tag{8}$$

The results given above also apply[3] if the constant c in Eq. (3) is replaced by a function $F(|k|)$ which is nonnegative and slowly varying near $k = 0$. In that case, c in Eqs. (4), (6), and (8) is replaced by $F(2/L)$. From these results, we see that the kernel $K^{(\alpha)}(\xi, \xi')$ defines an operator which approximates asymptotically the operator *inverse* to $(I - K)$, defined by Eqs. (1a) and (1b). {In obvious notation, I is the identity operator; the integral operator K corresponds to the kernel $L/2K[(L/2)(\xi - \xi')]$}. Hence, solving Eq. (1b) has been reduced to a quadrature [see Eq. (8)], and solving Eq. (1a) to solving Eq. (7). We may expect that the latter problem is simpler, since $K^{(\alpha)}(\xi, \xi')$ is determined not by the *full* Fourier transform $\hat{K}$, as would be the case with $(I - K)^{-1}$, but by its *local* properties near $k = 0$ only.

However, the expression Eq. (5) is so intricate that we may wonder whether anything has actually been gained. There are other difficulties connected with Eq. (5). Since the original integral operator [in Eqs. (1a) and (1b)] is symmetric, the integral operator with kernel $K^{(\alpha)}(\xi, \xi')$

Copyright © 1973 by the American Institute of Physics

is expected to be symmetric also. Though this can be proved,[2] it is by no means obvious from Eq. (5). In some physical situations, the solving of a decay problem with time variation $\exp(-\mu t)$ can be reduced to solving an equation like Eq. (1a). It then follows that the eigenvalues μ_j and hence the λ_j must be positive. It is therefore reasonable to presume that the integral operator with kernel $K^{(\alpha)}(\xi, \xi')$ is positive definite. Again, this is not obvious from Eq. (5), though it can be proved.[2] In this paper three alternative representations of $K^{(\alpha)}(\xi, \xi')$ will be derived. The corresponding integral operator is now easily seen to be symmetric and positive definite, and the kernel is much less difficult to handle than $K^{(\alpha)}(\xi, \xi')$ as defined by Eq. (5). The analog of this representation is given for the N-dimensional case, i.e. when we have equations like Eqs. (1a) or (1b) defined on a volume $V' = \{\mathbf{r}(\sum x_j^2)^{1/2} \leq L/2\}$, with a real convolution kernel dependent on the distance $|\mathbf{r} - \mathbf{r}'|$ only.

Integral equations of this type are frequently encountered in radiative transfer and neutron transport theory. They describe the transport in a slab, infinite cylinder, and sphere in the particular cases $N = 1, 2$, and 3.

Section II is introductory and heuristic. The basic formulas are derived which are used in Secs. III and IV for the derivation of the required integral equations. The results of Sec. II have been derived rigorously by Widom.[3] In Sec. III we deal with the one-dimensional and in Sec. IV with the N-dimensional cases. The conclusions are stated in Sec. V, where the applicability of the theory in radiative transfer is also discussed. In Appendix B the equivalence of our representation of $K^{(\alpha)}(\xi, \xi')$ and of that given in Eq. (5) will be proved directly.

II. BASIC EQUATIONS

Let us consider the following eigenvalue problem of the convolution type in N-dimensional Euclidean space E_N:

$$f(\mathbf{r}) - \int_{V'(L)} K(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')d\mathbf{r}' = \mu f(\mathbf{r}), \mathbf{r} \in V'(L). \quad (9)$$

In this formula $\mathbf{r}$ is the N-dimensional vector $(x_1, x_2, \ldots, x_N)$, and $d\mathbf{r}$ an N-dimensional volume element $dx_1 dx_2 \cdots dx_N$. The integration extends over the finite volume $V'(L) = \{\mathbf{r}, (\sum_{j=1}^N x_j^2)^{1/2} \leq L/2\}$. The following conditions are imposed on the function K in Eq. (9). If $N = 1$, K will be a real and even function. If $N \geq 2$, it will be a real function, dependent on $|\mathbf{r}|$ only. Furthermore, K will belong to $L_1(E_N)$, so that $\int |K(\mathbf{r})| d\mathbf{r}$ exists. Here, as in the following, integration over the entire space is to be understood if the integration sign is not accompanied by V'. The Fourier transform of K, denoted by $\hat{K}$, is defined as follows:

$$\hat{K}(\mathbf{k}) = \int e^{i\mathbf{k}\cdot\mathbf{r}} K(\mathbf{r}) d\mathbf{r}. \quad (10)$$

Here $\mathbf{k} \equiv (k_1, k_2, \ldots, k_N)$ and $\mathbf{k}\cdot\mathbf{r} = \sum_{j=1}^N k_j x_j$. Note that $\hat{K}$ depends only on $k = (\sum_{j=1}^N k_j^2)^{1/2}$. We shall consider only functions $\hat{K}$ such that $\hat{K}$ assumes its maximum for $k = 0$ and only for $k = 0$, and, for simplicity, we shall take $M = 1$. Finally, it is required that, for certain real, nonnegative values of the parameters c and α, $\hat{K}$ behaves near $k = 0$ like

$$\hat{K}(\mathbf{k}) \sim 1 - ck^\alpha, \quad k \to 0. \quad (11)$$

Compare Eqs. (9)–(11) for $N = 1$ with Eqs. (1a) through (3) in the Introduction, and note the analogy. Since the kernel $K(\mathbf{r} - \mathbf{r}')$ is real and depends on $|\mathbf{r} - \mathbf{r}'|$ only (is even, if $N = 1$), the corresponding integral operator is symmetric. It can be proved that, since K belongs to $L_1(E_N)$, this integral operator[4,5] is completely continuous on $L_2(V')$, the space of functions square-integrable on $V'(L)$. Consequently the eigenvalues $1 - \mu$ in Eq. (9) are real and discrete. Their only possible limit point is zero. The eigenfunctions corresponding to different eigenvalues are mutually orthogonal on $V'(L)$.[4]

We are interested in the asymptotic behavior for $L \to \infty$ of the solutions of Eq. (9). In the investigation of this behavior, it is convenient not to let $V'(L)$ increase indefinitely but to perform a change of variables, so that the volume can be kept fixed, and to consider the asymptotic behavior of the kernel. We introduce $\mu_L = \mu, \boldsymbol{\rho} = 2\mathbf{r}/L$, $\boldsymbol{\rho} = (\xi_1, \xi_2, \ldots, \xi_N)$, $d\boldsymbol{\rho} = d\xi_1 d\xi_2 \cdots d\xi_N, f_L(\rho) = f[(L/2)\rho]$, and the volume $V = \{\rho, (\sum_{j=1}^N \xi_j^2)^{1/2} \leq 1\}$. For ρ in V Eq. (9) takes the form

$$f_L(\boldsymbol{\rho}) - \left(\frac{L}{2}\right)^N \int_V K[(L/2(\boldsymbol{\rho} - \boldsymbol{\rho}')]f_L(\boldsymbol{\rho}')d\boldsymbol{\rho}' = \mu_L f_L(\boldsymbol{\rho}). \quad (12)$$

The domain of definition of Eq. (12) can be extended to all $\boldsymbol{\rho}$ in E_N, by putting the kernel and f_L equal to zero outside V. We introduce a projection operator P. P multiplies a function f_L with the characteristic function of V, denoted by $P(\boldsymbol{\rho})$. We have

$$\begin{aligned} Pf_L = P(\boldsymbol{\rho})f_L(\boldsymbol{\rho}) &= f_L, \quad \text{for all } \boldsymbol{\rho} \in V; \\ &= 0, \quad \text{for all } \boldsymbol{\rho} \notin V. \end{aligned} \quad (13)$$

For all $\boldsymbol{\rho}$ in E_N, Eq. (12) becomes

$$\begin{aligned} P(\boldsymbol{\rho})f_L(\boldsymbol{\rho}) - (L/2)^N P(\boldsymbol{\rho}) \int K[(L/2)(\boldsymbol{\rho} - \boldsymbol{\rho}')] \\ \times P(\boldsymbol{\rho}')f_L(\boldsymbol{\rho}')d\boldsymbol{\rho}' = \mu_L P(\boldsymbol{\rho})f_L(\boldsymbol{\rho}). \end{aligned} \quad (14)$$

The main intention of the manipulation leading from Eq. (12) to Eq. (14) is to extend the integration over the entire space so that Fourier techniques can be easily applied.

We define the following Fourier transforms:

$$\begin{aligned} \hat{f}_L(\mathbf{k}) &= (2\pi)^{-N/2} \int e^{i\mathbf{k}\cdot\boldsymbol{\rho}} P(\boldsymbol{\rho}) f_L(\boldsymbol{\rho}) d\boldsymbol{\rho} \\ &= (2\pi)^{-N/2} \int_V e^{i\mathbf{k}\cdot\boldsymbol{\rho}} f_L(\boldsymbol{\rho}) d\boldsymbol{\rho}, \\ \hat{P}_N(\mathbf{k}) &= (2\pi)^{-N/2} \int e^{i\mathbf{k}\cdot\boldsymbol{\rho}} P(\boldsymbol{\rho}) d\boldsymbol{\rho} \\ &= (2\pi)^{-N/2} \int_V e^{i\mathbf{k}\cdot\boldsymbol{\rho}} d\boldsymbol{\rho}. \end{aligned} \quad (15)$$

Now multiply both sides of Eq. (14) by $(2\pi)^{-N/2} e^{i\mathbf{k}\cdot\boldsymbol{\rho}}$ and integrate over all $\boldsymbol{\rho}$. The Fourier transform of a product of two functions is the convolution of their Fourier transforms.[6] The Fourier transform of a convolution of two functions is the product of their Fourier transforms.[6] Using the definitions in Eq. (15), we have

$$\hat{f}_L(\mathbf{k}) - (2\pi)^{-N/2} \int \hat{P}_N(\mathbf{k} - \mathbf{k}')\hat{K}(2\mathbf{k}'/L)\hat{f}_L(\mathbf{k}') = \mu_L \hat{f}_L(\mathbf{k}). \quad (16)$$

Widom[3] has proved that, for large L, the solutions of Eq. (16) approach in mean square the weak solutions of the equation obtained from Eq. (16) by replacing $\hat{K}(2\mathbf{k}/L)$ by the first two terms of its asymptotic expansion for $2k/L \to 0$, given in Eq. (11). Noting that[7]

$$\hat{f}_L(\mathbf{k}) = (2\pi)^{-N/2} \int \hat{P}_N(\mathbf{k} - \mathbf{k}')f_L(\mathbf{k}')d\mathbf{k}' \quad (17)$$

and combining Eqs. (16) and (17), we have for j fixed, $L \to \infty$,

$$\mu_{j,L} \sim c(2/L)^{\alpha}\lambda_j^{-1}(\alpha),$$
$$\hat{f}_{j,L} \sim \hat{f}_j^{(\alpha)} \quad \text{(in mean square)}, \tag{18}$$

where the $\lambda_j(\alpha)$ are the eigenvalues and the $f_j^{(\alpha)}$ are the weak eigenfunctions of the eigenvalue problem

$$\hat{f}^{(\alpha)}(\mathbf{k}) = \lambda(\alpha)(2\pi)^{-N/2}\int \hat{P}_N(\mathbf{k}-\mathbf{k}')k'^{\alpha}\hat{f}^{(\alpha)}(\mathbf{k}')d\mathbf{k}'. \tag{19}$$

The crucial problem is therefore to solve Eq. (19). However, first, we have to specify the space of functions in which solutions of Eq. (19) are to be found. The solutions of Eq. (14) are among the funtions Pf_L, square-integrable on V. The corresponding functional space is denoted by $L_2(V)$. Their Fourier transforms are square-integrable and fulfill Eq. (17). This functional space is designated by $\hat{L}_2(V)$. By a theorem of Titchmarsh,[1] there exists a one-to-one correspondence between the spaces $L_2(V)$ and $\hat{L}_2(V)$. Since the convergence of $\hat{f}_L$ to $\hat{f}^{(\alpha)}$ is in mean square, $\hat{f}^{(\alpha)}$ is also in $\hat{L}_2(V)$. However, this space of functions is too large. It is not hard to see that, due to the factor k'^{α}, $\alpha > 0$, in Eq. (19), the (symmetric) integral operator corresponding to the kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k}-\mathbf{k}')k'^{\alpha}$ is not bounded. By a theorem of Hellinger and Toeplitz,[8] it cannot be defined everywhere in $\hat{L}_2(V)$ [i.e., the integral in Eq. (19) does not exist for all $\hat{f}$ in $\hat{L}_2(V)$]. We shall, therefore, first define a dense subspace of $\hat{L}_2(V)$ on which the operator has a meaning. On this subspace—which does not contain the eigenfunctions—we derive a symmetric and completely continuous operator which is the inverse of the integral operator with kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k}-\mathbf{k}') \times k'^{\alpha}$. Since this new operator is bounded, its domain of definition can be extended to the entire $\hat{L}_2(V)$. It is, therefore, the inverse of the operator with kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k}-\mathbf{k}')k'^{\alpha}$ for all $\hat{f}$ for which the latter can be defined, and we can formulate the eigenvalue problem in terms of this new operator.[9] The operator also furnishes the solution of the inhomogeneous problem [Eq. (1b) and its analog for general N] if $g(r)$ is sufficiently well behaved. The idea is essentially due to Widom.[3] We shall need explicit expressions for the function $\hat{P}_N(\mathbf{k})$, Eq. (15). The reader can easily verify that

$$\hat{P}_N(\mathbf{k}) = (2\pi)^{-N/2}\int_V e^{i\mathbf{k}\cdot\rho}d\rho = k^{-N/2}J_{N/2}(k), \tag{20}$$

where $J_{N/2}$ denotes a Bessel function of order $N/2$.

III. THE ONE-DIMENSIONAL CASE

In the special case $N = 1$, $P_1(k_1)$ in Eq. (20) can be expressed in elementary functions. Substituting the result in Eq. (19), we have for a fixed value of $\alpha > 0$,

$$\hat{f}^{(\alpha)}(k_1) = \frac{\lambda(\alpha)}{\pi}\int_{-\infty}^{+\infty}\frac{\sin(k_1-k_1')}{k_1-k_1'}|k'|^{\alpha}\hat{f}^{(\alpha)}(k_1')dk_1'. \tag{21}$$

In this section, we shall henceforth omit the index 1 in k_1 and k_1'. For $N = 1$, the space $\hat{L}_2(V)$ consists of the functions $\hat{f}$ which are quadratically integrable on $(-\infty, +\infty)$ and obey the one-dimensional version of Eq. (17),

$$\hat{f}(k) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\sin(k-k')}{k-k'}\hat{f}(k')dk'. \tag{22}$$

As has been argued at the end of Sec. II, Eq. (21) cannot apply to all $\hat{f}$ in $\hat{L}_2(V)$ [i.e., the integral does not exist for all $\hat{f}$ in $\hat{L}_2(V)$]. The purpose of this section is, first, to find a dense subspace of $\hat{L}_2(V)$ belonging to the domain of definition of the integral operator defined by Eq. (21). We shall then derive a symmetric and completely continuous operator, with kernel $\hat{K}^{(\alpha)}(k,k')$, which is the inverse of that integral operator for all $\hat{f}$ in the subspace, i.e.,

$$\frac{1}{\pi}\int_{-\infty}^{+\infty}\hat{K}^{(\alpha)}(k'',k)\int_{-\infty}^{+\infty}\frac{\sin(k-k')}{k-k'}|k'|^{\alpha}\hat{f}(k')dk'dk = \hat{f}(k''). \tag{23}$$

The operator with kernel $\hat{K}^{(\alpha)}(k,k')$ is bounded, and its domain of definition can thus be extended to the entire $\hat{L}_2(V)$ by continuity.[9] For $\hat{f}^{(\alpha)}$ in $\hat{L}_2(V)$, the eigenvalue problem [Eq. (21)] then becomes

$$\int_{-\infty}^{+\infty}\hat{K}^{(\alpha)}(k,k')\hat{f}^{(\alpha)}(k')dk' = \lambda(\alpha)\hat{f}^{(\alpha)}(k). \tag{24}$$

Define for real values of $\sigma > 0$ and $n = 0, 1, \ldots,$

$$\hat{g}_{n,\sigma}(k) = (n+\sigma)^{1/2}k^{-\sigma}J_{n+\sigma}(k), \tag{25}$$

where $J_{n+\sigma}(k)$ is a Bessel function of order $n+\sigma$. Note that $\hat{g}_{n,\sigma}$ is an entire function of k, which is even or odd for n even or odd. The functions $\hat{g}_{n,\sigma}$ are quadratically integrable,[10] obey Eq. (22),[11] and therefore belong to $\hat{L}_2(V)$. Define also the inner product for some real $\tau \geq 0$ and for all $\hat{f}$, for which it has a meaning

$$(\hat{f},\hat{g}_{n,\sigma})_{\tau} = \int_{-\infty}^{+\infty}\hat{f}(k)\hat{g}_{n,\sigma}(k)|k|^{2\sigma-1+\tau}dk. \tag{26}$$

The $\hat{g}_{n,\sigma}$ are orthogonal[12] for $\tau = 0$, i.e.,

$$(\hat{g}_{n,\sigma},\hat{g}_{m,\sigma})_{\tau=0} = \delta_{n,m}. \tag{27}$$

First we shall take $0 < \alpha < 2$. Our required subspace will be the set of functions which are in $\hat{L}_2(V)$, are of order $O(k^{-2})$, $|k| \to \infty$, and which, for a particular value of σ (dependent on α), permit of the expansion

$$\hat{f}(k) = \sum_{n=0}^{\infty}\hat{g}_{n,\sigma}(k)(\hat{f},\hat{g}_{n,\sigma})_{\tau=0}. \tag{28}$$

The subspace is designated by $\hat{L}_2^{(2\sigma-1)}(V)$. The value of σ will be determined below, where it will be seen that $2\sigma - 1 = \alpha$. The expansion Eq. (28) is known as a Neumann series.[13] The conditions to be imposed on the functions $\hat{f}$ if they are to belong to $\hat{L}_2^{(\alpha)}(V)$ will be considered in Appendix A. It will be proved there that $\hat{L}_2^{(\alpha)}(V)$ is everywhere dense in $\hat{L}_2(V)$. It can easily be verified that for $0 < \alpha < 2$, $\hat{L}_2^{(\alpha)}(V)$ belongs to the domain of definition of the operator defined by Eq. (21)—i.e., the integral converges. The case $\alpha \geq 2$ calls for separate treatment. Again the subspace $\hat{L}_2^{(\alpha)}(V)$ will consist of functions $\hat{f}$ which permit of the expansion (28), but we impose somewhat different conditions on their behavior for $k \to \infty$. This question is further dealt with in Appendix A. It will be shown there that the expression for the kernel $\hat{K}^{(\alpha)}(k,k')$, which we are now going to derive for $0 < \alpha < 2$, is actually valid for all $\alpha > 0$.

A. The inverse operator

Consider

$$\frac{1}{\pi}\hat{g}_{n,\sigma}(k'')\int_{-\infty}^{+\infty}\hat{g}_{n,\sigma}(k)\int_{-\infty}^{+\infty}\frac{\sin(k-k')}{k-k'}|k'|^{\alpha}\hat{f}(k')dk'dk.$$

We may interchange the integrations for $0 < \alpha < 2$ by Fubini's theorem. Since the $\hat{g}_{n,\sigma}$ fulfill Eq. (22), we have for $n = 0, 1, \cdots$

$$= \hat{g}_{n,\sigma}(k'')(\hat{g}_{n,\sigma}, \hat{f})_{\tau=\alpha+1-2\sigma}.$$

Sum this result over all n, put $2\sigma - 1 = \alpha$, and compare the expression thus obtained with Eq. (28). It is seen that this expression is equal to $\hat{f}(k'')$. In other words, we have derived a representation of the kernel $\hat{K}^{(\alpha)}(k, k')$, defined in Eq. (23):

$$\hat{K}^{(\alpha)}(k, k') = \sum_{n=0}^{\infty} \hat{g}_{n,\sigma}(k)\hat{g}_{n,\sigma}(k'), \qquad \sigma = \tfrac{1}{2} + \tfrac{1}{2}\alpha. \quad (29)$$

It will be shown in Appendix A that Eq. (29), derived for $0 < \alpha < 2$, is actually valid for all $\alpha \geq 0$. Taking this for granted, we derive some alternative representations for $\hat{K}^{(\alpha)}(k, k')$. The sum in Eq. (29) is reminiscent of the addition theorem for Bessel functions,[10]

$$\hat{g}_{0,\sigma}(\omega) = 2^{\sigma}\sigma^{1/2}\Gamma(\sigma) \sum_{n=0}^{\infty} \hat{g}_{n,\sigma}(k)\hat{g}_{n,\sigma}(k')C_n^{\sigma}(\zeta). \quad (30)$$

Here $\omega = (k^2 - 2kk'\zeta + k'^2)^{1/2}$, $|\zeta| \leq 1$, and $C_n^{\sigma}(\zeta)$ is a Gegenbauer polynomial.[14]

From the inequalities[10,14]

$$|\hat{g}_{n,\sigma}(k)| \leq (n + \sigma)^{1/2}k^n/2^{n+\sigma}\Gamma(n + \sigma + 1)$$

and

$$|C_n^{\sigma}(\zeta)| \leq \Gamma(n + 2\sigma)/\Gamma(2\sigma)n!,$$

it follows that the series in Eq. (30) is absolutely and uniformly convergent in ζ. We now multiply both sides of Eq. (30) by $(1 - \zeta^2)^{\sigma-3/2}(1 + \zeta)$, integrate with respect to ζ from -1 to $+1$, and use the following identity, which can readily be proved by induction[15]:

$$\int_{-1}^{+1}(1 - \zeta^2)^{\sigma-3/2}(1 + \zeta)C_n^{\sigma}(\zeta)d\zeta = \pi^{1/2}\Gamma(\sigma - 1/2)/\Gamma(\sigma).$$

Equation (29) then takes the form

$$\hat{K}^{(\alpha)}(k, k') = \frac{2^{-\sigma}\sigma^{-1/2}}{\pi^{1/2}\Gamma(\sigma - 1/2)} \times \int_{-1}^{+1}(1 - \zeta^2)^{\sigma-3/2}(1 + \zeta)\hat{g}_{0,\sigma}(\omega)d\zeta.$$

In order to calculate the integral, we shall consider for k, k' fixed, s variable, $0 \leq s \leq 1$,

$$\frac{d}{ds}s^{2\sigma}\hat{K}^{(\alpha)}(sk, sk') = \frac{2^{-\sigma}\sigma^{-1/2}}{\pi^{1/2}\Gamma(\sigma - \frac{1}{2})} \times \frac{d}{ds}\int_{-1}^{+1}(1 - \zeta^2)^{\sigma-3/2}(1 + \zeta)s^{2\sigma}\hat{g}_{0,\sigma}(s\omega)d\zeta.$$

We next resubstitute Eq. (25), defining $\hat{g}_{0,\sigma}(s\omega)$. The differentiation may be carried out under the integral sign, both $s^{\sigma}J_{\sigma}(s\omega)$, $\sigma = \frac{1}{2} + \frac{1}{2}\alpha$, and its derivative being continuous in the rectangle $0 \leq s \leq 1$, $|\zeta| \leq 1$ for $\alpha > 0$. The result is

$$\frac{d}{ds}s^{1+\alpha}\hat{K}^{(\alpha)}(sk, sk') = \frac{2^{-\alpha/2-1/2}}{\pi^{1/2}\Gamma(\alpha/2)} \times \int_{-1}^{+1}(1 - \zeta^2)^{\alpha/2-1}(1 + \zeta)s^{1/2+\alpha/2}\omega^{1/2-\alpha/2}J_{\alpha/2-1/2}(s\omega)d\zeta.$$

The integral is calculated by again using the addition theorem, Eq. (30), but now for $\sigma = \alpha/2 - 1/2$. We interpret $(1 + \zeta)$ as the sum of two Gegenbauer polynomials of order $n = 0$ and $n = 1$. By applying the orthogonality relation[14] for these polynomials, we obtain

$$\frac{d}{ds}s^{1+\alpha}\hat{K}^{(\alpha)}(sk, sk') = (s/2)(kk')^{1/2-\alpha/2} \times [J_{\alpha/2-1/2}(sk)J_{\alpha/2-1/2}(sk') + J_{\alpha/2+1/2}(sk)J_{\alpha/2+1/2}(sk')]. \quad (31)$$

The equivalent expressions for $0 < \alpha < \infty$ are therefore, first Eq. (29) in terms of Bessel functions,

$$\hat{K}^{(\alpha)}(k, k') = (kk')^{-\alpha/2-1/2} \times \sum_{n=0}^{\infty}(n + \alpha/2 + 1/2)J_{n+\alpha/2+1/2}(k)J_{n+\alpha/2+1/2}(k'); \quad (32a)$$

secondly, the representation[16] found by integrating Eq. (31) between 0 and 1,

$$\hat{K}^{(\alpha)}(k, k') = [(kk')^{1/2-\alpha/2}/2]\int_0^1 s[J_{\alpha/2-1/2}(sk)J_{\alpha/2-1/2}(sk') + J_{\alpha/2+1/2}(sk)J_{\alpha/2+1/2}(sk')]ds; \quad (32b)$$

and thirdly, the expression found by calculating this integral[10] and simplifying with the aid of the recurrence formulas for Bessel functions,

$$\hat{K}^{(\alpha)}(k, k') = [(kk')^{1/2-\alpha/2}/2(k - k')] \times [J_{\alpha/2+1/2}(k)J_{\alpha/2-1/2}(k') - J_{\alpha/2-1/2}(k)J_{\alpha/2+1/2}(k')]. \quad (32c)$$

The kernel $\hat{K}^{(\alpha)}(k, k')$ supplies the solution to certain scattering problems. Its inverse (double) Fourier transform is also of interest. It is calculated by multiplying both the left- and right-hand sides of Eq. (23) by $(2\pi)^{-1/2}e^{-ik''\xi}$ and integrating over all k''. Parseval's relation is applied to the integral with respect to k. We then have

$$K^{(\alpha)}(\xi, \xi') = (2\pi)^{-1}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-ik\xi+ik'\xi'}\hat{K}^{(\alpha)}(k, k')dkdk'.$$

By substituting Eqs. (32a) and (32b) we obtain integrals which are all standard.[17] The result is that $K^{(\alpha)}(\xi, \xi') = 0$ for $|\xi|, |\xi'| > 1$. For $|\xi|, |\xi'| < 1$ we have, first, the expression corresponding to Eq. (32a)[18]:

$$K^{(\alpha)}(\xi, \xi') = \sum_{n=0}^{\infty} g_{n,\sigma}^{*}(\xi)g_{n,\sigma}(\xi'), \qquad \sigma = \tfrac{1}{2}\alpha + \tfrac{1}{2}; \quad (33a)$$

secondly, the expression corresponding to Eq. (32b):

$$K^{(\alpha)}(\xi, \xi') = \frac{2^{1-\alpha}}{\Gamma^2(\alpha/2)} \times \int_{\max(|\xi|,|\xi'|)}^{1} s^{-\alpha}(s^2 - \xi^2)^{\alpha/2-1}(s^2 - \xi'^2)^{\alpha/2-1}(s^2 + \xi\xi')ds; \quad (33b)$$

and thirdly, a representation obtained from Eq. (33b) by introducing the new variable $w = s^2 - 2\xi\xi' + \xi^2\xi'^2/s^2$ and identifying the resulting integral as a hypergeometric function[19]:

$$K^{(\alpha)}(\xi, \xi') = \frac{2^{-\alpha}}{\Gamma(\alpha/2)\Gamma(1 + \alpha/2)} \frac{(1 - \xi^2)^{\alpha/2}(1 - \xi'^2)^{\alpha/2}}{|\xi - \xi'|} \times {}_2F_1\left[\tfrac{1}{2}, \tfrac{1}{2}\alpha; 1 + \tfrac{1}{2}\alpha; -\frac{(1 - \xi^2)(1 - \xi'^2)}{(\xi - \xi')^2}\right]. \quad (33c)$$

In Eq. (33a) $g_{n,\sigma}(\xi)$ is defined as zero for $|\xi| > 1$ and as[17]

$$g_{n,\sigma}(\xi) = (2\pi)^{-1/2}\int_{-\infty}^{+\infty}e^{ik\xi}\hat{g}_{n,\sigma}(k)dk = 2^{\sigma-1/2}\Gamma(\sigma)\frac{(n + \sigma)^{1/2}n!i^n}{\pi^{1/2}\Gamma(n + 2\sigma)}(1 - \xi^2)^{\sigma-1/2}C_n^{\sigma}(\xi)$$

for $|\xi| < 1$, where $C_n^\sigma(\xi)$ is a Gegenbauer polynomial. The special case $\alpha = 1$ of Eq. (33a) [$C_n^1(\xi)$ is a Tchebicheff polynomial) and of Eq. (33c)—in the form to be given in Eq. (35) below—has been previously derived by Kac and Pollard.[20] In Appendix B we shall prove directly for $0 < \alpha < 2$ that Eq. (33a) is equivalent to the representation given by Widom, Eq. (5).

The integral operator with kernel $K^{(\alpha)}(\xi, \xi')$ is obviously symmetric. It can readily be shown that $K^{(\alpha)}(\xi, \xi')$ is continuous everywhere if $\alpha > 1$. It is also continuous for $\xi \neq \xi'$ if $0 < \alpha \leq 1$, but it has a logarithmic singularity for $|\xi - \xi'| \to 0$ if $\alpha = 1$ and a singularity of the type $|\xi - \xi'|^{\alpha-1}$ if $0 < \alpha < 1$. Hence, the corresponding integral operator is completely continuous on $L_2(V)$.[5] It is, therefore, bounded, and its domain of definition can be extended to the entire $L_2(V)$ by continuity. It should be remembered that, up to now, the domain of definition of the integral operator with kernel $\hat{K}^{(\alpha)}(k, k')$—and thus the one with $K^{(\alpha)}(\xi, \xi')$—was confined to a dense subspace (see Appendix A). We shall now prove that the integral operator is also positive definite. It is sufficient to show this for a set which is dense everywhere in $L_2(V)$. For this purpose we may take the space consisting of all finite and real linear combinations of Gegenbauer polynomials. Let

$$h(\xi) = \sum_{m=0}^{M} a_m C_m^\sigma(\xi), \qquad \sigma = \tfrac{1}{2}\alpha + \tfrac{1}{2}.$$

Using Eq. (33a) and the orthogonality relation of the Gegenbauer polynomials,[14] we obtain

$$\int_{-1}^{+1}\int_{-1}^{+1} h^*(\xi)K^{(\alpha)}(\xi, \xi')h(\xi')d\xi' d\xi = \frac{2^{2\sigma-1}\pi}{\Gamma^2(\sigma)} \sum_{m=0}^{M} (m + \sigma)^{-1}a_m^2 > 0.$$

In (ξ, ξ') space, the eigenvalue problem, Eq. (24), becomes

$$\lambda(\alpha)f^{(\alpha)}(\xi) = \int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')f^{(\alpha)}(\xi')d\xi'. \tag{34}$$

The reader is reminded that the solutions of Eq. (34) provide the asymptotic solutions of the original eigenvalue problem dealt with in Sec. II [see Eqs. (12) and (18)].

B. Special cases

For some values of α the hypergeometric function in Eq. (33c) can be expressed in elementary functions. For example, if $\alpha = 1$,

$$K^{(1)}(\xi, \xi') = \frac{1}{2\pi} \log\left(\frac{1 - \xi\xi' + (1 - \xi^2)^{1/2}(1 - \xi'^2)^{1/2}}{1 - \xi\xi' - (1 - \xi^2)^{1/2}(1 - \xi'^2)^{1/2}}\right). \tag{35}$$

The kernel occurs in air foil theory.[21]

If $\alpha = 2$,

$$K^{(2)}(\xi, \xi') = \frac{1}{2}[1 + \min(\xi, \xi')][1 - \max(\xi, \xi')]. \tag{36}$$

This is the familiar expression for the Green's function of the diffusion equation in one-dimensional space: $\lambda d^2f/d\xi^2 = -f$, with boundary condition $f(\xi = \pm 1) = 0$. Generally, if $\alpha = 2l$, $l = 1, 2, \ldots$, Eq. (33c) can be evaluated in a closed, though involved, form.

However, it is more interesting to notice that if $\alpha = 2l$, Eq. (33c) is the Green's function for the differential equation $\lambda d^{2l}f/d\xi^{2l} = (-1)^l f$, with boundary conditions $d^jf/d\xi^j = 0, \xi = \pm 1, j = 0, 1, \ldots, l - 1$. A calculation shows that

$$K^{(2l)}(\xi, \xi') = P_{2l-1}(\xi) + [(-1)^l/2(2l - 1)!](\xi - \xi')^{2l-2}|\xi - \xi'|.$$

$P_{2l-1}(\xi)$ is a polynomial of degree $(2l - 1)$ in ξ, of course, with coefficients dependent on ξ'. Differentiating $P_{2l-1}(\xi)$ $2l$ times with respect to ξ gives zero, and differentiating the second term $(2l - 1)$ times yields a step function, i.e., $(-1)^l/2$ for $\xi > \xi'$ and $(-1)^{l+1}/2$ for $\xi < \xi'$. If the step function is differentiated again, we obtain $(-1)^l\delta(\xi - \xi')$. It follows directly from Eq. (33c) that the eigenfucntions behave like $(1 - \xi^2)^l$, $\xi \to \pm 1$. Hence, the above-mentioned boundary conditions are fulfilled. This proves the statement. Some integrals of $K^{(\alpha)}(\xi, \xi')$ have been obtained in the past by probabilistic arguments without knowledge of the precise form of the kernel.[22]

We shall now calculate these for the sake of comparison. The zeroth moment of $K^{(\alpha)}(\xi, \xi')$ for $|\xi| < 1$ is

$$\int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')d\xi' = \frac{(1 - \xi^2)^{\alpha/2}}{\Gamma(1 + \alpha)}. \tag{37}$$

Finally, the zeroth moment of the second iterate of $K^{(\alpha)}(\xi, \xi')$ for $|\xi| < 1$ is

$$\int_{-1}^{+1}\int_{-1}^{+1} K^{(\alpha)}(\xi, \xi')K^{(\alpha)}(\xi', \xi'')d\xi' d\xi'' = \lim_{k''\to 0} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{ik\xi}\hat{K}^{(\alpha)}(k, k')\hat{K}^{(\alpha)}(k', k'')dk'dk = \frac{\alpha}{\Gamma^2(1 + \alpha)} \int_{|\xi|}^{1} s(s^2 - \xi^2)^{\alpha/2-1}{}_2F_1(-\tfrac{1}{2}\alpha, \tfrac{1}{2}; \tfrac{1}{2} + \tfrac{1}{2}\alpha; s^2)ds. \tag{38}$$

IV. THE N-DIMENSIONAL CASE

It is now our purpose to carry out in N-dimensional space, $N \geq 2$, the procedure described at the end of Sec. II: namely, to find a dense subspace of $\hat{L}_2(V)$, belonging to the domain of definition of the integral operator with kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k} - \mathbf{k}')k'^\alpha$ [see Eq. (19)], and then to derive on this subspace its symmetric and completely continuous inverse with kernel $\hat{K}_N^{(\alpha)}(\mathbf{k}, \mathbf{k}')$. For all $\hat{f}$ in this subspace, $\hat{K}_N^{(\alpha)}(\mathbf{k}, \mathbf{k}')$ by definition satisfies the equation

$$(2\pi)^{-N/2}\int \hat{K}_N^{(\alpha)}(\mathbf{k}'', \mathbf{k})\int \hat{P}_N(\mathbf{k} - \mathbf{k}')k'^\alpha \hat{f}(\mathbf{k}')d\mathbf{k}'d\mathbf{k} = \hat{f}(\mathbf{k}''). \tag{39}$$

The domain of definition of the integral operator with kernel $\hat{K}_N^{(\alpha)}(\mathbf{k}, \mathbf{k}')$ can be extended to the entire $\hat{L}_2(V)$. The eigenvalue problem then becomes

$$\int \hat{K}_N^{(\alpha)}(\mathbf{k}, \mathbf{k}')\hat{f}^{(\alpha)}(\mathbf{k}')d\mathbf{k}' = \lambda(\alpha)\hat{f}^{(\alpha)}(\mathbf{k}). \tag{40}$$

The reader is reminded that in all these equations $\mathbf{k}$ and $\mathbf{k}'$ are N-dimensional vectors. The N-dimensional volume element is $d\mathbf{k} = k^{N-1}dk\, d\Omega(\mathbf{k}_e)$, where $d\Omega(\mathbf{k}_e)$ is a surface element of the unit hypersphere $\Omega(\mathbf{k}_e)$ in N-dimensional space. The unit vector pointing in the direction of a certain vector is denoted by adding the subscript e. For instance, $\mathbf{k}_e = \mathbf{k}/k$.

A. Preliminary calculations

We evaluate some integrals that are required below:

$$(2\pi)^{-N/2}\int k^{-\sigma}J_{m+\sigma}(sk)G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e)e^{i\mathbf{k}\cdot\rho}d\mathbf{k} = F_m(\mathbf{k}_e, \rho, s)P(\rho/s),$$

$$F_m(\mathbf{k}_e'', \rho, s) = \frac{2^{(N/2)-\sigma}}{\Gamma(\sigma + 1 - N/2)} i^m \frac{\rho^m}{s^{m+\sigma}} \times (s^2 - \rho^2)^{\sigma-(N/2)} G_m^{N/2-1}(\mathbf{k}_e'' \cdot \boldsymbol{\rho}_e),$$

$$G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e) = \Gamma(N/2 - 1)(m + N/2 - 1) C_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e),$$

$$P(\rho/s) = 1, \quad 0 \leq \rho/s < 1, \qquad = 0, \quad \rho/s > 1. \tag{41}$$

In Eq. (41) σ is a real number, $\sigma > N/2 - 1$, $m = 0, 1, \ldots$, s is a real variable, $s \geq 0$; $C_m^{N/2}$ is a Gegenbauer polynomial; $\mathbf{k}_e''$ is a fixed unit vector. In the case $N = 2$ we define $G_m^{N/2-1}$ as the limit for $N \to 2$ of its defining relation, using

$$\lim \Gamma(N/2 - 1)(m + N/2 - 1) C_m^{N/2-1}(\cos\varphi) = 2\cos m\varphi, \quad m \geq 1, \qquad = 1, \quad m = 0.$$

The proof of Eq. (41) runs as follows. First, the absolutely and uniformly convergent series[10]

$$e^{ik\cdot\rho} = 2^{N/2-1} \sum_{l=0}^{\infty} i^l (\rho k)^{1-(N/2)} J_{l+N/2-1}(\rho k) G_l^{N/2-1}(\mathbf{k}_e \cdot \boldsymbol{\rho}_e)$$

is substituted in the left-hand side of Eq. (41). By using the orthogonality relation[23]

$$\int_\Omega G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e) G_l^{N/2-1}(\mathbf{k}_e \cdot \boldsymbol{\rho}_e) d\Omega(\mathbf{k}_e) = 2\pi^{N/2} \delta_{m,l} G_m^{N/2-1}(\mathbf{k}_e \cdot \boldsymbol{\rho}_e), \tag{42}$$

the result then becomes

$$= i^m \rho^{1-N/2} G_m^{N/2-1}(\mathbf{k}_e'' \cdot \boldsymbol{\rho}_e) \times \int_0^\infty k^{N/2-\sigma} J_{m+\sigma}(sk) J_{m+(N/2)-1}(k\rho) dk.$$

The integral is a well-known Hankel transform and Eq. (41) follows directly from this.[24]

It can be verified that Eq. (41) also holds true for $N = 1$, all inner products $(\mathbf{k}_e'' \cdot \mathbf{k}_e)$, etc., being defined as 1.

Another integral for $0 \leq s \leq 1$ is

$$(2\pi)^{-N/2} \int k^{-\sigma} J_{m+\sigma}(sk) G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e) \hat{P}_N(\mathbf{k} - \mathbf{k}') d\mathbf{k} = k'^{(-\sigma)} J_{m+\sigma}(sk') G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e'). \tag{43}$$

Equation (43) states that the function $k^{-\sigma} J_{m+\sigma}(sk) \times G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e)$ does not "see" the projection operator for $0 \leq s \leq 1$. The proof is simple. Obviously, the left-hand side of Eq. (43) is the inverse Fourier transform of its Fourier transform. Furthermore, it is a convolution. Hence, its Fourier transform is the product of the Fourier transforms of $k^{-\sigma} J_{m+\sigma}(sk) G_m^{(N/2)-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e)$ and $\hat{P}_N(\mathbf{k})$. The first is given in Eq. (41) and the second is $P(\boldsymbol{\rho})$; see Eqs. (13) and (15). The left-hand side of Eq. (43) is therefore equal to

$$(2\pi)^{-N/2} \int e^{-i\mathbf{k}'\cdot\rho} F_m(\mathbf{k}_e'', \rho, s) P(\rho/s) P(\boldsymbol{\rho}) d\boldsymbol{\rho}. \tag{44}$$

If $0 \leq s \leq 1$, $P(\rho/s)P(\boldsymbol{\rho}) = P(\rho/s)$. Therefore, for $0 \leq s \leq 1$ the expression in Eq. (44) is just the inverse Fourier transform of the right-hand side of Eq. (41). Hence, Eq. (43) follows. Equation (43) is also valid for $N = 1$, all inner products $(\mathbf{k}_e'' \cdot \mathbf{k}_e)$, etc., being defined as 1. We shall not need the integral in Eq. (43) for $s \geq 1$. The following relation is a direct consequence of Eq. (43):

$$\sum_{m=0}^{\infty} G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e') = 2\pi^{N/2} \delta(\mathbf{k}_e'' - \mathbf{k}_e'). \tag{45}$$

Finally, we shall need Hankel's theorem,[25] which is valid with appropriate conditions imposed on $\hat{g}$ and with $\tau > -\frac{1}{2}$:

$$\hat{g}(k'') = \int_0^\infty (sk'')^{1/2} J_\tau(sk'') \int_0^\infty (sk') J_\tau(sk') \hat{g}(k') dk' ds. \tag{46}$$

We shall now derive $\hat{K}_N^{(\alpha)}(\mathbf{k}, \mathbf{k}')$, and during the course of the proof we shall define the required dense subspace $\hat{L}_2^{(\alpha)}(V)$ of $\hat{L}_2(V)$. In order to demonstrate clearly the essential point in the proof, stripped of the nonessential complications introduced by the geometry, the one-dimensional result is rederived. It will be seen that the new method of proof almost invariably carries over to the N-dimensional case.

B. The one-dimensional case reconsidered

Surprisingly, it can be verified directly that $\hat{K}^{(\alpha)}(k, k')$ given by Eq. (32b) fulfills Eq. (23). Consider first for $\sigma = -\frac{1}{2} + \frac{1}{2}\alpha$,

$$\frac{k''^{-\sigma}}{2(2\pi)^{1/2}} \int_0^1 s J_\sigma(sk'') \int_{-\infty}^{+\infty} k^{-\sigma} J_\sigma(sk) \times \int_{-\infty}^{+\infty} \hat{P}_1(k - k') |k'|^\alpha \hat{f}(k') dk' dk ds.$$

By Eq. (43) this becomes

$$\frac{k''^{-\sigma}}{2} \int_0^1 s J_\sigma(sk'') \int_{-\infty}^{+\infty} J_\sigma(sk') k'^{-\sigma} |k'|^\alpha \hat{f}(k') dk' ds = k''^{-\sigma} \int_0^1 s J_\sigma(sk'') \int_0^\infty J_\sigma(sk') k'^{\alpha-\sigma} \hat{f}_+(k') dk' ds. \tag{47}$$

Here $\hat{f}_+(k') = [\hat{f}(k') + \hat{f}(-k')]/2$. Note that $k^{-\sigma} J_\sigma(sk)$ is an even function of k. We have assumed that the integrations with respect to k and k' may be interchanged, a matter which will be considered further below. Now if for $\sigma = -\frac{1}{2} + \frac{1}{2}\alpha$ the integral

$$\int_0^\infty J_\sigma(sk') k'^{(\alpha-\sigma)} \hat{f}_+(k') dk' = 0, \quad s > 1, \tag{48}$$

then in Eq. (47) the integration with respect to s may be extended to infinity. By Hankel's theorem [Eq. (46)], taking $\tau = \sigma > -\frac{1}{2}$ and $\hat{g}(k') = k'^{(\alpha/2)} \hat{f}_+(k')$, the result is just $\hat{f}_+(k'')$. Replacing J_σ by $J_{\sigma+1}$ in all these formulas, and noting that $k^{-\sigma} J_{\sigma+1}(sk)$ is an odd function of k, it can similarly be proved that the result is $\hat{f}_-(k'') = [\hat{f}(k'') - \hat{f}(-k'')]/2$. Because $\hat{f} = \hat{f}_+ + \hat{f}_-$, we have therefore verified directly that $\hat{K}^{(\alpha)}(k, k')$ is given by Eq. (32b). Hankel's theorem is not applicable to all functions in $\hat{L}_2(V)$. But there exists a dense subspace of $\hat{L}_2(V)$, belonging to the domain of definition of the operator with kernel $(2\pi)^{-1/2} \hat{P}_1(k - k') k'^\alpha$, on which it holds true, Eq. (48) is fulfilled, and the interchange of integrations leading to Eq. (47) is allowed. This subspace turns out to be $\hat{L}_2^{(\alpha)}(V)$, introduced in Sec. III and studied in Appendix A.

C. Derivation of the N-dimensional case

Consider for $\sigma = \frac{1}{2}N - 1 + \frac{1}{2}\alpha$ and for $m = 0, 1, 2, \ldots$,

$$\frac{k''^{-\sigma}}{(2\pi)^{N/2}} \int_0^1 s J_{m+\sigma}(sk'') \int k^{-\sigma} J_{m+\sigma}(sk) G_m^{N/2-1}(\mathbf{k}_e'' \cdot \mathbf{k}_e) \times \int \hat{P}_N(\mathbf{k} - \mathbf{k}') k'^\alpha \hat{f}(\mathbf{k}') d\mathbf{k}' d\mathbf{k} ds. \tag{49}$$

By Eq. (43) the expression becomes

$$k''^{-\sigma}\int_0^1 sJ_{m+\sigma}(sk'') \times\int k'^{(\alpha-\sigma)}J_{m+\sigma}(sk')G_m^{N/2-1}(\mathbf{k}_e''\cdot\mathbf{k}_e')\hat{f}(\mathbf{k}')d\mathbf{k}'\,ds, \quad (50)$$

again assuming that the interchange of integrations is allowed. Now if for $\sigma = \frac{1}{2}N - 1 + \frac{1}{2}\alpha$ and for all m, the integral

$$\int_0^\infty k'^{N-1+\alpha-\sigma}J_{m+\sigma}(sk')\int_\Omega G_m^{N/2-1}(\mathbf{k}_e''\cdot\mathbf{k}_e')\hat{f}(\mathbf{k}')d\Omega(\mathbf{k}_e')dk' = 0, \quad s>1, \quad (51)$$

then in Eq. (50) the integration with respect to s may be extended to infinity. Application of Eq. (46) yields, for $m = 0, 1, \ldots,$

$$= \int_\Omega G_m^{N/2-1}(\mathbf{k}_e''\cdot\mathbf{k}_e')\hat{f}(k''\mathbf{k}_e')d\Omega(\mathbf{k}_e). \quad (52)$$

Now sum Eq. (49)—and hence Eqs. (50) and (52)—over all m and apply Eq. (45). The result is precisely $2\pi^{N/2}\hat{f}(\mathbf{k}'')$. Comparing this with Eq. (39), we see that $\hat{K}_N^{(\alpha)}(\mathbf{k},\mathbf{k}')$ has been derived, namely,

$$\hat{K}_N^{(\alpha)}(\mathbf{k},\mathbf{k}') = \frac{(kk')^{-\sigma}}{2\pi^{N/2}} \times\sum_{m=0}^\infty G_m^{N/2-1}(\mathbf{k}_e\cdot\mathbf{k}_e')\int_0^1 s\,J_{m+\sigma}(sk)J_{m+\sigma}(sk')ds. \quad (53a)$$

The defining relation for $G_m^{(N/2)-1}$ in Eq. (41) is now resubstituted. The expression can be somewhat simplified with the aid of a relation for the Gegenbauer polynomials and by calculating the integral in Eq. (53a).[26] We then have

$$\hat{K}_N^{(\alpha)}(\mathbf{k},\mathbf{k}') = \frac{\Gamma(N/2)}{\pi^{N/2}}(kk')^{-N/2-\alpha/2}\sum_{m=0}^\infty (m+\tfrac{1}{2}N+\tfrac{1}{2}\alpha) \times C_m^{N/2}(\mathbf{k}_e\cdot\mathbf{k}_e')J_{m+N/2+\alpha/2}(k)J_{m+N/2+\alpha/2}(k'). \quad (53b)$$

Equation (53b) is also valid for $N = 1$, $\mathbf{k}_e\cdot\mathbf{k}_e'$ being defined as 1. Since $C_m^{1/2}(1) = 1$, Eq. (53b) then reduces to Eq. (32a). The series in Eq. (53b) can by summed in the (uninteresting) case $\alpha = 0$, $N \neq 1$ by the addition theorem for Bessel functions [Eq. (30)]. By Eq. (20) the result is $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k}-\mathbf{k}')$, which is as it should be. The calculations leading to Eq. (53) are formal and cannot be justified for all $\hat{f}$ in $\hat{L}_2(V)$. However, there again exists a dense subspace $\hat{L}_2^{(\alpha)}(V)$, belonging to the domain of definition of the operator with kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k}-\mathbf{k}')\,k'^\alpha$, such that for $\hat{f}$ in $\hat{L}_2^{(\alpha)}(V)$ Hankel's theorem may be applied, Eq. (51) is valid, and the interchange of integrations in Eq. (49) can be justified. The discussion of this problem is given in Appendix A, The kernel $\hat{K}_N^{(\alpha)}(\mathbf{k},\mathbf{k}')$ is transformed to ordinary $(\boldsymbol{\rho},\boldsymbol{\rho}')$ space by multiplying Eq. (39) by $(2\pi)^{-N/2}e^{-i\mathbf{k}''\cdot\boldsymbol{\rho}}$, integrating over all $\mathbf{k}''$, and applying Parseval's relation to the integral with respect to $\mathbf{k}$. We then have

$$K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = (2\pi)^{-N}\iint e^{-i\mathbf{k}\cdot\boldsymbol{\rho}+i\mathbf{k}'\cdot\boldsymbol{\rho}'}\hat{K}_N^{(\alpha)}(\mathbf{k},\mathbf{k}')d\mathbf{k}\,d\mathbf{k}'.$$

By substituting Eq. (53a), taking $\sigma = \frac{1}{2}N - 1 + \frac{1}{2}\alpha$, interchanging the summation and integrations,[27] and carrying out the integrations with the aid of Eq. (41), we find that $K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = 0$ for $\rho,\rho' > 1$ and that for $\rho,\rho' < 1$, $\alpha > 0$,

$$K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = \frac{2^{1-\alpha}}{\Gamma^2(\alpha/2)}\,\frac{1}{\pi^{N/2}}\sum_{m=0}^\infty G_m^{N/2-1}(\boldsymbol{\rho}_e\cdot\boldsymbol{\rho}_e') \times\int_{\max(\rho,\rho')}^1 (s^2-\rho^2)^{\alpha/2-1}(s^2-\rho'^2)^{\alpha/2-1}\left(\frac{\rho\rho'}{s^2}\right)^m s^{3-N-\alpha}ds.$$

The summation is carried out by the following generating function,[28] valid for $\rho\rho'/s^2 < 1$:

$$\sum_{m=0}^\infty\left(\frac{\rho\rho'}{s^2}\right)^m G_m^{N/2-1}(\boldsymbol{\rho}_e\cdot\boldsymbol{\rho}_e') = \Gamma(N/2)\frac{s^{2N-4}(s^4-\rho^2\rho'^2)}{(s^4-2s^2\boldsymbol{\rho}\cdot\boldsymbol{\rho}'+\rho^2\rho'^2)^{N/2}}$$

We thus have $K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = 0$ for $\rho,\rho' > 1$, and for $\rho,\rho' < 1$ first the representation

$$K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = \frac{2^{1-\alpha}}{\Gamma^2(\alpha/2)}\,\frac{\Gamma(N/2)}{\pi^{N/2}}\int_{\max(\rho,\rho')}^1 s^{N-1-\alpha}(s^2-\rho^2)^{\alpha/2-1}(s^2-\rho'^2)^{\alpha/2-1}\,\frac{(s^4-\rho^2\rho'^2)ds}{(s^4-2s^2\boldsymbol{\rho}\cdot\boldsymbol{\rho}'+\rho^2\rho'^2)^{N/2}}, \quad (54a)$$

and secondly, a representation obtained from Eq. (54a) by introducing the new variable $w = s^2 - 2\boldsymbol{\rho}\cdot\boldsymbol{\rho}' + \rho^2\rho'^2/s^2$ and identifying the resulting integral as a hypergeometric function,[19]

$$K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}') = \frac{2^{-\alpha}\Gamma(N/2)(1-\rho^2)^{\alpha/2}(1-\rho'^2)^{\alpha/2}}{\Gamma(\alpha/2)\Gamma(1+\alpha/2)\pi^{N/2}|\boldsymbol{\rho}-\boldsymbol{\rho}'|^N} \times{}_2F_1\left[\tfrac{1}{2}N,\tfrac{1}{2}\alpha;1+\tfrac{1}{2}\alpha;-\frac{(1-\rho^2)(1-\rho'^2)}{|\boldsymbol{\rho}-\boldsymbol{\rho}'|^2}\right]. \quad (54b)$$

By comparing Eqs. (54a) and (54b) with Eqs. (33b) and (33c), we see that they are also valid in the case $N = 1$, the inner product $2\boldsymbol{\rho}\cdot\boldsymbol{\rho}'$ being defined as $2\xi\xi'$.

It is clear that the operator with kernel $K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}')$ is symmetric. It is also completely continuous on $L_2(V)$, by essentially the same arguments[5] as those given in Sec. III for the one-dimensional case. Hence, it is bounded and its domain of definition can be extended to the entire $L_2(V)$ by continuity. In ordinary space, the eigenvalue problem [Eq. (40)] takes the form

$$\lambda(\alpha)f^{(\alpha)}(\boldsymbol{\rho}) = \int_V K_N^{(\alpha)}(\boldsymbol{\rho},\boldsymbol{\rho}')f^{(\alpha)}(\boldsymbol{\rho}')d\boldsymbol{\rho}'. \quad (55)$$

The solutions of Eq. (55) provide the asymptotic solutions of the original eigenvalue problem, Eq. (12).

D. Special cases

In a number of cases the hypergeometric function in Eq. (54b) can be expressed in elementary functions. The most interesting cases seem to be the following.

First, $N = 2$, $\boldsymbol{\rho} = (\xi_1,\xi_2)$, $\alpha = 2$:

$$K_2^{(2)}(\boldsymbol{\rho},\boldsymbol{\rho}') = (1/2\pi)[\log|(\boldsymbol{\rho}/\rho)-\rho\boldsymbol{\rho}'| - \log|\boldsymbol{\rho}-\boldsymbol{\rho}'|].$$

This is the well-known expression for the Green's function of the two-dimensional diffusion equation

$$\lambda\nabla_2^2 f \equiv \lambda\left(\frac{\partial^2}{\partial\xi_1^2}+\frac{\partial^2}{\partial\xi_2^2}\right)f = -f, \quad \boldsymbol{\rho}\text{ in }V,$$

with boundary condition $f(\rho = 1) = 0$, by which, of course, f is uniquely determined. Generally, take $N = 2$ and $\alpha = 2l$, $l = 1, 2, \cdots$. A straightforward calculation then shows that

$$K_2^{(2l)}(\rho,\rho') = P_{2l-2}(\rho) + 2^{2-2l}[(-1)^{l+1}/\Gamma^2(l)]|\rho-\rho'|^{2l-2}K_2^{(2)}(\rho,\rho').$$

$P_{2l-2}(\rho)$ is a polynomial in ξ_1 and ξ_2 (with coefficients dependent on ρ'), the sum of their degrees being at most $2l-2$. It vanishes upon application of $(\nabla_2^2)^l$. Because $\nabla_2^2|\rho-\rho'|^{2l-2} = 2^2(l-1)^2|\rho-\rho'|^{2l-4}$, if the second term is operated on $(l-1)$ times by ∇_2^2, we obtain $(-1)^{l+1}K_2^{(2)}(\rho,\rho')$. Again, operating with ∇_2^2 gives a δ-function. Hence, for $l = 1, 2, \ldots$, $K_2^{(2l)}(\rho,\rho')$ is the Green's function of the elliptic partial differential equation $\lambda(\nabla_2^2)^l f = (-1)^l f$, with boundary condition $\partial f^{j-1}/\partial\rho^{j-1} = 0$, $j = 1, \ldots, l$, at $\rho = 1$.

If $N = 3$, $\rho = (\xi_1, \xi_2, \xi_3)$, $\alpha = 2$,

$$K_3^{(2)}(\rho,\rho') = (1/4\pi)[|\rho-\rho'|^{-1} - |(\rho/\rho - \rho\rho'|^{-1}].$$

This is the familiar expression for the Green's function of the three-dimensional diffusion equation

$$\lambda\nabla_3^2 f \equiv \lambda\left(\frac{\partial^2}{\partial\xi_1^2} + \frac{\partial^2}{\partial\xi_2^2} + \frac{\partial^2}{\partial\xi_3^2}\right) f = -f, \quad \rho \text{ in } V,$$

with boundary condition $f(\rho = 1) = 0$. The reader will have no difficulty in proving that for $\alpha = 2l$, $l = 1, 2, \ldots$, $K_3^{(2l)}(\rho,\rho')$ is the Green's function of the elliptic partial differential equation $\lambda(\nabla_3^2)^l f = (-1)^l f$. Compare these results with their analogues in the one-dimensional case dealt with in Sec. III. These Green's functions for $\alpha = 2$ provide the solutions of the inhomogeneous Laplace equations in two- and three-dimensional space. This is in accordance with the statement made at the end of Sec. II that the kernels $K_N^{(\alpha)}(\rho,\rho')$ supply the asymptotic solution of the inhomogeneous problem for the integral equations dealt with.

Finally, we shall calculate some integrals for $\rho < 1$ which have been obtained in the literature by probabilistic arguments, without knowledge of the precise form of the kernel[22]:

$$\int_V K_N^{(\alpha)}(\rho,\rho')d\rho' = \frac{2^{-\alpha}\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}N+\frac{1}{2}\alpha)\Gamma(1+\frac{1}{2}\alpha)}(1-\rho^2)^{\alpha/2}, \quad (56)$$

$$\int_V\int_V K_N^{(\alpha)}(\rho,\rho')K_N^{(\alpha)}(\rho',\rho'')d\rho'd\rho'' = \frac{\alpha 2^{-2\alpha}\Gamma^2(\frac{1}{2}N)}{\Gamma^2(\frac{1}{2}N+\frac{1}{2}\alpha)\Gamma^2(1+\frac{1}{2}\alpha)} \times \int_\rho^1 s(s^2-\rho^2)^{\alpha/2-1}\,{}_2F_1(\tfrac{1}{2}N, -\tfrac{1}{2}\alpha; \tfrac{1}{2}N+\tfrac{1}{2}\alpha; s^2)ds. \quad (57)$$

In certain problems in radiative transfer, Eq. (56) is of interest because it occurs in the calculation of the mean number of scatterings which a photon undergoes before leaving the volume V.

V. SUMMARY AND CONCLUSIONS

The results obtained in the foregoing sections are also valid under slightly more general assumptions[3] relating to the behavior of the Fourier transform $\hat{K}(\mathbf{k})$ in Eq. (11). They are included in the following recapitulation.

Theorem I: Let $M > 0$, and let the following integral equations be defined for $\rho = (\xi_1, \ldots, \xi_N)$ in $V = \{\rho, (\sum_{j=1}^N \xi_j^2)^{1/2} \leq 1\}$, $N = 1, 2, \ldots$:

(A) $$Mf_L(\rho) - (L/2)^N\int_V K[(L/2)(\rho-\rho')]f_L(\rho')d\rho' = \mu_L f_L(\rho);$$

(B) $$Mf_L(\rho) - (L/2)^N\int_V K[(L/2)(\rho-\rho')]f_L(\rho')d\rho' = g(\rho).$$

Let the following Fourier transform assume its maximum M for $|\mathbf{k}| = 0$ and only for $|\mathbf{k}| = 0$, and let

$$\hat{K}(\mathbf{k}) \equiv \int e^{i\mathbf{k}\cdot\mathbf{r}}K(\mathbf{r})d\mathbf{r} \sim M - c|\mathbf{k}|^\alpha F(|\mathbf{k}|); \quad |\mathbf{k}| \to 0,$$

for certain real, nonnegative constants c and α. $F(|\mathbf{k}|)$ is a nonnegative slowly varying function of $|\mathbf{k}|$ near $|\mathbf{k}| = 0$. Then, in the case of problem (A) for $L \to \infty$, j fixed,

$$\mu_{j,L} \sim c(2/L)^\alpha\lambda_j^{-1}(\alpha)F(2/L),$$

$$f_{j,L} \sim f_j^{(\alpha)} \text{ (in mean square)},$$

where the functions $f_j^{(\alpha)}$ and the numbers $\lambda_j(\alpha)$ are the solutions of the eigenvalue problem for ρ in V:

$$\lambda(\alpha)f^{(\alpha)}(\rho) = \int_V K_N^{(\alpha)}(\rho,\rho')f^{(\alpha)}(\rho')d\rho',$$

with

$$K_N^{(\alpha)}(\rho,\rho') = \frac{2^{-\alpha}\Gamma(N/2)(1-\rho^2)^{\alpha/2}(1-\rho'^2)^{\alpha/2}}{\Gamma(\alpha/2)\Gamma(1+\alpha/2)\pi^{N/2}|\rho-\rho'|^N} \times {}_2F_1\left[\frac{N}{2}, \frac{\alpha}{2}; 1+\frac{\alpha}{2}; -\frac{(1-\rho^2)(1-\rho'^2)}{|\rho-\rho'|^2}\right].$$

The solution of the inhomogeneous problem (B) for a sufficiently well-behaved g is asymptotically for $L \to \infty$

$$f_L(\rho) \sim f(\rho) = \int_V K_N^{(\alpha)}(\rho,\rho')g(\rho')d\rho'.$$

A brief comment on the application of these results in radiative transfer and neutron transport theory might be useful. Suppose that the medium is inhomogeneous and source-free, that the scattering is isotropic and conservative, and that the problem is time-independent. The equation of radiative transfer is then[29]

$$\mathbf{s}\cdot\nabla I_\nu(\mathbf{r},\mathbf{s}) + k(\nu)I_\nu(\mathbf{r},\mathbf{s}) = (1/4\pi)\int\int P(\nu,\nu')I_{\nu'}(\mathbf{r},\mathbf{s})d\Omega(\mathbf{s}')d\nu',$$

with, of course, appropriate boundary conditions added. Here $\mathbf{s}\cdot\nabla$ is the directional derivative and all other quantities have their common meaning. Suppose first that $P(\nu,\nu') = \delta(\nu-\nu')k(\nu')$ ("grey" approximation). As is well-known, the equation can then be transformed into an integral equation of the type dealt with in Sec. II. The Fourier transform of the kernel is of the type of Eq. (11) with $\alpha = 2$. It follows that at large optical thickness, the problem can be described by the diffusion approximation. This has already been known for a long time. Suppose, however, that $P(\nu,\nu')$ is separable, i.e., $P(\nu,\nu') = \mathcal{L}(\nu)k(\nu')$, as is the case in good approximation in the theory of radiative transfer in a spectral line.[30] Then, again, the transport equation can be converted into an integral equation of the type considered in Sec. II. Again, the Fourier transform of the kernel is of the type of Eq. (11), but now with $0 < \alpha \leq 1$. The application of the results of this paper in that case has been given elsewhere.[31]

ACKNOWLEDGMENTS

The author is indebted to Professor N. G. de Bruijn, Professor N. G. van Kampen, and Professor C. Bouwkamp for valuable criticism.

APPENDIX A

1. The one-dimensional case

Our purpose is to investigate the conditions to be imposed on functions $\hat{f}$ in $\hat{L}_2(V)$ in order that they belong to the subspace $\hat{L}_2^{(\alpha)}(V)$ for $0 < \alpha < 2$, i.e., the space of functions $\hat{f}$ which are $O(k^{-2})$, $k \to \infty$, and can be written as

$$\hat{f}(k) = \sum_{n=0}^{\infty} (\hat{f}, \hat{g}_{n,\sigma})_{\tau=0} \hat{g}_{n,\sigma}(k), \qquad \sigma = \tfrac{1}{2} + \tfrac{1}{2}\alpha. \tag{A1}$$

We introduce the even and odd parts of $\hat{f}$: $\hat{f}_+(k) = [\hat{f}(k) + \hat{f}(-k)]/2$ and $\hat{f}_-(k) = [\hat{f}(k) - \hat{f}(-k)]/2$. By substituting the defining relations of $\hat{g}_{n,\sigma}$ [Eq. (25)] and of the inner product, Eq. (A1), in terms of $\hat{f}_+$ and $\hat{f}_-$, takes the more familiar form

$$\hat{f}_+(k) = 2k^{-\sigma} \sum_{n=0}^{\infty} (2n + \sigma) J_{2n+\sigma} \int_0^{\infty} k'^{\sigma-1} J_{2n+\sigma}(k') \hat{f}_+(k') dk', \tag{A2}$$

$$\hat{f}_-(k) = 2k^{-\sigma} \sum_{n=0}^{\infty} (2n + 1 + \sigma) J_{2n+1+\sigma}(k) \times \int_0^{\infty} k'^{\sigma-1} J_{2n+1+\sigma}(k') \hat{f}_-(k') dk'. \tag{A3}$$

The requirements for Eqs. (A2) and (A3) to apply are similar. We therefore deal only with Eq. (A2). Sufficient conditions for Eq. (A2) to hold true have been given by Wilkins.[13] They are: (i) $\hat{f}_+$ is continuous: (ii) $\hat{f}_+(k) = O(k^{-\alpha/2})$, $k \to \infty$; and (iii) [see also Eq. (48)] the following integral exists and has the property

$$\int_0^{\infty} J_{\alpha/2-1/2}(sk') k'^{1/2+\alpha/2} \hat{f}_+(k') dk' = 0, \qquad s > 1. \tag{A4}$$

Condition (ii) is weaker than the requirement $\hat{f}_+(k) = O(k^{-2})$. We replace it by the latter. Then also the integral in Eq. (A4) will exist. We now show that these conditions are met if the following demands are imposed on the inverse Fourier transform f_+ of $\hat{f}_+$, related by

$$\hat{f}_+(k) = (2\pi)^{-1/2} \int_{-1}^{+1} \cos(k\xi) f_+(\xi) d\xi. \tag{A5}$$

They are, for $0 < \alpha < 2$, $f_+(1) = 0$; f_+ is differentiable and its derivative is absolutely continuous. In that case, $\hat{f}_+$ is continuous—condition (i). By partial integration of Eq. (A5) and application of the Riemann–Lebesque theorem,[32] it is readily shown that $\hat{f}_+(k) = O(k^{-2})$, $k \to \infty$—condition (ii). The result of the partial integration is substituted in Eq. (A4). Upon application of Fubini's theorem, condition (iii) becomes

$$\int_{-1}^{+1} f_+'(\xi) \int_0^{\infty} J_{\alpha/2-1/2}(sk) k^{\alpha/2-1/2} \sin(k\xi) dk d\xi = 0, \qquad s > 1.$$

The integral with respect to k is well known.[33] It is equal to zero for $0 < |\xi| < s$, $0 < \alpha < 2$. Therefore, condition (iii) is also fulfilled. Note that explicit use has been made of the fact that f_+ vanishes outside $[-1, +1]$. The set of functions f_+ is dense in $L_2(V)$, for among them are the solutions of the diffusion equation with boundary condition $f_+(1) = 0$. They constitute a complete set. Hence, $\hat{L}_2^{(\alpha)}(V)$ is dense in $\hat{L}_2(V)$. The arguments for $0 < \alpha < 2$ can readily be extended to other values of α. For instance, take $2 \leq \alpha < 3$. We now impose of f_+ that f_+ is two times differentiable, that $f_+(1) = f_+'(1) = 0$, and that f_+'' is absolutely continuous. It is easily verified that then, again, Eq. (A2) applies and that $\hat{f}_+(k) = O(k^{-3})$, so that the set belongs to the domain of definition of the operator with kernel $(2\pi)^{-1/2}\hat{P}_1(k - k')|k'|^{\alpha}$, $\alpha < 3$. The set is of course dense in $L_2(V)$. The reader will also have no difficulty in verifying that in the rederivation of the one-dimensional result in Sec. IVB our conditions on $\hat{f}_+$ ensure in addition that Hankel's theorem is valid and that the interchange of the integrations is justified.

2. The N-dimensional case

In the case $N \geq 2$, $0 < \alpha < 2$, we shall define the space $\hat{L}_2^{(\alpha)}(V)$ as the set of (i) continuous functions $\hat{f}$ which are (ii) $O(k^{-N/2-3/2})$, $k \to \infty$, and (iii) have the property that the following integral (which exists) vanishes for $s > 1, m = 0, 1, \cdots$, $\sigma = \frac{1}{2}N - 1 + \frac{1}{2}\alpha$:

$$\int_0^{\infty} J_{m+\sigma}(sk') k'^{N/2+\alpha/2} \times \int_{\Omega} G_m^{N/2-1}(\mathbf{k}_e \cdot \mathbf{k}_e') \hat{f}(\mathbf{k}') d\Omega(\mathbf{k}_e') d\mathbf{k}' = 0, \qquad s > 1. \tag{A6}$$

It is easily verified in Sec. IVC, that $\hat{L}_2^{(\alpha)}(V)$ belongs to the domain of definition of the operator with kernel $(2\pi)^{-N/2}\hat{P}_N(\mathbf{k} - \mathbf{k}')\mathbf{k}'^{\alpha}$, that Hankel's theorem is valid, and that the interchange of integrations in Eq. (49) is justified. To $\hat{L}_2^{(\alpha)}(V)$ there corresponds a space $L_2^{(\alpha)}(V)$ of functions f related to $\hat{f}$ by

$$\hat{f}(\mathbf{k}) = (2\pi)^{-N/2} \int_V e^{i\mathbf{k}\cdot\boldsymbol{\rho}} f(\rho) d\boldsymbol{\rho}. \tag{A7}$$

We shall show that the above-mentioned requirements imposed on $\hat{f}$ are met if certain smoothness conditions are fulfilled for f. We first consider only complex functions f which can be written as the following finite sum:

$$f(\boldsymbol{\rho}) = \sum_{n,l} a_{n,l} f_n(\rho) Y_n^l(\boldsymbol{\rho}_e). \tag{A8}$$

The real radial function $f_n(\rho)$ in Eq. (A8) will be considered below; $Y_n^l(\boldsymbol{\rho}_e)$ is a complex surface harmonic of degree n in $(N-1)$ variables[34]; l stands for an array of indices $l = (l_1, \ldots, l_{N-1})$. The summation with respect to l is over all independent surface harmonics of degree n. For instance, if $N = 2$, $Y_n^l(\varphi) = e^{il\varphi}$, $l = \pm n$. If $N = 3$, $Y_n^l(\vartheta, \varphi) = (-1)^{n+l} 2^n n!\, e^{-il\varphi} P_n^l(\cos\vartheta)/(n+1)!$, $l = 0, \pm 1, \ldots, \pm n$.

By substituting in Eq. (A7) the expansion of $e^{i\mathbf{k}\cdot\boldsymbol{\rho}}$ given under Eq. (41) and Eq. (A8), and by calculating the integrals with respect to the angular variables with the aid of an orthogonality relation[35] for the surface harmonics, we have

$$\hat{f}(\mathbf{k}) = k^{1-N/2} \sum_{n,l} a_{n,l} i^n Y_n^l(\mathbf{k}_e) \times \int_0^1 J_{n+(N/2)-1}(k\rho) \rho^{N/2} f_n(\rho) d\rho. \tag{A9}$$

We require for $0 < \alpha < 2$ that $f_n(1) = 0$, that $\rho^{-n} f_n(\rho)$ is differentiable, and that its derivative is absolutely continuous. Then $\hat{f}$ is continuous—condition (i). Partial integration of Eq. (A9) yields

$$\hat{f}(\mathbf{k}) = -k^{-N/2} \sum_{n,l} a_{n,l} i^n Y_n^l(\mathbf{k}_e) \times \int_0^1 J_{n+N/2}(k\rho) \rho^{n+N/2} \frac{d}{d\rho} [\rho^{-n} f_n(\rho)] d\rho. \tag{A10}$$

Substitution of the familiar asymptotic expression for $J_{n+N/2}(k\rho)$, $k \to \infty$, and application of the Riemann–Lebesgue theorem[32] shows that $\hat{f}(\mathbf{k}) = O(k^{-N/2-3/2})$, $k \to \infty$—condition (ii). Now substitute Eq. (A10) in Eq. (A6). The angular integrals are readily evaluated[35] by an orthogonality relation for the surface harmonics. For $s > 1$ condition (iii) takes the form

$$2\pi^{N/2} \left[\sum_l a_{m,l} Y_m^l(\mathbf{k}_e)\right] i^m \int_0^1 \rho^{m+N/2} \frac{d}{d\rho} [\rho^{-m} f_m(\rho)]$$

$$\times \int_0^\infty J_{m+\sigma}(sk)J_{m+N/2}(k\rho)k^{\alpha/2}dkd\rho = 0, \quad s > 1,$$

if the sum over n contains m. If this is not the case, then the left-hand side of this equation is zero, and condition (iii) is fulfilled. The integral with respect to k is a familiar Hankel transform.[36] The result vanishes for $0 < \rho < s$, $0 < \alpha < 2$. Hence, condition (iii) is met. The finite linear combinations of functions $f_n(\rho)Y_n^l(\rho_e)$, with the above-mentioned conditions imposed on $f_n(\rho)$, are dense in $L_2(V)$. Among them are the solutions of the diffusion equation in N-dimensional space with boundary condition $f_n(1) = 0$. They constitute a complete set. Hence, $\hat{L}_2^{(\alpha)}(V)$ is dense in $\hat{L}_2(V)$. The analysis can be extended to $\alpha \geq 2$ in complete analogy with the one-dimensional case. This is left to the reader.

APPENDIX B

The purpose of this appendix is to prove directly the equivalence of the representation of $K^{(\alpha)}(\xi, \xi')$ given in Eq. (5) with the form given in Eq. (33a). Our starting point is Eq. (5) valid for $|\xi|, |\xi'| < 1, 0 < \alpha < 2, \xi \neq \xi'$ if $0 < \alpha \leq 1$,

$$K^{(\alpha)}(\xi, \xi') = \frac{\cos(\alpha\pi/2)}{\Gamma(\alpha)} |\xi - \xi'|^{\alpha-2} \max(0, \xi - \xi') - \frac{\sin(\alpha\pi/2)}{\pi\Gamma(\alpha)} (1 - \xi'^2)^{\alpha/2} \int_{-1}^{\xi} \frac{(\xi - \zeta)^{\alpha-1}}{(1 - \zeta^2)^{\alpha/2}} \frac{d\zeta}{\zeta - \xi'}. \quad \text{(B1)}$$

As has been mentioned already, the integral in Eq. (B1) has to be interpreted as a principal value if $\xi > \xi'$. Equation (B1) may also be written as

$$K^{(\alpha)}(\xi, \xi') = \mathrm{Re}\frac{ie^{i(\pi\alpha/2)}}{\pi\Gamma(\alpha)} (1 - \xi'^2)^{\alpha/2} \times \int_{-1}^{\xi} d\zeta \frac{(\xi - \zeta)^{\alpha-1}}{(1 - \zeta^2)^{\alpha/2}} \left[\frac{P}{\zeta - \xi'} - i\pi\delta(\zeta - \xi')\right];$$

P denotes principal value and $\delta(\zeta - \xi')$ the common δ function. We replace the principal value and the δ function by using the representation ($i0 \doteq i\epsilon$, $\epsilon > 0$)

$$P/(\zeta - \xi') - i\pi\delta(\zeta - \xi') = -i\int_0^\infty \exp[i(\zeta + i0 - \xi')s]ds.$$

Equation (B1) thus becomes

$$K^{(\alpha)}(\xi, \xi') = \mathrm{Re}\frac{e^{i\alpha(\pi/2)}}{\pi\Gamma(\alpha)} (1 - \xi'^2)^{\alpha/2} \int_{-1}^{\xi} d\zeta \frac{(\xi - \zeta)^{\alpha-1}}{(1 - \zeta^2)^{\alpha/2}} \times \int_0^\infty \exp[i(\zeta + i0 - \xi')s]ds. \quad \text{(B2)}$$

By Cauchy's theorem, Eq. (B2) is equivalent to Eq. (B1) if the eigenfunction $f^{(\alpha)}(\xi)$ is analytic in a neighborhood of the part of the line Im $\xi = 0$, $|\xi| < 1$.

In Eq. (B2), consider first

$$\frac{e^{i\alpha(\pi/2)}}{(1 - \zeta^2)^{\alpha/2}} \int_0^\infty \exp[i(\zeta + i0 - \xi')s]ds.$$

We use the following expansion,[10] which is uniformly convergent for $|\xi'| < 1$ in any closed domain in the (complex) s plane.

$$e^{-i\xi's} = s^{-1/2-\alpha/2} 2^{1/2+\alpha/2} \Gamma(\tfrac{1}{2} + \tfrac{1}{2}\alpha) \times \sum_{n=0}^{\infty} (-i)^n (n + \tfrac{1}{2} + \tfrac{1}{2}\alpha) C_n^{1/2+\alpha/2}(\xi') J_{n+1/2+\alpha/2}(s). \quad \text{(B3)}$$

Upon substitution of Eq. (B3) and interchange of the summation and integration, we have to calculate[10,37]

$$\frac{e^{i\alpha\pi/2}}{(1 - \zeta^2)^{\alpha/2}} (-i)^n \int_0^\infty e^{i(\zeta+i0)s} J_{n+1/2+\alpha/2}(s)s^{-1/2-\alpha/2}ds$$
$$= ie^{i\alpha\pi} \frac{2^{-n-1/2-\alpha/2}}{(\zeta + i0)^{n+1+\alpha}} \frac{n!}{\Gamma(n + 3/2 + \alpha/2)} \times {}_2F_1\left[\frac{n + 1 + \alpha}{2}, \frac{n + 2 + \alpha}{2}; n + \frac{3}{2} + \frac{\alpha}{2}; (\zeta + i0)^{-2}\right]$$
$$= \left(\frac{2}{\pi}\right)^{1/2} \frac{n!}{\Gamma(n + 1 + \alpha)} \frac{ie^{i\pi\alpha/2}}{[(\zeta + i0)^2 - 1]^{\alpha/4}} Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0);$$

$Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0)$ is a Legendre function of the second kind.[37] The result is

$$\frac{e^{i\alpha(\pi/2)}}{(1 - \zeta^2)^{\alpha/2}} \int_0^\infty e^{i(\zeta+i0-\xi')s}ds$$
$$= \frac{ie^{i(\pi/2)\alpha} 2^{1+\alpha/2}}{\pi^{1/2}[(\zeta + i0)^2 - 1]^{\alpha/4}} \Gamma\left(\frac{1}{2} + \frac{\alpha}{2}\right) \sum_{n=0}^{\infty} \frac{(n + 1/2 + \alpha/2)n!}{\Gamma(n + 1 + \alpha)} \times C_n^{1/2+\alpha/2}(\xi') Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0). \quad \text{(B4)}$$

We now take the real part of Eq. (B4)—See Eq. (B2). The Gegenbauer polynomial is real, while the Legendre function $Q_{n+\alpha/2}^{\alpha/2}$ is complex. Furthermore, the latter is somewhat difficult to handle. We therefore transform it to the Legendre functions of the first kind[37] $P_{n+\alpha/2}^{\alpha/2}$ and $P_{n+\alpha/2}^{-\alpha/2}$, which are easier:

$$\frac{ie^{i\alpha\pi/2}}{[(\zeta + i0)^2 - 1]^{\alpha/4}} Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0) = (-1)^{n+1} \frac{ie^{-i\alpha\pi/4}}{(1 - \zeta^2)^{\alpha/4}} \times Q_{n+\alpha/2}^{\alpha/2}(-\zeta - i0) = \frac{\pi}{2} \frac{(-1)^{n+1}}{(1 - \zeta^2)^{\alpha/4}} \frac{ie^{i\alpha\pi/4}}{\sin(\alpha\pi/2)} \times \left[P_{n+\alpha/2}^{\alpha/2}(-\zeta - i0) - \frac{\Gamma(n + 1 + \alpha)}{n!} P_{n+\alpha/2}^{-\alpha/2}(-\zeta - i0)\right]. \quad \text{(B5)}$$

The appropriate hypergeometric equations are substituted for the Legendre functions of the first kind in Eq. (B5). Since the hypergeometric functions take real values for real values of the argument, we can easily separate off the real part. We have[37]

$$\mathrm{Re}\{ie^{i\alpha\pi/2}/[(\zeta + i0^2 - 1]^{\alpha/4}\} Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0)^2 = \tfrac{1}{2}\pi[(-1)^n/\Gamma(1 - \alpha/2)](1 + \zeta)^{-\alpha/2} \times {}_2F_1(-n - \tfrac{1}{2}\alpha, n + 1 + \tfrac{1}{2}\alpha; 1 - \tfrac{1}{2}\alpha; \tfrac{1}{2}(1 + \zeta)). \quad \text{(B6)}$$

We reconsider Eq. (B2), substitute Eq. (B4), and interchange the summation and integration. Each term of the series is equal to

$$\frac{2^{1+\alpha/2}}{\pi^{1/2}} \Gamma(\tfrac{1}{2} + \tfrac{1}{2}\alpha)(1 - \xi'^2)^{\alpha/2} \times [(n + \tfrac{1}{2} + \tfrac{1}{2}\alpha)n!/\Gamma(n + 1 + \alpha)]C_n^{1/2+\alpha/2}(\xi')$$

multiplied by

$$\mathrm{Re}\frac{ie^{i\pi\alpha/2}}{\pi\Gamma(\alpha)} \times \int_{-1}^{\xi} (\xi - \zeta)^{\alpha-1} Q_{n+\alpha/2}^{\alpha/2}(\zeta + i0) \frac{d\zeta}{[(\zeta + i0)^2 - 1]^{\alpha/4}}. \quad \text{(B7)}$$

By applying Eq. (B6), the integral becomes

$$\frac{1}{2}\frac{(-1)^n}{\Gamma(\alpha)\Gamma(1-\alpha/2)}\int_{-1}^{\xi}\frac{(\xi-\zeta)^{\alpha-1}}{(1+\zeta)^{\alpha/2}} \times {}_2F_1\left(-n-\frac{\alpha}{2}, n+1+\frac{\alpha}{2}; 1-\frac{\alpha}{2}; \frac{1+\zeta}{2}\right)d\zeta$$

$$= \frac{1}{2}(-1)^n\frac{(1+\xi)^{\alpha/2}}{\Gamma(1+\alpha/2)}{}_2F_1\left(-n-\frac{\alpha}{2}, n+1+\frac{\alpha}{2}; 1+\frac{\alpha}{2}; \frac{1+\xi}{2}\right)$$

$$= (-1)^n\frac{2^{-n-(\alpha/2)-1}}{\Gamma(n+1+\alpha/2)}\frac{d^n}{d\xi^n}(1-\xi^2)^{n+\alpha/2}.$$

By Rodrigues' formula[38] the last expression is precisely

$$\frac{2^{\alpha/2-1}}{\pi^{1/2}}\Gamma(1/2+\alpha/2)\frac{n!}{\Gamma(n+1+\alpha)}(1-\xi^2)^{\alpha/2}C_n^{1/2+\alpha/2}(\xi). \tag{B8}$$

Taking together the results of Eqs. (B7) and (B8), we have

$$K^{(\alpha)}(\xi,\xi') = (2^\alpha/\pi)\Gamma^2(\tfrac{1}{2}+\tfrac{1}{2}\alpha)(1-\xi^2)^{\alpha/2}(1-\xi'^2)^{\alpha/2} \times \sum_{n=0}^{\infty}(n+\tfrac{1}{2}+\tfrac{1}{2}\alpha)\frac{\Gamma^2(n+1)}{\Gamma^2(n+1+\alpha)}C_n^{1/2+\alpha/2}(\xi)C_n^{1/2+\alpha/2}(\xi'). \tag{B9}$$

Therefore, the equivalence of Eqs. (5) and (33a) for $0<\alpha<2$ has been established directly. The justification of the interchange of summation and integration leading to Eq. (B7) appear to present no difficulties for any $i0=i\epsilon$, $\epsilon>0$, because, in that case, the series in Eq. (B4) turns out to be uniformly convergent. However, the passage to the limit $\epsilon\to 0$, as tacitly done above, is justified for $0<\alpha\leq 1$ only if $\xi\neq\xi'$.

Obviously, the analytic continuation[2] of $K^{(\alpha)}(\xi,\xi')$ in Eq. (B1) to $\mathrm{Re}\alpha\geq 2$ coincides with the expression in Eq. (B8), since the latter is an analytic function of α for ξ,ξ' fixed, $\mathrm{Re}\alpha>0$ ($\xi\neq\xi'$ if $0<\mathrm{Re}\alpha\leq 1$).

---

[1] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon, Oxford, 1948).

[2] H. Widom, Trans. Am. Math. Soc. **98,** 430 (1961); Trans. Am. Math. Soc. **100,** 252 (1961). In the latter paper, in Eq. (6) read $|x-y|^{\alpha-2}$ instead of $|x-y|^{\alpha-1}$. See also: M. Kac, *Proc. of the Sec. Berkeley Symp. on Math. Statistics and Probability,* edited by J. Neyman (University of California Press, Berkely, (1951); H. Kesten, Illinois J. Math. **5,** 267 (1961).

[3] H. Widom, Trans. Am. Math. Soc. **106,** 391 (1963).

[4] F. Riesz and B. Sz-Nagy, *Leçons d'analyse fonctionelle* (Académie des Sciences de Hongrie, Budapest, 1955).

[5] A kernel $A(\mathbf{r},\mathbf{r}')$ defined on a measurable, bounded set U in N-dimensional space, such that $|\mathbf{r}-\mathbf{r}'|^{N-\beta}A(\mathbf{r},\mathbf{r}'),\beta>0$, is bounded on U, is sometimes called a kernel with a weak singularity. For a proof that the corresponding integral operator is completely continuous on $L_2(U)$, see S. G. Mikhlin, *Mathematical Physics* (North-Holland, Amsterdam, 1970). In this reference it is also shown that the integral operator is completely continuous on $C(U)$, the space of continuous functions on U. It follows that the eigenfunctions are continuous. Consequently, if $K(\mathbf{r})$ in $L_1(E_N)$ is bounded everywhere except at $r=0$, where it has a singularity of the type mentioned above, then the integral operator in Eq. (9) is completely continuous on $L_2(V')$. If $K(\mathbf{r}-\mathbf{r}')$ has also other singularities, it can be approximated by kernels of this type, and is therefore also completely continuous on $L_2(V')$. It can also be shown by a direct argument that if $K(\mathbf{r})$ is in $L_1(E_N)$, the integral operator is completely continuous on $C(V')$.

[6] I. Sneddon, *Fourier Transforms* (McGraw-Hill, New York, 1951).

[7] Proof: Eq. (17) is in operator notation $f=P*f$, where $*$ means convolution; by applying inverse Fourier transformation we obtain $Pf=P^2f$; this is an elementary identity for P.

[8] Reference 4, p. 293.

[9] In particular, the domain of definition of the inverse operator comprises the set of functions for which Eq. (19) can be defined only in the weak sense. The weak eigenfunctions of Eq. (19) are eigenfunctions in the usual sense of the completely continuous inverse operator.

[10] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1966).

[11] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol, II, p. 345 (44). The formula holds true for odd values of n too.

[12] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II, p. 64.

[13] J. E. Wilkins Jr., Trans. Am. Math. Soc. **64,** 359 (1948); Trans. Am. Math. Soc. **69,** 55 (1950).

[14] G. Szegö, *Orthogonal Polynomials*, AMS Colloquium Publications, Vol. XXIII (Am. Math. Soc., Providence, 1939).

[15] The identity is easily seen to apply for $n=0$ and $n=1$ and for all positive α. Consider the equation for $n+1$, $n\geq 1$. Use Ref. 12, Vol. II, p. 175 (15) to remove in the integral the polynomial — and apply the orthogonality relation, Ref. 11, Vol. II, p. 281 (8). The derivative of the nth order polynomial is removed by Ref. 12, Vol. II, p. 176 (23). Apply now the induction hypothesis.

[16] The equivalence of Eqs. (32a) and (32b) is of basic importance in the theory of Neuman series; see Ref. 13. The present proof seems to be new and is much simpler than the one in Ref. 13.

[17] Reference 11, Vol. I, pp. 44, 100.

[18] We have to prove that the integration and summation may be interchanged. Because of finite interval, it is easier to justify the reverse case, namely that $\hat{K}^{(\alpha)}(k,k')$ in Eq. (32a) is the Fourier transform of $K^{(\alpha)}(\xi,\xi')$ in Eq. (33a). It is proved further on in the paper that Eq. (33a) is singular at $\xi=\xi'$ for $0<\alpha\leq 1$. Therefore we take first $\alpha>1$. By Ref. 14, p. 197, the series in Eq. (33a) is uniformly convergent for $|\xi|,|\xi'|<1$, and thus bounded everywhere. If it is multiplied by $(2\pi)^{-1/2}e^{-ik'\xi'}$, then the integration with respect to may be interchanged with the summation. The resulting series is absolutely and uniformly convergent by the inequalities under Eq. (30), In the Fourier transformation the summation and integration with respect to ξ may be interchanged. The result is analytically continued to $\alpha>0$, $\xi\neq\xi'$, the sum being an analytic function of α for $\mathrm{Re}\alpha>0$, $\xi\neq\xi'$. The justification of Eq. (33b) is similar.

[19] Reference 12, Vol. I, p. 59.

[20] M. Kac and H. Pollard, Can. J. Math. **2,** 375 (1950).

[21] H. Söhngen, Math. Z. **45,** 245 (1939).

[22] J. Elliott, Ill. J. Math. **3,** 200 (1959); R. K. Getoor, Trans. Am. Math. Soc. **101,** 75 (1961). Equations (38) and (57) should be multiplied by 2 to be comparable.

[23] Reference 12, Vol. II, p. 245. The case $N=2$ requires separate verification.

[24] Reference 11, Vol. II, p. 48.

[25] Reference 12, Vol. II, p. 73.

[26] Differentiate ref. 12, Vol. II, p. 176 (26) in order to split up Eq. (53a) into the difference of two series. Calculate the integral and apply two times the recursion relation in Ref. 12, Vol. II, p. 12 (56).

[27] The justification is essentially the same as that given for the one-dimensional case; see Ref. 18.

[28] Use Ref. 12, Vol II, p. 176 (26), differentiated with respect to x, in order to reduce this formula to the more familiar generating function of Ref. 12, Vol. II, p. 177 (29).

[29] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960); B. Davidson, *Neutron Transport Theory* (Clarendon, Oxford, 1958); K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).

[30] This Markof assumption has been introduced by T. Holstein, Phys. Rev. **72,** 1212 (1947); Phys. Rev. **83,** 1159 (1951), and by L. M. Biberman, Zh. Eksp. Teor. Fiz. **17,** 416 (1947); Zh. Eksp. Teor. Fiz. **19,** 584 (1949).

[31] C. van Trigt, Phys. Rev. **181,** 97 (1969); Phys. Rev. A **1,** 1298 (1970); Phys. Rev. A **4,** 1303 (1971).

[32] E. T. Whitaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1962), p. 172.

[33] Reference 11, Vol. I, p. 100.

[34] Reference 12, Vol. II, p. 240. The case $N=2$ requires separate verification.

[35] Reference 12, Vol. II, p. 247.

[36] Reference 11, Vol. II, p. 48.

[37] Reference 12, Vol. I, Chap. 3.

[38] Reference 12, Vol. II, p. 175.

A space-time calculus based on pairs of null directions*

R. Geroch

Department of Physics, University of Chicago, Chicago, Illinois 60637

A. Held

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

R. Penrose

Department of Mathematics, Birkbeck College, London, England

(Received 21 July 1972)

A formalism is presented for the treatment of space-times, which is intermediate between a fully covariant approach and the spin-coefficient method of Newman and Penrose. With the present formalism, a pair of null directions only, rather than an entire null tetrad, is singled out at each point. The concept of a spin- and boost-weighted quantity is defined, the formalism operating entirely with such quantities. This entails the introduction of modified differentiation operators, one of which represents a natural extension of the definition of the operator $\eth$ which had been introduced earlier by Newman and Penrose. For suitable problems, the present formalism should lead to considerable simplifications over that achieved by the standard spin-coefficient method.

1. INTRODUCTION

Certain types of calculation in general relativity are conveniently carried our using a tetrad formalism[1]: four vector fields are introduced which are linearly independent at each point. By taking components with respect to these vector fields any tensor field or tensor equation may be replaced by a system of scalar fields or scalar equations. In fact, ordinary calculations using tensor components according to some coordinate system are really a special case, since here we may use the gradients of the four scalar fields x^0, x^1, x^2, x^3 as the (covariant) tetrad vectors. (We should be careful to distinguish an actual tensor from its set of components in some coordinate or tetrad system.) For definiteness we shall adopt the "abstract index" conventions[2] according to which $\chi_{ab}{}^{c}{}_{d}$ denotes a tensor whereas $\chi_{\mathfrak{ab}}{}^{\mathfrak{c}}{}_{\mathfrak{d}}$ denotes its set of components in a tetrad or coordinate system.[3] For example, a single vector field ξ_a may be described in terms of four scalar fields $\xi_0, \xi_1, \xi_2, \xi_3$, where $\xi_{\mathfrak{a}} = \xi_a \delta^a_{\mathfrak{a}}$, the tetrad system being denoted by $\delta^a_0, \delta^a_1, \delta^a_2, \delta^a_3$. Similarly, a tensor equation such as $\nabla_{(a}\xi_{b)} = 0$, for example, can be replaced, using this tetrad system, by the ten separate scalar equations

$$\nabla_{(\mathfrak{a}}\xi_{\mathfrak{b})} = \Gamma^{\mathfrak{c}}_{(\mathfrak{a}\mathfrak{b})}\xi_{\mathfrak{c}}, \tag{1.1}$$

where $\nabla_0, \nabla_1, \nabla_2, \nabla_3$ are the tetrad components ("intrinsic derivatives") $\nabla_{\mathfrak{a}} = \delta^a_{\mathfrak{a}}\nabla_a$ of the covariant derivative operator ∇_a and where

$$\Gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}} = -\delta^a_{\mathfrak{a}}\delta^b_{\mathfrak{b}}\nabla_b\delta^{\mathfrak{c}}_a. \tag{1.2}$$

Here, the *dual* tetrad $\delta^{\mathfrak{a}}_a = \delta^0_a, \delta^1_a, \delta^2_a, \delta^3_a$, of $\delta^a_{\mathfrak{a}}$, is being employed (defined by $\delta^{\mathfrak{a}}_a\delta^a_{\mathfrak{b}} = \delta^{\mathfrak{a}}_{\mathfrak{b}}$, that is, $\delta^{\mathfrak{a}}_a\delta^a_{\mathfrak{b}}$ is unity if $\mathfrak{a} = \mathfrak{b}$ and zero otherwise). For a general tetrad there will be *64* algebraically unrelated scalar fields $\Gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}}$. In the case of tetrads arising from coordinate systems we have the symmetry $\Gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}} = \Gamma^{\mathfrak{c}}_{\mathfrak{b}\mathfrak{a}}$ (Christoffel symbol symmetry), and we have just *40* algebraically unrelated scalar fields $\Gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}}$. On the other hand, if the tetrads are normalized according to some scheme:

$$\delta^a_{\mathfrak{a}}\delta^b_{\mathfrak{b}}g_{ab} = g_{\mathfrak{a}\mathfrak{b}}, \tag{1.3}$$

where the $g_{\mathfrak{a}\mathfrak{b}}$ are constants, being the elements of some numerical symmetric matrix [e.g., diag(1, − 1, − 1, − 1) for an orthonormal tetrad in a locally Minkowskian space], then we have the skew-symmetry

$$g_{\mathfrak{c}\mathfrak{b}}\Gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}} = -g_{\mathfrak{c}\mathfrak{a}}\Gamma^{\mathfrak{c}}_{\mathfrak{b}\mathfrak{b}}. \tag{1.4}$$

Thus we have only *24* independent scalar fields. These are called *Ricci rotation coefficients* (sometimes written $\gamma^{\mathfrak{c}}_{\mathfrak{a}\mathfrak{b}}$).

Since tetrad equations such as (1.1) are scalar equations, theorems on the existence and uniqueness of differential equations can be used directly. With a normalized tetrad, we have only 24 Γ's so we may expect some simplification over the use of coordinate components. Also, if one wishes to solve the gravitational equations explicitly, it is sometimes convenient to introduce normalized tetrads. Occasionally the equations become more transparent when written in terms of normalized tetrads. This may be the case when one or more of the vectors is *preferred;* for example, when a tetrad vector is chosen to be a Killing vector or, say, a multiple principal null vector of the Weyl tensor. But often equations become more difficult to interpret when written out explicitly in terms of tetrads. In the completely covariant formalism, every expression has a definite interpretation (in principle, at least) in terms of the geometry of the space-time, while in an explicit tetrad formalism, many of the equations are simply expressing relations between (perhaps arbitrarily chosen) tetrad vectors.

The spin-coefficient formalism[4] (referred to henceforth as the NP formalism) may be regarded, in the present context, as arising from a special case of a normalized tetrad formalism. As a consequence, there are 24 independent rotation coefficients; but it turns out that these may be combined in pairs to give just *12* independent *complex* scalar fields called *spin coefficients*. Real equations combine together in a natural way and it becomes a feasible proposition to assign a different letter to each spin coefficient and write out all equations explicitly in terms of components.

A common feature, then of the spin coefficient and tetrad formalisms is that they both involve a choice of a complete tetrad at each point. Often this choice has to be made arbitrarily to some extent since the geometry of the situation may not define that much structure at each point in a natural way. However, in many problems there may be some smaller amount of structure naturally singled out at each point. In particular, it is frequently the case that, say, one or two vectors or directions are naturally defined at each point, but not four vectors. In radiation problems, or when one is dealing with space-

Copyright © 1973 by the American Institute of Physics

like (or timelike) 2-surfaces, it is often the case that two (or one) *null directions* are singled out at each point by the conditions of the problem. The formalism that we set forth in this paper is designed specifically to deal with such situations.

We suppose that two of our future-pointing null directions are assigned at each point of the space–time. We can then choose, as two of our tetrad vectors, two null vectors l^a and n^a which point in these two directions. We can choose a normalization $l^a n_a = 1$ and take our remaining tetrad vectors to be unit spacelike vectors X^a and Y^a orthogonal to each of l^a and n^a and to each other. However, there is a two-dimensional "gauge" freedom remaining at each point, namely the 2-parameter subgroup of the Lorentz group at each point which preserves these two null directions. This group is generated by the boosts

$$l^a \to r l^a, \qquad n^a \to r^{-1} n^a \tag{1.5}$$

and the spatial rotations

$$m^a \to e^{i\theta} m^a, \tag{1.6}$$

where the complex vector m^a is defined by $m^a = 2^{-1/2}(X^a + iY^a)$. It is convenient to combine r and θ together in the form of the complex number λ where $\lambda^2 = r e^{i\theta}$. Then the "gauge group" at each point is seen to be the multiplicative group of complex numbers λ. Our formalism will deal with scalars associated with such a tetrad $(l^a, m^a, \bar{m}^a, n^a)$ where the scalars undergo transformations[5]

$$\eta \to \lambda^p \bar{\lambda}^q \eta \tag{1.7}$$

whenever the tetrad is changed according to (1.5) and (1.6). Such a scalar will be called a *spin- and boost-weighted scalar of type* $\{p, q\}$ (or simply a scalar of type $\{p, q\}$). The *spin-weight* is $\frac{1}{2}(p - q)$ and the *boost weight* is $\frac{1}{2}(p + q)$.

The formalism may be though of as a compromise between the fully covariant formalism and the spin coefficient formalism. It exhibits, to a certain extent, the advantages (and some disadvantages) of each. The use of indices is avoided; equations frequently have direct geometrical interpretations. Since the space–times considered in practice often have one or two null directions singled out, it is hoped that the modification of the spin-coefficient formalism considered here will serve as a useful computational tool.

2. THE FORMALISM

Let o^A and ι^A be a pair of spinor fields on the space–time normalized according to

$$o_A \iota^A = 1. \tag{2.1}$$

Such a pair of spinor fields is called a *dyad* or *spin frame*. It is well known[4] that any dyad defines a unique null tetrad $(l^a, m^a, \bar{m}^a, n^a)$ at each point and, conversely, that any null tetrad defines a dyad uniquely up to sign. The relationship is as follows[3]:

$$l^a = o^A \bar{o}^{A'}, \qquad m^a = o^A \bar{\iota}^A, \qquad \bar{m}^a = \iota^A \bar{o}^{A'}, \qquad n^a = \iota^A \bar{\iota}^{A'}. \tag{2.2}$$

This is clearly unaffected by the replacement $(o^A, \iota^A) \to (-o^A, -\iota^A)$. [A choice of null tetrad is equivalent to a choice of orthonormal tetrad. Canonically, we can set $T^a = 2^{-1/2}(l^a + n^a)$, $X^a = 2^{-1/2}(m^a + \bar{m}^a)$, $Y^a = -i2^{-1/2}(m^a - \bar{m}^a)$, and $Z^a = 2^{-1/2}(l^a - n^a)$ to define an orthonormal tetrad.] The twelve spin coefficients and their relation to the Ricci rotation coeefficients for the null tetrad are as follows[4]:

$$\begin{aligned}
\kappa &= o^A \bar{o}^{A'} o^B \nabla_{AA'} o_B = m^b l^a \nabla_a l_b, \\
\sigma &= o^A \bar{\iota}^{A'} o^B \nabla_{AA'} o_B = m^b m^a \nabla_a l_b, \\
\rho &= \iota^A \bar{o}^{A'} o^B \nabla_{AA'} o_B = m^b \bar{m}^a \nabla_a l_b, \\
\tau &= \iota^A \bar{\iota}^{A'} o^B \nabla_{AA'} o_B = m^b n^a \nabla_a l_b, \\
\kappa' &= -\iota^A \bar{\iota}^{A'} \iota^B \nabla_{AA'} \iota_B = \bar{m}^b n^a \nabla_a n_b, \\
\sigma' &= -\iota^A \bar{o}^{A'} \iota^B \nabla_{AA'} \iota_B = \bar{m}^b \bar{m}^a \nabla_a n_b \\
\rho' &= -o^A \bar{\iota}^{A'} \iota^B \nabla_{AA'} \iota_B = \bar{m}^b m^a \nabla_a n_b, \\
\tau' &= -o^A \bar{o}^{A'} \iota^B \nabla_{AA'} \iota_B = \bar{m}^b l^a \nabla_a n_b
\end{aligned} \tag{2.3}$$

and

$$\begin{aligned}
\beta &= o^A \bar{\iota}^{A'} \iota^B \nabla_{AA'} o_B = \tfrac{1}{2}(n^b m^a \nabla_a l_b - \bar{m}^b m^a \nabla_a m_b), \\
\beta' &= -\iota^A \bar{o}^{A'} o^B \nabla_{AA'} \iota_B = \tfrac{1}{2}(l^b \bar{m}^a \nabla_a n_b - m^b \bar{m}^a \nabla_a \bar{m}_b), \\
\epsilon &= o^A \bar{o}^{A'} \iota^B \nabla_{AA'} o_B = \tfrac{1}{2}(n^b l^a \nabla_a l_b - \bar{m}^b l^a \nabla_a m_b), \\
\epsilon' &= -\iota^A \bar{\iota}^{A'} o^B \nabla_{AA'} \iota_B = \tfrac{1}{2}(l^b n^a \nabla_a n_b - m^b n^a \nabla_a \bar{m}_b).
\end{aligned} \tag{2.4}$$

The above notation differs from that originally given by Newman and Penrose in that we have used only six different Greek letters rather than twelve. The use of primed letters brings out the close relation between six of the spin coefficients and the remaining six. The other six Greek symbols of the NP formalism are related to ours by

$$\begin{aligned}
\nu &= -\kappa', \quad \lambda = -\sigma', \quad \mu = -\rho', \\
\pi &= -\tau', \quad \alpha = -\beta', \quad \gamma = -\epsilon'.
\end{aligned} \tag{2.5}$$

We shall make use of the prime systematically here to denote the operation of effecting the replacement:

$$o^A \to i\iota^A, \quad \iota^A \to i o^A, \quad \bar{o}^{A'} \to -i\bar{\iota}^{A'}, \quad \bar{\iota}^{A'} \to -i\bar{o}^{A'}, \tag{2.6}$$

so that

$$(l^a)' = n^a, \qquad (m^a)' = \bar{m}^a, \qquad (\bar{m}^a)' = m^a, \qquad (n^a)' = l^a. \tag{2.7}$$

This preserves the normalization (2.1) and the relationship between a quantity and its complex conjugate. Since the bar and prime operations commute, one can write $\bar{\eta}'$ without ambiguity. Furthermore the prime operation is involutory up to sign:

$$(\eta')' = (-1)^{p+q}\eta. \tag{2.7a}$$

(For all quantities explicitly defined in this paper, $p + q$ is in fact even, so this sign will play no role here.) This use of the prime not only halves the number of Greek letters needed, but also effectively halves the number of equations.[6]

The role of the (vector) covariant derivative operator ∇_a is taken over in the NP formalism by four scalar operators:

$$\begin{aligned}
D &= o^A \bar{o}^{A'} \nabla_{AA'} = l^a \nabla_a, \qquad D' = \iota^A \bar{\iota}^{A'} \nabla_{AA'} = n^a \nabla_a, \\
\delta &= o^A \bar{\iota}^{A'} \nabla_{AA'} = m^a \nabla_a, \qquad \bar{\delta} = \delta' = \iota^A \bar{o}^{A'} \nabla_{AA'} = \bar{m}^a \nabla_a.
\end{aligned} \tag{2.8}$$

Tensor or spinor equations on the space–time can now be written out explicitly in terms of the components with

respect to the null tetrad or the dyad, and the spin coefficients (2.3), (2,4), and the derivative operators (2.8).

So far this is just standard NP formalism. The innovation that we shall introduce here, however, is to work entirely with spin- and boost-weighted quantities, since these are the quantities which are apropriately "covariant" under the group of tetrad transformations which leave the two null directions of l^a and n^a invariant. Now the most general spin transformation which leaves these two null directions invariant [preserving the normalization (2.1)] is

$$o^A \to \lambda o^A, \qquad \iota^A \to \lambda^{-1}\iota^A, \tag{2.9}$$

where λ is an arbitrary (nowehere vanishing) complex scalar field. In terms of the null tetrad, the transformation (2.9) takes the form

$$l^a \to \lambda\bar{\lambda} l^a, \qquad m^a \to \lambda\bar{\lambda}^{-1} m^a,$$
$$\bar{m}^a \to \lambda^{-1}\bar{\lambda}\bar{m}^a, \qquad n^a \to \lambda^{-1}\bar{\lambda}^{-1} n^a. \tag{2.10}$$

These are just the transformations (1.5) and (1.6) considered earlier.

Recall that a scalar η of type $\{p, q\}$ was to be a scalar quantity which transformed according to

$$\eta \to \lambda^p \bar{\lambda}^q \eta, \tag{2.11}$$

whenever the dyad (o^A, ι^A) transformed according to (2.9) or, equivalently, whenever the null tetrad $(l^a, m^a, \bar{m}^a, n^a)$ transformed according to (2.10). Strictly speaking, we should think of η as a *function* which assigns a complex scalar field $\eta(o^A, \iota^A)$ to each ordered normalized pair of spinor fields o^A, ι^A, for which the null directions defined by o^A and by ι^A are the given pair of null directions. Alternatively, we can think of η as a function of the null tetrad. In any case, to be a proper spin- and boost-weighted quantity, the function η must be of a very special type, namely one which satisfies (2.11) when the dyad or null tetrad is changed according to (2.9) or (2.10), i.e. $\eta(\lambda o^A, \lambda^{-1}\iota^A) = \lambda^p\bar{\lambda}^q\eta(o^A, \iota^A)$. (Note that we may regard o^A and ι^A, themselves as *spinors* of type $\{1, 0\}$ and $\{-1, 0\}$, respectively, and $l^a, m^a, \bar{m}^a, n^a$ as *vectors* of type $\{1, 1\}, \{1, -1\}, \{-1, 1\}$ and $\{-1, -1\}$, respectively.)

The spin coefficients may now be divided into two classes according to whether or not they are proper spin- and boost-weighted quantities. In fact, the spin coefficients in the list (2.3) are all such quantities whereas those in the list (2.4) are not. Let us illustrate this with two examples:

$$\sigma \to (\lambda o^A)(\bar{\lambda}^{-1}\bar{\iota}^{A'})(\lambda o^B)\nabla_{AA'}(\lambda o_B) = \lambda^3\bar{\lambda}^{-1}\sigma, \tag{2.12a}$$

but

$$\beta \to (\lambda o^A)(\bar{\lambda}^{-1}\bar{\iota}^{A'})(\lambda^{-1}\iota^B)\nabla_{AA'}(\lambda o_B)$$
$$= \lambda\bar{\lambda}^{-1}\beta + \bar{\lambda}^{-1} o^A \bar{\iota}^{A'}\nabla_{AA'}\lambda. \tag{2.12b}$$

The types of the spin- and boost-weighted quantities (2.3) are as follows:

$$\kappa:\{3, 1\}, \sigma:\{3, -1\}, \rho:\{1, 1\}, \tau:\{1, -1\},$$
$$\kappa':\{-3, -1\}, \sigma':\{-3, 1\}, \rho':\{-1, -1\}, \tau':\{-1, 1\}. \tag{2.13}$$

With any spinor field or tensor field on the space–time there is associated a collection of spin- and boost-weighted scalars of various types $\{p, q\}$ which together define the spinor or tensor. These are obtained from the spinor by transvecting with the various combinations of o^A, ι^A, $\bar{o}^{A'}$, and $\bar{\iota}^{A'}$, or the tensor with the various combinations of $l^a, m^a, \bar{m}^a$, and n^a. Any tensor field, may, in fact, be interpreted as a spinor field if desired; but we get precisely the same set of scalars whichever way we do it because of the definition (2.2) of the null tetrad in terms of the dyad.

Evidently the product of a scalar of type $\{p, q\}$ with a scalar of type $\{u, v\}$ is a scalar of type $\{p + u, q + v\}$. On the other hand, sums are allowed only when the summands have the same type. The type of the sum is the same as that of each summand.

We next wish to introduce derivative operators into the formalism. Unfortunately the operators (2.8) of the NP formalism are not suitable for this purpose, for, when applied to a scalar of type $\{p, q\}$, with p, q not both zero, they do not in general produce a spin- and boost-weighted scalar. We, therefore, modify the derivative operators (2.8) by the inclusion of further terms involving the spin coefficients in the list (2.4). For η of type $\{p, q\}$ we define[7,8]

$$\text{Þ}\eta = (D - p\epsilon - q\bar{\epsilon})\eta, \qquad \text{Þ}'\eta = (D' + p\epsilon' + q\bar{\epsilon}')\eta,$$
$$\eth\eta = (\delta - p\beta + q\bar{\beta}')\eta, \qquad \eth'\eta = (\delta' + p\beta' - q\bar{\beta})\eta. \tag{2.14}$$

These combination have been so chosen that, under (2.9), the terms involving derivatives of λ cancel exactly. It should be noted that these operators are *derivations*, that is they are linear, and when operating on products, they satisfy the Leibniz rule. The operator $\eth$ as defined here reduces, in the appropriate circumstances, to that of Newman and Penrose.[9,10] The spin weights of the operators (2.14) are as follows:

$$\text{Þ}:\{1, 1\}, \qquad \eth:\{1, -1\},$$
$$\text{Þ}':\{-1, -1\}, \qquad \eth':\{-1, 1\}. \tag{2.15}$$

(To say that a differential operator has type $\{p, q\}$ is to say that when acting on a scalar of type $\{u, v\}$ it produces a scalar of type $\{p + u, q + v\}$.)

Alternatively the operators may be defined in terms of the type $\{0, 0\}$ operator (acting on a quantity of type $\{p, q\} = \{r + s, r - s\}$)

$$\Theta_{AA'} = \nabla_{AA'} - p\iota^B\nabla_{AA'}o_B - q\bar{\iota}^{B'}\nabla_{AA'}\bar{o}_{B'}$$
$$= \nabla_a - rn^b\nabla_a l_b + s\bar{m}^b\nabla_a m_b \tag{2.14a}$$

by the equation

$$\Theta_a = l_a\text{Þ}' + n_a\text{Þ} - m_a\eth' - \bar{m}_a\eth. \tag{2.14b}$$

In Eq. (2.14a), s and r are the spin and boost weights, respectively [cf. following 1.7)]. The original definitions (2.14) may be recovered by transvecting (2.14b) with l^a, n^a, m^a and $\bar{m}^a$.

Since we wish to operate only with quantities with a well-defined $\{p, q\}$, we are not able to use the spin coefficients in the list (2.4) directly. Instead, the role of these spin coefficients is to be found within the definitions of the operators (2.14).

The basic quantities with which we shall work are the eight spin coefficients $\kappa, \sigma, \rho, \tau; \kappa', \sigma', \rho', \tau'$ and the four differential operators Þ, $\eth$, Þ', $\eth'$; In addition, there is the operation of complex conjugation. We may also con-

sider the prime as effectively an allowable operation on the system. With the introduction of various tensor or spinor fields, such as the electromagnetic field tensor or Riemann tensor, we shall be able to combine the above elements with the tetrad components of the tensor fields and the dyad components of the spinor fields to obtain a self-contained calculus.

The effect of the derivative operators (2.14) is shown in Fig. 1. We associate with a scalar η of type $\{p, q\}$, the point with coordinates $\{p, q\}$ in the plane. Each of the derivative operators (2.14) has a characteristic effect on the type, which can be represented as a displacement in this diagram. Note that when two elements are multiplied together this corresponds to a vector sum in the diagram. If two elements are to be added together, then they must be represented by the same point in the diagram. The operation of complex conjugation is represented by a reflection in the line $p = q$, since the complex conjugate of an element of type $\{p, q\}$ is an element of type[18] $\{\bar{q}, \bar{p}\}$. We define, in fact,

$$\bar{\text{Þ}} = \text{Þ}, \quad \bar{\text{Þ}}' = \text{Þ}', \quad \bar{\eth} = \eth', \quad \bar{\eth}' = \eth. \tag{2.16}$$

Then the operation of complex conjugation will satisfy

$$\overline{\text{Þ}\eta} = \bar{\text{Þ}}\bar{\eta}, \quad \overline{\eth\eta} = \bar{\eth}\bar{\eta}. \tag{2.17}$$

Finally, the prime operation is represented in the diagram by a reflection in the origin, since if we prime an element of type $\{p, q\}$ we get an element of type $\{-p, -q\}$. The prime will commute with addition, multiplication, and complex conjugation [but note (2.7a)]. Furthermore, we have

$$(\text{Þ}\eta)' = \text{Þ}'\eta', \quad (\text{Þ}'\eta)' = \text{Þ}\eta', \quad (\eth\eta)' = \eth'\eta', \quad (\eth'\eta)' = \eth\eta'. \tag{2.18}$$

The various spin-coefficient formulae will now be given explicitly in the form allowable within our present formalism. As an example, consider Eq. (4.2k) of Ref. 4:

$$\delta\rho - \delta'\sigma = \rho(\beta - \bar{\beta}') + \sigma(\bar{\beta} + 3\beta') + \tau(\rho - \bar{\rho}) + \kappa(\bar{\rho}' - \rho') - \Psi_1 + \Phi_{01}. \tag{2.19}$$

Rewriting (2.19) as

$$(\delta - \beta + \bar{\beta}')\rho - (\delta' + 3\beta' + \bar{\beta})\sigma = (\rho - \bar{\rho})\tau + (\bar{\rho}' - \rho')\kappa - \Psi_1 + \Phi_{01} \tag{2.20}$$

and noting that ρ and σ are types $\{1, 1\}$ and $\{3, -1\}$, respectively, we see that the equation may be reexpressed, using (2.14) as

$$\eth\rho - \eth'\sigma = (\rho - \bar{\rho})\tau + (\bar{\rho}' - \rho')\kappa - \Psi_1 + \Phi_{01}. \tag{2.21}$$

Similarly, Eqs. (4.2a), (4.2b), (4.2c), (4.2p). and (4.2q) of Ref. 4 may be rewritten as

$$\text{Þ}\rho - \eth'\kappa = \rho^2 + \sigma\bar{\sigma} - \bar{\kappa}\tau - \tau'\kappa + \Phi_{00}, \tag{2.22}$$

$$\text{Þ}\sigma - \eth\kappa = \sigma(\rho + \bar{\rho}) - \kappa(\tau + \bar{\tau}') + \Psi_0, \tag{2.23}$$

$$\text{Þ}\tau - \text{Þ}'\kappa = \rho(\tau - \bar{\tau}') + \sigma(\bar{\tau} - \tau') + \Psi_1 + \Phi_{01}, \tag{2.24}$$

$$\eth\tau - \text{Þ}'\sigma = -\rho'\sigma - \bar{\sigma}'\rho + \tau^2 + \kappa\bar{\kappa}' + \Phi_{02}, \tag{2.25}$$

and

$$\text{Þ}'\rho - \eth'\tau = \rho\bar{\rho}' + \sigma\sigma' - \tau\bar{\tau} - \kappa\kappa' - \Psi_2 - 2\Lambda. \tag{2.26}$$

Applying the prime operation to each of these six equations, we obtain six more equations, these being equivalent to (4.2m), (4.2n), (4.2j), (4.2i), (4.2g), and (4.2h) of Ref. 4, respectively. We recall the definitions of the dyad components of the Weyl spinor Ψ_{ABCD} and the trace-free Ricci spinor

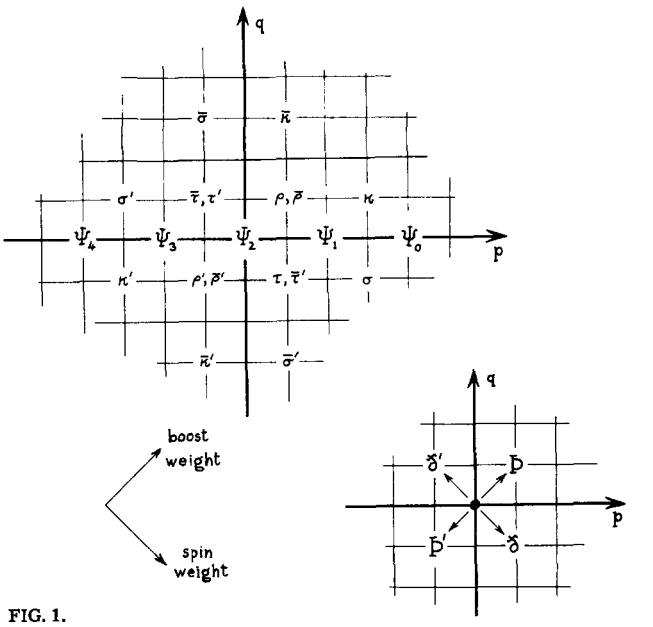

FIG. 1.

$\Phi_{ABC'D'}$:

$$\begin{aligned}
\Psi_0 &= o^A o^B o^C o^D \Psi_{ABCD} = \Psi_4' && \{4, 0\},\\
\Psi_1 &= o^A o^B o^C \iota^D \Psi_{ABCD} = \Psi_3' && \{2, 0\},\\
\Psi_2 &= o^A o^B \iota^C \iota^D \Psi_{ABCD} = \Psi_2' && \{0, 0\},\\
\Psi_3 &= o^A \iota^B \iota^C \iota^D \Psi_{ABCD} = \Psi_1' && \{-2, 0\},\\
\Psi_4 &= \iota^A \iota^B \iota^C \iota^D \Psi_{ABCD} = \Psi_0' && \{-4, 0\};
\end{aligned} \tag{2.27}$$

$$\begin{aligned}
\Phi_{00} &= o^A o^B \bar{o}^{A'} \bar{o}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{00} = \Phi_{22}' && \{2, 2\},\\
\Phi_{01} &= o^A o^B \bar{o}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{10} = \Phi'_{21} && \{2, 0\},\\
\Phi_{02} &= o^A o^B \bar{\iota}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{20} = \Phi'_{20} && \{2, -2\},\\
\Phi_{10} &= o^A \iota^B \bar{o}^{A'} \bar{o}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{01} = \Phi'_{12} && \{0, 2\},\\
\Phi_{11} &= o^A \iota^B \bar{o}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{11} = \Phi'_{11} && \{0, 0\},\\
\Phi_{12} &= o^A \iota^B \bar{\iota}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{21} = \Phi'_{10} && \{0, -2\},\\
\Phi_{20} &= \iota^A \iota^B \bar{o}^{A'} \bar{o}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{02} = \Phi_{02}' && \{-2, +2\},\\
\Phi_{21} &= \iota^A \iota^B \bar{o}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{12} = \Phi_{01}' && \{-2, 0\},\\
\Phi_{22} &= \iota^A \iota^B \bar{\iota}^{A'} \bar{\iota}^{B'} \Phi_{ABA'B'} = \bar{\Phi}_{22} = \Phi_{00}' && \{-2, -2\}.
\end{aligned} \tag{2.28}$$

In each case, the type is given on the right. Furthermore, the scalar curvature is defined by

$$\Lambda = \bar{\Lambda} = \Lambda' = \tfrac{1}{24}R \quad \{0, 0\}. \tag{2.29}$$

The list (2.21)–(2.26), together with the corresponding list of primed equations does not completely exhaust the NP equations (4.2) of Ref. 4. The remaining equations (4.2d), (4.2e), (4.2f), (4.2l), (4.2o), and (4.2r) in Ref. 4 refer to derivatives of spin coefficients which are not spin- and boost-weighted quantities. They cannot, therefore, be written explicitly in our present formalism as equations like (2.21). Instead, they play their role as part of the commutator equations for the differential operators Þ, Þ′, $\eth$ and $\eth'$. These commutators, when applied to a spin- and boost-weighted scalar η of type $\{p, q\}$, are

$$(\text{Þ}\text{Þ}' - \text{Þ}'\text{Þ})\eta = [(\bar{\tau} - \tau')\eth + (\tau - \bar{\tau}')\eth' - p(\kappa\kappa' - \tau\tau' + \Psi_2 + \Phi_{11} - \Lambda) - q(\bar{\kappa}\bar{\kappa}' - \bar{\tau}\bar{\tau}' + \bar{\Psi}_2 + \Phi_{11} - \Lambda)]\eta, \quad (2.30)$$

$$(\text{Þ}\eth - \eth\text{Þ})\eta = [\bar{\rho}\eth + \sigma\eth' - \bar{\tau}'\text{Þ} - \kappa\text{Þ}' - p(\rho'\kappa - \tau'\sigma + \Psi_1) - q(\bar{\sigma}'\bar{\kappa} - \bar{\rho}\bar{\tau}' + \Phi_{01})]\eta, \quad (2.31)$$

$$(\eth\eth' - \eth'\eth)\eta = [(\bar{\rho}' - \rho')\text{Þ} + (\rho - \bar{\rho})\text{Þ}' + p(\rho\rho' - \sigma\sigma' + \Psi_2 - \Phi_{11} - \Lambda) - q(\bar{\rho}\bar{\rho}' - \bar{\sigma}\bar{\sigma}' + \Psi_2 - \Phi_{11} - \Lambda)]\eta, \quad (2.32)$$

together with the remaining commutator equations obtained by applying prime, complex conjugation, and both to (2.31). Note that the spin- and boost-weights of η enter explicitly on the right-hand side. We must be careful, when applying primes and bars to these equations, to remember that η', $\bar{\eta}$, and $\bar{\eta}'$ have types which are not quite those of η. Thus, under the prime, p becomes $-p$ and q becomes $-q$; under the bar, p becomes q and q becomes p; under both bar and prime p becomes $-q$ and q becomes $-p$.

The commutator equations are the one place where our present formalism yields more complicated formulas than the original NP formalism. This seems to be the price we pay for the very considerable formal simplification that we obtain for the other equations. But we must bear in mind that our commutators are actually combining information which comes from two different places in the NP formalism. There is, however, a gain as regards geometric content of the commutators with our present formulation. The extra terms which arise when p or q is nonzero may sometimes be interpreted as curvature quantities referring to submanifolds in the space-time. We shall see this explicitly for Eq. (2.32) in the next section.

The full Bianchi identities consist of[11]

$$\text{Þ}\Psi_1 - \eth'\Psi_0 - \text{Þ}\Phi_{01} + \eth\Phi_{00} = -\tau'\Psi_0 + 4\rho\Psi_1 - 3\kappa\Psi_2 + \bar{\tau}'\Phi_{00} - 2\bar{\rho}\Phi_{01} - 2\sigma\Phi_{10} + 2\kappa\Phi_{11} + \bar{\kappa}\Phi_{02}, \quad (2.33)$$

$$\text{Þ}\Psi_2 - \eth'\Psi_1 - \eth'\Phi_{01} + \text{Þ}'\Phi_{00} + 2\text{Þ}\Lambda = \sigma'\Psi_0 - 2\tau'\Psi_1 + 3\rho\Psi_2 - 2\kappa\Psi_3 + \bar{\rho}'\Phi_{00} - 2\bar{\tau}\Phi_{01} - 2\tau\Phi_{10} + 2\rho\Phi_{11} + \bar{\sigma}\Phi_{02}, \quad (2.34)$$

$$\text{Þ}\Psi_3 - \eth'\Psi_2 - \text{Þ}\Phi_{21} + \eth\Phi_{20} - 2\eth'\Lambda = 2\sigma'\Psi_1 - 3\tau'\Psi_2 + 2\rho\Psi_3 - \kappa\Psi_4 - 2\rho'\Phi_{10} + 2\tau'\Phi_{11} + \bar{\tau}'\Phi_{20} - 2\bar{\rho}\Phi_{21} + \bar{\kappa}\Phi_{22}, \quad (2.35)$$

$$\text{Þ}\Psi_4 - \eth'\Psi_3 - \eth'\Phi_{21} + \text{Þ}'\Phi_{20} = +3\sigma'\Psi_2 - 4\tau'\Psi_3 + \rho\Psi_4 - 2\kappa'\Phi_{10} + 2\sigma'\Phi_{11} + \bar{\rho}'\Phi_{20} - 2\bar{\tau}\Phi_{21} + \bar{\sigma}\Phi_{22} \quad (2.36)$$

together with their primed versions, the contracted Bianchi identities

$$\text{Þ}\Phi_{11} + \text{Þ}'\Phi_{00} - \eth\Phi_{10} - \eth'\Phi_{01} + 3\text{Þ}\Lambda = (\rho' + \bar{\rho}')\Phi_{00} + 2(\rho + \bar{\rho})\Phi_{11} - (\tau' + 2\bar{\tau})\Phi_{01} - (2\tau + \bar{\tau}')\Phi_{10} - \bar{\kappa}\Phi_{12} - \kappa\Phi_{21} + \sigma\Phi_{20} + \bar{\sigma}\Phi_{02} \quad (2.37)$$

$$\text{Þ}\Phi_{12} + \text{Þ}'\Phi_{01} - \eth\Phi_{11} - \eth'\Phi_{02} + 3\eth\Lambda = (\rho' + 2\bar{\rho}')\Phi_{01} + (2\rho + \bar{\rho})\Phi_{12} - (\tau' + \bar{\tau})\Phi_{02} - 2(\tau + \bar{\tau}')\Phi_{11} - \bar{\kappa}'\Phi_{00} - \kappa\Phi_{22} + \sigma\Phi_{21} + \bar{\sigma}'\Phi_{10} \quad (2.38)$$

and the primed versions of these equations. [More generally, equations similar to (2.37) and (2.38) express the equations of an arbitrary conserved symmetric two index tensor.]

The content of Einstein's vacuum equations is obtained by putting all the Φ's and Λ's equal to zero in (2.21)–(2.26) and in (2.30)–(2.32). The Bianchi identities (2.33)–(2.36) (with $\Phi = \Lambda = 0$) become considerably simpler in this case and closely resemble Maxwell's source-free equations. In the present formalism, Maxwell's source-free equations are

$$\text{Þ}\phi_1 - \eth'\phi_0 = -\tau'\phi_0 + 2\rho\phi_1 - \kappa\phi_2, \quad (2.39)$$

$$\text{Þ}\phi_2 - \eth'\phi_1 = \sigma'\phi_0 - 2\tau'\phi_1 + \rho\phi_2, \quad (2.40)$$

together with their primed versions. Here we have

$$\begin{aligned} \phi_0 &= o^A o^B \phi_{AB} = -\phi_2' \quad \{2,0\}, \\ \phi_1 &= o^A \iota^B \phi_{AB} = -\phi_1' \quad \{0,0\}, \\ \phi_2 &= \iota^A \iota^B \phi_{AB} = -\phi_0' \quad \{-2,0\}, \end{aligned} \quad (2.41)$$

where the symmetric spinor ϕ_{AB} is related to Maxwell's field tensor by

$$F_{ab} = \phi_{AB}\epsilon_{A'B'} + \epsilon_{AB}\bar{\phi}_{A'B'},$$

Finally, we remark on the existence of an additional symmetry possessed by the spin-coefficient formalism which was noticed some time ago by Sachs.[12] We denote the Sachs symmetry by an asterisk (*), defining

$$(o^A)^* = o^A, \quad (\iota^A)^* = \iota^A, \quad (\bar{o}^{A'})^* = \bar{\iota}^{A'}, \quad (\bar{\iota}^{A'})^* = -\bar{o}^{A'}; \quad (2.42)$$

so that

$$(l^a)^* = m^a, \quad (m^a)^* = -l^a, \quad (\bar{m}^a)^* = n^a, \quad (n^a)^* = -\bar{m}^a. \quad (2.43)$$

This preserves the required normalizations and orthogonality relations). Clearly the Sachs symmetry operation does not commute with complex conjugation, since the relation between an object and its complex conjugate is destroyed. However, we do have

$$(n^*)^* = (-1)^q\eta, \quad (\eta')^* = (-1)^q(\eta^*)', \quad \text{and } \overline{\eta^*} = (-i)^{p+q}(\bar{\eta}')^*, \quad (2.44)$$

where η is the quantity of type $\{p, q\}$. Note for such an η, the quantity η^* is of type $\{p, -q\}$.

From (2.3) we get

$$\kappa^* = \sigma, \quad \sigma^* = -\kappa, \quad \rho^* = \tau, \quad \tau^* = -\rho, \quad \kappa'^* = -\sigma', \quad \sigma'^* = \kappa', \quad \rho'^* = -\tau', \tau'^*, = \rho',$$
$$\bar{\kappa}^* = -\bar{\sigma}', \quad \bar{\sigma}^* = -\bar{\kappa}', \quad \bar{\rho}^* = \tau', \quad \bar{\tau}^* = \bar{\rho}', \quad \bar{\kappa}'^* = \bar{\sigma}, \quad \bar{\sigma}'^* = \bar{\kappa}, \quad \bar{\rho}'^* = -\bar{\tau}, \quad \bar{\tau}'^* = -\bar{\rho}, \quad (2.45)$$

$$\text{Þ}^* = \eth, \quad \eth^* = -\text{Þ}, \quad \text{Þ}'^* = -\eth', \quad \eth'^* = \text{Þ}', \quad (2.46)$$

$$\Psi_0{}^* = \Psi_0, \quad \Psi_1{}^* = \Psi_1, \quad \Psi_2{}^* = \Psi_2,$$

$$\Psi_3^* = \Psi_3, \quad \Psi_4^* = \Psi_4,$$
$$\bar{\Psi}_0^* = \bar{\Psi}_4, \quad \bar{\Psi}_1^* = -\bar{\Psi}_3, \quad \bar{\Psi}_2^* = \bar{\Psi}_2,$$
$$\bar{\Psi}_3^* = -\bar{\Psi}_1, \quad \bar{\Psi}_4^* = \bar{\Psi}_0, \tag{2.47}$$

$$\Phi_{00}^* = \Phi_{02}, \quad \Phi_{01}^* = -\Phi_{01}, \quad \Phi_{02}^* = \Phi_{00}, \Phi_{10}^* = \Phi_{12},$$
$$\Phi_{11}^* = -\Phi_{11}, \quad \Phi_{12}^* = \Phi_{10} \quad \Phi_{20}^* = \Phi_{22},$$
$$\Phi_{21}^* = -\Phi_{21}, \quad \Phi_{22}^* = \Phi_{20}, \quad \Lambda^* = \Lambda. \tag{2.48}$$

Under the Sachs symmetry operation, the equations in our list (2.21)–(2.26) are permuted among themselves and so are those of the lists (2.33)–(2.36) or (2.37), (2.38); so also the commutators [cf. (2.30)-(2.32)]. The Sachs symmetry, together with the prime operation, can be used to simplify the generation of equations; alternatively it provides a useful check of equations obtained by other means.

3. APPLICATION TO 2-SURFACES

In order to establish the connection between our present formalism and existing work which employs an ð operator, we must show that our ð operator does in fact reduce to the one defined previously,[9,10] under the appropriate circumstances. Let S be a spacelike or timelike 2-surface in the space–time M. Then at each point of S there are two preferred null directions, defined by the property that they are orthogonal to S if S is spacelike and tangent to S if S is timelike. We shall discuss only the spacelike case below. The situation for a timelike S is analogous. (Essentially, to pass to the timelike case from the spacelike case we must let l^a and n^a take over the roles of m^a and $\bar{m}^a$; hence, Þ and Þ′ will take over the roles of ð and ð′ below.)

Let us suppose, then, that S is spacelike and that a null tetrad and dyad system has been set up in M (in the neighborhood of S at least), so that l^a and n^a are perpendicular to S at each point of S. (For this to be possible *globally* over S would require some topological restrictions on S and M—but we shall ignore such matters here.) The freedom in choosing such a tetrad system, lies partly in the fact that it is only the two null *directions* which are uniquely singled out at each point of S, not a complete null tetrad, and partly in the fact that the choice of null tetrad in the remainder of M is (apart from smoothness considerations) quite arbitrary. This latter freedom is of no concern for us here since we shall be interested only in quantities defined at points of S and in differentiations which act within S itself (i.e., in directions tangent to S). The remiaining freedom, is, of course, precisely that which the formalism of this paper is designed to handle.

Consider any tensor field defined on M. At each point of S we shall have a uniquely defined tensor within S, which is obtained by projecting the original tensor orthogonally into the surface. Conversely, any tensor defined within the 2-surface S can be obtained by such a projection. In particular, some tensor fields defined on M will have the property that they are completely tangent to S at each point of S. Such fields we regard as being *unaffected*, at points of S, by the above orthogonal projection. We may interpret any tensor field defined *within* S as such a tensor field in M modulo its behavior off S.

To be more explicit about the nature of this orthogonal projection, define

$$E^b_a = -m_a\bar{m}^b - \bar{m}_a m^b \quad \text{and} \quad F^b_a = l_a n^b + n_a l^b. \tag{3.1}$$

Then we have

$$E^b_a E^c_b = E^c_a, \quad F^b_a F^c_b = F^c_a, \quad F^b_a E^c_b = 0 = E^b_a F^c_b,$$
$$\text{and} \quad E^b_a + F^b_a = \delta^b_a. \tag{3.2}$$

Furthermore, if

$$\xi^a = \eta^a + \zeta^a, \tag{3.3}$$

where η^a is the part of ξ^a linearly dependent on m^a and $\bar{m}^a$ and where ζ^a is the part of ξ^a linearly dependent on l^a and n^a, then we have

$$\eta^a = \xi^b E^a_b, \quad \zeta^a = \xi^b F^a_b, \quad \eta_a = \xi_b E^b_a, \quad \zeta_a = \xi_b F^b_a. \tag{3.4}$$

Thus, E^b_a is the orthogonal projector which can be used to project a vector or tensor at a point of S *into* S, while F^b_a can be used to pick out the parts of a tensor which are *perpendicular* to S. For example, if the tensor $T_{ab}{}^c$ on M is defined at a point of S; then the tensor $U_{ab}{}^c = T_{pq}{}^r E^p_a E^q_b E^c_r$, at this point, is the orthogonal projection of $T_{ab}{}^c$ into S. If $U_{ab}{}^c = T_{ab}{}^c$, then $T_{ab}{}^c$ is completely tangent to S at this point. In this case, the projections of $T_{ab}{}^c$ which involve components perpendicular to S, such as $V_{ab}{}^c = T_{pq}{}^r F^p_a E^q_b E^c_r, \ldots, Z_{ab}{}^c = T_{pq}{}^r F^p_a F^q_b F^c_r$, must all vanish. Thus we may interpret $T_{ab}{}^c (= U_{ab}{}^c)$ at the point as a 2-tensor defined within the 2-surface S. For a general $T_{ab}{}^c$ which is not completely tangent to S, there will be some nonvanishing projections among $V_{ab}{}^c, \ldots, Z_{ab}{}^c$. The information contained in $T_{ab}{}^c$ will be shared among all of $U_{ab}{}^c, V_{ab}{}^c, \ldots, Z_{ab}{}^c$; but only $U_{ab}{}^c$ is interpretable as a 2-tensor defined within S.

If we restrict attention for the moment only to tensors, such as $U_{ab}{}^c$, which are completely tangent to S, then since we may regard them as 2-tensors within S (at points of S) we may ask that the definition of *covariant derivative* within S be defined in terms of the four-dimensional covariant derivative operator ∇_a. To do this, we merely apply the four-dimensional operator ∇_a to any tensor which is completely tangent to S and then project the resulting tensor back into S. The resulting tensor, when interpreted as a 2-tensor defined within S is precisely the (two-dimensional) covariant derivative of the original tensor interpreted as a 2-tensor.[13]

For example, in the case of $U_{ab}{}^c$ above, the two-dimensional covariant derivative of $U_{ab}{}^c$ may be interpreted as

$$E^p_a E^q_b E^c_r (E^s_d \nabla_s) U_{pq}{}^r. \tag{3.5}$$

We can also define a "covariant derivative" operation (analogous to Fermi or Fermi–Walker transport,[14] but where the 2-surface S replaces the curve) for quantities defined at points of S which need not be completely tangent to S. For example, in the case of $V_{ab}{}^c, \ldots, Z_{ab}{}^c$ defined above, the respective "covariant derivatives" would be

$$F^p_a E^q_b E^c_r (E^s_d \nabla_s) V_{pq}{}^r, \ldots, F^p_a F^q_b F^c_r (E^s_d \nabla_s) Z_{pq}{}^r. \tag{3.6}$$

Let us now return to consideration of the set of spin- and boost–weighted quantities associated with a tensor. Take the tensor $T_{ab}{}^c$, for example. We can construct a total of 64 spin- and boost-weighted scalars from $T_{ab}{}^c$, namely those quantities obtained from $T_{ab}{}^c$ by transvection with the various vectors of the null tetrad. If we transvect only with the vectors m^a and $\bar{m}^a$ then we obtain eight quantities referring to the part of $T_{ab}{}^c$ which is completely tangent to S. For example, $m^a\bar{m}^b m_c T_{ab}{}^c = m^a\bar{m}^b m_c U_{ab}{}^c$. These quantities may be thought

of as arising from a 2-tensor defined within S. Since m^a and $\bar{m}^a$ are of respective types $\{1,-1\}$ and $\{-1,1\}$, it follows that the type of any such quantity will be of the form $\{s,-s\}$. Thus, the boost weight of such a quantity vanishes and the spin-weight s alone serves to characterize the type.

We have remarked above that the two-dimensional covariant derivative of $U_{ab}{}^c$ may be obtained by projecting the four-dimensional derivative of $U_{ab}{}^c$ back into the surface S. We may achieve this projection instead by contracting with m^a and $\bar{m}^a$. For example,

$$m^a\bar{m}^b m_c(m^d\nabla_d)U_{ab}{}^c = m^p m^q\bar{m}^r m_s E^a_q E^b_r E^s_c(E^d_p\nabla_d)U_{ab}{}^c. \tag{3.7}$$

Expression (3.7) can be rewritten in terms of the $\eth$ operator as follows:

$$m^d m^a\bar{m}^b m_c\nabla_d U_{ab}{}^c = \eth(m^a\bar{m}^b m_c U_{ab}{}^c). \tag{3.8}$$

Perhaps the easiest way to see this is to write the left-hand side as $m^a\bar{m}^b m_c\eth U_{ab}{}^c$—where we allow $\eth$ to act on tensorial quantities (here $U_{ab}{}^c$ has type $\{0,0\}$)—and observe that $E^b_a\eth m^a = 0$ (m^a having type $\{1,-1\}$) and $E^b_a\eth\bar{m}^a = 0$ ($\bar{m}^a$ having type $\{-1,1\}$). In fact, we have quite generally (allowing $\eth$ and Þ to act on spinor quantities)[15]

$$\begin{aligned} &\text{Þ}o^A = -\kappa\iota^A, \quad \eth o^A = -\sigma\iota^A, \quad \eth' o^A = -\rho\iota^A, \\ &\qquad\qquad \text{Þ}' o^A = -\tau\iota^A; \\ &\text{Þ}\iota^A = -\tau' o^A, \quad \eth\iota^A = -\rho' o^A \\ &\qquad\qquad \eth'\iota^A = -\sigma' o^A, \quad \text{Þ}'\iota^A = -\kappa' o^A. \end{aligned} \tag{3.9}$$

Thus we obtain particularly,

$$\eth m^a = -\bar{\sigma}'l^a - \sigma n^a, \quad \eth\bar{m}^a = -\rho' l^a - \bar{\rho} n^a. \tag{3.10}$$

Expression (3.8) shows that $\eth$ is, in effect, just a two-dimensional covariant differentiation operator acting within the surface S. In fact (3.8) is a simple extension, apart from the absence of their factor $\sqrt{2}$ on the right-hand side, of an expression given by Goldberg *et al.*[10] [cf. their Eq. (2.13)] for the operator $\eth$ as had been originally defined by Newman and Penrose.[9] We observe that any spin-weighted quantity defined on S (of integral spin weight) can be expressed in the form $m^a\cdots m^c\bar{m}^d\cdots\bar{m}^f T_{a\ldots cd\ldots f}$, so (3.8) fully characterizes $\eth$ on S. To get complete agreement with the Newman–Penrose (θ,ϕ) expressions for $\eth$, we should require S to be intrinsically a metric sphere of radius $2^{-1/2}$;

$$-ds^2 = \tfrac{1}{2}(d\theta^2 + \sin^2\theta d\phi^2) \tag{3.11}$$

on S. But for complete agreement in the case of a *general* 2-surface metric we would have (strictly speaking) to multiply up our $\eth$ operator by a factor $\sqrt{2}$. (It is perhaps unfortunate to have to introduce a discrepancy of this kind into our definition of $\eth$. However, the formulas of this paper would look unnecessarily cumbersome were we to retain complete *numerical* agreement with the original Newman–Penrose definition.)

Note that, as applied to spin-weighted scalars defined on the surface S, the operator $\eth$ is really completely *intrinsic* to S; that is to say, its effect depends only on the intrinsic metric of S and not on the way that S is embedded in the space-time M. When applied to boost-weighted scalars, the effect of $\eth$ *does* depend on the embedding. We can illustrate both these facts if we examine the geometrical meaning of the commutator (2.32). In our present situation the quantities ρ and ρ' are both necessarily real at points of S. This is because l^a and n^a are the null tangents (i.e., normal) to the two null hypersurfaces which intersect in S, so the curl of both l^a and n^a must vanish. (This is just the condition that the plane-elements spanned by the real and imaginary parts of m^a are surface-forming.) Thus, the commutator (2.32) becomes

$$(\eth\eth' - \eth'\eth)\eta = -(pK - q\bar{K})\eta, \tag{3.12}$$

where

$$K = \sigma\sigma' - \rho\rho' - \Psi_2 + \Lambda + \Phi_{11}. \tag{3.13}$$

The complex scalar K (type $\{0,0\}$) is a kind of *complex* curvature for the surface S and has been studied earlier in connection with the characteristic initial value problem.[16] The Gaussian curvature of the surface S is twice[17] the real part of K:

$${}^{(2)}R = (K + \bar{K}).$$

This can be seen by choosing $p = 1$, $q = -1$ in (3.12); then η is defined by a real tangent vector y^a given by

$$y^a = -\bar{\eta} m^a - \eta\bar{m}^a, \qquad \text{so } \eta = m^a y_a. \tag{3.14}$$

Equation (3.12) expresses the rotation of y^a as the vector y^a is parallel-transported within S around a small loop on S. In a similar way, if we choose $p = q = 1$ in (3.12), then η is defined by a vector z^a normal to S given by

$$z^a = \xi l^a + \eta m^a, \qquad \text{so } \eta = l^a z_a. \tag{3.15}$$

As the vector z^a is transported around a small loop on S [this is the generalized Fermi transport of (3.6)—the vector must be continually projected out perpendicular to S as it is carried around], then it undergoes a boost of magnitude given by the product of $i(\bar{K} - K)$ with the small area enclosed by the loop. Thus the imaginary part of K describes an extrinsic curvature invariant of S in M.

To sum up, we see that our formalism provides an effecive means of studying the intrinsic geometry of Riemannian 2-surfaces. This is when we restrict attention merely to spin-weighted quantities (type $\{s,-s\}$) and employ a tetrad of vectors related to a surface S as described above. This is essentially the same $\eth$ formalism as has been employed previously. However, if we allow spin- and boost-weighted quantities (general type $\{p,q\}$) then we can also use our formalism for the study of the extrinsic geometry of 2-surface embedded in a space-time. (This is still employing only the $\eth$ and $\eth'$ operators at points of a 2-surface S.) But our formalism is, of course, much more general than this since we can use the Þ and Þ′ operators to study the relation between different embedded 2-surfaces, and also in circumstances in which the plane-elements spanned by the real and imaginary parts of m^a need not be surface-forming.

4. CONCLUSIONS

We have presented, here, a formalism which should have many applications to the type of situation in which the NP formalism has proved useful in the past. It is likely that the present approach will provide considerable formal simplifications in most cases. In particular, it should yield asymptotic formulas directly in terms of the fami-

liar ð operator, obviating any necessity of having to pass through a complicated intermediate stage of explicit dependence on angular coordinates. We have contented ourselves here with presenting only the basic formalism. Detailed applications will be given elsewhere.

ACKNOWLEDGMENT

The authors are indebted to Dr. M. Walker and Dr. J. Stewart for pointing out errors in the original manuscript

---

*Supported in part by the National Science Foundation under Grant GP-34721X with the University of Chicago.

[1]Cf. L. Eisenhart, *Riemannian Geometry* (Princeton, U. P., Princeton, N. J., 1966).

[2]R. Penrose, *Battelle Rencontres*, edited by C. DeWitt, J. Wheeler (Benjamin, New York, 1968).

[3]Whenever a Latin letter such as $a,b,...,a_0,...$ or $A,B,...,A_0,...A',...$ is used as an index, it is to be read merely as a *label* indicating the tensor or spinor nature of a quantity. Such indices are treated formally and are not assigned values such as 0,1,2,3. Thus ξ^a is to be thought of *as* a vector, not as a set of components. Similarly, $\psi_A{}^{BC'}$ will denote an actual spinor and not a set of spinor components. When we wish to look at individual components of a vector, or tensor, or spinor we require a basis frame to be introduced. This basis frame could, in the case of tensors, be that defined by some coordinate system (holonomic frame) or it could be introduced independently as a tetrad system, orthonormal or otherwise (nonholonomic frame). For spinors we would require a spinor dyad system. Components in a basis frame are labeled in Ref. 2 by German letters $\mathfrak{a},\mathfrak{b},...,\mathfrak{a}_0,...$ (tensor components) or $\mathfrak{A},\mathfrak{B},...,\mathfrak{A}_0,...\mathfrak{A}',...$ (spinor components) and for consistency's sake we shall stick to this here. Then $\mathfrak{a},\mathfrak{b},...$ range over 0,1,2,3, and $\mathfrak{A},\mathfrak{B},...$ range over 0,1 etc. One significant advantage of this type of formulation is that it enables the spinor–tensor correspondence to be expressed without the use of Infeld–van der Waerden connection symbols $\sigma_{\mathfrak{a}}{}^{\mathfrak{A}\mathfrak{A}'}$. It is now legitimate to *equate* a tensor with its spinor equivalent, so we can correctly write $\chi_a{}^{bc} = \chi_{AA'}{}^{BB'CC'}$. The tensor label a may thus be regarded as a *shorthand* for the pair of spinor labels AA'; similarly b is a shorthand for BB', etc. (The symbols $\sigma_{\mathfrak{a}}{}^{\mathfrak{A}\mathfrak{A}'}$ would merely express the relation between tensor and spinor frames.)

[4]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962). [In using equations from this paper, we have taken account of the published errata: J. Math. Phys. **4**. 998 (1963)].

[5]R. Penrose, "An Analysis of Space–Time," Adam's Prize Essay, 1966.

[6]Although we shall treat the prime as effectively an allowable operation within the formalism, it is possible in certain circumstances to have quantities to which the prime could not be applied (e.g., in an algebraically special vacuum solution, which is not type D, if l^a were chosen in the direction of the repeated principal null direction, then quantities could be definable from l^a having no analog in terms of n^a). When applied to the quantities defined in this paper, however, the prime will denote a well-defined operation.

[7]The symbol Þ is pronounced "thorn" and ð is pronounced "e(d)th"; Þ and ð are the phonetic symbols for the soft and hard "th", respectively.

[8]J. Winicour, private communication, has defined a similar operator ð having some, but not all of the properties of our ð operator.

[9]E. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).

[10]J. N. Goldberg, A. J. Macfarlane, F. Rohrlich, E. C. G. Sudarshan and E. T. Newman, J. Math. Phys. **8**, 2155 (1967).

[11]F. A. E. Pirani, *Brandeis Summer Institute, 1964* (Prentice-Hall, Englewood Cliffs, N. J., 1965).

[12]R. K. Sachs, private communication (1961).

[13]One may verify that this operation is indeed covariant differentiation within S by observing that the various axioms for covariant differentiation are satisfied, with regard to tensors completely tangent to S.

[14]Fermi–Walker transport (A. G. Walker, Proc. R. Soc. Edinb. **52**, 345 (1932)) of a vector v^a with respect to a (timelike) curve γ may be defined in the following way. First we decompose v^a into two parts, $v^a = p^a + t^a$, where p^a is perpendicular to γ and t^a is tangent to γ. The part t^a is to undergo "parallel transport" within γ (i.e., to have merely its length $t^a t_a$ remain constant), while p^a is to undergo Fermi transport [E. Fermi, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Nat. Rend. **31**, 21 (1922)] along γ. The definition of Fermi transport is such that p^a is carried along γ always remaining perpendicular to γ. This is achieved by demanding that the projection orthogonal to γ, of the covariant derivative of p^a in the tangent direction to γ, should vanish. In a similar way, we can define a means of transporting a vector v^a in directions lying within a 2-surface S. Again we form a decomposition $v^a = p^a + t^a$, where p^a is now perpendicular to S and t^a is tangent to S. The vector t^a is to be carried by (intrinsic) parallel transport within S, while p^a undergoes a form of "Fermi transport" on S. Thus, we demand that the projection orthogonal to S, of the covariant derivative of p^a in directions tangent to S, should vanish. However, since S is two-dimensional, this analog of Fermi(–Walker) transport is not normally "integrable." That is to say, the transport of v^a is dependent upon the path within S along which it is carried.

[15]Equations (3.9), (3.10), etc. can be used quite generally to translate covariant equations into our formalism as follows: $m^a\bar{m}^b o^E(\nabla_a T_{bE})$
$= \bar{m}^b o^E \eth T_{bE} = \eth(T_{bE}\bar{m}^b o^E) - T_{bE} o^E \eth\bar{m}^b - T_{bE}\bar{m}^b \eth o^E$
$= \eth(T_{bE}\bar{m}^b o^E) + \rho' T_{bE}\, l^b o^E + \bar{\rho} T_{bE} n^b o^E + \sigma T_{bE}\bar{m}^b l^E$.

[16]R. Penrose, "Null Hypersurface Initial Data for Classical Fields of Arbitrary Spin and for General Relativity", in P. G. Bergmann's ARL Tech. Documentary Report 63-56, (1963).

[17]The numerical factor in the equality of $K + \bar{K}$ with the Gaussian curvature of S can be checked by choosing S to be a sphere of radius $2^{-1/2}$ and comparing with Refs. 9 and 10.

[18]For normal applications, p and q will be integers, so $p = \bar{p}$ and $q = \bar{q}$. However, the formalism will still work if p and q are any pair of complex numbers for which $p - q$ is an integer. See, for example, M. A. Naimark, *Linear Representation of the Lorentz Group* (Pergamon, Oxford, 1964).

Isospinors

M. D. Maia

Departamento de Matemática, Universidade de Brasília, Brasília, Brasil
(Received 15 September 1972)

A spinor structure which is covariant under a representation of the groups of isometries in general relativity is defined. The relationship of isospinors with space–time objects is investigated and their use in describing elementary particles in general relativity is discussed.

I. INTRODUCTION

In 1967 Penrose introduced a spinor formalism in special relativity based on a representation of the group of conformal motions in Minkowski space.[1] One of the motivations for the introduction of that spinor formalism is the fact that the two-component spinor field equations in special relativity show different behavior when submitted to homogeneous Lorentz transformations and to translations. In the new spinor formalism (twistors) the covariance under rotations and translations appear in a single expression.

The twistor formalism suggested that a similar spinor structure (isospinors) can be constructed both in special and in general relativity but which transforms according to a representation of the group of isometries of the space–time. In the case of general relativity the covariance of the isospinors has the same expression for rotations and translations but in special relativity two distinct expressions appear.

As it may have already become clear, isospinors will have physical significance only when the space–time in question admit a group of isometries. For this reason we will restrict ourselves to the set of space–times with at least one Killing vector field.

The isospinors form a natural vehicle to describe structures in general relativity which are covariant under isometries.

Therefore, one of the possible applications of isospinors is the description of elementary particles in general relativity. In a subsequent paper we shall present a similar formalism which shows covariance under a spinor representation of the groups of conformal motions in general relativity.

II. EMBEDDING BUNDLES

The two-component spinor fields in general relativity are defined on the tangent bundle to a space–time $R4$. The relation between the tangent spaces and the space–time $R4$ can be realized by the use of a tetrad field which transforms partially by the group of propagation of the tetrad field (the tetrad group) and partially by the Lorentz group. Therefore, the two-component formalism defined on the tangent bundle is suitable to describe the Lorentz covariance. If we assume that $R4$ has a group of isometries, then part of this group is absorbed by the tetrad group, and consequently the isometric covariance is not fully represented by the (flat) spinor group. To obtain a flat spinor structure which is covariant under the group of isometries of $R4$, we replace the tangent bundle by the local isometric embedding bundle of $R4$.

Suppose $R4$ has a minimal embedding space $M(p,r,s)$ with dimension $p = 4 + t$ and signature $r + s$. The isometric embedding of $R4$ is the manifold $\beta = (M(p,r,s), R4, \pi)$, where $M(p,r,s)$ is the fibre and π is the projection map.

Let X^μ be the Cartesian coordinates of a point in $M(p,r,s)$, and let $x^{\underline{\alpha}}$ be the Gaussian coordinates based on $R4$ (all Greek indices run from 1 to p). The coordinates of a point in $R4$ are $x^{\underline{i}}$ and the coordinates measured on the directions orthogonal to $R4$ are $x^{\underline{A}}$ (lower case Latin indices run from 1 to 4 while capital Latin indices run from 5 to p. The tilde under the Gaussian indices are used to distinguish them from Cartesian indices when numerical values are given). The embedding gives the coordinates transformations $X^\mu = X^\mu(x^{\underline{\alpha}})$ and its derivative map gives the transformation of tensors referred to the two systems. We use the notation

$$x^{\underline{\alpha}}{}_\mu = \frac{\partial x^{\underline{\alpha}}}{\partial X^\mu}, \quad X^\mu{}_{\underline{\alpha}} = \frac{\partial X^\mu}{\partial x^{\underline{\alpha}}}, \quad X_\mu{}^{\underline{\alpha}} = \frac{\partial X^\mu}{\partial x^{\underline{\alpha}}},$$
$$x_{\underline{\alpha}}{}^\mu = \frac{\partial x^{\underline{\alpha}}}{\partial X^\mu}. \tag{II. 1}$$

Thus if U^μ are the Cartesian components of a vector in $M(p,r,s)$, its Gaussian components are $\xi^{\underline{\alpha}} = x_\mu{}^{\underline{\alpha}} U^\mu$.

On the other hand, if $\eta^{\mu\nu}$ are the Cartesian components of the metric tensor of $M(p,r,s)$ and $g^{\underline{\alpha}\underline{\beta}}$ are its Gaussian components, they are related by

$$g^{\underline{\alpha}\underline{\beta}} = x_\mu{}^{\underline{\alpha}} x_\nu{}^{\underline{\beta}} \eta^{\mu\nu}. \tag{II. 2}$$

Conversely, we have the relations

$$\eta^{\mu\nu} = X_{\underline{\alpha}}{}^\mu X_{\underline{\beta}}{}^\nu g^{\underline{\alpha}\underline{\beta}}, \qquad \eta_{\mu\nu} = x_\mu{}^{\underline{\alpha}} x_\nu{}^{\underline{\beta}} g_{\underline{\alpha}\underline{\beta}},$$

such that $g^{\underline{\alpha}\underline{\beta}} g_{\underline{\beta}\underline{\gamma}} = \delta^{\underline{\alpha}}{}_{\underline{\delta}}$.

Now we can show that if $R4$ has a group of isometries, then this group is locally induced by the homogeneous fibre group of the bundle β. This fibre group is the group of isometries of $M(p,r,s)$ which is the group of pseudo rotations and reflections $O(r,s)$ in that space. Let

$$X'^\mu = X^\mu + U^\mu, \qquad U^\mu = \epsilon^\mu_\nu X^\nu, \tag{II. 3}$$

be one infinitesimal transformation of $O(r,s)$, where the $p(p-1)/2$ infinitesimals ϵ^μ_ν are constants and satisfy $\epsilon^{(\mu\nu)} = 0$. The isometric character of this transformation is expressed by the vanishing of the Lie derivative of $\eta^{\mu\nu}$ respect to U^μ:

$$\pounds_U \eta^{\mu\nu} = U^{(\mu,\nu)} = 0. \tag{II. 4}$$

(Round brackets on indices mean complete symmetrization.) On the other hand, the Gaussian components of the generators are $\xi^{\underline{\alpha}} = x^{\underline{\alpha}}_\mu U^\mu$, so that we obtain the infinitesimal transformation of the Gaussian coordinates,

$$x'^{\underline{\alpha}} = x^{\underline{\alpha}} + \xi^{\underline{\alpha}}. \tag{II. 5}$$

Copyright © 1973 by the American Institute of Physics

Since (II. 4) is a tensor expression, we have also

$$g^{\underline{\alpha}\underline{\beta}} = \xi^{(\underline{\alpha};\underline{\beta})}, \tag{II. 6}$$

where the semicolon denotes covariant derivatives. Expression (II. 5) can be divided in two parts:

$$x'^{\underline{i}} = x^{\underline{i}} + \xi^{\underline{i}}, \quad x'^{\underline{A}} = x^{\underline{A}} + \xi^{\underline{A}}, \tag{II. 7}$$

and (II. 6) gives

$$\xi^{(\underline{i};\underline{j})} = 0, \quad \xi^{(\underline{i};\underline{A})} = 0, \quad \xi^{(\underline{A};\underline{B})} = 0.$$

Now, in the Gaussian system the space–time $R4$ is simply defined by $x^{\underline{A}} = 0$. If $f(x^{\underline{\alpha}})$ is any real function defined in $M(p, r, s)$, its space–time 'projection' is $\lim f(x^{\underline{\alpha}}) = f(x^{\alpha})|_{R4}$ as $x^{\underline{A}} \to 0$. Sometimes a function is only defined on the surface $R4$; we denote this fact by $f(R4)$. In particular, we have[2]

$$g^{\underline{i}\underline{j}}|_{R4} = x^{\underline{i}}_{\mu} x^{\underline{j}}_{\nu} \eta_{\mu\nu}|_{R4} = g^{\underline{i}\underline{j}}(R4),$$

$$g^{\underline{i}\underline{A}}|_{R4} = 0,$$

$$g^{\underline{A}\underline{B}}|_{R4} = x^{\underline{A}}_{\mu} x^{\underline{B}}_{\nu} \eta^{\mu\nu}|_{R4} = \pm\, \delta^{\underline{A}\underline{B}}.$$

The transformations induced in $R4$ by (II. 7) are

$$x'^{\underline{i}} = x^{\underline{i}} + \xi^{\underline{i}}|_{R4} = 0, \quad x'^{\underline{A}} = \xi^{\underline{A}}|_{R4} = 0 \tag{II. 8}$$

and are subjected to the conditions

$$\xi^{(\underline{i};\underline{j})}|_{R4} = 0, \quad \xi^{(\underline{i},\underline{A})}|_{R4} = 0, \quad \xi^{(\underline{A};\underline{B})}|_{R4} = 0 \tag{II. 9}$$

The covariant derivative of a covariant vector in the Gaussian system is given by

$$\xi^{\underline{\alpha};\underline{\beta}} = \xi^{\underline{\alpha},\underline{\beta}} + g^{\underline{\beta}\underline{\delta}} \Gamma_{\underline{\gamma}\underline{\delta}}^{\underline{\alpha}} \xi^{\underline{\gamma}},$$

where the Gaussian components of the connection are

$$\Gamma_{\underline{\beta}\underline{\gamma}}^{\underline{\alpha}} = g^{\underline{\alpha}\underline{\delta}} \Gamma_{\underline{\beta}\underline{\gamma},\underline{\delta}} = \tfrac{1}{2} g^{\underline{\alpha}\underline{\delta}} (g_{\underline{\beta}\underline{\delta},\underline{\gamma}} + g_{\underline{\gamma}\underline{\delta},\underline{\beta}} - g_{\underline{\beta}\underline{\gamma},\underline{\delta}}).$$

As it follows, the expression $\xi^{\underline{i};\underline{j}}|_{R4}$ does not coincide with the expression of covariant derivative in $R4$ unless the condition

$$\xi^{\underline{A}}|_{R4} = 0, \tag{II. 10}$$

is imposed. Under this condition, (II. 8) becomes

$$x'^{\underline{i}} = x^{\underline{i}} + \xi^{\underline{i}}|_{R4}, \quad x'^{\underline{A}} = 0.$$

Consequently, the condition (II. 10) is such that the transformation does not alter the definition of $R4$. Thus points in space–time are mapped into points of space–time. Therefore, (II. 10) defines a subgroup of $O(r, s)$ which depends on the space–time embedded. Together with (II. 10) the first equation (II. 9) gives part of Killing's equations of $R4$, corresponding to 'rotations' in space–time. The second equation corresponds to rotations in the planes $[\underline{i}, \underline{A}]$ projected in $R4$. These projections give the translations in the neighborhood of the embedding point of $R4$. (As can be deduced from the process of group contraction of Inönü and Wigner.[3] Finally the last equation is identically zero. Thus, adding the reflections, we obtain from $O(r, s)$ and the condition (II. 10) the local group of isometries of $R4$. Conversely, given the transformation

$$x'^{\underline{i}} = x^{\underline{i}} + \xi^{\underline{i}}(R4),$$

then, obviously, $\xi^{\underline{A}}|_{R4} = 0$, and we get a subgroup of $O(r, s)$.

III. CLIFFORD ALGEBRAS

The formal process to introduce a spinor structure in a Euclidean space is through a representation of the Clifford algebra defined on the same space. A Clifford algebra of dimension 2^n, defined on the Euclidean space E_n, can be defined as the quotient

$$C_n = T(E_n)/I,$$

where $T(E_n)$ is the tensor algebra of E_n and I is the ideal generated by the elements of the form

$$x \otimes x - f(x) \cdot 1, \quad x \in E_n,$$

where $\otimes$ denotes the tensor product and $f(x)$ a quadratic form in E_n (Chevalley[4]). The algebra has n generators e_μ such that

$$e_{(\mu} e_{\nu)} = \delta_{\mu\nu}.$$

In our case we have a pseudo-Euclidean space $M(p, r, s)$ with metric $n_{\mu\nu}$ so that the generators of C_p satisfy

$$e_{(\mu} e_{\nu)} = \eta_{\mu\nu}. \tag{III. 1}$$

One important property is that the group of automorphisms of the Clifford algebra C_p defined on $M(p, r, s)$ is isomorphic to $O(r, s)$ (Brauer and Weyl[5]; see also Boerner[6]). The subset of the algebra generated by the elements

$$M_{\mu\nu} = \tfrac{1}{2} e_{[\mu} e_{\nu]} \tag{III. 2}$$

(square brackets on indices mean complete antisymmetrization) satisfy the commutation relation

$$[M_{\mu\nu} M_{\rho\sigma}] = (\eta_{\mu\rho} M_{\nu\sigma} + \eta_{\nu\sigma} M_{\mu\rho} - \eta_{\mu\sigma} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\sigma}).$$

Therefore, this subset is isomorphic to the Lie algebra of $O(r, s)$. Let

$$e'_\mu = S e_\mu S^{-1} \tag{III. 3}$$

be an inner automorphism of C_p. Then it can be shown that the operators S are generated by

$$\delta S = 1 + \tfrac{1}{2} \epsilon^{\mu\nu} M_{\mu\nu}. \tag{III. 4}$$

Therefore, the subset of the algebra generated by $M_{\mu\nu}$ generates the group of inner automorphisms of the Clifford algebra. Combining this result with the result of the proceeding section it follows that under the condition (II. 10) the group of automorphisms of the Clifford algebra C_p defined on $M(p, r, s)$ induces locally the group of isometries of the space–time $R4$ embedded locally in $M(p, r, s)$.

IV. ISOSPINORS

The Clifford algebra C_2 (quaternion algebra) has a matrix representation given by the Pauli matrices:

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

This representation can be used to obtain the matrix representation of any Clifford algebra.[5] We use the following matrices:

$$P_\alpha = \sigma_2 \otimes \cdots \otimes \sigma_2 \otimes \sigma_1 \otimes \sigma_0 \otimes \cdots \otimes \sigma_0,$$
$$Q_\alpha = \sigma_2 \otimes \cdots \otimes \sigma_2 \otimes \sigma_3 \otimes \sigma_0 \otimes \cdots \otimes \sigma_0,$$
$$P_0 = \sigma_2 \otimes \cdots \otimes \sigma_2, \quad \alpha = 1 \cdots \nu \tag{IV. 2}$$

where we have ν factors, $p = 2\nu$ or $p = 2\nu + 1$ and the matrices σ_1, σ_3 occupy the αth place. The Kronecker product above indicated is once for all defined as from right to left. Thus, for example, if A, B are 2×2 matrices, we have

$$A \otimes B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes B = \begin{pmatrix} aB & bB \\ cB & dB \end{pmatrix}.$$

The Weyl representation of the Clifford algebra is obtained by choosing a particular matrix P_α, Q_α, or P_0 to represent e_μ. Thus by renaming the matrices, we obtain many equivalent representations.

The spinor structure is derived from the above representation as one ideal of the algebra generated by a certain algebraic element.[4] Equivalently we can regard the matrices (IV. 2) as operators acting on a 2^ν-dimensional complex vector space, called the spinor space of the algebra. The matrices (IV. 2) themselves belong to the tensor algebra of the spinor space. The spinors transform according to the corresponding matrix representation of the group of automorphisms of the Clifford algebra. Let $DM_{\mu\nu}$ be the matrix representation of (III. 2). Then we get from (III. 4) the matrices

$$\delta D = I + \tfrac{1}{2} \epsilon^{\mu\nu} DM_{\mu\nu}, \tag{IV. 3}$$

which are the generators of the spinor group. The covariant spinors transform as

$$\psi' = D\psi \tag{IV. 4}$$

and the contravariant spinors transform as

$$\chi' = \chi D^{-1}. \tag{IV. 5}$$

We use capital Latin letters to label spinor, and unless otherwise stated they run from 1 to 2^ν. Thus, in terms of components, (IV. 4) and (IV. 5) read

$$\psi'^A = D^A{}_B \psi^B, \quad \chi'_A = D^B{}_A \chi_B.$$

In general ψ, χ, and D are complex matrices so that we may have also complex conjugate transformations. We use the dot notation to indicate the transformation of the spinors by $\overline{D}$. Thus

$$\psi'^{\dot A} = \overline{D}^{\dot A}{}_{\dot B} \psi^{\dot B}, \quad \chi'_{\dot A} = \overline{D}^{\dot B}{}_{\dot A} \chi_{\dot B}.$$

Let Δ be the spinor group generated by (IV. 3). As it is a representation of the group of automorphisms of the Clifford algebra, then it follows that its subgroup defined by the condition (II. 10), $\Delta(R4)$ is a representation of the group of isometries of the space-time $R4$ embedded in $M(p, r, s)$. The 2^ν-component spinors defined in $M(p, r, s)$ and which transform according to $\Delta(R4)$ are called the isospinors of $R4$ and $\Delta(R4)$ is called the isospinor group of $R4$. On the other hand, since $M(p, r, s)$ is the embedding space for a class of space-times, Δ is called the class isospinor group. Any isospinor group of a space-time embedded in $M(p, r, s)$ can be obtained from Δ by imposing the corresponding condition (II. 10). Thus, for example, in $M(6, 4, 2)$ we have the Schwarzschild spinor group Δ(Schwarzschild) which, as we shall see is the same as $SU(2, 2)$(Schwarzschild).

The number of isospinor components can be as high as 32. However, in certain cases, when $p = 2\nu$ and ν is an odd number, we can split the spinors into two equivalent halves (Cartan[7]: semispinors). In the case $p = 8$ we can also deal only with half the number of spinor components by using the triality principle of the Clifford algebras (Chevalley[4]; see also Gamba[8]).

V. CLASS ISOSPINOR GROUPS

As the Clifford algebras are defined on a pseudo-Euclidean space, it follows that the matrices which represent a Clifford number are not necessarily Hermitian. Instead of the Hermitian condition they satisfy the relation

$$X^\dagger = \sigma \mu X \mu^{-1}, \tag{V. 1}$$

where $X^\dagger$ is the Hermitian conjugate of X, μ is a constant matrix, and σ is $+1$ when s is even and -1 when s is odd. The matrices which represent the generators e_μ are either Hermitian or anti-Hermitian. Let $e_{\mu i}$ ($i = 1 \cdots m$) be the anti-Hermitian matrices. From $-e_{\mu i} = \sigma \mu e_{\mu i} \mu^{-1}$ we get

$$\mu = \alpha(s) e_{\mu 1} e_{\mu 2} \cdots e_{\mu m}, \tag{V. 2}$$

where we choose $\alpha(s) = i$ or $\alpha(s) = 1$ such that μ will be always Hermitian. Since (V. 1) must be true for any basis of the algebra, we must have also

$$X'^\dagger = \sigma \mu X' \mu, \tag{V. 3}$$

with $X' = D \times D^{-1}$. By comparing with (V. 1), it follows that

$$\mu = \pm D^\dagger \mu D, \tag{V. 4}$$

where the plus sign holds for proper transformations [rotations in $M(p, r, s)$] and the minus holds for improper transformations in $M(p, r, s)$. Therefore, the constant matrix μ is a characteristic of the class isospinor group. From (IV. 2) we have that the matrices e_μ are of either type:

$$e_\mu = \left(\begin{array}{c|c} g_\mu & 0 \\ \hline 0 & h_\mu \end{array}\right) \quad \text{or} \quad e_\mu = \left(\begin{array}{c|c} 0 & g_\mu \\ \hline h_\mu & 0 \end{array}\right). \tag{V. 5}$$

Thus according to the parity of p and s we have four cases to consider.

(a) p even, s even: In this case we can always choose the Weyl representation so that the anti-Hermitian matrices have the shape

$$e_{\mu i} = \left(\begin{array}{c|c} 0 & g_{\mu i} \\ \hline h_{\mu i} & 0 \end{array}\right).$$

(In other words the matrices $e_{\mu i}$ are different from Q_1.) With this choice of representation we have from (V. 2) that μ has the shape

$$\mu = \left(\begin{array}{c|c} M & \\ \hline & \pm M \end{array}\right), \tag{V. 6}$$

where

$$M = g_{\mu 1} h_{\mu 2} \cdots g_{\mu s}$$

are Hermitian matrices.

Suppose we are given a rotation matrix:

$$D = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}. \tag{V.7}$$

Then we get from (V.4)

$$M = a^{\dagger} M a. \tag{V.8}$$

We have two subcases: If ν is even and reflections are considered, then the isospinor group is the group of $2^{\nu} \times 2^{\nu}$ complex matrices satisfying (V.4) with μ given by (V.6). If ν is odd, then we have semi-isospinors which transform by the group of $2^{\nu-1} \times 2^{\nu-1}$ matrices which satisfy (V.8). The matrix which characterizes the semi-isospinor group is M. We shall denote by $SU(A, A, A, \ldots; A)$ the spinor group characterized by the matrix μ of the form

$$\mu = \begin{pmatrix} I_A & & & \\ & -I_A & & \\ & & \ddots & \\ & & & I_A \end{pmatrix}, \tag{V.9}$$

where I_A is the $A \times A$ unit matrix.

One example is the isospinor group of the Schwarzschild solution. The Schwarzschild solution belongs to the class $M(6, 4, 2)$ and the class isospinor group is $SU(2, 2)$. By imposing the condition (II.10) we obtain the Schwarzschild isospinor group $SU(2, 2)$ (Schwarzschild), which is a subgroup of $SU(2, 2)$.

(b) p even, s odd: Again choosing the matrices not so that $e_{\mu i}$ is different from Q_1, we get from (V.2)

$$\mu = \left(\begin{array}{c|c} 0 & M \\ \hline \pm M & 0 \end{array}\right).$$

Considering the rotation matrices similar to (V.7), one gets

$$M = a M^{\dagger} b. \tag{V.10}$$

There are two subcases: If ν is even, the group of isometries of $M(p, r, s)$ is $O(r, s)$ and the isospinor group is generally denoted by $SL(2^{\nu}, C)$. This spinor group may be characterized by the presence of the little groups which correspond to the cases in which $a = b$ in (V.7). A typical example is given by the Minkowski space $M(4, 3, 1)$. The isospinors are the Dirac spinors with the isospinor group $SL(4, C)$ which contain the little groups $SU(2)$ and $SU(1, 1)$. If ν is odd, we have semi-isospinors with $2^{\nu-1}$ components. In this case the group of semi-isospinors are not characterized directly by the matrices M and N. We denote the isospinors by $SL(2^{\nu-1}, C)$ and this group is characterized by the collection of little groups which satisfy (V.10). Examples of classes where this may occur are $M(6, 5, 1)$ and $M(6, 3, 3)$.

(c) p odd, s even: In this case we have

$$\mu = \left(\begin{array}{c|c} M & 0 \\ \hline 0 & \pm M \end{array}\right).$$

And the spinor groups is, as in the case (a), denoted by $SU(A, A, \ldots, A)$. Since p is odd, semispinors cannot occur.

A typical example is the class of the de Sitter space-time $M(5, 3, 2)$ whose isospinor group is $SU(1, 1, 1, 1)$. We notice that in the cases of spaces with constant curvature the condition (II.10) is trivially satisfied. Therefore, the isospinors of de Sitter space are the four component spinors which transform according to $SU(1, 1, 1, 1)$.

(d) p odd, s odd: The form of μ is

$$\mu = \left(\begin{array}{c|c} 0 & M \\ \hline \pm M & 0 \end{array}\right),$$

so that the isospinor group is denoted by $SL(2^{\nu}, C)$ and it is specified by its collection of little groups as in the case (b). As p is odd, we cannot have semispinors. A typical example is the class $M(5, 4, 1)$ which is the class of the anti de Sitter space-time.

In the Appendix we list all the 22 possible isospinor class groups which may occur in general relativity.

VI. THE GEOMETRY OF ISOSPINORS

In order to establish the relation between isospinors and tensors in space-time we construct the p spinor tensors $e_{\mu}{}_{B}{}^{A}$ which are the rank-2 mixed spinors associated with each generator e_{μ} of the Clifford algebra. If X^{μ} is a vector in $M(p, r, s)$, then it corresponds to the algebraic element $X = X^{\mu} e_{\mu}$ and also to the rank-2 spinor with components

$$X^{A}{}_{\dot{B}} = X^{\mu} e_{\mu}{}^{A}{}_{\dot{B}}. \tag{VI.1}$$

Conversely, from (III.1) we have

$$e_{(\mu}{}^{A}{}_{|\dot{B}|} e_{\nu)}{}^{\dot{B}}{}_{C} = \eta_{\mu\nu} \delta^{A}{}_{C}. \tag{VI.2}$$

Applying this expression in (VI.1), we get the components of the vector which is associated with a rank-2 mixed spinor

$$X^{\mu} = e^{\mu A}{}_{\dot{B}} \chi_{A}{}^{\dot{B}}.$$

If $R4$ is a space-time belonging to the class $M(p, r, s)$, we may construct the Gaussian system of coordinates and define the condition (II.10) under which the above spinors become isospinors. In this Gaussian system we define the isospinor tensors $e_{\underline{\alpha}}{}^{A}{}_{\dot{B}} = X^{\mu}{}_{\underline{\alpha}} e_{\mu}{}^{A}{}_{\dot{B}}$, so that if $\chi_{A}{}^{B}$ is a given isospinor, we get the Gaussian components of the associated vector,

$$\xi_{\underline{\alpha}} = e_{\underline{\alpha}}{}^{A}{}_{\dot{B}} \chi_{A}{}^{\dot{B}},$$

and the space-time components of this vector are

$$\xi_{\underline{i}}\,|_{R4} = e_{\underline{i}}{}^{A}{}_{\dot{B}} \chi_{A}{}^{\dot{B}}.$$

Conversely given a vector in $R4$ with Gaussian components $\xi_{\underline{i}}$ ($\xi_{\underline{A}} = 0$) the corresponding isospinor is

$$\chi^{A}{}_{\dot{B}} = \xi^{\underline{i}} e_{\underline{i}}{}^{A}{}_{\dot{B}}.$$

A metric isospinor can be introduced naturally from the theory of spinors. This metric isospinor is intrinsically associated with the equivalence between the spinor representation of the group of isometries of $M(p, r, s)$ and its adjoint representation. If the matrices of the spinor representation are generated by (IV.3), the matrices of the adjoint representation are generated by

$$\delta D = \pm\,(1 - \tfrac{1}{2}\,\epsilon^{\mu\nu}\,DM_{\mu\nu}) = \det D(\delta D)^{-1}.$$

To see the equivalence between these representations we notice that the transposed matrices e_μ^T are also the generators of the same Clifford algebra

$$e^T_{(\mu}\,e^T_{\nu)} = \eta_{\mu\nu}.$$

Now, consider the matrix ϵ such that $e_\mu^T = \epsilon e_\mu \epsilon^{-1}$. From the Weyl representation we have that either $e_\mu^T = e_\mu$ or $e_\mu^T = -e_\mu$.

Therefore,

$$\epsilon = e_{\mu 1} e_{\mu 2} \cdots e_{\mu r}. \tag{VI. 3}$$

where now $e_{\mu i}$ are such that $e^T_{\mu i} = -e_{\mu i}$. When ν is even, we have $\epsilon^T = \epsilon$ if $\nu/2$ is even and $\epsilon^T = -\epsilon$ if $\nu/2$ is odd, and the form of the matrix ϵ is

$$\epsilon = \left(\begin{array}{c|c} \epsilon' & 0 \\ \hline 0 & \epsilon'' \end{array}\right). \tag{VI. 4}$$

On the other hand, when ν is odd, we have $\epsilon^T = -\epsilon$ if $(\nu-1)/2$ is even and $\epsilon^T = \epsilon$ if $(\nu-1)/2$ is odd, and the form of ϵ is

$$\epsilon = \left(\begin{array}{c|c} 0 & \epsilon' \\ \hline \epsilon'' & 0 \end{array}\right). \tag{VI. 5}$$

Consider now the transpose of $e'_\mu = De_\mu D^{-1}$. Using the fact that $\det(D) = \pm 1$, we get, from $D = \epsilon D \epsilon^{-1}$, that $\epsilon = D\epsilon D^T$. Therefore, ϵ is a rank-2 contravariant isospinor whose components we denote by ϵ_{AB}. It follows that ϵ^{-1} is a covariant isospinor with components ϵ^{AB}. These isospinors can be used to rise and lower isospinor indices. A covariant isospinor transforms as $\phi' = D\phi$. It follows that $\epsilon\phi = \epsilon D\phi = D\epsilon\phi = D^{-1}\epsilon\phi$. Defining $\psi^T = \epsilon\phi$, we get $\psi' = \epsilon\phi' = \psi D^{-1}$ which shows that ψ is a contravariant isospinor. In terms of components we set

$$\phi_A = \phi^B \epsilon_{BA} \quad \text{and} \quad \psi^B = \epsilon^{BA}\psi_A.$$

In the cases where semi-isospinors occur, the relation to space-time tensors follows without difficulty.

VII. ELEMENTARY PARTICLES IN GENERAL RELATIVITY

As we mentioned in the Introduction, the isospinors form a suitable vehicle to describe the properties of space-time which have isometric covariance. In particular the isospinors may be used to describe elementary particles in general relativity. At present there are various suggestions for a definition of elementary particles in general relativity from the group theoretic point of view.[9] The main difficulty in this direction lies in the choice of a group of isometries in general relativity. Among the possible candidates there is the BMS group,[10] whose representations have been studied by Cantoni[11] and McCarthy.[12] Another candidate is the fiber group of the tangent bundle, which is the Poincaré group. This group is justified mainly by the weakness of the gravitational force. However, the group theoretical approach to elementary particles is a purely kinematical problem, and the use of such argument would be somewhat artificial.

Now we shall try to envisage the fiber group of the embedding bundles as group of isometries in general relativity. We can regard the curved space-times of general relativity as deformations of Minkowski space

TABLE 1.

p	$M(p,r,s)$	$L(p,r,s)$	Spinor group	Main little groups	Important subgroups	Notes
4	$M(4,3,1)$	$SO(3,1)$	$SL(2,C)$	$SU(2)$, $SU(1,1)$		f
4	$M(4,2,2)$	$O(2,2)$	$SU(1,1) \times SU(1,1)$			g
5	$M(5,4,1)$	$SO(4,1)$	$SL(4,C)$	$SU(2)$, $SU(1,1)$	$SU(2) \times SU(2)$	g
5	$M(5,3,2)$	$SO(3,2)$	$SU(1,1,1,1)$		$SU(1,1) \times SU(1,1)$	g
6	$M(6,5,1)$	$O(5,1)$	$SL'(4,C)$	$O(5)$, $SO(4,1)$		a, e, g
6	$M(6,4,2)$	$O(4,2)$	$SU(2,2)$		$SU(2) \times SU(2)$	a, g
6	$M(6,3,3)$	$O(3,3)$	$SL''(4,C)$	$SU(1,1,1,1)$	$SU(1,1) \times SU(1,1)$	a
7	$M(7,6,1)$	$SO(6,1)$	$SL(8,C)$	$SU(4)$, $SO(4,1)$	$SU(4)$	c, e
7	$M(7,5,2)$	$SO(5,2)$	$SU(2,2,2,2)$		$SU(2,2)$, $SU(1,1,1,1)$	g
7	$M(7,4,3)$	$SO(4,3)$	$S'L(8,C)$	$SU(1,1,1,1)$	$SU(2) \times SU(2)$	g
8	$M(8,7,1)$	$O(7,1)$	$S''L(8,C)$	$O(7)$, $SU(4,4)$	$SU(4)$	a, c, e, b
8	$M(8,6,2)$	$O(6,2)$	$SU(1,1) \times SU(4,4)$		$SU(4)$	b, c
8	$M(8,5,3)$	$O(5,3)$	$S'''L(8,C)$	$SU(2,2,2,2)$	$SU(2) \times SU(2)$	a, b
8	$M(8,4,4)$	$O(4,4)$	$SU(1,1) \times SU(2222)$		$SU(2) \times SU(2)$	b
9	$M(9,8,1)$	$SO(8,1)$	$SL(16,C)$	$SU(8)$, $SU(4,4)$	$SU(4)$	c
9	$M(9,7,2)$	$SO(7,2)$	$SU(4,4,4,4)$		$SU(4,4)$	c
9	$M(9,6,3)$	$SO(6,3)$	$S'L(16,C)$	$SU(2,2,2,2)$	$SU(4)$	c, g
9	$M(9,5,4)$	$SO(5,4)$	$SU(22222222)$		$SU(2) \times SU(2)$	g
10	$M(10,9,1)$	$O(9,1)$	$S''L(16,C)$	$O(9)$, $O(8,1)$		a
10	$M(10,8,2)$	$O(8,2)$	$SU(8,8)$		$SU(8)$	a, c
10	$M(10,7,3)$	$O(7,3)$	$S'''L(16,C)$	$SU(4,4,4,4)$, $SU(2,2,2,2,2,2,2,2)$		a, c
10	$M(10,6,4)$	$O(6,4)$	$SU(4,4,4,4)$		$SU(4)$	
10	$M(10,5,5)$	$O(5,5)$	$S^{iv}(16,C)$	$SU(2,2,2,2,2,2,2,2)$	$SU(2) \times SU(2)$	a, g

[a] Semispinor groups.

[b] Semispinor groups may occur with the use of the triality principle: $M(8,6,2) \Rightarrow SU(4,4); M(8,4,4) \Rightarrow SU(2,2,2,2)$.

[c] $O(6) \sim SU(4) \supset SU(3)$.

[d] $SU(8) \supset SU(4) \supset SU(3)$.

[e] $O(5)$ as subgroups of $SU(4)$, $SO(4,1)$ as subgroup of $SU(2,2)$, $O(7)$ as subgroup of $SU(8)$.

[f] nonphysical example.

[g] Spaces which contain known space-times embedded. For example, $M(5,4,1)$, de Sitter; Einstein space-times, $M(5,3,2)$, anti de Sitter space-time, $M(6,5,1)$, Kruskal space-time, $M(6,4,2)$, Schwarzschild space-time, $M(7,5,2)$, Petrov space $T_2/G4/4$,* $M(7,4,3)$, Petrov space $T_1/G4/5$, 6,* $M(9,6,3)$, Robinson-Trautman space-time, $c \leq 0$, $M(9,5,4)$, Robinson-Trautman space-time, $c \geq 0$, $M(10,6,4)$, Weyl axisymmetric space-time,* also Osvath anti-Mach space-time,* $M(10,5,5)$, Gödel space-time.*

* As in the paper by Rosen.[13] There is no proof that these isometric embeddings are minimal.

by the introduction of a gravitational field. As this happens the Minkowski space-time is replaced locally by the fiber of the embedding bundle β and at the same time the Poincaré group is replaced by the fibre group of β. General relativity does not deal with a single space-time but with a collection of space-times which are solutions of Einstein's equations. Therefore, it is reasonable to think of the group of isometries of general relativity as a group of isometries of a class of space-times, each class being determined by the curvature of the space-times. As we have just seen the fiber group of each embedding bundle induce locally the isometries of each space-time in the class defined by the bundle. Therefore, we can regard the 22 possible fiber groups of the 22 embedding bundles as natural extensions of the Lorentz group. By the classification of the unitary irreducible representations of these groups, we obtain 22 structures for elementary particles in general relativity. It remains to see which one of these structures agree with the phenomenology of elementary particles, which in turn may provide a new experimental basis for general relativity.

APPENDIX

Table 1 contains the 22 isospinor class groups which occur in general relativity.

[1]R. Penrose, J. Math. Phys. **8,** 345 (1967).
[2]P. Szekeres, Nuovo Cimento **43,** 1062 (1966).
[3]E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39,** 510 (1953).
[4]C. Chevalley, *The Construction and Study of Certain Important Algebras* (Math Soc. of Japan, Tokyo, 1955).
[5]R. Brauer and H. Weyl, Am. J. Math. **57,** 425 (1935).
[6]H. Boerner, *Representations of Groups* (1963).
[7]E. Cartan, *The Theory of Spinors* (MIT Press, Cambridge, Mass., 1966).
[8]A. Gamba, J. Math. Phys. **8,** 775 (1967).
[9]E. T. Newman, Nature **206**, 811 (1965).
[10]H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. R. Soc. A **269,** 21 (1962).
[11]V. Cantoni, J. Math. Phys. **8,** 1700 (1967).
[12]P. J. McCarthy, thesis (King's College, University of London, 1971).
[13]J. Rosen, Rev. Mod. Phys. **37,** 215 (1965).

Multiple scattering theory of radiative transfer in inhomogeneous atmospheres

Madhoo Kanal

Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01002
(Received 8 November 1972)

In this paper we treat the multiple scattering theory of radiative transfer in plane-parallel inhomogeneous atmospheres. The treatment presented here may be adopted to model atmospheres characterized by any optical depth dependent coherent scattering phase function. For the purpose of illustration we consider the semi-infinite medium in which the absorption property of the atmosphere is characterized by an exponential function. The methodology employed here is the extension of the case treated previously by the author for homogeneous atmospheres.

1. INTRODUCTION

In an earlier paper[1] the multiple scattering theory of radiative transfer[2] was discussed for plane-parallel atmospheres which were assumed to be homogeneous. Clearly the realistic treatment of model atmospheres, such as those of planets, must take into account the inhomogeneities. In particular, if the albedo for single scattering is a function of optical depth, the reflection spectra can not be correctly interpreted in terms of homogeneous models. In this paper we present the multiple scattering theory for plane-parallel inhomogeneous atmospheres under the assumption that the phase function is separable; i.e., the phase function can be written as a product of two functions, one containing the argument τ (the optical depth) and the other the relative angle $\Omega'\cdot\Omega$ (the angle between the incident and the scattered radiation). However, we keep the formulation sufficiently general so that it is readily extended to situations where the phase function is degenerate in τ and $\Omega'\cdot\Omega$; i.e., it can be written as a finite sum of separable functions.

It is worthwhile to mention here that the multiple scattering theory provides a convenient tool for computing the absorption line shape of the emergent radiation (say from planetary atmospheres) by separating the effects of multiple scattering from that of true absorption. For homogeneous atmospheres such a separation is effective uniformly for the entire region, while for inhomogeneous atmospheres it is effective locally.

As a consequence the former situation lends itself to a power series representation in the albedo for single scattering (frequently called the Neumann expansion[2,3]) in which the coefficients (Neumann coefficients) involve the radiation frequency only via the total optical thickness of the atmosphere. In the latter situation we shall see that, due to the local character of separation between scattering and true absorption, the analogous Neumann coefficients will depend on the functional form of the local albedo for single scattering. This should not be construed as a disadvantage, but merely a complication which was to be expected in dealing with inhomogeneous atmospheres.

In this paper we treat the semi-infinite atmosphere in which the albedo for single scattering is exponentially dependent on the optical depth.

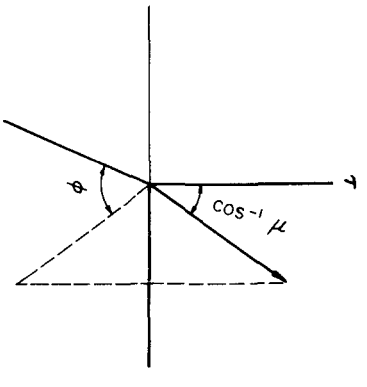

FIG. 1. Angular coordinates in the semi-infinite atmosphere.

2. GENERAL FORMULATION

Equation of radiative transfer

Let us consider the equation of radiative transfer and the Neumann expansion of it

$$\mu\frac{\partial I(\tau,\Omega)}{\partial\tau}+I(\tau,\Omega)=\frac{1}{4\pi}\int d\Omega' P(\tau,\Omega\cdot\Omega')I(\tau,\Omega') \tag{2.1}$$

and

$$\mu\frac{\partial I_n(\tau,\Omega)}{\partial\tau}+I_n(\tau,\Omega)=\frac{1}{4\pi}\int d\Omega' P(\tau,\Omega\cdot\Omega')I_{n-1}(\tau,\Omega'), \tag{2.2}$$

where $P(\tau,\Omega\cdot\Omega)$ is the coherent scattering phase function which is local in τ and

$$I(\tau,\Omega)=\sum_{n=0}^{\infty}I_n(\tau,\Omega) \tag{2.3}$$

with $I_{-1}(\tau,\Omega)\equiv 0$. We assume that $P(\tau,\Omega\cdot\Omega')$ is degenerate, i.e.,

$$P(\tau,\Omega\cdot\Omega')=\sum_{j=0}^{N}\omega_j(\tau)p_j(\Omega\cdot\Omega'), \tag{2.4}$$

where N, the index of degeneracy is arbitrary but finite. Further, if we write

$$p_j(\Omega\cdot\Omega')=\sum_{m=-\infty}^{\infty}e^{im(\phi-\phi')}S_{j,m}(\mu,\mu') \tag{2.5}$$

where $\Omega=(\mu,\phi)$ and $\Omega'=(\mu',\phi')$ (see Fig. 1) and $S_{j,m}(\mu,\mu')$ is the scattering indicatrix, then the Neumann expansion equation (2.2), using the orthogonality property of the phase factors $e^{im\phi}$, becomes

$$\mu\frac{\partial I_{n,m}(\tau,\mu)}{\partial\tau}+I_{n,m}(\tau,\mu)=\tfrac{1}{2}\sum_{j=0}^{N}\omega_j(\tau)\int_{-1}^{1}d\mu' S_{j,m}(\mu,\mu')I_{n-1,m}(\tau,\mu'), \tag{2.6}$$

where

$$I_{n,m}(\tau,\mu)=\int_0^{2\pi}d\phi\, e^{im\phi}I_n(\tau,\Omega) \tag{2.7}$$

so that

Copyright © 1973 by the American Institute of Physics

$$I_n(\tau,\Omega)=\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}e^{-im\phi}I_{n,m}(\tau,\mu). \tag{2.8}$$

We note that in all the equations given above the frequency of radiation appears as a parameter; this is due to the assumption that the scattering is coherent. Hence that dependence is not explicitly indicated.

Integral representation of $I_{n,m}(\tau,\mu)$ for semi-infinite atmospheres

Now following the procedure of Ref. 1, we consider the time reversed adjoint equation for a purely absorbing medium, i.e.,

$$-\mu\frac{\partial}{\partial\tau}G(\tau,-\mu;\tau_0,-\mu_0)+G(\tau,-\mu;\tau_0,-\mu_0)=\delta(\mu-\mu_0)\delta(\tau-\tau_0). \tag{2.9}$$

We note that this Green's function satisfies the reciprocity relation

$$G(\tau,-\mu;\tau_0,-\mu_0)=G(\tau_0,\mu_0;\tau,\mu). \tag{2.10}$$

We also note that $G(\tau,\mu;\tau_0,\mu_0)$ has the following Fourier integral representation:

$$G(\tau,\mu;\tau_0,\mu_0)=\frac{\delta(\mu-\mu_0)}{2\pi}\int_{-\infty}^{\infty}dK\,\frac{e^{iK(\tau-\tau_0}}{1+iK\mu}.$$

If we now multiply Eq. (2.6) by $G(\tau,-\mu;\tau_0,-\mu_0)$, Eq. (2.9) by $I_{n,m}(\tau,\mu)$, subtract one equation from the other and integrate with respect to μ from -1 to 1 with respect to τ from 0 to ∞ and use the reciprocity relation (2.10) and the Fourier integral representation of $G(\tau,\mu;\tau_0,\mu_0)$, we obtain the Fourier integral representation of $I_{n,m}(\tau,\mu)$ which is given below:

$$I_{n,m}(\tau,\mu)\Theta(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dK\,\frac{e^{iK\tau}}{1+iK\mu}\times\left[\mu I_{n,m}(0,\mu)+\tfrac{1}{2}\sum_{j=0}^{N}R^{(K;\mu)}_{n-1,j,m}\right], \tag{2.11}$$

where $\Theta(\tau)$ is the Heaviside step function and, in Eq. (2.11),

$$R_{n,j,m}(K;\mu)=\int_0^{\infty}d\tau\,e^{-iK\tau}\omega_j(\tau)\int_{-1}^{1}d\mu' S_{j,m}(\mu;\mu')I_{n,m}(\tau,\mu'). \tag{2.12}$$

We shall refer to the quantities $R_{n,j,m}$ defined by (2.12) as the "restricted Fourier amplitudes."

We note that once the restricted amplitudes $R_{n,j,m}(K;\mu)$ in Eq. (2.11) are determined (subject to the appropriate boundary condition at $\tau=0$), then the components $I_{n,m}(\tau,\mu)$ are also determined everywhere. It is clear from the same equation that in principle one can, for any value of n (and j), always write down $R_{n,j,m}(K;\mu)$ in terms of the known quantities. For instance, for $n=0$, we have

$$I_{0,m}(\tau,\mu)\Theta(\tau)=I_{0,m}(0,\mu)e^{-\tau/\mu}\Theta(\mu) \tag{2.13}$$

Since $I_{0,m}(0,\mu)$ for $\mu>0$ is known from the boundary condition at $\tau=0$, one may then merely use (2.13) to substitute for $I_{0,m}(\tau,\mu)$ in (2.12) and obtain $R_{0,j,m}(K;\mu)$ for all values of j. Physically, Eq. (2.13) describes the situation in which the radiation entering at $\tau=0$ has suffered no scattering in the atmosphere. In the next iteration, upon computing $R_{0,j,m}(K;\mu)$, one then determines $I_{1,m}(\tau,\mu)$ [subject to the boundary condition $I_{1,m}(0,\mu)=0$ for $\mu>0$] which corresponds to the component of the radiation which has been scattered once and so on. This procedure is obviously cumbersome and restricted in usefulness to media which are highly absorbing.

Furthermore, the task of obtaining $R_{n,j,m}(K;\mu)$ for arbitrary values of n (and j) is too much and unrevealing. For that reason we appeal to the procedure discussed in Ref. 1. Our point of view will be to reexpress Eq. (2.11) as a spectral integral[4] in which the unknown coefficients obey a relatively simple recurrence relation. We shall find that such coefficients are precisely the difference of the boundary values, about the branch cuts, of $R_{n,j,m}(K;\mu)$ in the complex K plane. Therefore, in order to obtain the appropriate spectral representation of $I_{n,m}(\tau,\mu)$, our first step would be to investigate the properties of the analytic continuation of $R_{n,j,m}(K;\mu)$ in the complex K plane. Said in another way we must first determine the spectrum of the iterated transport operator of Eq. (2.6). We remark here that, as pointed out by Case and Zweifel,[5] the main difficulty one encounters in treating the inhomogeneous atmosphere is in the determination of the spectrum of the transport operator. However, in several cases of interest, we shall show that it is possible to obtain the appropriate spectra.[6]

Singular properties of $R_{n,j,m}(K;\mu)$

Let us first introduce the quantities $T_{n,l,m}(K;\mu)$ defined by

$$T_{n,l,m}(K;\mu)=\int_0^{\infty}d\tau\,e^{-iK\tau}\int_{-1}^{1}d\mu' S_{l,m}(\mu,\mu')I_{n,m}(\tau,\mu'). \tag{2.14}$$

For reasons which will become clear soon, we shall refer to $T_{n,l,m}(K;\mu)$ as the "generators of $R_{n,l,m}(K;\mu)$". Now a relation between $T_{n,l,m}$ and $R_{n,l,m}$ is readily derived by taking the Fourier transform of Eq. (2.11) with respect to τ, then multiplying both sides by $S_{l,m}(\mu,\mu')$, and integrating with respect to μ. The result is

$$T_{n,l,m}(K;\mu)=\int_{-1}^{1}d\mu'\,\frac{S_{l,m}(\mu,\mu')}{1+iK\mu'}\times\left[\mu' I_{n,m}(0,\mu')-\tfrac{1}{2}\sum_{j=1}^{N}R_{n-1,j,m}(K;\mu')\right]. \tag{2.15}$$

Further, by comparing the definitions (2.12) and (2.14) of $R_{n,l,m}(K;\mu)$ and $T_{n,l,m}(K;\mu)$, respectively, we also conclude that

$$R_{n,l,m}(K;\mu)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dK'\,W_l(K'-K)T_{n,l,m}(K';\mu) \tag{2.16}$$

where

$$W_l(K'-K)=\int_0^{\infty}d\tau\,e^{i\tau(K'-K)}\omega_l(\tau). \tag{2.17}$$

It is because of the convolution integral relation (2.16) we wish to call $T_{n,l,m}(K;\mu)$ the generators of $R_{n,l,m}(K;\mu)$. Here we have assumed that the "correlation functions" $W_l(K'-K)$, defined by Eq. (2.17), exist. In fact Eq. (2.17) is actually an identity. The most important point, however, is that Eqs. (2.15) and (2.16) will determine the complete spectrum of the iterated transport operator of Eq. (2.2). We shall now examine that spectrum as far as it is possible.

First of all it is clear from Eq. (2.15) that the $T_{n,l,m}(K;\mu)$ are sectionally holomorphic in the complex K plane cut from $K=i$ to $i\infty$ and $K=-i\infty$ to $-i$. Further

$T_{n,l,m}(K;\mu)$ may have additional branch cuts and poles (possibly none) which will be determined by the correlation functions $W_l(K'-K)$ [see Eq. (2.17)]. Equation (2.16) will then determine the analyticity properties of $R_{n,l,m}(K;\mu)$. Thus, in order to determine the analytical structure of $T_{n,l,m}(K;\mu)$ and $R_{n,l,m}(K;\mu)$, it is important to know the corresponding properties of $W_l(K'-K)$. To illustrate these ideas, let us consider a simple example in which the index of degeneracy N of the phase function [see Eq. (2.4)] is zero and $\omega_0(\tau)$ is exponentially decaying.

3. APPLICATION ($N = 0$)

Exponential albedo

In the integral representation (2.11) for $I_{n,m}(\tau,\mu)$ we set $N = 0$ and let

$$\omega_0(\tau) = \omega e^{-\alpha\tau}, \tag{3.1}$$

where ω and α are nonnegative constants for a given frequency. Then, we have

$$I_{n,m}(\tau,\mu)\Theta(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dK\, \frac{e^{iK\tau}}{1+iK\mu}\left[\mu I_{n,m}(0,\mu) + \tfrac{1}{2} R_{n-1,m}(K;\mu)\right], \tag{3.2}$$

where now

$$R_{n,m}(K;\mu) = \omega\int_0^{\infty} d\tau\, e^{-\tau(\alpha+iK)}\int_{-1}^{1} d\mu'\, S_m(\mu,\mu')I_{n,m}(\tau,\mu'). \tag{3.3}$$

For the sake of convenience we have omitted the subscript $j = 0$ in $R_{n-1,j,m}$ and $S_{j,m}(\mu,\mu')$. Corresponding to functional relations (2.16) and (2.17),

$$T_{n,m}(K;\mu) = \int_{-1}^{1} d\mu'\, \frac{S_m(\mu,\mu')\mu' I_{n,m}(0,\mu')}{1+iK\mu'} - \tfrac{1}{2}\int_{-1}^{1} d\mu'\, \frac{S_m(\mu,\mu')R_{n-1,m}(K;\mu')}{1+iK\mu'} \tag{3.4}$$

and

$$R_{n,m}(K;\mu) = -\frac{\omega}{2\pi i}\int_{-\infty}^{\infty} dK'\, \frac{T_{n,m}(K';\mu)}{K'-(K-i\alpha)} \tag{3.5}$$

where, in Eq. (3.5), we have used the fact that

$$W_1(K'-K) = \int_0^{\infty} d\tau\, e^{i\tau(K'-K)}\omega_0(\tau) = -\frac{\omega}{i(K'-K+i\alpha)}. \tag{3.6}$$

Now, by substituting $T_{n,m}(K;\mu)$ in Eq. (3.5) by (3.4), we obtain

$$R_{n,m}(K;\mu) = -i\omega\int_0^1 d\mu'\, \frac{S_m(\mu,\mu')I_{n,m}(0,\mu')}{K-i(\alpha+1/\mu')} + \frac{\omega}{4\pi i}\int_{-1}^{1} d\mu'\, S_m(\mu,\mu')\int_{-\infty}^{\infty} dK'\, \frac{R_{n-1,m}(K';\mu')}{(1+iK'\mu')(K'-K+i\alpha)}. \tag{3.7}$$

In principle Eq. (3.7) will determine all the restricted amplitudes $R_{n,m}(K;\mu)$ in terms of known quantities. For instance, under the boundary condition

$$I_{n,m}(0,\mu) = \delta_{n0} I_{0,m}(0,\mu), \qquad \mu > 0, \tag{3.8}$$

where $I_{0,m}(0,\mu)$ is known, the first two amplitudes, for $n = 0$ and $n = 1$, are

$$R_{0,m}(K;\mu) = -i\omega\int_0^1 d\mu'\, \frac{S_m(\mu,\mu')I_{0,m}(0,\mu')}{K-i(\alpha+1/\mu')} \tag{3.9}$$

and

$$R_{1,m}(K;\mu) = \frac{i\omega^2}{2}\int_{-1}^{1} d\mu_1\int_0^1 d\mu_2\, S_m(\mu,\mu_1)S_m(\mu_1,\mu_2)I_{0,m}(0,\mu_2) \times\left[\frac{\Theta(\mu_1)}{\mu_1[K-i(\alpha+1/\mu_1)][K-i(2\alpha+1/\mu_2)]} - \frac{\Theta(-\mu_1)}{\mu_1[K-i(2\alpha+1/\mu_2)][\alpha+1/\mu_2-1/\mu_1]}\right]. \tag{3.10}$$

However, as mentioned previously, the resulting Fourier integral representation of $I_{n,m}(\tau,\mu)$, given by Eq. (3.2), is not suitable for the purposes of numerical computations. For that reason we shall examine the singular properties of the analytic continuation of $R_{n,m}(K;\mu)$ in the complex K plane. In particular, since the exponential term in the integrand of Eq. (3.2) is convergent in the upper half of the complex K plane, we shall analytically continue $R_{n,m}(K;\mu)$ in that part.

Singular properties of $R_{n,m}(K;\mu)$

(1) From Eq. (3.7) one may readily conclude that $R_{n,m}(K;\mu)$ has a sequence of branch cuts at $K = i[(j+1)\alpha + 1]$ to $i\infty, j = 0, 1, 2, \ldots, n$.

(2) $R_{n,m}(K;\mu)$ has no poles for any value of n.

(3) For real values of K and any value of n, $R_{n,m}(K;\mu)$ is regular.

(4) From property (1) and Eq. (3.4) we conclude that $T_{n,m}(K;\mu)$ has a sequence of branch cuts at $k = i(j\alpha + 1)$ to $i\infty, j = 0, 1, 2, \ldots, n$.

(5) From properties (1) and (4) given above and the relation (3.5) between $R_{n,m}(K;\mu)$ one may readily show that the difference of boundary values across their respective branch cuts bear the following relationship to each other:

$$R^{+}_{n,m}(K;\mu) - R^{-}_{n,m}(K;\mu) = \omega[T^{+}_{n,m}(K-i\alpha;\mu) - T^{-}_{n,m}(K-i\alpha;\mu)] \tag{3.11}$$

where K takes any value on the ray $K = i(\alpha+1)$ to $i\infty$ in the complex K plane and + (−) represents the boundary value as K approaches the branch cut (the ray) from the left (right). In particular if we shift the arguments in Eq. (3.11) by replacing K by $K - ij\alpha$, then

$$R^{+}_{n,m}(K-ij\alpha;\mu) - R^{-}_{n,m}(K-ij\alpha;\mu) = \omega[T^{+}_{n,m}(K-i(j+1)\alpha;\mu) - T^{-}_{n,m}(K-i(j+1)\alpha;\mu)] \tag{3.12}$$

where now, for fixed value of j, K takes any value on the ray $K = i[(j+1)\alpha + 1]$ to $i\infty$. In the complex $z = (i/K)$ plane the identity (3.12) becomes

$$R^{+}_{n,m}\left(\frac{\nu}{1-\nu j\alpha};\mu\right) - R^{-}_{n,m}\left(\frac{\nu}{1-\nu j\alpha};\mu\right) = \omega\left[T^{+}_{n,m}\left(\frac{\nu}{1-\nu(j+1)\alpha};\mu\right) - T^{-}_{n,m}\left(\frac{\nu}{1-\nu(j+1)\alpha};\mu\right)\right] \tag{3.13}$$

for $0 \le \nu \le 1/[1 + \alpha(j + 1)]$, where + (—) represents the boundary value as z approaches the branch cut (the ray $0 \le \nu \le 1/[1 + \alpha(j + 1)]$ from the top (bottom). In other words, on the semi-infinite positive real axis in the complex z plane, $R_{n,m}(1/(1 - \nu\alpha j); \mu)$ and $T_{n,m}(1/[1 - \alpha(j + 1)]; \mu)$ are regular for $\nu > 1/[1 + \alpha(j + 1)]$.

(6) Finally from the analytic continuation of Eq. (3.4) in the upper half of the complex K plane, or more conveniently in the corresponding part of the $z(= i/K)$ plane, one obtains the following relation between the boundary values of $R_{n,m}(z; \mu)$ and $T_{n,m}(z; \mu)$:

$$R^{+}_{n-1,m}(\nu; \nu) + R^{-}_{n-1,m}(\nu; \nu) = -4\nu I_{n,m}(0, \nu) - \frac{2}{\pi i \nu} \frac{T^{+}_{n,m}(\nu; \mu) - T^{-}_{n,m}(\nu; \mu)}{S_m(\nu, \mu)} + \frac{1}{i\pi} \mathcal{P} \int_{-1}^{1} \frac{d\mu'}{\nu - \mu'} \frac{S_m(\mu'; \mu)}{S_m(\nu, \mu)} [R^{+}_{n-1,m}(\nu; \mu') - R^{-}_{n-1,m}(\nu; \mu')], \tag{3.14}$$

where $\mathcal{P}$ represents the principle value and + (—) represents the boundary value from the top (bottom) in the complex $z(= i/K)$ plane cut from 0 to 1.

It is worthwhile to recall here that, in the complex z plane, $R_{n,m}(z; \mu)$, for a given value of n, has a sequence of branch cuts at $z = 0$ to $1/[(j + 1)\alpha + 1]$, $j = 0, 1, \ldots, n$ [see property (1) above] while $T_{n,m}(z; \mu)$ for a fixed n has a sequence of cuts at $z = 0$ to $1/(j\alpha + 1)$, $j = 0, 1, \ldots, n$ [see property (5) above]. In consequence $R_{n,m}(z; \mu)$ are regular for $z > 1/(1 + \alpha)$.

We shall now obtain the spectral representation of $I_{n,m}(\tau, \mu)$, the procedure for which is exactly parallel to the one discussed in Ref. 1.

Spectral representation of $I_{n,m}(\tau, \mu)$

Let us consider the Fourier integral representation (3.2) of $I_{n,m}(\tau, \mu)$. We note that the integrand in the second term on the right hand side has a pole at $K = i/\mu$ and, by property (1) of $R_{n-1,m}(K; \mu)$, a sequence of branch cuts in the upper half of the complex K-plane starting at $K = i(\alpha + 1)$ to $i\infty$. Since μ can vary continuously between -1 and $+1$, we distort the path of integration, as shown in Fig. 2, surrounding the cut starting at $K = i$, and obtain

$$I_{n,m}(\tau, \mu)\Theta(\tau) = I_{n,m}(0, \mu) e^{-\tau/\mu}\Theta(\mu) - \frac{1}{4\pi i} \mathcal{P} \int_0^{1/(1+\alpha)} \frac{d\nu}{\nu^2} e^{-\tau/\nu} \frac{\nu}{\nu - \mu} [R^{+}_{n-1,m}(\nu; \mu) - R^{-}_{n-1,m}(\nu; \mu)] + \frac{1}{4\mu} e^{-\tau/\mu}\Theta(\mu)[R^{+}_{n-1,m}(\mu; \mu) + R^{-}_{n-1,m}(\mu; \mu)], \tag{3.15}$$

where we have set $K = i/z$ and noted the fact that $R_{n-1,m}(z; \mu)$ is regular for $z > 1/(1 + \alpha)$, hence the upper limit of $1/(1 + \alpha)$ instead of unity. We may eliminate the last term on the right-hand side of (3.15), which involves the sum of boundary values of $R_{n-1,m}(z; \mu)$, by means of the boundary relation (3.14). Further use of Eq. (3.11) then yields

$$I_{n,m}(\tau, \mu)S_m(\mu, \nu)\Theta(\tau) = \Gamma_{n,m}(\mu; \nu)e^{-\tau/\mu}\Theta(\mu) + \omega S_m(\mu, \nu) \int_0^{1/(1+\alpha)} d\nu' \phi_{\nu'}(\mu) e^{-\tau/\nu'} \times \Gamma_{n-1,m}\left(\frac{\nu'}{1 - \alpha\nu'}; \mu\right) \Big/ (1 - \alpha\nu')^2$$

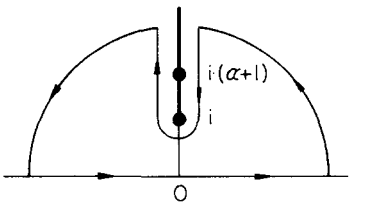

FIG. 2. Contour in the complex K plane.

$$- \omega e^{-\tau/\mu}\Theta(\mu) \int_{-1}^{1} d\nu' \phi_{\mu}(\nu') S_m(\nu'; \nu) \times \Gamma_{n-1,m}\left(\frac{\mu}{1 - \alpha\mu}; \nu'\right) \Big/ (1 - \alpha\mu)^2, \tag{3.16}$$

where

$$\Gamma_{n,m}(\mu; \nu) \equiv -(1/2\pi i\mu^2)[T^{+}_{n,m}(\mu; \nu) - T^{-}_{n,m}(\mu; \nu)] \tag{3.17}$$

and we recall by virtue of property (5) given above that

$$\Gamma_{n,m}(\mu/[1 - \alpha(j + 1)\mu]; \nu) = 0 \quad \text{for} \quad \mu > 1/[1 + \alpha(j + 1)], \quad j = 0, 1, \cdots. \tag{3.18}$$

Also the functions $\phi_{\nu}(\mu)$ are defined by

$$\phi_{\nu}(\mu) = \frac{\nu}{2} \mathcal{P} \frac{1}{\nu - \mu} + \frac{\Lambda^{+}(\nu) + \Lambda^{-}(\nu)}{2} \delta(\nu - \mu) \tag{3.19}$$

with

$$\Lambda(z) = 1 - \frac{z}{2} \int_{-1}^{1} \frac{d\mu}{z - \mu}. \tag{3.20}$$

We remark here that $\phi_{\nu}(\mu)$, given by (3.19), are the familiar Case's continuum normal modes for the medium which is homogeneous and isotropically scattering.

Equation (3.16) is what we previously referred to as the spectral representation of $I_{n,m}(\tau, \mu)$. The unknown coefficients $T_{nm}(\mu; \nu)$ may be determined from the boundary condition at $\tau = 0$. That equation, under the arbitrary boundary condition

$$I_{n,m}(0, \mu) = \delta_{n0} I_{0,m}(0, \mu), \quad \mu > 0, \tag{3.21}$$

is obtained by letting $\tau \to 0$ in Eq. (3.16), i.e.,

$$\Gamma_{n,m}(\mu; \nu) = \delta_{n0} I_{0,m}(0, \mu) S_m(\mu, \nu) + \omega \int_{-1}^{1} d\nu' \phi_{\mu}(\nu') S_m(\nu', \nu) \times \Gamma_{n-1,m}\left(\frac{\mu}{1 - \alpha\mu}; \nu'\right) \Big/ (1 - \alpha\mu)^2 - \omega S_m(\mu, \nu) \times \int_0^{1/(1+\alpha)} d\nu' \phi_{\mu'}(\mu) \times \Gamma_{n-1,m}\left(\frac{\nu'}{1 - \alpha\nu'}; \mu\right) \Big/ (1 - \alpha\nu')^2, \quad \mu > 0. \tag{3.22}$$

This is an integro-recurrence relation for the unknown coefficients $\Gamma_{n,m}(\mu, \nu)$ occurring in Eq. (3.16). If we define the quantity

$$\Psi^{(m)}_{\nu_1}(\nu_0; \nu) \equiv \phi_{\nu_1}(\nu_0) S_m(\nu_0, \nu), \tag{3.23}$$

then we may rewrite Eq. (3.22) as

$$\Gamma_{n,m}(\mu; \nu) = \omega^n \int_0^{1/(1+\alpha)} \frac{d\nu'}{(1 - \alpha\nu')^2} I_{0,m}\left(0, \frac{\nu'}{1 - \alpha\nu'}\right) \times B_{n,m}(\nu'; \mu; \nu) \tag{3.24}$$

where the $B_{n,m}(\nu';\mu;\nu)$ are independent of the boundary condition at $\tau = 0$ and are determined by

$$B_{n+1,m}(\nu_1;\mu;\nu) = \int_{-1}^{1} d\nu''\ \Psi_{\mu}^{(m)}(\nu'';\nu) \times B_{n,m}\left(\nu_1;\frac{\mu}{1-\alpha\mu};\nu''\right)\Big/(1-\alpha\mu)^2 - \int_0^{1/(1+\alpha)} \frac{d\nu'}{(1-\alpha\nu')^2}\Psi_{\nu'}^{(m)}(\mu;\nu)\,B_{n,m}\left(\nu_1;\frac{\nu'}{1-\alpha\nu'};\mu\right), \quad n = 0,1,\cdots, \tag{3.25}$$

with

$$B_{0,m}(\nu_1;\mu;\nu) = \delta(\nu_1 - \mu/(1+\alpha\mu))S_m(\mu;\nu). \tag{3.26}$$

We shall refer to the quantities $B_{n,m}$ as the singular cluster integrals. Once these cluster integrals are determined from the integro-recurrence relation (3.25), the unknown coefficients $\Gamma_{n,m}$ occurring in Eq. (3.16) are determined by Eq. (3.24) subject to the appropriate boundary condition. The total specific intensity is then given by [see Eqs. (3.16), (2.8) and (2.3)]

$$I(\tau,\Omega) = \frac{1}{2\pi}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty} e^{-im\phi}\left[e^{-\tau/\mu}\Theta(\mu)\frac{\Gamma_{n,m}(\mu;\nu)}{S_m(\mu,\nu)} + \omega\int_0^{1/(1+\alpha)} d\nu'\phi_{\nu'}(\mu)e^{-\tau/\nu'}\Gamma_{n-1,m}\left(\frac{\nu'}{1-\alpha\nu'};\mu\right)\Big/(1-\alpha\nu')^2 - \omega e^{-\tau/\mu}\Theta(\mu)\int_{-1}^{1} d\nu'\,\phi_{\mu}(\nu')\frac{S_m(\nu',\nu)}{S_m(\mu,\nu)} \times \Gamma_{n-1,m}\left(\frac{\mu}{1-\alpha\mu};\nu'\right)\Big/(1-\alpha\mu)^2\right]. \tag{3.27}$$

In particular, the reflected intensity ($\mu < 0$) at $\tau = 0$ is

$$I(0,\Omega) = \omega\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty} e^{-im\phi}\int_0^{1/(1+\alpha)}\frac{d\nu'}{(1-\alpha\nu')^2} \times \phi_{\nu'}(\mu)\Gamma_{n,m}\left(\frac{\nu'}{1-\alpha\nu'};\mu\right), \quad \mu < 0. \tag{3.28}$$

For the albedo problem the boundary condition is

$$I_{0,m}(0,\mu) = \delta(\mu-\mu_0)e^{im\phi_0}, \tag{3.29}$$

where $(\mu_0,\phi_0) \equiv \Omega_0$ represents the incident angle. Using (3.29) in (3.24), we get

$$\Gamma_{n,m}(\mu;\nu) = \omega^n B_{n,m}(\mu_0/(1+\alpha\mu_0);\mu;\nu)e^{im\phi_0}. \tag{3.30}$$

The emergent intensity is then given by

$$I(0,\Omega) = \sum_{n=0}^{\infty}\omega^{n+1}\sum_{m=-\infty}^{\infty} e^{im(\phi-\phi_0)}\int_0^{1/(1+\alpha)}\frac{d\nu'}{(1-\alpha\nu')^2} \times \phi_{\nu'}(\mu)B_{n,m}\left(\frac{\mu_0}{1+\alpha\mu_0};\mu;\nu\right), \quad \mu < 0. \tag{3.31}$$

We restate here that the singular cluster integrals $B_{n,m}$ are independent of ω and Eq. (3.31) represents the Neumann expansion for inhomogeneous atmospheres in which the albedo for single scattering decays exponentially. We also note that the $B_{n,m}$ in the same equation depend on α, which characterizes the absorption properties of the medium under consideration.

ACKNOWLEDGMENTS

The author wishes to thank Professor William Irvine for many enlightening discussions and useful suggestions. This work was supported in part by NASA Grant NGL 22-010-023 and NSF Grant GP 22742.

[1] M. Kanal, J. Math. Phys. **12**, 2187 (1971).
[2] H. C. Van de Hulst, Astrophys. J. **107**, 220 (1948).
[3] W. M. Irvine, Astrophys. J. **152**, 823 (1968).
[4] By spectral integral here we mean the integral representation in which the integral is over the spectrum of the transport operator [here the operator corresponding to Eq. (2.6)].
[5] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Massachusetts, 1967).
[6] P. F. Zweifel, private communication. In order to not mislead the reader, one must point out that the resolution of the identity for the transport operator has not been proved. The method used in this paper essentially corresponds to the determination of the resolution of the identity corresponding to a maximal symmetric operator. The spectrum is then obtained from the points of nonconstancy of the spectral family. Thus, the procedure bears an analogy with the one described by Riesz and Sz. Nagy in their book on *Functional Analysis* (Ungar, New York, 1965).

Local observables in the Dirac theory

David Hestenes

Physics Department, Arizona State University, Tempe, Arizona 85281
(Received 3 January 1973)

By a new method, the Dirac electron theory is completely reexpressed as a set of conservation laws and constitutive relations for local observables, describing the local distribution and flow of mechanical quantities. The coupling of the electromagnetic field to the electron is shown to be determined by the definitions of the observables rather than by the Dirac equation. Planck's constant appears in the equations only in connection with the electron spin. The equations are most readily interpreted by assuming that the electron is a structureless point charge, the spin and magnetic moment arising from the dynamics of electron motion.

INTRODUCTION

Most experimental tests of the Dirac electron theory pertain only to "global observables" such as average energy, average momentum, and average angular momentum. The theory actually contains much more detailed implications about "local observables" which describe a spacial distribution and flow of charge, mass, energy-momentum, and angular momentum. The local observables of the Dirac theory have some peculiar properties of which Dirac was undoubtedly unaware when the theory was first formulated. These peculiarities deserve careful study; first, because they can be expected to lead to particularly sensitive tests of the theory if they can be subjected to experimental scrutiny; and second, because they furnish valuable theoretical clues about the interpretation and significance of the theory.

A systematic analysis of the properties of local observables happens to be surprisingly difficult by conventional methods. This paper uses an unconventional mathematical formalism developed in Refs. 1 and 2 to achieve a compact reformulation of the complete Dirac theory in terms of local observables. The new method greatly simplifies the derivation of conservation laws and "relations" among local observables. The results obtained are *complete* in the sense that the relations found among the observables determine the theory uniquely. It is believed that *all* significant relations among local observables found by previous authors are derived here, though they usually appear in quite different form and often as only part of a more general relation. Because of the considerable difference in method and viewpoint, comparison of results is sometimes quite tedious, though it can always be carried out by the method explained in Appendix A. Since the work of Takabaysi[3] is much more extensive than anything else in the literature on local observables, special effort was made to compare it with the present work; exact agreement has been found even on comparing some of the more complicated and esoteric formulas, though the two approaches have not been compared in every detail.

The difficult problem of subjecting the theory of local observables to experimental test will not be broached in this paper, though it is hoped that the present formulation of the theory will facilitate the task. What the theory needs most is an experimental test of the peculiar noncollinearity of velocity and momentum predicted by eq. (5. 8) below. The problem of devising such a test has engaged the attention of Costa de Beauregard for many years; he has reviewed the problem in Ref. 4, including some significant progress that has recently been made. Reasoning by analogy with the Dirac theory, Costa de Beauregard was lead to suggest tests for asymmetry in the free space electromagnetic energy-momentum tensor for circularly polarized light. This asymmetry implies that the energy flux in a polarized light beam is not collinear with the momentum density, the difference being due to the "photon spin." In a clever experiment, Imbert[5] showed that this difference manifests itself as a lateral displacement of a reflected light beam with a magnitude and direction in complete agreement with theoretical predictions. This success makes it all the more likely that the analogous asymmetry in the Dirac theory can somehow be tested experimentally.

It should be emphasized that the asymmetry just alluded to does not reveal itself in the Dirac theory when observables are defined in terms of operators in the usual manner. The operator definitions refer only to global observables, the local features being suppressed by averaging (i.e., integrating over all space). The local theory is more detailed, defining mechanical quantities such as momentum and angular momentum densities as definite functions of position and time. One local observable, the probability density determined by the Dirac wave function, is already widely used to predict the electron charge distribution in an atom. The local distributions of other mechanical quantities have not as yet been associated with any experimentally accessible effects.

To achieve a complete theory of local observables it is necessary to go beyond the original assumptions of Dirac. It is important to realize that the definitions of observables in Dirac theory have far-reaching consequences quite independent of the exact form of the Dirac equation. The specification of observables in Diracs original paper [6] was incomplete in several respects. His crucial assumption about observables was made by adopting, without comment, the operator definition of energy already used in Schrödinger's theory. This was all he needed to predict the energy levels of stationary atomic states. Dirac's initial assumptions were not sufficient to prove local conservation of probability. But this deficiency was soon rectified[7] by defining a 'probability current', which, in accordance with the Dirac equation, has vanishing divergence and so describes a locally conserved quantity. This definition is actually a new physical assumption, as the current specifies a local flow (or at least, a "probable" or "local average" flow) of electron charge and mass, and tacitly attributes a constant ratio of charge to mass density for the "smeared out" electron.

After the local distribution and flow of electron charge was specified by defining a conserved current, the local distribution and flow of energy-momentum was specified by Tetrode's definition[8] of an energy-momentum tensor. It is remarkable that these definitions above completely determine a local distribution and flow of angular momentum, since in general continuum theory

Copyright © 1973 by the American Institute of Physics

the angular momentum tensor is not determined by the energy–momentum tensor. Thus no further assumptions are needed to determine a complete set of local observables for the Dirac theory.

Once a complete set of local observables has been defined, the wave function can be eliminated and the Dirac theory completely reformulated as a set of equations for local observables. Reformulation proved to be a surprisingly difficult task. Many authors achieved partial results. But it was not until 1957 that Takabayasi[3] developed a systematic approach and carried the reformulation to completion. Unfortunately, physicists have derived little benefit from his work, most probably because of its great complexity. However, much of this complexity is unnecessary, because it arises from redundancy in the admixture of matrix and tensor algebras in the mathematical formalism usually used to express the Dirac theory. Such redundancy is eliminated in this paper by employing the "space–time algebra" developed in Refs. 1 and 2.

Because of the much simpler method employed here, it has proved possible to work out the relations among local observables in considerably more detail than was done previously, especially in Sec. 6 below. It should be mentioned that the many relations among local observables derived below are not mutually independent. No attempt has been made to select one particular complete set of relations as more fundamental than another, because the desirability of any particular selection is determined by the use to which it will be put. However, a careful distinction has been made between those relations which are determined by the Dirac equation and those which are not.

The particular properties of the energy–momentum tensor which are determined by the Dirac equation are ascertained in Sec. 2. However, the main aim of Secs. 1 through 4 is to determine those properties of local observables which follow from the definitions of the probability current by (1.3) and (1.4) and the energy–momentum tensor (2.3) as specific functions of the wave function (1.1), without appeal to the Dirac equation. Since local conservation is essential to the notion of probability density, the eq. (1.4) which expresses local conservation of probability is taken as part of the definition of the probability current even though it can be derived from the Dirac equation (e.g., see Appendix B). Of course, the mere fact that the Dirac equation implies that the "current" (1.3) has vanishing divergence does not entail that that particular quantity describes the local flow of probability. It seems best, then, to say merely that the Dirac equation is consistent with the identification of (1.3) as probability current.

In Sec. 5 the Dirac equation is completely reformulated in terms of local observables. It will be noted that the electromagnetic potential does not appear in the resulting eq. (5.7). Therefore, the Dirac equation by itself implies nothing about electromagnetic interactions of the electron. Moreover, the reformulated Dirac equation does not yield equations of motion for local observables. Rather, it functions as a "constitutive equation", expressing, as shown by (5.8), the local momentum as a definite function of the local velocity and spin. The equations of motion for the local observables are just the general conservation laws ascertained in Secs. 1 and 2. But these equations are underdetermined and cannot be solved without the additional "constitutive relations" provided by Dirac equation. Since these "constitutive relations" do not involve the electromagnetic field, it can be concluded that the coupling of the electron to the electromagnetic field is already completely determined by the identification of (2.26) as the equation for energy–momentum conservation.

The relation of the spin density to the local circulation of charge in the Dirac theory is studied in Secs. 6 and 7. Section 8 discusses the bearing of local observables on the interpretation of quantum mechanics.

1. WAVE FUNCTION

This paper continues the reformulation and analysis of Dirac's theory begun in Ref. 1. For the most part definitions, conventions, and results appearing in Ref. 1 are adopted here with little or no comment. The relation to the usual formulation of Dirac theory is discussed in Appendix A.

In Ref. 1 it was established that the Dirac "wave function" $\psi = \psi(x)$ can be written in the canonical form

$$\psi = \rho^{1/2} e^{i(1/2)\beta} R. \tag{1.1}$$

The wave function does not have a direct physical significance, and a crucial part of the Dirac theory is to relate ψ to observable quantities. The canonical decomposition (1.1) greatly facilitates this task and, in addition, makes the geometric content of the theory explicit.

The spinor $R = R(x)$ describes a Lorentz rotation at every point x of space–time which takes an orthonormal frame of constant vectors γ_α into the frame $e_\alpha = e_\alpha(x)$ according to the equation

$$e_\alpha = R\gamma_\alpha \tilde{R}, \quad (\alpha = 0, 1, 2, 3). \tag{1.2}$$

The vectors e_0 and e_3 have direct physical interpretations in the Dirac theory. To emphasize its interpretation the vector $v \equiv e_0 = v(x)$ is called the *(local) particle velocity* (at x). In agreement with established parlance, the quantity

$$\psi\gamma_0\tilde{\psi} = \rho R\gamma_0\tilde{R} = \rho e = \rho v \tag{1.3}$$

is called the *probability current* and ρ is called the *proper probability density*. Local conservation of probability is expressed by the equation

$$\square \cdot (\rho v) = 0. \tag{1.4}$$

With the identification of $m\rho$ as *proper mass density*, (1.4) expresses local conservation of mass. Equation (1.4) is a consequence of the Dirac equation, but it must be emphasized that the physical interpretation given to ρv is an independent postulate of the Dirac theory.

The vector

$$s \equiv \frac{\hbar}{2} e_3 = s(x) \tag{1.5}$$

is called the (local) *spin (vector)*, and the bivector

$$S \equiv isv = \frac{\hbar}{2} ie_3e_0 = \frac{\hbar}{2} e_2e_1 = \frac{\hbar}{2} R\gamma_2\gamma_1\tilde{R} \tag{1.6}$$

is called the (local) spin (bivector). Either s or S can be used to describe the spin of the electron; S is preferable because angular momentum is fundamentally a bivector quantity, but s has the advantage that vectors are sometimes easier to manipulate than bivectors. At any rate, (1.6) makes it easy to switch from one to the other. The physical interpretation of S as angular momentum need not be introduced into the Dirac theory as an inde-

pendent postulate; it follows from the definition of the energy–momentum tensor given in the next section.

The e_α describe intrinsic properties of the electron and so are independent of any coordinate system associated with an observer. In contrast, the choice of the γ_α in (1. 2) is a mere convention, though once that choice is made, the R that gives the e_α is uniquely determined. The choice of the γ_α is disguised in the usual matrix formulation of the Dirac theory as a choice of matrix representation. A change in the matrix representation is equivalent to a change in the choice of the γ_α. The γ_α can always be related to a set of inertial (Cartesian) coordinates $\{x^\alpha\}$ by the equations $\gamma_\alpha = \partial_\alpha x$ or $\gamma^\alpha = \Box x^\alpha$.

Equations (1. 2) can be solved for R (see Sec. 17 of Ref. 2). One obtains

$$R = (\tilde{A}A)^{-1/2}A, \tag{1. 7}$$

where

$$A = e_\alpha \gamma^\alpha. \tag{1. 8}$$

This is one way of exhibiting the dependence of R on the e_α, but no further use of it will be made in this paper. The important point is that once the convenient arbitrary choice of the γ_μ has been made, then the Lorentz rotation R is determined by eight (scalar) parameters. Five of these are determined by specifying the velocity and the spin directions. This much determines the plane of e_1 and e_2; the additional parameter χ, which is needed to fix the directions of e_1 and e_2 in the plane, is the *phase* of the Dirac wave function.

The Dirac wave function ψ is completely determined by specifying the set of eight independent quantities:

$$\{\rho, \beta, v, S \text{ or } s, \chi\} \tag{1. 9}$$

Of these, ρ, v, S have straightforward physical interpretations. Comments on the interpretation of β will be deferred until the role of β in the Dirac theory has been studied in some detail. The observable significance of χ is indirect, and it will not be convenient to make explicit use of χ in equations for local observables. Rather, the energy–momentum vector, to be introduced later, will contain an implicit dependence on the gradient of χ.

2. ENERGY-MOMENTUM TENSOR

One of the key assumptions in Dirac's initial paper 6 is that the total energy E of the electron in a stationary state is to be obtained from the equation

$$(-1)^{1/2}\hbar\partial_t\Psi = E\Psi, \tag{2. 1}$$

where Ψ is a column spinor. Superficially, (2. 1) appears to be identical to an assumption made by Schrödinger in his first papers on quantum theory. But something new is present because of the spinor character of Ψ. This becomes apparent when, by using the spinor Ψ, (2. 1) is reexpressed in a form which is independent of any matrix representation:

$$\partial_t\psi\gamma_2\gamma_1\hbar = E\psi. \tag{2. 2}$$

The equivalence of (2. 2) and (2. 1) is easily established by the method of Appendix A. Comparison of (2. 1) and (2. 2) then reveals that the root of minus one on the left of (2. 1) has a geometrical significance; it may be regarded as a representation of the spacelike bivector $\gamma_2\gamma_1 = i\gamma_3\gamma_0 = i\sigma_3$, which itself is a particular root of minus one since $(\gamma_2\gamma_1)^2 = -1$. Moreover, this root of minus one has a physical significance, for according to (1. 6) it determines the direction of the electron spin. In fact, it is only through (2. 1) or (2. 2) and its generalizations that electron spin finds its way into the Dirac theory.

To complete the formulation of the Dirac theory, (2. 2) must be generalized to give an expression for the energy even when the electron is not in a stationary state, if only because the wavefunction cannot be an eigenfunction of the energy in all inertial systems unless it is a plane wave. Besides, (2. 2) determines only the energy density, which is but one component of a complete energy–momentum tensor.

The most straightforward guess at the required generalization of (2. 2) is given by

$$\begin{aligned} T_{\mu\nu} &\equiv \{\gamma_0\tilde{\psi}\gamma_\mu(\partial_\nu\psi\gamma_2\gamma_1\hbar - eA_\nu\psi)\}_S \\ &= \hbar(\gamma_\mu\partial_\nu\psi i\gamma_3\tilde{\psi})_S - e\rho v_\mu A_\nu. \end{aligned} \tag{2. 3}$$

The second term on the right has been added to remove the contribution of the electromagnetic field to the electron's energy–momentum and so (hopefully) to produce a tensor that describes only intrinsic properties of the electron. To get the last line of (2. 3), the definition of velocity (1. 3) has been used.

That (2. 3) is a reasonable generalization of (2. 2) can be shown by computing the average energy in "the inertial system of γ_0", namely

$$\begin{aligned} \langle E\rangle &= \langle E(t)\rangle = \int d^3x(T_{00} + e\rho v_0 A_0) \\ &= \int d^3x(\gamma_0\tilde{\psi}\gamma_0\partial_0\psi\gamma_2\gamma_1\hbar)_S, \end{aligned} \tag{2. 4}$$

where the integral is taken over the spacelike hypersurface of points x satisfying the equation $x\cdot\gamma_0 = x_0 = t$. If (2. 2) is satisfied, then

$$\langle E\rangle = \int d^3xE(\gamma_0\tilde{\psi}\gamma_0\psi)_S = E\int d^3x\rho_0 = E, \tag{2. 5}$$

where

$$\rho_0 \equiv \rho v_0 = \rho v\cdot\gamma_0 = (\psi\gamma_0\tilde{\psi}\gamma_0)_S \tag{2. 6}$$

is the particle probability density in the inertial system of γ_0. It is very important to note the appearance of v in (2. 5) and (2. 3), for it shows that consistency of (2. 3) with (2. 2) requires the conservation law (1. 4) and the interpretation previously given to v.

Appendix A shows that (2. 3) is equivalent to Tetrode's definition of the electron's energy–momentum tensor. Though this tensor has been much studied with matrix methods, something may yet be learned by analyzing its properties with the methods of space–time algebra.

An energy–momentum tensor is a linear vector function of a vector variable. To be more specific, let $T(n, x)$ denote a flux of energy–momentum through a hypersurface with normal n at the space–time point x. Suppressing the x dependence, one says that $T(n)$ is an energy–momentum tensor. Since T is a linear function of $n = n^\mu\gamma_\mu$,

$$T(n) = n^\mu T(\gamma_\mu) = n^\mu T_\mu. \tag{2. 7}$$

So an energy–momentum tensor is completely characterized by the vectors $T_\mu \equiv T(\gamma_\mu)$, which specify the flux in four independent directions. The components of the T_μ are

$$T_{\mu\nu} \equiv T_\mu\cdot\gamma_\nu. \tag{2. 8a}$$

Inversely,

$$T_\mu = T_{\mu\nu}\gamma^\nu. \tag{2.8b}$$

With this, the T_μ for the Tetrode tensor can be determined from (2.3).

The conservation law (1.4) implies the existence of *velocity streamlines*, timelike trajectories with tangents $v = v(x)$ describing the local flow of the "probability fluid." The *proper energy–momentum density* given by

$$\rho p = T(v) = v^\mu T_\mu \tag{2.9}$$

describes the flow of energy–momentum along the velocity streamline. The vector p deserves a special name, because it is one of the most fundamental quantities of the Dirac theory. The name *local energy–momentum* or simply *momentum* is appropriate. Now T_μ can be decomposed in the form

$$T_\mu = \rho v_\mu p + N_\mu. \tag{2.10}$$

Because of (2.9), $N(v) = v^\mu N_\mu = 0$, so the N_μ describe the flow of energy momentum normal to the velocity streamline.

Certain special properties of the Tetrode tensor are determined by the Dirac equation. These properties are most easily ascertained by studying the "transposed" tensor defined from (2.3) by

$$\begin{aligned}\overline{T}_\mu \equiv \gamma^\nu T_{\mu\nu} &= \gamma^\nu\{\hbar\,\gamma_\nu\cdot(\partial_\mu\psi i\gamma_3\tilde{\psi})_V - e\rho v_\nu A_\mu\}\\ &= \hbar\,(\partial_\mu\psi i\gamma_3\tilde{\psi})_V - e\rho v A_\mu.\end{aligned} \tag{2.11}$$

First observe that

$$\hbar\,(\partial_\mu\psi i\gamma_3\tilde{\psi})_V = \frac{\hbar}{2}\,(\partial_\mu\psi i\gamma_3\tilde{\psi} - \psi i\gamma_3\partial_\mu\tilde{\psi}), \tag{2.12}$$

Also, note that the definition of the spin vector (1.5) implies $\frac{1}{2}\hbar\psi\gamma_3\tilde{\psi} = \rho s$, so

$$\partial_\mu(i\rho s) = \frac{\hbar}{2}(\partial_\mu\psi i\gamma_3\tilde{\psi} + \psi i\gamma_3\partial_\mu\tilde{\psi}). \tag{2.13}$$

Hence

$$\hbar\,(\Box\psi)i\gamma_3\tilde{\psi} = \hbar\,\gamma^\mu(\partial_\mu\psi i\gamma_3\tilde{\psi})_V + \Box(i\rho s). \tag{2.14}$$

Next, recall the form of Dirac equation given in Ref. 5. With a conveninet change in sign convention and explicit introduction of Planck's constant, it can be written

$$\hbar\Box\psi i\gamma_3\gamma_0 = m\psi\gamma_0 + eA\psi. \tag{2.15}$$

Multiply this on the right by $\gamma_0\tilde{\psi}$ and use $\psi\tilde{\psi} = \rho e^{i\beta}$ as well as (1.3) to get

$$\hbar\Box\psi i\gamma_3\tilde{\psi} = m\rho e^{i\beta} + eA\rho v. \tag{2.16}$$

Finally, combine (2.9), (2.11), (2.14), and (2.16) to obtain

$$\gamma^\mu\,\overline{T}_\mu = T_\mu\,\gamma^\mu = m\rho e^{i\beta} + i\Box\rho s. \tag{2.17}$$

The pseudoscalar part of (2.17) yields

$$\Box\cdot(\rho s) = -m\rho\,\sin\beta. \tag{2.18}$$

This says nothing, at least directly, about the Tetrode tensor. But the scalar part of (2.17) gives the trace of the Tetrode tensor:

$$T^\mu_\mu = T_\mu\cdot\gamma^\mu = m\rho\,\cos\beta, \tag{2.19}$$

and the bivector part of (2.17) is

$$\gamma^\mu\wedge\overline{T}_\mu = T_\mu\wedge\gamma^\mu = i(\Box\wedge\rho s) = -\Box\cdot(i\rho s). \tag{2.20}$$

This specifies the antisymmetric part of the Tetrode tensor, as can be seen by expressing it in component form:

$$\begin{aligned}(\gamma_\mu\wedge\gamma_\nu)\cdot(T_\beta\wedge\gamma^\beta) = (\gamma_\mu\wedge\gamma_\nu)\cdot(\gamma^\alpha\wedge\gamma^\rho)\,T_{\rho\alpha} &= T_{\mu\nu} - T_{\nu\mu}\\ &= i\gamma_\mu\wedge\gamma_\nu\wedge\Box\wedge(\rho s) = -\epsilon_{\mu\nu\alpha\beta}\partial^\alpha(\rho s^\beta),\end{aligned} \tag{2.21}$$

where $\epsilon_{\mu\nu\alpha\beta} = -i\gamma_\mu\wedge\gamma_\nu\wedge\gamma_\alpha\wedge\gamma_\beta$.

By definition, the divergence of the correct energy–momentum tensor must equal the density of force acting on the electron. The divergence of the tetrode tensor is determined by the Dirac equation. To evaluate it, first note that

$$\partial_\mu T^\mu = \partial_\mu\overline{T}^\mu, \tag{2.22}$$

because the divergence of the antisymmetric part of the Tetrode tensor vanishes, as follows immediately from (2.16) or (2.18). With the help of (2.12), the divergence of (2.11) can be written

$$\partial_\mu\overline{T}^\mu = \hbar\,(\Box^2\psi i\gamma_3\tilde{\psi})_V - e\partial_\mu(\rho v A^\mu). \tag{2.23}$$

To express the first term on the right on (2.23) in terms of observables, take the gradient of the Dirac equation (2.15) and multiply on the right by $\gamma_0\tilde{\psi}$ to get

$$\hbar\,(\Box^2\psi)i\gamma_3\tilde{\psi} = m(\Box\psi)\tilde{\psi} + e(\Box A\psi)\gamma_0\tilde{\psi}.$$

Again using the Dirac equation, rewrite the right side of this expression to get

$$\begin{aligned}\hbar\Box^2\psi i\gamma_3\tilde{\psi} = \hbar^{-1}(e^2A^2 - m^2)i\rho s + e(\Box A)\rho v\\ + 2e(A\cdot\Box\psi)\gamma_0\tilde{\psi}.\end{aligned} \tag{2.24}$$

The vector part of this equation is

$$\begin{aligned}(\hbar\Box^2\psi i\gamma_3\tilde{\psi})_V &= \frac{\hbar}{2}\,(\Box^2\psi i\gamma_3\tilde{\psi} - \psi i\gamma_3\Box^2\tilde{\psi})\\ &= \rho e(\Box\wedge A)\cdot v + e\rho v\Box\cdot A + eA\cdot\Box(\rho v)\\ &= \rho eF\cdot v + e\partial_\mu(\rho v A^\mu),\end{aligned} \tag{2.25}$$

where $F = \Box\wedge A$. Substitution into (2.23) and use of (2.10) yields

$$\partial_\mu T^\mu = \rho\dot{p} + \partial_\mu N^\mu = \rho eF\cdot v = \rho f, \tag{2.26}$$

where a dot has been used to represent the "proper time" derivative along the streamline by writing $v\cdot\Box p = \dot{p}$. The local force $f = eF\cdot v$ is just the familiar Lorentz force. This further confirms the compatibility of interpretations given to T^μ and v. The striking fact is that (2.26) has exactly the form that classical electrodynamics gives for the effect of an electromagnetic field F acting on a charged current. Note that there are no multipole force terms, such as would arise if the electron had some complicated local structure. Thus, according to (2.26) the effect of external electromagnetic forces acting on the electron is exactly what one would expect from classical electrodynamics. The peculiar features of the Dirac theory reside in the specific nature of the T^μ.

The results above already suffice to show that the Dirac

electron possesses an angular momentum which does not appear to be induced by external forces. Observe that, since $\partial_\mu x = \gamma_\mu$, (2.26) implies

$$\partial_\mu (T^\mu \wedge x) = T^\mu \wedge \gamma_\mu + \rho f \wedge x. \tag{2.27}$$

And note that, by virtue of (2.20),

$$T^\mu \wedge \gamma_\mu = -\partial_\mu S^\mu, \tag{2.28}$$

where

$$\begin{aligned} S^\mu &= \rho i s \wedge \gamma^\mu = \rho (is)\cdot\gamma^\mu = \rho (S\wedge v)\cdot\gamma^\mu \\ &= \rho v^\mu S + \rho S\cdot\gamma^\mu v. \end{aligned} \tag{2.29}$$

So, with the definition

$$J^\mu \equiv T^\mu \wedge x + S^\mu, \tag{2.30}$$

(2.27) can be written

$$\partial_\mu J^\mu = \rho f \wedge x. \tag{2.31}$$

The right side of this equation can be identified as the local torque, so J^μ can be interpreted as the flux of angular momentum through a hypersurface with normal γ^μ. Moreover, J^μ consists of an orbital part $T^\mu \wedge x$ and an "intrinsic part" S^μ. The angular momentum flow along the velocity streamline is described by the *proper angular momentum density*,

$$J(v) = v_\mu J^\mu = \rho(p \wedge x + S), \tag{2.32}$$

where $S = \rho^{-1} v_\mu S^\mu = is \wedge v$ is seen to be precisely the local spin, as advertised in Sec. 1.

The right side of (2.29) is a decomposition of the spin angular momentum density S^μ into a part $\rho v^\mu S$ which describes angular momentum flow along the streamline and a part

$$M^\mu \equiv \rho S\cdot\gamma^\mu v = \rho \tfrac{1}{2}[S, \gamma^\mu \wedge v], \tag{2.33}$$

which describes angular momentum flow normal to the streamline. Using this, and the corresponding decomposition (2.10) for T_μ, (2.28) can be written in the form

$$\rho \dot{S} + \rho p \wedge v = \gamma_\mu \wedge N^\mu - \partial_\mu M^\mu, \tag{2.34}$$

where, of course, $\dot{S} = v\cdot\Box S$.

Equations (1.4), (2.26), and (2.34) are local conservation laws for mass, energy–momentum and "intrinsic angular momentum," respectively. But they constitute a determinate set of equations only when "constitutive equations" have been specified which relate the basic local observables ρ, v, S and p and expresses the fluxes N_μ and M_μ in terms of them. In the general form given above with a more general form for the local force than is given in (2.26), the local conservation laws hold for any classical relativistic theory of continuous media as well as for the Dirac theory. The peculiar features of the Dirac theory are found not in the conservation laws, but in the form it gives to the constitutive relations and boundary conditions.

3. LOCAL MOMENTUM AND ANGULAR VELOCITY

From (1.2) it follows that

$$\gamma_\mu\cdot\Box e_\alpha = \partial_\mu e_\alpha = \tfrac{1}{2}[\Omega_\mu, e_\alpha] = \Omega_\mu\cdot e_\alpha, \tag{3.1}$$

where

$$\Omega_\mu \equiv 2(\partial_\mu R)\tilde{R}. \tag{3.2}$$

Equation (3.1) says that on displacement in the γ_μ direction, the frame $\{e_\mu\}$ rotates with "angular velocity" Ω_μ. To see what physical significance such a rotation might have, the angular velocity must be expressed as a function of local observables. To this end, introduce quantities P_μ and q_μ by the equation

$$P_\mu + iq_\mu = \frac{\hbar}{2}(\partial_\mu R\gamma_2\gamma_1\tilde{R} - R\gamma_2\gamma_1\partial_\mu\tilde{R}). \tag{3.3}$$

Also, use (1.6) to obtain

$$\partial_\mu S = \frac{\hbar}{2}(\partial_\mu R\gamma_2\gamma_1\tilde{R} + R\gamma_2\gamma_1\partial_\mu\tilde{R}) \tag{3.4}$$

and

$$\hbar\partial_\mu R\gamma_2\gamma_1\tilde{R} = (\partial_\mu R\tilde{R})(\hbar R\gamma_1\gamma_2\tilde{R}) = \Omega_\mu S. \tag{3.5}$$

The sum of (3.3) and (3.4) yields

$$P_\mu + iq_\mu + \partial_\mu S = \Omega_\mu S = \hbar\partial_\mu R\gamma_2\gamma_1\tilde{R}. \tag{3.6}$$

The scalar part of (3.6) is

$$P_\mu = \Omega_\mu\cdot S = \hbar(\partial_\mu R\gamma_2\gamma_1\tilde{R})_S, \tag{3.7}$$

which shows that P_μ measures the component of the angular velocity in the local spin plane.

The pseudoscalar part of (3.6) yields

$$q_\mu = -i(\Omega_\mu \wedge S) = -\Omega_\mu\cdot(iS) = \hbar(\partial_\mu R\gamma_0\gamma_3\tilde{R})_S \tag{3.8}$$

which shows that q_μ measures the component of the angular velocity in the plane orthogonal to the spin plane. Finally, the bivector part of (3.6) is

$$\partial_\mu S = \tfrac{1}{2}(\Omega_\mu S - S\Omega_\mu) \equiv \tfrac{1}{2}[\Omega_\mu, S], \tag{3.9}$$

which measures the rate at which the spin plane changes direction on displacement along γ_μ.

To find an expression for Ω_μ in terms of observables, solve (3.6).

$$\begin{aligned} \Omega_\mu &= (\partial_\mu S + P_\mu + iq_\mu)S^{-1} \\ &= s(\partial_\mu v)vs^{-1} + (\partial_\mu s)s^{-1} + q_\mu v s^{-1} + P_\mu S^{-1}, \end{aligned} \tag{3.10}$$

where $S^{-1} = |S|^{-2}\tilde{S} = is^{-1}v$ and $s^{-1} = -|s|^{-2}s$. The quantity q_μ can be expressed in terms of the spin and velocity by relating (3.10) to the derivative of the velocity.

$$\partial_\mu v = \Omega_\mu\cdot v = s^{-1}(v\cdot\partial_\mu s) + \partial_\mu v - q_\mu s^{-1}.$$

Hence

$$q_\mu = v\cdot\partial_\mu s = -s\cdot\partial_\mu v \tag{3.11}$$

or

$$q = \gamma^\mu q_\mu = -v\cdot(\Box\wedge s) + v\cdot\Box s = -s\cdot\Box v + s\cdot(\Box\wedge v). \tag{3.12}$$

An expression similar to (3.11) can be found for P_μ from

$$\partial_\mu e_1 = \Omega_\mu\cdot e_1 = (e_1\cdot\partial_\mu v)v + (e_1\cdot\partial_\mu s)s^{-1} + P_\mu\frac{2}{\hbar}e_2.$$

Thus,

$$P_\mu = -\frac{\hbar}{2}e_2\cdot\partial_\mu e_1 = \frac{\hbar}{2}e_1\cdot\partial_\mu e_2. \tag{3.13}$$

Equation (3.11) says that q_μ measures the rate of rotation in the (vs) plane, while (3.13) says that P_μ measures the rate of rotation in the spin plane though no physical significance has been attributed to e_1 and e_2 separately. Since P_μ itself is an observable, (3.10) along with (3.11) gives the complete relation of Ω_μ to observables.

A change of "phase" of the wave function by an amount $\hbar^{-1}\Lambda$, i.e., the spinor transformation

$$R \to R e^{-\gamma_2\gamma_1\Lambda/\hbar}$$

induces, by (3.2), the change of angular velocity

$$\Omega_\mu \to \Omega_\mu + \partial_\mu\Lambda S^{-1},$$

and by (3.2), the transformation

$$P_\mu \to P_\mu + \partial_\mu\Lambda. \tag{3.14}$$

This shows that P_μ depends on the phase only through its gradient.

The physical significance of P_μ can be ascertained by relating it to the energy–momentum tensor. Use (1.1) and (3.2) to write

$$\partial_\mu\psi = \tfrac{1}{2}\{\partial_\mu \ln(\rho e^{i\beta}) + \Omega_\mu\}\psi. \tag{3.15}$$

Multiply this on the left by $\hbar i\gamma_3\tilde{\psi}$ and use $\tfrac{1}{2}\hbar\psi i\gamma_3\tilde{\psi} = \rho S v$ together with (3.6) to obtain

$$\begin{aligned}\hbar\partial_\mu\psi i\gamma_3\tilde{\psi} &= \{\partial_\mu \ln(\rho e^{i\beta}) + \Omega_\mu\}(\rho S v)\\ &= \{P_\mu + iq_\mu + \partial_\mu S + S\partial_\mu \ln(\rho e^{i\beta})\}\rho v\\ &= \{P_\mu + iq_\mu + W_\mu\}\rho v,\end{aligned} \tag{3.16}$$

where the bivector W_μ is defined by

$$W_\mu \equiv (\rho e^{i\beta})^{-1}\partial_\mu(\rho e^{i\beta}S) = \partial_\mu S + S(\partial_\mu \ln\rho + i\partial_\mu\beta). \tag{3.17}$$

The vector part of (3.16) is

$$\hbar(\partial_\mu\psi i\gamma_3\tilde{\psi})_V = \rho(vP_\mu - v\cdot W_\mu). \tag{3.18}$$

Hence,

$$\hbar\gamma_\nu\cdot(\partial_\mu\psi i\gamma_3\tilde{\psi})_V = \rho(v_\nu P_\mu + (v\Lambda\gamma_\nu)\cdot W_\mu). \tag{3.19}$$

Write

$$P_\mu = p_\mu + eA_\mu \tag{3.20}$$

and compare (3.19) with (2.3) and (2.10) to get

$$T_{\mu\nu} = \rho v_\mu p_\nu + N_{\mu\nu}, \tag{3.21}$$

where, with the help of (3.17),

$$\begin{aligned}N_{\mu\nu} &= N_\mu\cdot\gamma_\nu = \rho(v\wedge\gamma_\mu)\cdot W_\nu\\ &= \rho(v\wedge\gamma_\mu)\cdot\partial_\nu S - \rho s_\mu\partial_\nu\beta.\end{aligned} \tag{3.22}$$

This shows that the p_μ introduced in (3.20) are exactly the components of local momentum introduced by (2.9). Equations (3.7) and (3.20) thus show how the local momentum is related to rotations in the local spin plane. In addition, (3.22) expresses the components $N_{\mu\nu}$ of the momentum flux tensor in terms of local observables.

4. INTEGRABILITY CONDITIONS

The fundamental local observables v, S, and p are all determined by the single spinor field R and its derivatives. Since R itself is completely parametrized by v, S, and phase χ, p must be completely determined by these quantities and their derivatives. Since p depends on the phase only through its gradient, the curl of p must be a function only of the spin, the velocity, and their derivatives. This function can be found in the following systematic way.

Write (3.2) in the form

$$\partial_\mu R = \tfrac{1}{2}\Omega_\mu R. \tag{4.1}$$

Differentiate, to get

$$\partial_\nu\partial_\mu R = \tfrac{1}{2}(\partial_\nu\Omega_\mu + \tfrac{1}{2}\Omega_\mu\Omega_\nu)R. \tag{4.2}$$

But

$$\partial_\nu\partial_\mu R = \partial_\mu\partial_\nu R. \tag{4.3}$$

So

$$\partial_\nu\Omega_\mu - \partial_\mu\Omega_\nu = \tfrac{1}{2}[\Omega_\nu, \Omega_\mu]. \tag{4.4}$$

Thus the derivatives of the angular velocities are not mutually independent. These "integrability conditions" can be expressed as relations among observables by using (3.6). One obtains

$$\partial_\mu P_\nu - \partial_\nu P_\mu + i(\partial_\nu q_\mu - \partial_\mu q_\nu) = \tfrac{1}{2}[\partial_\nu S, \partial_\mu S]S^{-1}. \tag{4.5}$$

The bivector part of (4.5) gives nothing new since it is just a consequence of the fact that S^2 is a constant. The pseudoscalar part of (4.5) gives an expression for the curl of q_μ, but that is of little interest since the relation of q_μ to the spin and velocity is already completely exhibited by (3.11). However, the scalar part of (4.5) gives the following valuable relations, first derived in a different form by Takabayasi[3]:

$$\begin{aligned}\partial_\mu P_\nu - \partial_\nu P_\mu &= \tfrac{1}{2}[\partial_\nu S, \partial_\mu S]\cdot S^{-1}\\ &= (\partial_\nu S\partial_\mu S S^{-1})_S = S\cdot(\partial_\nu v\wedge\partial_\mu v + \partial_\nu s\wedge\partial_\nu s^{-1})\\ &= is\wedge v\wedge[\partial_\nu v\wedge\partial_\mu v + \partial_\mu s\wedge\partial_\mu s^{-1}].\end{aligned} \tag{4.6}$$

Since $P_\mu = p_\mu + eA_\mu$,

$$\partial_\nu p_\mu - \partial_\mu p_\nu + eF_{\mu\nu} = (\partial_\nu S\partial_\mu S)\cdot S^{-1}, \tag{4.7}$$

where

$$F_{\mu\nu} = (\gamma_\nu\wedge\gamma_\mu)\cdot F = (\gamma_\nu\wedge\gamma_\mu)\cdot(\square\wedge A) = \partial_\mu A_\nu - \partial_\nu A_\mu$$

are the components of the electromagnetic field.

It is worth emphasizing that (4.7) or (4.6) depends on the definitions of the local observables only and not at all on the Dirac equation. Since it relates different observables, (4.7) can be regarded as a kind of constitutive relation. A constitutive relation that gives the full dependence of p on v and S is obtained from the Dirac equation in the next section.

5. PHYSICAL CONTENT OF THE DIRAC EQUATION

The conclusions of Sec. 2 are worth repeating. Equations (1.4), (2.26), and (2.34) are precisely the conservation laws of mass, energy–momentum, and angular momentum which are expected to hold for all physical theories. The use of the Dirac equation to obtain these laws merely serves to show that the Dirac theory is consistent with general principles. It can now be shown that the real physical content of the Dirac equation is to be found in the fact that it provides *constitutive relations* among the local density, velocity, momentum, and spin.

When these relations are put along side those relations which follow from the definitions of the local observables in terms of the Dirac wave function, then the conservation laws become determinate differential equations describing the time evolution of the local observables.

To express the Dirac equation (2.15) as a relation among local observables, first multiply it on the right by $\tilde{\psi}$ to obtain

$$\hbar(\Box\psi)\gamma_2\gamma_1\tilde{\psi} = m\rho v + eA\rho e^{i\beta}. \tag{5.1}$$

Next use (3.15) and (3.6) to obtain

$$\begin{aligned}\hbar(\partial_\mu\psi)\gamma_2\gamma_1\tilde{\psi} &= \{\partial_\mu(\rho e^{i\beta}) + \Omega_\mu\rho e^{i\beta}\}\frac{\hbar}{2}R\gamma_2\gamma_1\tilde{R} \\ &= \partial_\mu(\rho e^{i\beta})S + (P_\mu + iq_\mu + \partial_\mu S)\rho e^{i\beta} \\ &= (P_\mu + iq_\mu)\rho e^{i\beta} + \partial_\mu(\rho e^{i\beta}S).\end{aligned} \tag{5.2}$$

This implies

$$\hbar(\Box\psi)\gamma_2\gamma_1\tilde{\psi} = (P + qi)\rho e^{i\beta} + \Box(\rho e^{i\beta}S). \tag{5.3}$$

Finally, equate (5.1) to (5.3) and use $p = P - eA$ to get

$$\rho e^{-i\beta}(p - iq) = \rho m v - \Box(\rho e^{i\beta}S). \tag{5.4}$$

The pseudovector part of (5.4) yields

$$\rho(p\sin\beta + q\cos\beta) = \Box\cdot(\rho e^{i\beta}iS). \tag{5.5}$$

The vector part of (5.4) is

$$\rho(p\cos\beta - q\sin\beta) = \rho m v - \Box\cdot(\rho e^{i\beta}S). \tag{5.6}$$

This quantity, multiplied by e/m, is commonly known as the *Gorden current* and assumed to be the source of the electromagnetic field produced by the electron. The last term in (5.6), then, is the divergence of a magnetic moment density. This is consistent with the identification of the magnetic moment in (6.28) below.

Equation (5.4) displays the physical content of the Dirac equation as a relation among local observables. The Dirac equation can be recovered from (5.4) by writing the local observables as functions of the Dirac wave function and its derivatives, but it should be remembered that these expressions are physical assumptions quite independent of the Dirac equation.

The "Dirac relations" among local observables are better expressed by multiplying (5.4) on the right to get

$$\rho(p - iq) = \rho m e^{i\beta}v - \Box(\rho S) + i(\Box\beta)\rho S. \tag{5.7}$$

The vector part of (5.7) gives the momentum density as a function of velocity and spin.

$$\rho p = \rho m v\cos\beta - \Box\cdot(\rho S) + \rho(iS)\cdot\Box\beta. \tag{5.8}$$

This is the simplest way to express the general noncollinearity of velocity and momentum in the Dirac theory. Note that a valid physical interpretation of β must account for the strange factor $\cos\beta$, which reduces the contribution of the "mass density" to the energy–momentum density. The last term $(iS)\cdot\Box\beta = (v\wedge s)\cdot\Box\beta = vs\cdot\Box\beta - sv\cdot\Box\beta$ shows a dependence of momentum on the rate of change of β in the $v\wedge s$ plane.

The trivector part of (5.7) yields

$$-\rho q = \rho m\sin\beta v + \Box\cdot(\rho s v) + (\Box\beta)\cdot(\rho S). \tag{5.9}$$

This can be reduced to simpler terms by using (2.18) and (3.12) to get

$$S\cdot\Box\beta = s\cdot(\Box\wedge v) - v\cdot\Box s = s\cdot\Box v - v\cdot(\Box\wedge s). \tag{5.10}$$

Since $S\cdot\Box\beta = (is\wedge v)\cdot\Box\beta = is\wedge v\wedge\Box\beta = -s\cdot(iv\wedge\Box\beta$, the first equation (5.10) can be written

$$v\cdot\Box s = \dot{s} = s\cdot(\Box\wedge v) + s\cdot(iv\wedge\Box\beta). \tag{5.11}$$

This is an equation for the rate of change of spin along a velocity streamline, and so exhibits explicitly the physical content of the trivector part of (5.7).

A number of important auxiliary formulas are easily obtained from (5.8) by utilizing algebraic properties of the velocity and spin:

$$\begin{aligned}p\cdot v &= m\cos\beta - (v\wedge\Box)\cdot S - (iS)\cdot(v\wedge\Box\beta) \\ &= m\cos\beta - S\cdot(\Box\wedge v + iv\wedge\Box\beta),\end{aligned} \tag{5.12}$$

$$v\cdot\Box\beta = \dot{\beta} = p\cdot s^{-1} + S\cdot(\Box\wedge s^{-1}) = p\cdot s^{-1} + is^{-1}\wedge v\wedge\Box s, \tag{5.13}$$

$$p\cdot S = \rho^{-1}S\cdot(\Box\cdot\rho S) = -S\cdot(S\cdot\Box\ln\rho) + S\cdot(\Box\cdot S), \tag{5.14}$$

$$\begin{aligned}v\wedge p &= \rho^{-1}[\Box\cdot(\rho S)]\wedge v - iS\dot{\beta} \\ &= \rho^{-1}\Box\cdot(\rho is) - \dot{S} + (S\cdot\Box)\wedge v - iS\dot{\beta}.\end{aligned} \tag{5.15}$$

Equation (5.12) is an expression for the *local energy* $p\cdot v$ that flows along a streamline. In the first term the rest mass is reduced by the factor $\cos\beta$. The remaining terms involve the "normal gradient" $v\wedge\Box$, which shows that their contribution to the local energy is determined by the flow of S and β onto the streamline.

By comparing (5.15) with (2.10) and (2.20) one finds

$$\rho^{-1}\gamma_\mu\wedge N^\mu = \dot{S} - (S\cdot\Box)\wedge v + iS\dot{\beta}. \tag{5.16}$$

The same result can be obtained with more effort by direct evaluation from (3.22).

6. PROPER FLOWS

The reformulation of the Dirac theory as a set of conservation laws and constitutive equations for local observables has already been completed. But further insight into the theory can be obtained by casting some of the equations into different forms. It is particularly interesting to study the flow of local observables along a streamline. This can be approached systematically by studying the *proper angular velocity* Ω, i.e., the angular velocity along a streamline:

$$\Omega \equiv 2\dot{R}\tilde{R} = 2(v\cdot\Box R)\tilde{R} = v^\mu\Omega_\mu. \tag{6.1}$$

Here the Ω_μ are just the angular velocities defined by (3.2)

An expression for Ω in terms of observables can be obtained directly from the Dirac equation by utilizing the identity

$$\Omega = 2\dot{R}\tilde{R} = \{(\Box R)\tilde{R}, v\} - \Box v. \tag{6.2}$$

The curly brackets denote anticommutator. The identity can be established by noting that

$$(\Box R)\tilde{R}vR = \Box R\gamma_0 = \Box(vR) = (\Box v)R - v\Box R + 2v\cdot\Box R.$$

Now write the Dirac equation (3.15) in the form

$$\hbar(\Box\psi)i\gamma_3\gamma_0\tilde{\psi} = 2(\Box\psi)\tilde{\psi}S = m\rho v + eA\rho e^{i\beta}. \tag{6.3}$$

Also note that

$$2(\Box\psi)\tilde{\psi} = (\Box\ln\rho + (\Box\beta)i + 2(\Box R)\tilde{R})\rho e^{i\beta}. \tag{6.4}$$

Hence, from (6.3) and (6.4)

$$2(\Box R)\tilde{R} = -\Box\ln\rho + i\Box\beta + (me^{i\beta}v + eA)S^{-1}. \tag{6.5}$$

So

$$\tfrac{1}{2}\{2(\Box R)\tilde{R}, v\} = \{2(\Box R)\tilde{R}\}\cdot v$$
$$= -v\cdot\Box\ln\rho + v\cdot(i\Box\beta) + v\cdot(mv\cos\beta + eA)S^{-1}. \tag{6.6}$$

Finally, substitute (6.6) into (6.2) to get

$$\Omega = -\Box\wedge v + v\cdot(i\Box\beta) + v\cdot(mv\cos\beta + eA)S^{-1}. \tag{6.7}$$

By (6.7), the proper time derivatives of the velocity and the spin are

$$\dot{v} = \Omega\cdot v = v\cdot(\Box\wedge v), \tag{6.8}$$

$$\dot{s} = \Omega\cdot s = s\cdot(\Box\wedge v) + s\cdot[v\cdot(\Box\beta i)], \tag{6.9}$$

$$\dot{S} = \tfrac{1}{2}[\Omega, S] = \tfrac{1}{2}[S, \Box\wedge v] + \tfrac{1}{2}[S, v\cdot(\Box\beta i)]. \tag{6.10}$$

Equation (6.8) is a mere identity, which depends only on the fact that v^2 is constant. Equation (6.9) is identical with (5.11), and, of course, (6.10) follows from (6.8) and (6.9), though it is handier to get it from (6.7). Clearly, these equations are not of much help unless a useful expression for $\Box\wedge v$ can be found.

Before proceeding further, it is worthwhile to examine the "classical limit", i.e., the limit in which the magnitude of the spin $|s| = \frac{1}{2}\hbar$ is regarded a negligibly small quantity. In that limit, the compatibility conditions (4.6) can be written

$$\Box\wedge P = 0. \tag{6.11}$$

This implies that

$$P = \Box\chi. \tag{6.12}$$

It is easy to see that the scalar χ is the phase of the Dirac wave function, so the classical limit amounts to a kind of "eikonal approximation" to the Dirac equation.

In the classical limit, Eq. (2.18) becomes $\sin\beta = 0$, which implies $\cos\beta = \pm 1$. In the same limit, the Dirac equation in the form (5.7) reduces to

$$p = \pm mv = \Box\chi - eA. \tag{6.13}$$

Clearly, the two signs correspond to limits describing particles with opposite charge.

The square of (6.13) is just the Hamilton–Jacobi equation for a classical "test charge":

$$(\Box\chi - eA)^2 = m^2. \tag{6.14}$$

Given the external potential A, one solves (6.14) to get χ. But equation (6.13) is still needed to get the velocity field from χ.

The curl of (6.13) is

$$\pm m\Box\wedge v = -e\Box\wedge A = -eF. \tag{6.15}$$

So, from (6.7) the proper angular velocity is just

$$\Omega = -\Box\wedge v = \pm\frac{e}{m}F. \tag{6.16}$$

When this is substituted into (6.8), one obtains the Lorentz force.

But, more generally, (6.16) gives a spinor form for the Lorentz force:

$$\dot{R} = \pm\frac{e}{2m}FR. \tag{6.17}$$

The solution of this equation is a one parameter family of Lorentz transformations $R = R(\tau)$ describing the rotation of the e_μ as they "move" along a streamline. In particular, this describes the rotation of the spin. So the spin does not simply disappear in the "classical limits" to the Dirac theory; only the effect of the spin on the motion of the particle disappears; an effect of the motion on the spin remains.

Now, to see what can be said about $\Box\wedge v$ without any approximation, write (5.7) in the form

$$(P + qi) - eA = mve^{-i\beta} - \gamma^\nu W_\nu, \tag{6.18}$$

where W_ν is defined by (3.17). The gradient of (6.18) is

$$(\Box P + \Box qi) - e\Box A = m[\Box v - i(\Box\beta)v]e^{-i\beta} + \gamma^\mu\gamma^\nu\partial_\mu W_\nu. \tag{6.19}$$

The bivector part of (6.19) is

$$m[\Box\wedge v + iv\wedge\Box\beta]e^{-i\beta} = -eF + \partial_\mu W^\mu + \tfrac{1}{2}[\gamma^\mu\wedge\gamma^\nu, \partial_\mu W_\nu] + (\Box\wedge P + \Box\wedge qi). \tag{6.20}$$

The dependence on momentum in the last term can be eliminated by using the integrability conditions (4.5) in the form

$$\Box\wedge P + \Box\wedge qi = \tfrac{1}{4}\gamma^\mu\gamma^\nu[\partial_\nu S, \partial_\mu S]S^{-1}. \tag{6.21}$$

Also, a little calculation shows that

$$\partial_\mu W_\nu - \partial_\nu W_\mu + \tfrac{1}{2}[\partial_\mu S, \partial_\nu S]S^{-1} = \tfrac{1}{2}[W_\mu, W_\nu]S^{-1}, \tag{6.22}$$

so the last two terms of (6.20) can be combined. When this has been done, (6.20) can be written

$$\Box\wedge v + iv\wedge\Box\beta = -\frac{e}{m}Fe^{i\beta} + C, \tag{6.23}$$

where

$$mC = e^{i\beta}(\partial_\mu W^\mu + \tfrac{1}{4}[\gamma^\mu\gamma^\nu, [W_\mu, W_\nu]S^{-1}]). \tag{6.24}$$

Equation (6.23) is very important because it *completely* describes the interaction of the external electromagnetic field F with the "Dirac observables" in a manifestly gauge invariant form. Section 5 shows that the electromagnetic field is absent from the basic relations among observables implied by the Dirac equation. The electromagnetic field is related to the momentum solely by the integrability conditions (4.8). Equation (6.22) expresses the implications of the integrability conditions in a form independent of p. And note that C in (6.24) is manifestly independent of velocity. The appropriate use of (6.23) is to eliminate $\Box\wedge v$ from equations to reveal explicitly the influence of the external field F on relations among observables.

The dependence of Ω on F is revealed by substituting (6.23) into (6.7) to get

$$\Omega = \frac{e}{m} F e^{i\beta} - C + v\cdot(mv\cos\beta + eA)S^{-1}. \tag{6.25}$$

This gives immediately equations for the velocity and spin which show the dependence on F explicitly:

$$\dot{v} = \frac{e}{m}(Fe^{i\beta})\cdot v + v\cdot C, \tag{6.26}$$

$$\dot{S} = \tfrac{1}{2}[F, \frac{e}{m} S e^{i\beta}] + \tfrac{1}{2}[S, C]. \tag{6.27}$$

In (6.27) the commutivity of $e^{i\beta}$ with bivectors was used to obtain a form that suggests that

$$\mu \equiv \frac{|e|}{m} S e^{i\beta} \tag{6.28}$$

be interpreted as the local magnetic moment. And, in fact,

$$|\mu| = \frac{|e|}{m}|S| = \frac{|e|\hbar}{2m} \tag{6.29}$$

is the famous magnitude of the electron's magnetic moment, which was the first striking consequence of Dirac's theory. But the interpretation of the magnetic moment is not so simple a matter as (6.27) and (6.28) suggest, for (6.25) and (6.26) indicate that the factor $e^{i\beta}$ belongs with the F and not with the S.

Other important relations can be derived from (6.25) by using (3.6) to get

$$\begin{aligned} v\cdot(p + eA) + iv\cdot q + \dot{S} &= \Omega S \\ &= \frac{e}{m} F e^{i\beta} S - CS + m\cos\beta + ev\cdot A \end{aligned} \tag{6.30}$$

The bivector part of (6.30) just gives equation (6.27). But, with the help of (3.11), the pseudoscalar part of (6.30) gives

$$\begin{aligned} v\cdot q = -v\cdot\dot{s} = s\cdot\dot{v} &= i[C\wedge S - \frac{e}{m}(Fe^{i\beta})\wedge S] \\ &= \frac{e}{m}(Fe^{i\beta})\cdot(s\wedge v) + C\cdot(v\wedge s). \end{aligned} \tag{6.31}$$

The scalar of (6.30) is

$$p\cdot v = m\cos\beta + \left(\frac{e}{m} S e^{i\beta}\right)\cdot F - C\cdot S. \tag{6.32}$$

The same result can be obtained by substituting (6.23) into (5.12).

Equation (6.32) explicitly shows the effect of an external field F on the local energy.

Equation (2.26) describes the rate of change of momentum along a streamline, but it does not reveal the full effect of the electromagnetic field F on the momentum flow, because it does not show how the momentum flux N^μ is affected by F. The influence of F on the flux can be found by beginning with the expression (3.22) for $N_{\mu\nu}$ and evaluating $\partial_\mu N^\mu$ with the help of (2.18) and the integrability conditions in the form (6.23). Thus,

$$\begin{aligned} \partial^\mu N_{\mu\nu} &= \rho S\cdot(\Box\wedge\partial_\nu v) - (\partial_\nu v\wedge\Box)\cdot(\rho S) - \rho s\cdot\Box\partial_\nu\beta \\ &\quad - (\partial_\nu\beta)\Box\cdot\rho s \\ &= -\rho S\cdot\partial_\nu\left(\frac{e}{m} F e^{i\beta} - C\right) + \rho(\partial_\nu v\wedge\Box\beta)\cdot(iS) \\ &\quad - (\partial_\nu v\wedge\Box)\cdot(\rho S) - m\partial_\nu\cos\beta \\ &= -\rho\left(S\cdot\partial_\nu\frac{e}{m}(Fe^{i\beta} - C) + (\partial_\nu v\wedge\gamma^\mu)\cdot W_\mu \right. \\ &\quad \left. + m\partial_\nu\cos\beta\right). \end{aligned} \tag{6.33}$$

Substitution of this expression into (2.26) yields

$$\begin{aligned} \dot{p} = \frac{e}{m} F\cdot v + \frac{e}{m}\Box(Fe^{i\beta})\cdot\not{S} - \Box C\cdot\not{S} \\ + \Box(v\wedge\gamma^\mu)\cdot\not{W}_\mu + m\Box\cos\beta. \end{aligned} \tag{6.34}$$

Here a slash has been used to denote quantities which are not differentiated by the gradient operating from the left.

The second term on the right of (6.34) is a force of the Stern–Gerlach type. But note that the β is differentiated whereas the S is not, so it is not quite the usual form for the force due to a magnetic moment given by (6.28), Moreover, this force is not a body force; rather, as the derivation shows, it expresses only the effect of the external field on the local momentum flux.

It is also important to realize that the influence of F on the spin is entirely due to its effect on the spin flux. This can be shown by beginning with (2.34) and evaluating $\partial_\mu M^\mu$ with the help of (6.23). Thus,

$$\begin{aligned} \partial_\mu M^\mu &= \rho\tfrac{1}{2}[S, \Box\wedge v] + \tfrac{1}{2}[v\wedge\Box, \rho S] \\ &= \rho\tfrac{1}{2}\left[F, \frac{e}{m} S e^{i\beta}\right] + \rho\tfrac{1}{2}[S, C] + \tfrac{1}{2}[v\wedge\gamma^\mu, W_\mu]. \end{aligned} \tag{6.35}$$

Substitution of this into (2.34) and comparison with (6.27) shows

$$\rho v\wedge p + \gamma_\mu\wedge N^\mu = \gamma_\mu\wedge T^\mu = \tfrac{1}{2}[v\wedge\gamma^\mu, W_\mu]. \tag{6.36}$$

It is not difficult to establish that this relation follows from the Dirac equation without appeal to the integrability conditions.

The above facts about the influence of external fields on the local observables have important bearing on the interpretation of the Dirac theory.

7. WEYSSENHOFF MOTION

The velocity streamlines of the Dirac theory compose a congruence of timelike curves in space–time, and Eqs. (2.26) and (2.34) describe the flow of momentum and angular momentum along these curves. The flow along one streamline is coupled to that along its neighbors by the flux of spin through the walls of a comoving volume element. To get some conception of the streamlines in the Dirac theory, suppose that along a particular stream line the net flux of momentum and angular momentum through the walls of a comoving volume element vanishes. This supposition can be stated mathematically by writing.

$$\partial_\mu N^\mu = 0, \tag{7.1}$$

$$\gamma_\mu\wedge N^\mu = \partial_\mu M^\mu. \tag{7.2}$$

The equations of motion for a streamline satisfying these conditions are decoupled from those of its neighbors. However, the extent to which such a decoupling can be justified either rigorously or as some approximation to the Dirac theory is a difficult and unsolved problem.

Along the "decoupled" streamline just described, Eqs. (2.26) and (2.34) take the simple form

$$\dot{p} = f, \tag{7.3}$$

$$\dot{S} = v \wedge p. \tag{7.4}$$

The relation between velocity can be found by multiplying (7.4) by v and solving to get

$$p = v(p \cdot v + \dot{S}) = (p \cdot v)v + v \cdot \dot{S}. \tag{7.5}$$

Of course these equations must be supplemented by the conditions that v and S are orthogonal and have constant magnitude.

Equations (7.3), (7.4), and (7.5) were obtained from a model of a fluid "with spin" by Wessenhoff[9]; they have been studied by many authors as a "classical approximation" to the Dirac theory; details can be found in books by Corben[10] and Halbwachs[11]. Of course there is nothing surprising in the fact that these equations can be obtained from a classical model; they require for their validity only general conservation laws and the assumptions that decouple the streamlines. The classical models become unphysical when they assume that a continuous system can be shrunk to a point particle obeying (7.3) and (7.4). For insight into the Dirac theory it is sufficient to suppose that the equations describe only a single streamline.

The interesting point is that the equations for the decoupled streamline can be solved. In the absence of external forces the streamline is a generalized helix. This helical motion persists in the presence of a constant magnetic field and, as Corben[10] has repeatedly emphasized, gives the correct gyromagnetic ratio for the electron, the simple number for which the Dirac theory is most famous. Thus, however dubious the decoupling assumptions, the decoupled equations retain some of the main features of the Dirac theory. And their solutions suggest that the magnetic moment of the electron is not due to any structure of the electron; it appears to be a dynamical effect, arising from the general tendency to execute helical motion because of the noncollinearity of velocity and momentum.

8. INTERPRETATION OF THE DIRAC THEORY

The set of local observables adopted in this paper is complete and thoroughly satisfactory in the sense that the entire *mathematical content* of the Dirac theory can be expressed as a system of determinate equations for these quantities. On the other hand, the *physical content* of the Dirac theory depends on the physical interpretation accorded to the local observables, that is, on the correspondence of the mathematical quantities called observables with quantities measured experimentally. Unfortunately, this correspondence is not so well established as to eliminate the possibility that the local observables have been incorrectly identified or that the Dirac theory is incorrect in some of its more detailed implications.

There does not seem to be any reason to doubt that the local velocity (1.3) has been correctly identified. The identification adopted here leads to a reasonable interpretation of the energy levels in hydrogen, which seems to be the reason it was originally accepted by the Dirac. To say that it has been universally adopted by practitioners of the Dirac theory would not be much of an exaggeration.

The correctness of identifying the Tetrode tensor (2.3) with the energy–momentum tensor can be questioned. It is easy to write down different tensors which yield identical values for global observables such as the energy levels of stationary atomic states. Each such tensor defines a different physical theory distinguished by the particular local distribution and flow of energy–momentum it predicts. No such distinction has as yet been tested experimentally. Nevertheless, there are good reasons to favor the Tetrode tensor.

Tetrode's definition of the energy–momentum tensor is adopted in this paper on the theoretical ground that it is the most straightforward generalization of the operator definitions of energy and momentum taken over by Dirac from Schrödinger theory. Perhaps a better reason for adopting the Tetrode tensor is the fact that it leads to the classical Lorentz force (2.26); this seems to have been Tetrode's original reason. Also, on close inspection it becomes clear that no alternative to the local momentum (2.9) determined by the Tetrode tensor could lead to simpler equations for local observables. Finally, it should be mentioned that the successful prediction of the Imbert effect[4,5] was made from a nonsymmetric electromagnetic energy–momentum tensor obtained by analogy from the Tetrode tensor. Thus the Tetrode tensor is supported by indirect experimental evidence. The challenge is to test it directly.

In the absence of any experimental or theoretical reason to doubt that the identification of local observables has been correctly made, it is necessary to develop a coherent interpretation of the Dirac theory which is consistent with the present identifications of local observables. Any satisfactory interpretation of the Dirac theory must explain the prominent role played by spin. In Dirac's original paper[6] the appearance of spin is rather mysterious, since no mention of spin was made in his basic assumptions. It is still widely believed that spin emerged as a consequence of relativity, though this has been refuted by many authors (e.g., Ref. 12). With a complete theory of local observables in hand, it is not difficult to ascertain the key assumption by which Dirac (implicitly) introduced spin into his theory. The derivation of the local conservation laws in Sec. 2 leads to the identification of the local spin in Eq. (2.23). Following the argument backwards, one discovers that spin was introduced by the definition of energy (2.1), in particular by the factor $(-1)^{1/2}\hbar$ in that definition. The usual matrix formulation of the Dirac theory hides the relation of $(-1)^{1/2}\hbar$ to spin very well indeed. But the relation was uncovered in Ref. 1 when matrices were dispensed with. The relation can be explained succinctly as follows: In the Dirac theory the spin is a skew symmetric tensor, or, equivalently, the bivector defined by (1.6); the usual matrix representation of this bivector has eigenvalue $\frac{1}{2}(-1)^{1/2}\hbar$, as is shown by Eq. (A1). Thus, the factor $(-1)^{1/2}\hbar$ which appears in the matrix formulation of the Dirac theory is a representation of the spin tensor by its eigenvalue; its true identity is revealed by reformulating the theory in terms of observables.

It is important to note that $(-1)^{1/2}$ and $\hbar$ always occur together as a factor $(-1)^{1/2}\hbar$ in the original basic equations of Dirac, specifically, in the definition of the energy operator (2.1) and in the Dirac equation (A5) of Appendix A. It follows that *Planck's constant is inseparably related to the spin in the Dirac theory*. So it should be no surprise that examination of Secs. 1 to 5 reveals that Planck's constant appears in the equations for local observables only as twice the magnitude of the local spin.

The ubiquitous connection between spin and Planck's constant obviously must persist in any nonrelativistic approximation to the Dirac theory, including the

Schrödinger theory. Indeed, as shown explicitly in Ref. 12, the Schrödinger equation is identical to the Pauli equation in the absence of magnetic fields, and spin appears in the Schrödinger theory as the innocent factor $(-1)^{1/2\hbar}$. It follows that every appearance of Planck's constant in the Schrödinger theory is directly related to the existence of spin. This fact is difficult to reconcile with the usual interpretation of uncertainty relations derived from Schrödinger's equation. Though Planck's constant has a prominent place in the uncertainty relations, none of the usual interpretations seem capable of accounting for its connection with spin. It is strange that uncertainty relations for position and momentum, which are presumed to be fundamental to the interpretation of quantum theory, are not derivable from the Dirac theory without suppressing spin.

The Dirac theory poses another difficulty for the usual interpretation of the uncertainty relations. The noncollinearity of local momentum and velocity suggests that uncertainty relations for momentum and position are not, as is usually assumed, equivalent to uncertainty relations for velocity and position. Indeed, when the nonrelativistic limit is correctly carried out it can be shown that the inequivalence of velocity and momentum persists even in the Schrödinger theory. Details will be given elsewhere.

Because of these problems with the interpretation of the uncertainty principle, it is difficult to reconcile the details of the Dirac theory with the socalled Copenhagen interpretation of quantum theory in general. But there is an alternative interpretation which appears to be more congenial to the Dirac theory. It is called the *statistical interpretation* of quantum mechanics in a review by Ballentine.[13] According to the statistical interpretation, an electron is always to be regarded as a particle, and the Dirac theory describes an ensemble average of its motion. The probability density in the Dirac theory specifies the relative probability that the electron is located at a given place at a given time.

The simplest model of an electron compatible with the statistical interpretation is a structureless point charge. Of course it is impossible to derive any such model from the mathematical formulation of the Dirac theory. But it is easy to see that the model is consistent with many properties of local observables implied by the mathematical structure of the theory. First, if the charge and mass of the electron are actually localized at a single point, then the "smearing out" of the electron by any kind of averaging process must produce a distribution with constant charge to mass ratio, in agreement with the Dirac theory. Second, if the charge of the electron were not localized at a point, then terms describing the coulomb interaction between different elements of charge would appear in the local conservation laws and constitutive equations derived in Sec. 2 and 5; no such terms exist; but this is to be expected if the density in the Dirac theory describes only the probable location of a single point charge. Third, the fact that the local electromagnetic interaction is described in the Dirac theory by the Lorentz force is just what would be expected for a point charge. Finally, if the electron is assumed to be a structureless point charge, then the electron spin and magnetic moments must arise from some peculiar dynamical property of the average electron motion. These last two points deserve elaboration.

The coupling of a Dirac electron to the electromagnetic field is completely described by Eq. (2.26). But (2.26) is exactly the "classical" expression for the electromagnetic force on a local distribution of charge, the socalled Lorentz force; moreover, the angular momentum conservation law (2.31) shows that the Dirac electron has no local intrinsic magnetic moment such as would appear in higher multipole moments of the local charge distribution if the spin of the electron were associated with some local structure of the electron. Thus, the electromagnetic interaction in the Dirac theory differs in no way from that given by the "classical" theory of a local charge distribution.

It follows that the magnetic moment of the electron arises from the circulation of the local charge distribution. But what about the well-known theorem that the "classical" circulation of a fluid with constant charge to mass ratio and with angular momentum equal to the known spin of the electron cannot give rise to the known magnetic moment of the electron? That theorem does not apply! Because it implicitly assumes that the local momentum flow is collinear with the local flow of charge. This assumption certainly does not hold in the Dirac theory, nor, in fact, is it required even in classical theories.

Exactly how in the Dirac theory the local spin is related to a local circulation of charge giving rise to the observed magnetic moment of the electron is difficult to ascertain. But the problem is attacked in Sec. 6 where a general expression for the curl of the local velocity is obtained [Eq. (6.23)]. Unfortunately, the physical significance of the complicated term (6.24) is difficult to fathom, injecting some uncertainty in the conclusions that can be drawn. Nevertheless, it seems that the contribution of the electromagnetic field to the circulation of charge is completely and explicitly revealed by Eq. (6.24). Indeed, the equation is shown to lead to the well-known value for the electron magnetic moment in Eqs. (6.26), (6.27), and (6.32). However, the relations among local observables in these equations are given in more detail than in corresponding equations in the literature, and it appears to be no simple matter to interpret them fully.

Valuable insight into the relation of charge circulation to magnetic moment is given by approximation to the Dirac theory briefly discussed in Sec. 7. The resulting "Weyssenhoff motion" explicitly shows the correct gyromagnetic ratio for the electron as arising from a generalized "helical motion" which is a consequence of the noncollinearity of velocity and momentum. It seems reasonable, therefore, to suppose that the electron magnetic moment is but one consequence of the general noncollinearity of local velocity and momentum. But a great deal more study will be necessary before firm conclusions can be drawn.

There seems to be no alternative to the point charge model of an electron which is capable of interpreting the details of the Dirac theory just mentioned. Therefore, the hypothesis that the Dirac theory describes some sort of average motion of a structureless point charge ought to be examined very carefully.

ACKNOWLEDGMENT

I am indebted to R. Gurtler for checking some of the calculations.

APPENDIX A: MATRIX FORM OF THE DIRAC THEORY

There are a number of ways to establish the equivalence of the formalism used in this paper to the matrix formalism usually used to express the Dirac theory. Though

this equivalence has already been established in Refs. 1 and 2, a brief discussion of how to translate expressions from one formalism to the other should be helpful. The simplest method is to replace the vectors γ_μ directly by their representations as 4×4 matrices.

One can represent the vector γ_0 by a hermitian matrix and the vectors $\gamma_i (i = 1, 2, 3)$ by antihermitan matrices. Some writing is saved by using the same symbols γ_μ for both the vectors and their matrix representations. But, when this is done, the symbol $i = \gamma_0\gamma_1\gamma_2\gamma_3$ for the unit pseudoscalar should be replaced by the symbol $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$ usually used for the antihermitian matrix which represents it, so as to avoid confusion with the uninterpreted unit imaginary usually symbolized by $i = \sqrt{-1}$ in matrix theory.

To express the Dirac wavefunction as a column spinor, introduce a unit column spinor u which is simultaneously an "eigenvector" of the matrices γ_0 and $\gamma_5\gamma_3\gamma_0 = \gamma_2\gamma_1$ with eigenvalues 1 and i respectively; i.e., write

$$\gamma_0 u = u \quad \text{and} \quad \gamma_2\gamma_1 u = iu. \tag{A1}$$

This can be done, for example, with the matrices

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \gamma_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$
$$\gamma_2\gamma_1 = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}. \tag{A2}$$

Now, considering the spinor ψ or Eq. (1. 1) as a matrix operator, one obtains a corresponding column spinor Ψ by operating on u:

$$\Psi = \psi u. \tag{A3}$$

The above specifications suffice to relate expressions in the "space-time algebra" to expressions in the matrix algebra. For example, in the space-time algebra the Dirac equation can be written

$$\hbar\Box\psi\gamma_2\gamma_1 - eA\psi = m\psi\gamma_0. \tag{A4}$$

Considering this as a matrix equation, multiplying by u on the right and using (A1) and (A3), one obtains the usual matrix form of the Dirac equation:

$$(i\hbar\Box - eA)\Psi = \gamma^\mu(i\hbar\partial_\mu - eA_\mu)\Psi = m\Psi \tag{A5}$$

Equation (A5) may look simpler than (A4), but actually it is not, because some of its properties depend implicitly on the choice of matrix representation. Equation (A4) is independent of any matrix representation.

Translation of expressions for observables from one language to the other requires an understanding of the role of hermitian conjugation in the theory. Observe that hermitian conjugation of γ_μ can be expressed as a multiplicative operation:

$$\gamma_\mu^\dagger = \gamma_0\gamma_\mu\gamma_0 \tag{A6}$$

Let M be a linear combination of the γ_μ and their products with *real coefficients*. It follows from (A6) that the hermitian conjugate of M is

$$M^\dagger = \gamma_0\tilde{M}\gamma_0, \tag{A7}$$

where the tilde means reverse order of multiplication of all products of the γ_μ. Thus, from (1. 1), one obtains

$$\tilde{\psi} = \tilde{R}e^{1/2\beta\gamma_5}\rho^{1/2} = \rho^{1/2}e^{1/2\beta\gamma_5}\tilde{R}, \tag{A8}$$

$$\psi^\dagger = \gamma_0\tilde{\psi}\gamma_0 = \rho^{1/2}e^{-1/2\beta\gamma_5}R^\dagger. \tag{A9}$$

The operation (A7) is essentially different and must be distinquished from the complex conjugation of a scalar in matrix algebra. The latter has no physical significance if it cannot be reduced to the former, as is shown by the fact that it has been completely eliminated from the geometric language used in this paper.

From (A2) it is easily ascertained that

$$uu^\dagger = \tfrac{1}{4}(1 + \gamma_0)(1 - i\gamma_2\gamma_1). \tag{A10}$$

So

$$\begin{aligned}\Psi\Psi^\dagger\gamma_0 &= \psi uu^\dagger\psi^\dagger\gamma_0 = \psi uu^\dagger\tilde{\psi} \\ &= \tfrac{1}{4}\{\psi\tilde{\psi} + \psi\gamma_0\tilde{\psi} - i\psi\gamma_2\gamma_1\tilde{\psi} - i\psi\gamma_5\gamma_3\tilde{\psi}\} \\ &= \tfrac{1}{4}\rho\{e^{\beta\gamma_5} + v + ie^{\beta\gamma_5}\hat{S} - i\gamma_5\hat{s}\}.\end{aligned} \tag{A11}$$

To get this last line, the canonical form (1. 1) for ψ and the definitions of Sec. 1 have been taken over into the matrix algebra; also, $S = R\gamma_1\gamma_2\tilde{R}, R\gamma_0\gamma_2\gamma_1\tilde{R} = R\gamma_5\gamma_3\tilde{R} = \gamma_5 R\gamma_3\tilde{R} = \gamma_5\hat{s}$. From (A11), it follows that for any matrix M

$$\begin{aligned}\Psi^\dagger\gamma_0 M\Psi &= \mathrm{Tr}\{\Psi^\dagger\gamma_0 M\Psi\} = \mathrm{Tr}\{M\Psi\Psi^\dagger\gamma_0\} \\ &= \tfrac{1}{4}\mathrm{Tr}\{M[\psi\tilde{\psi} + \psi\gamma_0\tilde{\psi} - i\psi\gamma_2\gamma_1\tilde{\psi} - i\psi\gamma_0\gamma_2\gamma_1\tilde{\psi}]\} \\ &= (M\psi\tilde{\psi})_S + (M\psi\gamma_0\tilde{\psi})_S - i(M\psi\gamma_2\gamma_1\tilde{\psi})_S \\ &\quad - i(M\psi\gamma_0\gamma_2\gamma_1\tilde{\psi})_S.\end{aligned} \tag{A12}$$

The trace of a matrix in the Dirac matrix algebra is equal to four times the scalar part of the corresponding multivector in the space-tine algebra, so with this understood, the last line of (A12) has the same form and value in both languages. This greatly facilitates translation from one language to the other. Thus, from (A12) and the last time of (A11), one easily obtains the following variety of equivalent expressions:

$$\begin{aligned}\Psi^\dagger\gamma_0\gamma_\mu\Psi &= \mathrm{Tr}\{\gamma_\mu\Psi\Psi^\dagger\gamma_0\} = \tfrac{1}{4}\,\mathrm{Tr}\{\gamma_\mu\psi\gamma_0\tilde{\psi}\} \\ &= \psi^\dagger\gamma_0\gamma_\mu\psi)_S = (\gamma_\mu\psi\gamma_0\tilde{\psi})_S = \rho(\gamma_\mu v)_S = \rho\gamma_\mu\cdot v = \rho v_\mu,\end{aligned} \tag{A13}$$

$$\begin{aligned}i\Psi^\dagger\gamma_0\gamma_\mu\gamma_5\Psi &= i\,\mathrm{Tr}\{\gamma_\mu\gamma_5\Psi\Psi^\dagger\gamma_0\} = \frac{i^2}{4}\,\mathrm{Tr}\{\gamma_\mu\gamma_5\psi\gamma_5\gamma_3\tilde{\psi}\} \\ &= \tfrac{1}{4}\,\mathrm{Tr}\{\gamma_\mu\psi\gamma_3\tilde{\psi}\} = \rho(\gamma_\mu\hat{s})_S = \rho\gamma\cdot\hat{s} = \rho\hat{s}_\mu.\end{aligned} \tag{A14}$$

This establishes the equivalence of the usual expressions for velocity and spin in the matrix language with those adopted in this paper. In going from the left to the right sides of (A13) and (A14), use has been made of the fact that the trace of an odd product of the matrices γ_μ vanishes, which is equivalent to the fact that an odd product of the vectors γ_μ has no scalar part. Note that the sole function of the i on the left side of (A14) is to cancel the i which (A11) shows to be hidden in the matrix representation.

In matrix language the components of the Tetrode tensor are

$$T_{\mu\nu} = \frac{i\hbar}{2}\{\Psi^\dagger\gamma_0\gamma_\mu\partial_\nu\Psi - \partial_\nu\Psi^\dagger\gamma_0\gamma_\mu\Psi\} - eA_\nu\Psi^\dagger\gamma_0\gamma_\mu\Psi. \tag{A15}$$

Using (A10) as before, one obtains

$$\begin{aligned} i\Psi^\dagger\gamma_0\gamma_\mu\partial_\nu\Psi &= i\,\mathrm{Tr}\{\gamma_\mu(\partial_\nu\Psi)\Psi^\dagger\gamma_0\} \\ &= \frac{i}{4}\,\mathrm{Tr}\{\gamma_\mu\partial_\nu\psi[1 + \gamma_0 - i\gamma_2\gamma_1 - i\gamma_5\gamma_3]\tilde{\psi}\} \\ &= i(\gamma_\mu\partial_\nu\psi\gamma_0\tilde{\psi})_S + (\gamma_\mu\partial_\nu\psi\gamma_5\gamma_3\tilde{\psi})_S. \end{aligned} \tag{A16}$$

Two terms vanished in proceding to the last line of (A7) because they are odd, which follows from the fact that both ψ and $\partial_\nu\psi$ are even. Similarly,

$$\begin{aligned} i\partial_\nu\Psi^\dagger\gamma_0\gamma_\mu\Psi &= i(\gamma_\mu\psi\gamma_0\partial_\nu\tilde{\psi})_S + (\gamma_\mu\psi\gamma_5\gamma_3\partial_\nu\tilde{\psi})_S \\ &= i(\gamma_\mu\partial_\nu\psi\gamma_0\tilde{\psi})_S - (\gamma_\mu\partial_\nu\psi\gamma_5\gamma_3\tilde{\psi})_S. \end{aligned} \tag{A17}$$

The last line of (A17) follows by using the fact that the scalar part of a product is unchanged by reversing the order of multiplication. Subtracting (A17) from (A16) and using (A13) one finds that (A15) can be written

$$T_{\mu\nu} = \hbar(\gamma_\mu\partial_\nu\psi\gamma_5\gamma_3\tilde{\psi})_S - e\rho v_\mu A_\nu, \tag{A18}$$

which is the form used in the test above.

APPENDIX B: DIVERGENCE OF THE J_μ

Equation (5.17) of Ref. 1 contains an error and should be amended to read

$$\Box\cdot J_\mu = -2m\,\sin\beta e_3\cdot J_\mu + 2ei(e_3\wedge e_0\wedge J_\mu\wedge A). \tag{B1}$$

The two sentences following that equation should be corrected accordingly.

In the interest of completeness, it may be worthwhile to give the simple derivation of (B1) directly from the Dirac equation. So multiply the Dirac equation (2.15) on the right by $i\gamma_0\gamma_3\gamma_\mu\tilde{\psi}$ to get

$$\begin{aligned} \hbar(\Box\psi)\gamma_\mu\tilde{\psi} &= -im\psi\gamma_3\gamma_\mu\tilde{\psi} - eiA\psi\gamma_0\gamma_3\gamma_\mu\tilde{\psi} \\ &= -im\rho e^{i\beta}R\gamma_3\gamma_\mu\tilde{R} - e\rho iAR\gamma_0\gamma_3\gamma_\mu\tilde{R} \\ &= -im\rho e^{i\beta}e_3e_\mu - e\rho iAe_0e_3e_\mu. \end{aligned}$$

Write $J_\mu = \psi\gamma_\mu\tilde{\psi}$ and take the scalar part of this expression

$$\begin{aligned} \Box\cdot J_\mu &= (\Box J_\mu)_S = 2[(\Box\psi)\gamma_\mu\tilde{\psi}]_S \\ &= -\frac{2}{\hbar}[im\rho e^{i\beta}e_3J_\mu + eAie_3e_0J_\mu]_S, \end{aligned}$$

$$\Box\cdot J_\mu = \frac{2m}{\hbar}\,\sin\beta e_3\cdot J_\mu - \frac{2e}{\hbar}(ie_3e_0J_\mu A)_S. \tag{B2}$$

The last term on the right can be written in several different ways:

$$\begin{aligned} (ie_3e_0J_\mu A)_S &= i(e_3\wedge e_0\wedge J_\mu\wedge A) = (e_2e_1J_\mu A)_S \\ &= (e_2e_1)\cdot(J_\mu\wedge A). \end{aligned}$$

Equation (B2) with $\hbar = 1$ is seen to agree with (B1) except for a sign which comes from using a different sign convention in the Dirac equation.

The divergences of the probability and spin currents given by Eqs. (1.4) and (2.18) are seen to agree with (B2) when $\mu = 0$ and 3, respectively.

---

[1] D. Hestenes, J. Math. Phys. **8**, 798 (1967).
[2] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966).
[3] T. Takabayasi, Prog. Theor. Phys. Suppl. **4**, 1 (1957). Summarized in Nuovo Cimento **3**, 233 (1956). Elaborated in Nuovo Cimento **7**, 118 (1958). Takabayasi and Costa de Beauregard, Ref. 4, give references to earlier work on local observables. The equivalence of the parameter β with Takabayasi's parameter θ was pointed out by G. Casanova, C.R. Acad. Sci. B **266**, 1551 (1968). Some of the mathematical expressions in Secs. 3, 5, and 6 of this paper were derived independently by R. Boudet, private communication and C.R. Acad. Sci. A **272**, 767 (1971); also R. Boudet and P. Quilichini, C.R. Acad. Sci. A **268**, 724 (1969).
[4] O. Costa de Beauregard, Found. Phys. **2**, 111 (1972).
[5] C. Imbert, Phys. Rev. D **5**, 787 (1972).
[6] P. A. M. Dirac, Proc. R. Soc. Lond. **117**, 610 (1928).
[7] G. Breit, Proc. Natl. Acad. Sci. USA **14**, 553 (1928).
[8] H. Tetrode, Z. Phys. **49**, 858 (1928).
[9] J. Wessenhoff and A. Raabe, Acta Phys. Pol. **9**, 7 (1947).
[10] H. Corben, *Classical and Quantum Theories of Spinning Particles* (Holden-Day, San Francisco, 1948).
[11] F. Halbwachs, *Theorie rélativiste des fluids à spin* (Gauthier-Villars, Paris, 1960).
[12] D. Hestenes and R. Gurtler, Am. J. Phys. **39**, 1028 (1971).
[13] L. Ballentine, Rev. Mod. Phys. **42**, 358 (1970).

On the solution of the partial differential equation $\Pi_{i=1}^{N}(\nabla^2+k_i^2)\psi=0$*

G. S. Agarwal†

Institut fur Theoretische Physik, Universitat Stuttgart, 7 Stuttgart 1, Germany

A. J. Devaney‡ **and D. N. Pattanayak**

Department of Physics and Astronomy, University of Rochester, New York 14627

(Received 15 September 1972; final revised manuscript received 22 December 1972)

It is shown that the solution of the partial differential equation $\Pi_{i=1}^{N}(\nabla^2 + k_i^2)\psi = 0$, $(k_i^2 \neq k_j^2)$ subject to appropriate boundary conditions may be written as $\psi = \Sigma_{i=1}^{N}\alpha_i\psi_i$ where the ψ_i's are the solutions of the Helmholtz equation $(\nabla^2+k_i^2)\psi_i = 0$ and the α_i's are constants. The explicit forms of the ψ_i's in terms of the boundary values are also given. It is also shown that the solution of the partial differential equation $(\nabla^2 + k^2)^N\psi = 0$ is obtained by means of a certain limiting procedure from the solution of the nondegenerate problem.

In this paper we determine (subject to appropriate boundary conditions) the solution of the partial differential equation

$$\prod_{i=1}^{N}(\nabla^2 + k_i^2)\psi(\mathbf{r}) = 0, \qquad k_i^2 \neq k_j^2, \ i \neq j, \tag{1}$$

in some three-dimensional domain V bounded by the surface Σ. The propagation constants k_i are, in general, complex numbers. In the case when $N = 1$, Eq. (1) reduces to the classical Helmholtz equation whose solution is, of course, well known.[1] In what follows we will show that the solutions (1) may be built from the solutions of the Helmholtz equation. We also treat the degenerate case (i.e., when all k_i are equal) and show that the solution in this case may be obtained from the solution to (1) by an appropriate limiting procedure.

The partial differential equation (1) is of the "elliptic" type and elliptic equations of higher order have also been studied extensively by mathematicians.[2] However, the results which we present here and which may be obtained rather simply appear to be new. Equations of the type (1) have occurred recently in the electrodynamic theory of spatially dispersive media.[3] Such equations arise rather naturally in the theory of integro-differential equations. For example, consider the integro-differential equation

$$\mathcal{L}_1\psi(\mathbf{r}) = \beta\int d^3r'\, g(\mathbf{r},\mathbf{r}')\psi(\mathbf{r}'), \tag{2}$$

where $g(\mathbf{r},\mathbf{r}')$ is a Green's function solution of the equation

$$\mathcal{L}_2 g(\mathbf{r},\mathbf{r}') = -4\pi\delta(\mathbf{r}-\mathbf{r}'), \tag{3}$$

and β is a constant. In Eqs. (2) and (3) $\mathcal{L}_1$ and $\mathcal{L}_2$ are operators that are polynomials in ∇^2, with constant coefficients.

Upon operating with $\mathcal{L}_2$ on both sides of (2), we find that ψ must satisfy the differential equation

$$\mathcal{L}_2\mathcal{L}_1\psi + 4\pi\beta\psi = 0, \tag{4}$$

which may be rearranged in the form (1) where, however, some k_i^2 may be identical. Hence, we may obtain the solution of (2) by choosing the solution of (4) that satisfies also Eq. (2).

We first obtain the solution of (1) for the special case of $N = 2$. For this case we take as appropriate boundary conditions the values of ψ and $\nabla^2\psi$ on the boundary of the region V.[2,4] (For the general case one requires the values of ψ and its derivatives $\nabla^{2r}\psi$ up to order $r = N-1$ on the boundary Σ). Thus, we require the solution to the equation

$$(\nabla^2 + k_1^2)(\nabla^2 + k_2^2)\psi(\mathbf{r}) = 0, \qquad k_1^2 \neq k_2^2, \tag{5}$$

with ψ and $\nabla^2\psi$ assuming prescribed values on Σ. In what follows we exclude the cases of resonance.

Let $G_1(\mathbf{r},\mathbf{r}')$ and $G_2(\mathbf{r},\mathbf{r}')$ be the Green's functions satisfying the equations

$$(\nabla^2 + k_i^2)G_i(\mathbf{r},\mathbf{r}') = -4\pi\delta(\mathbf{r}-\mathbf{r}'), \qquad i = 1, 2, \tag{6}$$

which vanish on the boundary Σ of the domain V. Let ψ_1 and ψ_2 be defined by

$$(\nabla^2 + k_2^2)\psi(\mathbf{r}) = \psi_1(\mathbf{r}), \tag{7a}$$

$$(\nabla^2 + k_1^2)\psi(\mathbf{r}) = \psi_2(\mathbf{r}). \tag{7b}$$

We have the identity

$$\psi(\mathbf{r}) \equiv \frac{[(\nabla^2 + k_1^2) - (\nabla^2 + k_2^2)]\psi(\mathbf{r})}{(k_1^2 - k_2^2)} = \frac{\psi_1(\mathbf{r})}{k_2^2 - k_1^2} + \frac{\psi_2(\mathbf{r})}{k_1^2 - k_2^2}, \tag{8}$$

which holds as long as $k_1^2 \neq k_2^2$. It is clear from (5) and (7) that

$$(\nabla^2 + k_1^2)\psi_1(\mathbf{r}) = 0, \tag{9a}$$

$$(\nabla^2 + k_2^2)\psi_2(\mathbf{r}) = 0. \tag{9b}$$

The solution to (9) may be expressed in the form

$$\psi_1(\mathbf{r}) = -\frac{1}{4\pi}\int_\Sigma ds'(\nabla'^2 + k_2^2)\psi(\mathbf{r}')\frac{\partial}{\partial n'}G_1(\mathbf{r},\mathbf{r}'), \tag{10a}$$

$$\psi_2(\mathbf{r}) = -\frac{1}{4\pi}\int_\Sigma ds'(\nabla'^2 + k_1^2)\psi(\mathbf{r}')\frac{\partial}{\partial n'}G_2(\mathbf{r},\mathbf{r}'), \tag{10b}$$

where Eqs. (7) has been used and $\partial/\partial n' = \hat{\mathbf{n}}\cdot\nabla'$, where $\hat{\mathbf{n}}$ is the outward unit vector normal to the surface Σ. In Eq. (10) the prime on ∇^2 and n indicates differentiation with respect to the primed coordinates. We have thus shown that the solution of the equation

$$(\nabla^2 + k_1^2)(\nabla^2 + k_2^2)\psi(\mathbf{r}) = 0, \qquad k_1^2 \neq k_2^2,$$

within some three-dimensional domain V bounded by the surface Σ is expressible in the form

Copyright © 1973 by the American Institute of Physics

$$\psi(\mathbf{r}) = \frac{\psi_1(\mathbf{r})}{k_2^2 - k_1^2} + \frac{\psi_2(\mathbf{r})}{k_1^2 - k_2^2},$$

where ψ_1 and ψ_2 are the solutions of the Helmholtz equation with propagation constants k_1 and k_2, respectively. The explicit forms of ψ_1 and ψ_2 in terms of the values of ψ and $\nabla^2\psi$ on the boundary Σ are given in (10).

We next discuss the generalization of this result corresponding to arbitrary values of N. By integrating the function $\Pi_{i=1}^{N}(z^2 + k_i^2)^{-1}(z - z_0)^{-1}$ with $k_i^2 \neq k_j^2$ $(i \neq j)$ around a closed contour at infinity, one can easily prove the identity

$$1 \equiv \sum_{i=1}^{N} \alpha_i \prod_{j\neq i} (z^2 + k_j^2), \tag{11}$$

where

$$\alpha_i = \prod_{j\neq i} (k_j^2 - k_i^2)^{-1}. \tag{12}$$

From (11) we obtain the identity

$$\psi(\mathbf{r}) \equiv \sum_{i=1}^{N} \alpha_i \prod_{j\neq i} (\nabla^2 + k_j^2)\psi(\mathbf{r}) = \sum_{i=1}^{N} \alpha_i \psi_i(\mathbf{r}), \tag{13}$$

where

$$\psi_i(\mathbf{r}) = \prod_{j\neq i} (\nabla^2 + k_j^2)\psi(\mathbf{r}). \tag{14}$$

Note, that, for the case when $N = 2$, (13) reduces to (8).

From (1) and (14) we have that

$$(\nabla^2 + k_i^2)\psi_i(\mathbf{r}) = 0. \tag{15}$$

Equation (15) is immediately solved to yield

$$\psi_i(\mathbf{r}) = -\frac{1}{4\pi} \int_\Sigma ds' \prod_{j\neq i} (\nabla'^2 + k_j^2)\psi(\mathbf{r}') \frac{\partial}{\partial n'} G_i(\mathbf{r}, \mathbf{r}'), \tag{16}$$

where $G_i(\mathbf{r}, \mathbf{r}')$ are the Green's functions defined in (6) with $i = 1, 2, \ldots, N$. Note that (16) involves the value of ψ and its derivatives $\nabla^{2r}\psi$ up to order $r = N - 1$ on the boundary Σ. As mentioned previously, these N boundary conditions are sufficient to solve the Eq. (1) uniquely. On combining (15) and (16) we obtain[5]

Theorem: The solution of the equation

$$\prod_{i=1}^{N} (\nabla^2 + k_i^2)\psi(\mathbf{r}) = 0, \qquad k_i^2 \neq k_j^2,\ i \neq j,$$

in some three-dimensional domain V bounded by the surface Σ may be expressed as $\psi(\mathbf{r}) = \sum_{i=1}^{N} \alpha_i \psi_i(\mathbf{r})$, where the ψ_i are the solutions of the Helmholtz equation [with explicit forms given by (16)] and the α_i are given by (12).

We now consider the equation

$$(\nabla^2 + k^2)^2 \psi(\mathbf{r}) = 0, \tag{17}$$

and show that its solution can be obtained from the solution of (5) be means of an appropriate limiting procedure. To this end, we use the following lemma:

Lemma: The solution of (17) subject to the boundary conditions that ψ and $\nabla^2\psi$ assume prescribed values on Σ is given by the limit as $k_2 \to k_1 = k$ of the solution to (5) that satisfies these same boundary conditions.

Lemma proof: Putting $k_1 = k$ and $k_2 = k + \epsilon$, we have from (5) and (17) that the difference $\delta\psi$ between the solutions to these two equations satisfies the equation

$$(\nabla^2 + k^2)[(\nabla^2 + k^2)\delta\psi(\mathbf{r}) - \epsilon(2k + \epsilon)\phi(\mathbf{r})] = 0, \tag{18}$$

where ϕ denotes the solution to (5). From (18) we conclude that

$$(\nabla^2 + k^2)\delta\psi(\mathbf{r}) = -\frac{1}{4\pi} \int_\Sigma ds'[(\nabla'^2 + k^2)\delta\psi(\mathbf{r}')] \times \frac{\partial}{\partial n'} G(\mathbf{r}, \mathbf{r}') + \epsilon\lambda(\mathbf{r}), \tag{19}$$

where $G(\mathbf{r}, \mathbf{r}')$ is the Green's function defined in (6) with $k_i^2 = k^2$ and $\lambda(\mathbf{r})$ is given by

$$\lambda(\mathbf{r}) = (2k + \epsilon)\left[\phi(\mathbf{r}) + \frac{1}{4\pi} \int_\Sigma ds'\phi(\mathbf{r}') \frac{\partial}{\partial n'} G(\mathbf{r}, \mathbf{r}')\right]. \tag{20}$$

Since $\delta\psi$ and $\nabla^2\delta\psi$ must (by hypothesis) vanish on Σ, (19) reduces to

$$(\nabla^2 + k^2)\delta\psi(\mathbf{r}) = \epsilon\lambda(\mathbf{r}). \tag{21}$$

From (21) we deduce that

$$\delta\psi(\mathbf{r}) = -\frac{1}{4\pi} \int_\Sigma ds'\delta\psi(\mathbf{r}') \frac{\partial}{\partial n'} G(\mathbf{r}, \mathbf{r}') - 4\pi\epsilon \int_V d^3r'\lambda(\mathbf{r}')G(\mathbf{r}, \mathbf{r}'). \tag{22}$$

The first term on the right-hand side of (22) vanishes since $\delta\psi = 0$ on Σ. Moreover, in the limit of $\epsilon \to 0$, the second term vanishes giving $\delta\psi = 0$, which establishes the lemma.

Making use of the above lemma, we have from (8) that the solution to (17) is given by

$$\psi(\mathbf{r}) = \lim_{k_2 \to k_1 = k} \left(\frac{\psi_2(\mathbf{r}) - \psi_1(\mathbf{r})}{k_1^2 - k_2^2}\right) = \lim_{k_2 \to k_1 = k} \left[\left(\frac{\partial\psi_2(\mathbf{r})}{\partial k_1} - \frac{\partial\psi_1(\mathbf{r})}{\partial k_1}\right) \Big/ 2k\right], \tag{23}$$

where we have applied l'Hospital's rule. On combining (23) and (10), we obtain

$$\psi(\mathbf{r}) = \frac{1}{8\pi k} \int_\Sigma ds' \left((\nabla'^2 + k^2)\psi(\mathbf{r}) \frac{\partial^2}{\partial k \partial n'} G(\mathbf{r}, \mathbf{r}') - 2k\psi(\mathbf{r}') \frac{\partial}{\partial n'} G(\mathbf{r}, \mathbf{r}')\right). \tag{24}$$

Finally, we mention that one may use a similar procedure for obtaining the solution of the equation

$$(\nabla^2 + k^2)^N \psi(\mathbf{r}) = 0. \tag{25}$$

In particular, by following a procedure similar to that used in proving the above lemma, it is not difficult to show that the solution to (25) is given by the limit as $k_1 \to k_2 \to \cdots \to k_N = k$ of the solution to (1). For example, for the case when $N = 3$, one obtains

$$\psi(\mathbf{r}) = -\frac{1}{4\pi} \int_\Sigma ds' \left(\psi(\mathbf{r}') \frac{\partial}{\partial n'} G(\mathbf{r}, \mathbf{r}') - \frac{1}{8k^3} (\nabla'^2 + k^2)^2 \psi(\mathbf{r}') \frac{\partial^2}{\partial k \partial n'} G(\mathbf{r}, \mathbf{r}') - \frac{1}{2k} (\nabla'^2 + k^2)\psi(\mathbf{r}') \frac{\partial^2}{\partial k \partial n'} G(\mathbf{r}, \mathbf{r}')\right.$$

$$+\frac{1}{8k^2}(\nabla'^2+k^2)^2\psi(\mathbf{r}')\frac{\partial^3}{\partial k^2\partial n'}G(\mathbf{r},\mathbf{r}')\Big). \tag{26}$$

The results obtained here appear to have widespread applicability. A particular application in the electrodynamics of spatially dispersive media will be discussed elsewhere.

ACKNOWLEDGMENT

The authors would like to express their deep gratitude to Professor E. Wolf for very many enlightening discussions.

---

*Research supported by the Army Research Office (Durham)

†On leave of absence from the Department of Physics and Astronomy, University of Rochester, Rochester, New York.

‡Now with the Eikonix Corporation, Burlington, Mass. 01803.

[1]See, for example, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York 1953), Vol. I, Chap. 7.

[2]See, for example, M. Schechter, Commun. Pure Appl. Math. **13**, 407 (1960) and the references quoted therein. A referee has drawn our attention to A. Weinstein, **88**, 15 (1937); A. Weinstein and W. Z. Cheng, Q. Appl. Math. **1**, 61 (1943) where use is also made of the identity (8) in connection with the Eq. (5). However, both of these papers are concerned with an eigenvalue problem whereas our objective is to obtain explicit solutions of the boundary value problem.

[3]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Phys. Rev. Lett. **27**, 1022 (1971); see also, D. N. Pattanayak, contribution in *Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf (Plenum, New York, 1973), p. 359.

[4]For a general discussion of the boundary conditions we refer to 2. Here we simply state that, as in the case of the classical Helmholtz equation, we have some freedom in the choice of the boundary conditions; for example, we could have prescribed $\partial\psi/\partial n$ and $(\partial/\partial n)(\nabla^2\psi)$ at the boundary. The case of these boundary conditions could be treated in a similar manner.

[5]An independent proof of this theorem, first conjectured by one of us (G. S. A.) in connection with the electrodynamics of spatially dispersive media, has also been recently obtained by F. Lobkowicz (private communication), who employed the theory of vector spaces.

Weyl systems for the $(\phi^4)_3$ model

James D. Fabrey

Department of Mathematics, University of North Carolina, Chapel Hill, North Carolina 27514
(Received 22 December 1972)

Two Weyl systems, developed independently by Hepp and the author in connection with the $(\phi^4)_3$ model, are shown to be equal. A cyclic vector is exhibited.

In Glimm's fundamental paper[1] on the $(\phi^4)_3$ model, a Hilbert $F_{\rm ren}$ is constructed during the renormalization of the Hamiltonian. Glimm conjectured that a representation of the CCR, or *Weyl system*, arises in a natural way, as a "weak limit" on $F_{\rm ren}$. A partial answer was provided by Hepp,[2] who enlarged $F_{\rm ren}$ via a Gel'fand–Naimark–Segal construction to obtain a Hilbert space K, on which a non-Fock Weyl system W_K is defined by weak limits. Fabrey[3] independently constructed a non-Fock Weyl system W_r by weak limits on a Hilbert space F_r, which is an inductive limit of Hilbert spaces containing $F_{\rm ren}$.

It is easy to identify K with a subspace of F_r and to show that W_K is a subrepresentation of W_r. At first glance, Hepp's representation might seem to be a closer and simpler answer to Glimm's conjecture. On the other hand, the construction on F_r makes stronger use of bounded operators, has a larger supply of explicitly described vectors, and hence is possibly a better setting in which to answer technical questions (e.g., Eckmann and Osterwalder[4,5]).

There is no physical reason to suppose that the representations differ, and Eckmann and Osterwalder[2] have conjectured that $K = F_r$ and $W_K = W_r$. We give a short proof of this by verifying a conjecture of Hepp (Ref. 5, p. 117) that a certain vector is cyclic.

Let us summarize the notation, which is primarily a mixture of that used by Hepp[2] and the author.[3] Let F denote the Fock space for free bosons with single particle space $L_2(R^2)$. Let $V = a^{*4}(v)$ create four particles in F with wavefunction $v \notin L_2(R^8)$, defined in Ref. 3. We write $V_\sigma = \Sigma_{j=0}^{\infty} V_{j\sigma}$, $\sigma \geq 0$, where $V_{j\sigma} = a^{*4}(v_{j\sigma})$ is the truncation of V defined by

$$v_{j\sigma}(k) = v(k) \quad \text{if} \max_{1\leq i\leq 4} |k_i| \in [\alpha(j), \min(\alpha(j+1), \sigma))$$
$$= 0 \quad \text{otherwise.}$$

Here $\alpha(j) = 2^j$, $j \geq 1$, and $\alpha(0) = 0$. A family of dressing transformations $\{T_{j\sigma}\}$, $j \geq 0$, is defined by $T_{j\sigma} = \exp_j(V_{j\sigma})$, where $\exp_j(x) = \Sigma_{i=0}^{j} x^i/i!$ is a truncated exponential. As a domain for these transformations we take D, the set of vectors in F with a finite number of particles and compact support.

The basic idea behind both representations is that the following limits exist for all $n, j, k \geq 0$ and $\phi, \psi \in D$:

$$W_{jk}(\phi, A_1, A_2, \ldots, A_n, \psi) = \lim_{\sigma\to\infty} (T_{j\sigma}\phi, A_1 A_2 \ldots A_n T_{k\sigma}\psi) e^{-\Lambda(\sigma)}.$$

Here $\Lambda(\sigma) = 4!\, \Sigma_{j=0}^{\infty} \|v_{j\sigma}\|^2$ and $A_0 = I$. The A_i belong to the algebra $\mathfrak{A}$ generated by the operators in the set

$$\{a(f), a^*(f), N(B_\rho), E_{m\rho} : f \in J,\ \rho < \infty,\ m \geq 0\},$$

where the test function space is

$$J = \{f : \mu^\epsilon f \in L_2(R^2), \text{some } \epsilon > 0\},$$

$N(B_\rho)$ is the number of particles with momentum less than ρ, and $N(B_\rho) = \Sigma_{m=0}^{\infty} m E_{m\rho}$ is its spectral decomposition.

The representation $W_K(f)$, $f \in J$, is defined as follows. Let $\mathfrak{L}$ be the linear space of elements $\Phi(A_1, A_2, \ldots, A_k, \phi)$, $A_i \in \mathfrak{A}$ and $\phi \in D$ ($\Phi = \Phi(\phi)$ if $k = 0$). Let $\mathfrak{N}$ denote the radical of the sesquilinear form $\langle \cdot, \cdot \rangle$ on $\mathfrak{L}$ defined by

$$\langle \Phi(A_1, \ldots, A_k, \phi), \Phi(A_{k+1}, \ldots, A_n, \psi) \rangle = W_{00}(\phi, A_k^*, \ldots, A_1^*, A_{k+1}, \ldots, A_n, \psi).$$

Then K is the completion of $\hat{\mathfrak{D}} = \mathfrak{L}/\mathfrak{N}$. We identify vectors Φ with their cosets $[\Phi]$ so that $\Phi_1 = \Phi_2$ if and only if $[\Phi_1 - \Phi_2] = 0$. Annihilation and creation operators are defined on $\hat{\mathfrak{D}}$ in a natural way by

$$\hat{A}\Phi(A_1, \ldots, A_k, \phi) = \Phi(A, A_1, \ldots, A_k, \phi)$$

for $A \in \mathfrak{A}$. Let $\hat{\mathfrak{A}}$ denote the algebra of all such $\hat{A}$. Hepp then shows that $\hat{\mathfrak{D}}$ is a dense set of analytic vectors for $\hat{\mathfrak{A}}$, especially for $\hat{\phi}(f)$, $f \in J$, where $\phi(f) = 2^{-1/2}[a^*(f) + a(f)]$ is a field operator in F. Finally, $W_K(f) = \exp[i\hat{\phi}(f)]$ is defined analytically on $\hat{\mathfrak{D}}$ and extended by continuity to K.

The "exponential" representation $W_r(f)$, $f \in J$, is defined as follows. It is easy to see that $T_{j\sigma}D \subset T_{k\sigma}D$ for $j < k$ since

$$T_{j\sigma}\psi = T_{k\sigma} \prod_{i=j+1}^{k} \exp_i(V_{i\sigma})\psi.$$

We may view $T_jD \subset T_kD$ as sets by identifying the element $T_j\psi$ with $T_k \Pi_{i=j+1}^{k} \exp_i(V_{i\infty})\psi$. Then F_r is the completion of the increasing union $\mathfrak{D} = \cup_{j=0}^{\infty} T_jD$ equipped with the sesquilinear form

$$(T_j\phi, T_k\psi) = W_{jk}(\phi, \psi).$$

The Weyl operators $W_r(f)$ are defined on the dense domain $\mathfrak{D}$ as "weak limits" of Weyl operators defined analytically on D:

$$(T_j\phi, W_r(f)T_k\psi) = \sum_{n=0}^{\infty} W_{jk}(\phi, \underbrace{i\phi(f), \ldots, i\phi(f)}_{i \text{ times}}/n!, \psi).$$

$W_r(f)$ then extends by continuity to F_r. It is shown[3] that annihilation and creation operators are defined in a natural way on $\mathfrak{D}$ by

$$(T_j\phi, \tilde{A}T_k\psi) = W_{jk}(\phi, A, \psi)$$

for $A \in B$. Let $\tilde{\mathfrak{A}}$ denote the algebra of all such $\tilde{A}$. Then $\mathfrak{D}$ is a dense set of analytic vectors for $\tilde{\mathfrak{A}}$, especially for $\tilde{\phi}(f)$, which generates $W_r(f)$.

Theorem: Let $\Omega \in F$ be the Fock vacuum. Then $\Phi(\Omega)$

Copyright © 1973 by the American Institute of Physics

and $T_0\Omega$ are cyclic vectors for $W_K(f)$, $f \in J$, and $W_r(f)$, $f \in J$, respectively. Moreover, $K = F_r$ and $W_K = W_r$.

Proof: We first prove that $T_0\Omega$ is cyclic for $W_r(f)$, $f \in J$. Note that $\tilde{\mathfrak{G}}$ is affiliated with the W^* algebra generated by $W_r(f), f \in J$, by Ref. 3, Theorem 3. Therefore, given $\epsilon > 0, k \geq 0$, and $\psi \in D$, it suffices to find $\tilde{A} \in \tilde{B}$ such that

$$\|T_k\psi - \tilde{A}T_0\Omega\| \leq \epsilon.$$

Now $\psi = B\Omega$, where $B = \Sigma_{m=0}^{M} a^{*m}(w_m)$ for some $M < \infty$ and w_m of compact support. Each w_m is the L_2 limit on a compact set of elementary tensor products. Hence, by Ref. 3, Lemma 3.6, and the fact that B commutes with $T_{k\sigma}$, there exists an operator $\tilde{C} \in \tilde{\mathfrak{G}}$ such that $T_k\psi = T_kB\Omega$ is within $\epsilon/2$ in norm of $\tilde{C}T_k\Omega$.

It then remains to find an $\tilde{A}$ for which

$$\|\tilde{C}T_k\Omega - \tilde{A}T_0\Omega\| \leq \epsilon/2. \tag{1}$$

Let $\rho = 2^l \geq 2^k$ and $R = \tilde{C}\tilde{T}_{k\rho}\tilde{E}_{0\rho}$, where $\tilde{T}_{k\sigma}$ is defined as the weak limit of $T_{k\rho}$. Note that $\tilde{T}_{k\rho} \notin \tilde{B}$ but that $T_{k\rho}$ is a finite sum of terms with kernels of compact support. Also, $\tilde{E}_{0\rho}T_0\Omega = T_0^\rho\Omega$, where the dressing transformation T_0^ρ is the truncation of T_0 in which there are no particles with momentum less than ρ in absolute value. By the above reasoning and Ref. 3, there exists a $\tilde{T} \in \tilde{\mathfrak{G}}$ such that

$$\|(R - \tilde{C}\tilde{T}\tilde{E}_{0\rho})T_0\Omega\| \leq \epsilon/4.$$

Let $\tilde{A} = \tilde{C}\tilde{T}\tilde{E}_{0\rho} \in \tilde{\mathfrak{G}}$. Then the left-hand side of (1) is bounded by

$$\epsilon/4 + \|(\tilde{C}T_k - RT_0)\Omega\| = \epsilon/4 + \|\tilde{C}\tilde{T}_{k\rho}(\tilde{T}_{\log_2\rho} - \tilde{E}_{0\rho}T_0)\Omega\|, \tag{2}$$

where there is at least one particle with momentum greater than ρ in absolute value in each term of $\tilde{T}_{\log_2\rho} - \tilde{E}_{0\rho}T_0$. By Ref. 5, Remark 3.1, (2) is bounded by

$$\epsilon/4 + O(\rho^{-6}) \leq \epsilon/2$$

for ρ sufficiently large. Thus $T_0\Omega$ is cyclic for $W_r(f)$, $f \in J$.

Let us now imbed $K \subset F_r$ as follows. We define a map θ: $\mathfrak{L} \to \mathfrak{D}$ by

$$\theta\Phi(A_1, \ldots, A_k, \phi) = A_1A_2 \cdots A_kT_0\phi. \tag{3}$$

Clearly $\theta\mathfrak{N} = 0$ so that θ induces a map, also called θ, from $\hat{\mathfrak{D}}$ to $\mathfrak{D}$. The map θ is obviously an isometry, which extends by continuity to an isometry from K into F_r. We may therefore view K as a subspace of F_r by identifying Φ with $\theta\Phi$.

Let us now prove that $K = F_r$. We rewrite (3) with $\phi = \Omega$:

$$\theta\hat{A}_1\hat{A}_2 \cdots \hat{A}_k\Phi(\Omega) = \tilde{A}_1\tilde{A}_2 \cdots \tilde{A}_kT_0\Omega. \tag{4}$$

Thus, $\theta\hat{A}\Phi(\Omega) = \hat{A}T_0\Omega$ for all $A \in \mathfrak{G}$, so that $\hat{\mathfrak{G}}\Phi(\Omega) \subset K$ is identified with the dense subset $\hat{\mathfrak{G}}T_0\Omega$ of F_r. Hence, $K = F_r$. Moreover, by (4), $\hat{A}$ is identified with $\tilde{A}$. Letting $A = (i\phi(f))^j/j!$ and summing over $j \geq 0$, we see that $W_K(f)$ is identified with $W_r(f)$. Finally, $\Phi(\Omega)$ is identified with $T_0\Omega$ and therefore is cyclic for $W_K(f), f \in J$.

Remark: This theorem establishes a conjecture of Glimm[6]: The cyclic subspaces and representations generated by $T_j\Omega$, $j \geq 0$, coincide. The result of this theorem is implicitly understood by Eckmann.[7]

---

[1]J. Glimm, Commun. Math. Phys. **10**, 1 (1968).
[2]K. Hepp, *Théorie de la renormalization* (Springer-Verlag, Berlin, 1969), p. 116.
[3]Fabrey, Commun. Math. Phys. **19**, 1 (1970).
[4]J.-P. Eckmann, Commun. Math. Phys. **25**, 1 (1972).
[5]J.-P. Eckmann and K. Osterwalder, Helv. Phys. Acta **44**, 884 (1971)
[6]J. Glimm, lecture at University of Toronto, 1970 (unpublished).
[7]J.-P. Eckmann, Strassburg lecture, 1972 (to appear).

Unidirectional energy transfer in nonlinear wave-wave interactions

P. K. C. Wang

School of Engineering and Applied Science, University of California, Los Angeles, California
(Received 15 January 1973)

In this paper, the wave-energy transfer in a system of three nonlinearly interacting waves is studied. Sufficient conditions for the unidirectional transfer of energy into any wave from the remaining waves and bounds for the wave amplitudes are established. Also, a relation governing the rates of energy transfer between the waves is derived. The results are applied to a magnetized plasma with three nonlinearly interacting waves.

I. INTRODUCTION

The nonlinear interaction of waves in various mediums such as plasmas has been investigated both theoretically and experimentally.[1-9] In most of these works, particular attention has been focused on the so-called explosive instabilities in which the amplitude of one or more waves tends to infinity in finite time. For conservative systems involving three-wave interactions with well-defined phases, conditions for the existence of explosive instabilities have been established for the case of weak interactions.[5] In a paper by Wilhelmsson, Stenflo and Engelmann[3], a necessary condition for explosive instability is derived taking into account the effect of linear damping or growth. Recently, a sufficient condition for the nonexistence of explosive instability is obtained by first deriving appropriate bounds for the solutions of the differential equations governing the dynamics of nonlinear wave-wave interactions.[7] A fundamental aspect in the study of instabilities in such interactions is the transfer of energy from one wave to another. Here, the wave-energy transfer in a nonlinear three-wave interacting system is studied. In particular, it is shown that under certain conditions, unidirectional transfer of energy into any wave from the remaining waves can be achieved. This implies that continuous extraction of energy from one wave or pumping of energy into one wave is achievable under suitable conditions. The significance of these conditions in the case of a magnetized plasma with three nonlinearly interacting waves is examined in detail.

II. SUFFICIENT CONDITIONS FOR UNIDIRECTIONAL ENERGY TRANSFER

Consider the nonlinear interaction of three monochromatic waves with complex amplitudes a_0, a_1, and a_2 governed by the following set of complex ordinary differential equations[10]:

$$\frac{da_0}{dt} = j\omega_0 a_0 + \mu_0^* a_1 a_2, \quad \frac{da_1}{dt} = j\omega_1 a_1 + \mu_1 a_0 a_2^*, \quad \frac{da_2}{dt} = j\omega_2 a_2 + \mu_2 a_0 a_1^* \tag{1}$$

with initial conditions at $t = 0$ given by

$$a_i(0) = a_{0i}, \quad i = 0, 1, 2, \tag{2}$$

where $j = \sqrt{-1}$ and $(\cdot)^*$ denotes complex conjugates, ω_i corresponds to the complex frequency of the ith wave, and the μ_i's are the coupling coefficients.

Let C_3 denote the complex vector space of ordered triplets of complex numbers $\mathbf{a} = (a_0, a_1, a_2)$. Consider the following equations obtained by multiplying the ith equation in (1) and its complex conjugate by a_i^* and a_i respectively and then adding the results:

$$\frac{d|a_i(t)|^2}{dt} = -2\mathrm{Im}(\omega_i)|a_i(t)|^2 + 2\eta_i(\mathbf{a}(t)), \quad i = 0, 1, 2, \tag{3}$$

where

$$\eta_i(\mathbf{a}(t)) \equiv \mathrm{Re}[\mu_i^* a_0^*(t) a_1(t) a_2(t)]. \tag{4}$$

Since the energy of the ith wave at time t is directly proportional to $|a_i(t)|^2$, the term $\eta_i(\mathbf{a}(t))$ is directly proportional to the rate of change of energy due to the interaction of the ith wave with the remaining waves. The constancy in sign of $\eta_i(\mathbf{a}(t))$ for all $t \geq 0$ along a solution $\mathbf{a}(t)$ of (1) implies unidirectional transfer of energy to the ith wave from the remaining waves. In particular, for the case where the energy of the ith wave at time t is equal to $C_i|a_i(t)|^2$ with $C_i > 0$, the positivity of $\eta_i(\mathbf{a}(t))$ for all $t \geq 0$ along a solution $\mathbf{a}(t)$ of (1) implies that the ith wave is gaining energy from the remaining waves for all $t \geq 0$. In the sequel, we shall establish sufficient conditions in terms of the parameters ω_i, μ_i, a_{0i}, $i = 0, 1, 2$, under which unidirectional energy transfer is possible.

To simplify the subsequent development, we introduce the following subsets of C_3:

$$\Omega_i^+ \equiv \{\mathbf{a}_0 \in C_3 : \eta_i(\mathbf{a}(t; \mathbf{a}_0, 0)) \geq 0 \text{ for all } t \geq 0\},$$
$$\Omega_i^- \equiv \{\mathbf{a}_0 \in C_3 : \eta_i(\mathbf{a}(t; \mathbf{a}_0, 0)) \leq 0 \text{ for all } t \geq 0\} \tag{5}$$

for $i = 0, 1, 2$, where $\mathbf{a}(t; \mathbf{a}_0, 0)$ denotes the solution of (1) at time t corresponding to initial condition $\mathbf{a}_0$ at $t = 0$. The significance of Ω_i^+ and Ω_i^- is that for a given set of parameters ω_j, μ_j, $j = 0, 1, 2$, $\Omega_i^+ \cup \Omega_i^-$ represents the set of all initial points $\mathbf{a}_0$ for which unidirectional energy transfer is possible. Also, Ω_i^+ and Ω_i^- correspond to sets which are invariant under the family of transformations defined by the solutions for $t \geq 0$ [i.e., for each fixed $t \geq 0$, $\mathbf{a}(t; \cdot, 0)$ is a transformation on C_3 into C_3; moreover, $\mathbf{a}(t; \Omega_i^+, 0) \subset \Omega_i^+$ for all $t \geq 0$). Thus the problem of establishing conditions for unidirectional energy transfer for given ω_j, μ_j, $j = 0, 1, 2$ is that of finding the largest invariant sets Ω_i^+ and Ω_i^-. The exact determination of these sets is not straightforward. However, we observe that if $\eta_i(\mathbf{a}_0) \geq 0$ and $d\eta_i(\mathbf{a}(t; \mathbf{a}_0, 0))/dt \geq 0$ for all $t \geq 0$, then $\eta_i(\mathbf{a}(t; \mathbf{a}_0, 0)) \geq 0$ for all $t \geq 0$ or $\mathbf{a}_0 \in \Omega_i^+$. On the other hand, if $\eta_i(\mathbf{a}_0) \leq 0$ and $d\eta_i(\mathbf{a}(t; \mathbf{a}_0, 0))/dt \leq 0$ for all $t \geq 0$ implies that $\eta_i(\mathbf{a}(t; \mathbf{a}_0, 0)) \leq 0$ for all $t \geq 0$ or $\mathbf{a}_0 \in \Omega_i^-$.

Now, using (1), we have

$$\frac{d\eta_i(\mathbf{a}(t; \mathbf{a}_0, 0))}{dt} = \mathrm{Re}\left[\mu_i^*\left(\frac{da_0^*(t)}{dt} a_1(t) a_2(t) + a_0^*(t)\frac{da_1(t)}{dt} a_2(t)\right.\right.$$

Copyright © 1973 by the American Institute of Physics

$$+ a_0^*(t)a_1(t)\frac{da_2(t)}{dt}\Big)\Big]$$

$$= \mathrm{Re}\{\mu_i^*[\mu_0|a_1(t)|^2|a_2(t)|^2 + \mu_1|a_0(t)|^2|a_2(t)|^2 + \mu_2|a_0(t)|^2|a_1(t)|^2 + j(\omega_1 + \omega_2 - \omega_0^*)a_0^*(t)a_1(t)a_2(t)]\}. \quad (6)$$

If we assume that the complex frequencies ω_j satisfy the matching condition

$$\omega_0^* = \omega_1 + \omega_2, \quad (7)$$

then (6) simplifies to

$$\frac{d\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))}{dt} = \mathrm{Re}(\mu_i^*\mu_0)|a_1(t)|^2|a_2(t)|^2 + \mathrm{Re}(\mu_i^*\mu_1)|a_0(t)|^2|a_2(t)|^2 + \mathrm{Re}(\mu_i^*\mu_2)|a_0(t)|^2|a_1(t)|^2. \quad (8)$$

If we assume further that $\eta_i(\mathbf{a}_0) \geq 0$ and

$$\mathrm{Re}(\mu_i^*\mu_j) \geq 0, \quad j = 0,1,2, \quad (9)$$

then $d\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))/dt \geq 0$ and $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \geq 0$ for all $t \geq 0$. Consequently, $\mathbf{a}(t;\mathbf{a}_0,0) \in \Omega_i^+$ for all $t \geq 0$. Note that one of the inequalities in (9) is $\mathrm{Re}(\mu_i^*\mu_i) \geq 0$, which is automatically satisfied. The remaining inequalities have the form

$$\mathrm{Re}(\mu_i)\mathrm{Re}(\mu_j) \geq -\mathrm{Im}(\mu_i)\mathrm{Im}(\mu_j). \quad (10)$$

On the other hand, if $\eta_i(a_0) \leq 0$ and

$$\mathrm{Re}(\mu_i^*\mu_0) \leq 0, \quad j = 0,1,2, \quad (11)$$

then $d\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))/dt \leq 0$ and $\mathbf{a}(t;\mathbf{a}_0,0) \in \Omega_i^-$ for all $t \geq 0$. However, since one of the inequalities in (11) is $\mathrm{Re}(\mu_i^*\mu_i) \leq 0$, it can be satisfied if and only if $\mu_i = 0 + j0$. The foregoing result can be summarized as follows:

Theorem 1: Assume that for some $i, i = 0,1,2$, Ω_i^+ contains at least one point $\mathbf{a}_0$ other than the zero vector in C_3 and the frequency matching condition (7) is satisfied. Then there exists at least one nontrivial solution $\mathbf{a}(t;\mathbf{a}_0,0)$ of (1), defined for all $t \geq 0$, along which the energy transfer to the ith wave is unidirectional for all $t \geq 0$ provided that $\eta_i(\mathbf{a}_0) \geq 0$ and (9) is satisfied.

Note that the zero vector in C_3 belongs to both Ω_i^+ and Ω_i^-, since it is an equilibrium point of (1). Also the condition in Theorem 1 regarding Ω_i^+ implicitly assumes the existence of a nontrivial solution $\mathbf{a}(t;\mathbf{a}_0,0)$ defined for all $t \geq 0$. The statement of Theorem 1 can be modified in a trivial way when the interval of definition of the solution is $0 \leq t \leq T < \infty$. Since the right-hand sides of (1) have continuous partial derivatives with respect to all a_j at every point in C_3, it is well known[11] that a unique solution $\mathbf{a}(t;\mathbf{a}_0,0)$ defined on some finite time interval $0 \leq t \leq T$ for any $\mathbf{a}_0 \in C_3$ with $|a_{0i}| < \infty$ for $i = 0,1,2$ always exists. Finally, we observe that the special resonant case where all the ω_i are real is included in condition (7).

III. BOUNDS FOR WAVE AMPLITUDES

In this sections, upper and lower bounds for the wave amplitudes will be derived. The results may be used to estimate the growth and decay rates of the energy of any wave. In what follows, we shall first establish a simple mathematical result which will be used later.

Lemma 1: Let f and g be specified real-valued functions of t defined on the interval $0 \leq t < \infty$. Moreover, they are integrable on every subinterval $0 \leq t \leq \tau < \infty$. Let z be a differentiable real-valued function of t defined on $0 \leq t < \infty$ satisfying the double differential inequalities

$$\beta z(t) + g(t) \leq \frac{dz(t)}{dt} \leq \alpha z(t) + f(t) \quad \text{for all } t > 0 \quad (12)$$

and $z(0) = z_0$, where α, β, and z_0 are given real numbers. Then, $z(t)$ satisfies

$$z_0 \exp(\beta t) + \int_0^t \exp[\beta(t-\tau)]g(\tau)d\tau \leq z(t) \leq z_0 \exp(\alpha t) + \int_0^t \exp[\alpha(t-\tau)]f(\tau)d\tau \quad (13)$$

for all $t \geq 0$.

Proof: Consider the upper bound for $dz(t)/dt$ in (12). Let

$$p(t) = \frac{dz(t)}{dt} - \alpha z(t) - f(t) \leq 0. \quad (14)$$

By treating p as a given function of t, the solution to (14) with $z(0) = z_0$ is

$$z(t) = \exp(\alpha t)z_0 + \int_0^t \exp[\alpha(t-\tau)]f(\tau)d\tau + \int_0^t \exp[\alpha(t-\tau)]p(\tau)d\tau \quad (15)$$

defined for all $t \geq 0$. The upper bound in (13) follows directly from the fact that the second integral in (15) is nonpositive. The lower bound in (13) can be established in a similar manner. This completes the proof.

Now, we shall establish a lower bound for $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))$ which will be used to derive a lower bound for $|a_i(t)|^2$.

Lemma 2: Let $(\omega_1 + \omega_2 - \omega_0^*)$ be a pure imaginary number. If condition (9) is satisfied, then

$$\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \geq \eta_i(\mathbf{a}_0)\exp[\mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)t] \quad (16)$$

for all $t \geq 0$.

Proof: Since $(\omega_1 + \omega_2 - \omega_0^*)$ is pure imaginary, (6) can be rewritten as

$$\frac{d\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))}{dt} = \mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) + \mathrm{Re}(\mu_i^*\mu_0)|a_1(t)|^2|a_2(t)|^2 + \mathrm{Re}(\mu_i^*\mu_1)|a_0(t)|^2|a_2(t)|^2 + \mathrm{Re}(\mu_i^*\mu_2)|a_0(t)|^2|a_1(t)|^2. \quad (17)$$

Now, if (9) is satisfied, then we have

$$\frac{d\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)}{dt} \geq \mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)), \quad t > 0. \quad (18)$$

Applying the lower bound in Lemma 1 with $g(t) \equiv 0$ leads directly to (16). Thus the proof is complete.

Note that the sign of the lower bound in (16) depends on the sign of $\eta_i(\mathbf{a}_0)$. In particular, if $\eta_i(\mathbf{a}_0) \geq 0$, then $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \geq 0$ for all $t \geq 0$ or $\mathbf{a}_0 \in \Omega_i^+$. Thus, the frequency matching condition (7) as required in Theorem

1 may be replaced by the requirement that $(\omega_1 + \omega_2 - \omega_0^*)$ be a pure imaginary number.

Theorem 2: Let $(\omega_1 + \omega_2 - \omega_0^*)$ be a pure imaginary number and $\mathbf{a}(t;\mathbf{a}_0, 0)$ be a solution of (1) and (2). If the coupling coefficients μ_j satisfy (9), then

$$|a_i(t;\mathbf{a}_0,0)|^2 \geq |a_{0i}|^2 \exp[-2\,\mathrm{Im}(\omega_i)t] + 2\eta_i(\mathbf{a}_0)P_i(t) \tag{19}$$

for all $t \geq 0$, where

$$P_i(t) = \begin{cases} \{\exp[\mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)t] \\ \qquad - \exp[-2\,\mathrm{Im}(\omega_i)t]\}/\mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*) \\ \qquad\qquad \text{if } \mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*) \neq 0, \\ t\,\exp[-2\,\mathrm{Im}(\omega_i)t] \quad \text{if } \mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*) = 0. \end{cases} \tag{20}$$

Proof: From (3) and (16), we have

$$\frac{d|a_i(t;\mathbf{a}_0,0)|^2}{dt} = -2\,\mathrm{Im}(\omega_i)|a_i(t;\mathbf{a}_0,0)|^2 + 2\eta_i(\mathbf{a}(t;\mathbf{a}_0,0))$$
$$\geq -2\,\mathrm{Im}(\omega_i)|a_i(t;\mathbf{a}_0,0)|^2 + 2\eta_i(\mathbf{a}_0)\exp[\mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)t] \tag{21}$$

for all $t > 0$. Applying the lower bound in Lemma 1 with $\beta = -2\,\mathrm{Im}(\omega_i)$ and $g(t) = 2\eta_i(\mathbf{a}_0)\exp[\mathrm{Im}(\omega_0^* - \omega_1 - \omega_2)t]$ leads to the estimate

$$|a_i(t;\mathbf{a}_0,0)|^2 \geq |a_{i0}|^2 \exp[-2\,\mathrm{Im}(\omega_i)t] + 2\eta_i(\mathbf{a}_0)\int_0^t \exp[-2\,\mathrm{Im}(\omega_i)(t-\tau) - \mathrm{Im}(\omega_1 + \omega_2 - \omega_0^*)\tau]d\tau. \tag{22}$$

The lower bound (19) is obtained by evaluating the integral in (22) whose value at time t is $P_i(t)$. This completes the proof.

Remarks:

(R − 1) From (19), it is evident that under the assumptions of Theorem 2, if the ith wave has linear growth [i.e., $\mathrm{Im}(\omega_i) < 0$], then the wave amplitude $|a_i(t;\mathbf{a}_0,0)| \to \infty$ as $t \to \infty$ in the nonlinear case also. An upper bound for the wave amplitude may be readily derived using the results of Ref. 7.

(R − 2) From (21) and Lemma 1, if $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \leq 0$ on some time interval $0 \leq t_0 \leq t \leq t_1$, then the amplitude of the ith wave with nonlinear interactions can be estimated by that in absence of interactions, i.e.,

$$|a_i(t)| \leq |a_i(t_0)|\,\exp[-2\,\mathrm{Im}(\omega_i)(t - t_0)] \tag{23}$$

for all t in the interval $t_0 \leq t \leq t_1$. Similarly, if $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \geq 0$ on some time interval $0 \leq t_0' \leq t \leq t_1'$, then

$$|a_i(t)| \geq |a_i(t_0')|\,\exp[-2\,\mathrm{Im}(\omega_i)(t - t_0')] \tag{24}$$

for all t in the interval $t_0' \leq t \leq t_1'$. Note that if $\eta_i(\mathbf{a}(t;\mathbf{a}_0,0)) \leq 0$ for all $t \geq 0$ and $\mathrm{Im}(\omega_i) > 0$, then $|a_i(t)|$ is bounded for all $t \geq 0$ and $|a_i(t)| \to 0$ as $t \to \infty$ as expected.

(R −3) In (19), if $\eta_i(\mathbf{a}_0) < 0$, there may exist a finite time $t_i > 0$ such that the right-hand side of (19) is equal to zero at t_i and becomes negative for all $t > t_i$. In fact, if $\mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*) > 0$ and $|a_{0i}| \neq 0$, such a t_i always exists and is given by

$$t_i = [\mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*)]^{-1} \times \ln\{1 + |a_{0i}|^2|\eta_i(\mathbf{a}_0)|^{-1}\,\mathrm{Im}(2\omega - \omega_1 - \omega_2 + \omega_0^*)/2\}. \tag{25}$$

If $\mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*) < 0$ and $|a_{0i}| \neq 0$, a positive t_i given by (25) exists provided that

$$|a_{0i}|^2|\eta_i(\mathbf{a}_0)|^{-1}|\mathrm{Im}(2\omega_i - \omega_1 - \omega_2 + \omega_0^*)| < 2. \tag{26}$$

In both cases, the right-hand side (19) may be replaced by zero for all $t > t_i$.

IV. RELATION FOR WAVE-ENERGY TRANSFER RATES

Now, a relation for the η_j or the rates of wave-energy transfer will be derived. Consider the pair of equations in (3) given explicitly by

$$\begin{bmatrix} \dfrac{d|a_0(t)|^2}{dt} + 2\,\mathrm{Im}(\omega_0)|a_0(t)|^2 \\ \dfrac{d|a_1(t)|^2}{dt} + 2\,\mathrm{Im}(\omega_1)|a_1(t)|^2 \end{bmatrix} = \begin{bmatrix} \mu_0^* & \mu_0 \\ \mu_1^* & \mu_1 \end{bmatrix} \begin{bmatrix} a_0^* a_1 a_2 \\ a_0 a_1^* a_2^* \end{bmatrix}. \tag{27}$$

Assuming that $\mathrm{Im}(\mu_0^*\mu_1) \neq 0$, we can solve for $a_0^* a_1 a_2$:

$$a_0^* a_1 a_2 = \left[\mu_1\left(\frac{d|a_0|^2}{dt} + 2\,\mathrm{Im}(\omega_0)|a_0|^2\right) - \mu_0\left(\frac{d|a_1|^2}{dt} + 2\,\mathrm{Im}(\omega_1)|a_1|^2\right)\right]/2j\,\mathrm{Im}(\mu_0^*\mu_1). \tag{28}$$

Substituting the above expression into the equation in (3) with $i = 2$ leads to

$$\frac{d|a_2|^2}{dt} + 2\,\mathrm{Im}(\omega_2)|a_2|^2 = \mu_2^* a_0^* a_1 a_2 + \mu_2 a_0 a_1^* a_2^* = \frac{1}{\mathrm{Im}(\mu_0^*\mu_1)}\left[\mathrm{Im}(\mu_1\mu_2^*)\left(\frac{d|a_0|^2}{dt} + 2\,\mathrm{Im}(\omega_0)|a_0|^2\right) + \mathrm{Im}(\mu_0^*\mu_2)\left(\frac{d|a_1|^2}{dt} + 2\,\mathrm{Im}(\omega_1)|a_1|^2\right)\right]. \tag{29}$$

It is apparent from (3) that an alternate form for (29) is

$$\mathrm{Im}(\mu_0^*\mu_1)\eta_2(\mathbf{a}(t)) = \mathrm{Im}(\mu_1\mu_2^*)\eta_0(\mathbf{a}(t)) + \mathrm{Im}(\mu_0^*\mu_2)\eta_1(\mathbf{a}(t)), \tag{30}$$

which gives the relationship between the wave energy transfer rates. Now, suppose that $\mathrm{Im}(\mu_1\mu_2^*) \neq 0$ and the zeroth wave has unidirectional energy transfer, in particular, $\eta_0(\mathbf{a}(t)) \geq 0$ for all $t \geq 0$. Then, from (30), we have

$$0 \leq \eta_0(\mathbf{a}(t)) = [\mathrm{Im}(\mu_0^*\mu_1)\eta_2(\mathbf{a}(t)) - \mathrm{Im}(\mu_0^*\mu_2)\eta_1(\mathbf{a}(t))]/\mathrm{Im}(\mu_1\mu_2^*) \tag{31}$$

or

$$[\mathrm{Im}(\mu_0^*\mu_1)/\mathrm{Im}(\mu_1\mu_2^*)]\eta_2(\mathbf{a}(t)) \geq [\mathrm{Im}(\mu_0^*\mu_2)/\mathrm{Im}(\mu_1\mu_2^*)]\eta_1(\mathbf{a}(t)) \tag{32}$$

for all $t \geq 0$. Similar inequalities for the wave-energy transfer rates can be established for the case where $\eta_0(\mathbf{a}(t)) \leq 0$ for all $t \geq 0$ and for other waves.

V. UNIDIRECTIONAL ENERGY TRANSFER IN A MAGNETIZED PLASMA

We shall now apply the results developed in Sec. III to the nonlinear interaction between two transverse waves (with propagation constants k_0, k_1 and frequencies $\omega_0\omega_1$) and a longitudinal electron plasma wave (with propagation constant k_2 and frequency ω_2) in a homogeneous unbounded electron plasma with a constant magnetic field B_0. It is assumed that all the waves propagate along the magnetic field lines so that only the variations in the direction of B_0 need to be considered. For this case, it has been shown by Sjölund and Stenflo[9] that under the matching conditions

$$k_2 = k_0 - k_1 \quad \text{and} \quad \omega_2 \simeq \omega_0 - \omega_1 \tag{33}$$

the nonlinear interaction between three waves is describable by a set of differential equations in the form of (1) with the coupling coefficients given by

$$\mu_0^* = (1 + S_0S_1)\frac{jk_2u^2\omega_p^2\omega_0}{8N_0\omega_2^2}\frac{[k_2 + \omega_c(S_0k_0/\omega_0 - S_1k_1/\omega_1)]}{[(\omega_1 - S_1\omega_c)^2 + S_1\omega_c\omega_p^2/2\omega_1]}, \tag{34}$$

$$\mu_1 = (1 + S_0S_1)\frac{jk_2u^2\omega_p^2\omega_1}{8N_0\omega_2^2}\frac{[k_2 + \omega_c(S_1k_1/\omega_1 - S_0k_0/\omega_0)]}{[(\omega_0 - S_0\omega_c)^2 + S_0\omega_c\omega_p^2/2\omega_0]}, \tag{35}$$

$$\mu_2 = \frac{N_0^2\omega_p^2\omega_2^3\mu_1}{2k_2^2u^4\omega_1[(\omega_1 - S_1\omega_c)^2 + S_1\omega_c\omega_p^2/2\omega_1]}, \tag{36}$$

where

$$S_j = \begin{cases} 1 & \text{if the } (\omega_j, k_j) \text{ wave is right-hand polarized,} \\ -1 & \text{if the } (\omega_j, k_j) \text{ wave is left-hand polarized.} \end{cases}$$

The parameters ω_p and ω_c are the electron plasma and cyclotron frequencies respectively. N_0 is the electron density. The frequency of the longitudinal wave ω_2 is given by the following relation:

$$\omega_2^2 = \omega_p^2 + k_2^2u^2, \tag{37}$$

where u is the thermal velocity of the electrons.

First, we shall investigate the conditions as established in Theorem 1 for the unidirectional energy transfer into the transverse (ω_0, k_0) wave. It is assumed that the polarizations of the two transverse waves are identical (i.e., $S_0 = S_1$ and $S_0S_1 = 1$) so that the coupling coefficients μ_i do not vanish. Now, consider condition (9) in Theorem 1. By elementary calculations, we have

$$\mathrm{Re}(\mu_0^*\mu_1) = \frac{-k_2^2\mu^4\omega_p^4\omega_0\omega_1}{2^4N_0^2\omega_2^4}\frac{[k_2^2 - \omega_c^2(k_0/\omega_0 - k_1/\omega_1)^2]}{[(\omega_1 - S_0\omega_c)^2 + S_0\omega_c\omega_p^2/2\omega_1][(\omega_0 - S_0\omega_c)^2 + S_0\omega_c\omega_p^2/2\omega_0]}, \tag{38}$$

$$\mathrm{Re}(\mu_0^*\mu_2) = \frac{-\omega_p^6\omega_0}{2^5\omega_2}\frac{[k_2^2 - \omega_c^2(k_0/\omega_0 - k_1/\omega_1)^2]}{[(\omega_1 - S_0\omega_c)^2 + S_0\omega_c\omega_p^2/2\omega_1]^2[(\omega_0 - S_0\omega_c)^2 + S_0\omega_c\omega_p^2/2\omega_0]}. \tag{39}$$

It is apparent that if both the transverse waves are right-hand polarized (i.e., $S_0 = S_1 = 1$), then

$$\mathrm{sgn}[\mathrm{Re}(\mu_0^*\mu_1)] = \mathrm{sgn}[\mathrm{Re}(\mu_0^*\mu_2)] = \mathrm{sgn}[\omega_c^2(k_0/\omega_0 - k_1/\omega_1)^2 - k_2^2]. \tag{40}$$

Thus, condition (9), in view of (33), implies

$$\omega_c|k_0/\omega_0 - k_1/\omega_1| \geq |k_0 - k_1|. \tag{41}$$

On the other hand, if both transverse waves are left-hand polarized ($S_0 = S_1 = -1$), then condition (9) is satisfied if and only if

$$\mathrm{sgn}[\omega_c^2(k_0/\omega_0 - k_1/\omega_1)^2 - k_2^2] = \mathrm{sgn}[(\omega_0 + \omega_c)^2 - \omega_c\omega_p^2/2\omega_0] \tag{42}$$

and

$$(\omega_1 + \omega_c)^2 > \omega_c\omega_p^2/2\omega_1,$$

or

$$\omega_c|k_0/\omega_0 - k_1/\omega_1| = |k_0 - k_1|. \tag{43}$$

The latter condition corresponds to the case where $\mathrm{Re}(\mu_0^*\mu_1) = \mathrm{Re}(\mu_0^*\mu_2) = 0$. From (33), the frequency matching condition (7) is satisfied. Thus, for right-hand polarized (ω_0, k_0) and (ω_1, k_1) waves, we conclude from Theorem 1 that unidirectional energy transfer into the (ω_0, k_0) wave is possible if

$$\mathrm{Re}[\mu_0^*a_0^*(0)a_1(0)a_2(0)] \geq 0$$

and

$$\omega_c|k_0/\omega_0 - k_1/\omega_1| \geq |k_0 - k_1|. \tag{44}$$

The corresponding condition for left-hand polarized (ω_0, k_0) and (ω_1, k_1) waves is that $\mathrm{Re}[\mu_0^*a_0(0)a_1(0)a_2(0)] \geq 0$ and (42) or (43) are satisfied. Finally, since the ω_i are real, the bound for the amplitude of the (ω_0, k_0) wave as given by (19) reduces trivially to $|a_0(t;\mathbf{a}_0, 0)|^2 \geq |a_0(0)|^2 + 2\eta_0(\mathbf{a}_0)t$ for all $t \geq 0$.

In a similar manner, conditions for the unidirectional energy transfer into the (ω_1, k_1) or (ω_2, k_2) wave can be established. In fact, if $S_0 = S_1 = 1$, then under (41), condition (9) is satisfied for both $i = 1$ and 2. For $S_0 = S_1 = -1$, condition (9) is satisfied for both $i = 1$ and 2 if (42) holds.

VI. CONCLUDING REMARKS

Although the results of this paper are limited to three-wave interactions, one may consider the unidirectional energy transfer in multiwave interactions. However, general necessary and sufficient conditions for uni-

directional wave-energy transfer are not readily obtainable. Physically speaking, the existence of conditions for unidirectional energy transfer implies the continuous energy extraction from one wave or energy pumping into one wave is possible. Further applications of the results obtained here to other types of plasmas will be discussed elsewhere.

ACKNOWLEDGMENT

This work was supported by AFOSR Grant No. AFOSR 72-2303.

[1]R. Z. Sadeev and A. A. Galeev, *Nonlinear Plasma Theory* (Benjamin, New York, 1969).
[2]V. N. Tsytovich, *Nonlinear Effects in Plasma* (Plenum, New York, 1970).
[3]H. Wilhelmsson, J. Stenflo, and F. Engelmann, J. Math. Phys. **11**, 1738 (1970).
[4]F. Engelmann and H. Wilhelmsson, Z. Naturforsch. A **24**, 206 (1969).
[5]J. Fukai, S. Krishan, and E. G. Harris, Phys. Fluids **13**, 3031 (1970).
[6]R. C. Davidson, *Methods in Nonlinear Plasma Theory* (Academic, New York, 1972).
[7]P. K. C. Wang, J. Math. Phys. **13**, 943 (1972).
[8]E. A. Jackson, J. Math. Phys. **13**, 1189 (1972).
[9]A. Sjölund and L. Stenflo, Z. Phys. **204**, 211 (1967).
[10]In Ref. 7, the coupling coefficient μ_0 in all equations except (46) should be replaced by μ_0^*. In (13) and (14), μ_i and μ_i^* should be interchanged. Equation (47) is valid only if $(\mu_i\mu_j^*)$ is real. Also, the correct expressions for A_2^2 and Δ_0 are: $A_2^2 = \omega_0\omega_1/\mu_0^*\mu_1$ and $\Delta_0 = [2\omega_0/(\mu_0^* \operatorname{Im} \mu_1\mu_2^*)]$ $[(\omega_1/\mu_1)\operatorname{Im}(\omega_2)\operatorname{Im}\{\mu_0\mu_1^*\} - (\omega_2/\mu_2)\operatorname{Im}(\omega_1)\operatorname{Im}\{\mu_0\mu_2^*\}]$.
[11]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

On discrete inverse scattering problems. II*

K. M. Case

The Rockefeller University, New York, New York 10021
(Received 23 January 1973)

A discrete version of the inverse scattering problem of the Schrödinger equation with a potential is discussed. The approach is via the Marchenko equation. Interest is primarily pedagogical. All steps are elementary and relatively obvious. Passage to the continuous case as a limit is heuristically straightforward. An example shows "how" the formalism does produce the potential from scattering data.

I. INTRODUCTION

In a previous paper[1] with a similar title we considered a discrete version of the inverse scattering problem of the Schrödinger equation with a potential. The essential point was one of pedagogy. In the discrete version the motivation for various steps becomes obvious. At any point an appropriate limiting process leads to the classical results.

The construction of the potential from the scattering data proceeded in two steps. First, following Jost and Kohn[2] the scattering data is used to construct the spectral function. Then, following Gel'fand and Levitan,[3] the potential is constructed from the spectral function. The second of the steps is completely elementary in the discrete problem. However, the first step still demanded a Wiener–Hopf type factorization—and could still not be considered elementary.

An alternate solution of the classical inverse scattering problem is associated with among others, the names of Agranovich, Krein, and Marchenko.[4] Here we consider this attack on the discrete version of the inverse scattering problem. It will be seen that all steps are now elementary, obvious, and almost trivial. Passage to the continuous limit can be made at any point. Furthermore, certain relations and the origin of particular terms become very clear.

Our program is as follows: In Sec. II the discrete version of the Schrödinger potential problem is formulated. Relevant properties of such equations are briefly recapitulated. The main portion of the work is Sec. III where the formal solution of the inverse scattering problem is given. An example is worked out in Sec. IV and the continuous limit is obtained in Sec. V.

II. BASIC EQUATIONS AND PROPERTIES

Consider the eigenvalue problem associated with the equation

$$\tfrac{1}{2}[\psi(\lambda, n+1) + \psi(\lambda, n-1)] = \lambda g(n)\psi(\lambda, n), \qquad n \geq 1. \tag{1}$$

We ask for those λ for which bounded solutions to Eq. (1) exist subject to the boundary conditions

$$\psi(\lambda, 0) = 0, \qquad \psi(\lambda, 1) = \text{const} \neq 0. \tag{2}$$

If we write

$$\lambda = 1 - E\Delta^2, \qquad g(n) = e^{q(n)\Delta^2} \tag{3}$$

and pass to the limit $\Delta \to 0$, $n \to \infty$, $n\Delta = x$ (fixed) this is the Schrödinger equation

$$\frac{1}{2}\frac{d^2\psi(E,x)}{dx^2} - q(x)\psi(E,x) = -E\psi(E,x), \tag{4}$$

with potential $q(x)$. Choosing the const in Eq. (2) as Δ the boundary conditions become

$$\psi(E,0) = 0, \qquad \left.\frac{d\psi}{dx}\right|_{x=0} = 1. \tag{5}$$

Now if $q(x)$ is suitably well behaved the spectrum corresponding to Eq. (4) consists of a finite number of discrete, simple eigenvalues on the negative real E axis and a continuum $0 < E < \infty$. For large x the continuum eigenfunctions are $\sim \sin[\sqrt{2E}x + \delta(E)]$. The classical problem is to determine $q(x)$ given the scattering date, i.e., the phase shift $\delta(E)$ for $0 < E < \infty$, the positions of the bound states $(-\epsilon_i)$ and the bound state normalization constants C_i (which are the integrals of the squares of the bound state wave functions). Here we treat the analogous problem for the discrete version.

Let us suppose that $g(n)$ in Eq. (1) is always finite, positive and such that $\lim_{n\to\infty} g(n) = 1$. [A strong requirement would be that $g(n) = 1$, $n > N$ for some N. It is convenient to keep this case in mind. However, a weaker sufficient condition is that $\sum_{n=1}^{\infty} |g(n) - 1|\, n < \infty$. Unless otherwise stated, our conclusions will be true for this weaker condition]. Then the spectrum of our discrete eigenvalue problem consists of:

(i) A finite number of simple real eigenvalues occurring in equal and opposite pairs such that $|\lambda_i| > 1$

(ii) A continuum corresponding to $-1 < \lambda < 1$. For large n these continuum functions are $\sim \sin(n\theta + \delta)$, where $\cos\theta = \lambda$. The question is to determine $g(n)$ given $\delta(\lambda)$, the λ_i, and the constants $C_i = \sum_{n=1}^{\infty} \psi^2(\lambda_i, n)g(n)$.

Greater symmetry in the final formulas results if we make the transformation

$$\phi(\lambda, n) = \sqrt{g(n)}\,\psi(\lambda, n). \tag{6}$$

Then Eq. (1) becomes:

$$a(n+1)\phi(\lambda, n+1) + a(n)\phi(\lambda, n-1) = \lambda\phi(\lambda, n), \qquad n \geq 1, \tag{7}$$

where

$$a(n) = 1/2\sqrt{g(n)g(n-1)}\,. \tag{8}$$

[*Notice:* In our problem $g(0)$ does not really occur. It multiplies $\phi(\lambda, 0) = 0$, and hence is arbitrary. For convenience we take it as $1/\sqrt{g(1)}$.] Our problem is now to find the $a(n)$.

The various quantities involved are conveniently expressed in terms of the analog of the Jost function. Thus, let $z = \lambda - (\lambda^2 - 1)^{1/2}$ [and then $z^{-1} = \lambda + (\lambda^2 - 1)^{1/2}$].

Then Eq. (7) becomes

$$a(n+1)\phi(\lambda, n+1) + a(n)\phi(\lambda, n-1) = [(z + z^{-1})/2]\phi(\lambda, n), \qquad n \geq 1. \tag{9}$$

Copyright © 1973 by the American Institute of Physics

Two linearly independent solutions of Eq. (9) are defined by the boundary conditions

$$\lim_{n\to\infty} |\phi_\pm(z,n) - z^{\pm n}| = 0. \tag{10}$$

Choosing the initial constant so that $\phi(\lambda,1)=1$, it is readily shown that

$$\phi(\lambda,n) = \frac{[\phi_-(z,0)\phi_+(z,n) - \phi_+(z,0)\phi_-(z,n)]}{z - z^{-1}} \quad \text{for } |z| = 1. \tag{11}$$

The phase shift is given by $e^{2i\delta} = S(z) = \psi_-(z,0)/\psi_+(z,0)$.

Note some analytic properties:

(i) $\phi(\lambda,n)$ is analytic for z within the unit circle except for a pole of order $n-1$ at $z=0$. [Indeed $\phi(\lambda,n)$ is a polynomial of order $n-1$ in $\lambda = (z+1/z)/2$].

(ii) $\phi_+(z,0)$ is analytic for z within the unit circle. If $g(n)=1$, $n>N$ this is obvious since it is then a polynomial in z. However, it is also true under the weaker condition mentioned above

(iii) $\phi_-(z,0)$ is analytic for z outside the unit circle. It can be analytically continued within the unit circle at least if $g(n)-1$ vanishes faster than any exponential.[5]

The bound states are given by the zeros of ϕ_+ within the unit circle, i.e.,

$$\phi_+(z_i,0) = 0. \tag{12}$$

(These zeros are real, simple, and occur in equal and opposite pairs). The bound state wave functions are given by

$$\phi(\lambda_i,n) = \phi_+(z_i,n)/\phi_+(z_i,1). \tag{13}$$

When ϕ_- can be continued within the unit circle this can also be written as

$$\phi(\lambda_i,n) = \frac{\phi_-(z_i,0)\phi_+(z_i,n)}{z_i - z_i^{-1}}. \tag{14}$$

III. THE MARCHENKO EQUATION AND SOLUTION OF THE PROBLEM

Consider functions $\phi_\pm^0(z,n)$ defined as above but satisfying

$$\tfrac{1}{2}[\phi_\pm^0(z,n+1) + \phi_\pm^0(z,n-1)] = \lambda\phi_\pm^0(z,n), \qquad n \geq 1. \tag{15}$$

Clearly,

$$\phi_\pm^0(z,n) = z^{\pm n}. \tag{16}$$

The Marchenko equation is obtained by noting that there exist constants $K(n,m)$ (independent of z) such that

$$\phi_\pm(z,n) = \sum_{n'=n}^{\infty} K(n,n')\phi_\pm^0(z,n'). \tag{17}$$

It is simplest to see this by considering the case when $g(n)=1$, $n>N$. Then the statement is trivial for sufficiently large n. Assuming this true from some n we see from Eq. (7) written in the form

$$a(n)\phi_\pm(z,n-1) = \frac{(z+1/z)}{2}\phi_\pm(z,n) - a(n+1)\phi_\pm(z,n+1), \tag{18}$$

and using Eq. (15), that it is true for $n-1$. Indeed, we read off that

$$a(n) = \frac{1}{2}\frac{K(n,n)}{K(n-1,n-1)} \tag{19}$$

[That this still holds for the weak convergence of $g(n)$ to 1 requires but a little analysis].

Let us rewrite Eq. (11) in the form

$$\frac{(z-z^{-1})\phi(\lambda,n)}{\phi_+(z,0)} = -\phi_-(z,n) + \frac{\phi_-(z,0)}{\phi_+(z,0)}\phi_+(z,n) = \phi_+(z,n) - \phi_-(z,n) - [1-S(z)]\phi_+(z,n). \tag{20}$$

Inserting the expansion of Eq. (17) gives

$$\frac{(z-z^{-1})\phi(\lambda,n)}{\phi_+(z,0)} = \sum_{n'=n}^{\infty} K(n,n')[\phi_+^0(z,n') - \phi_-^0(z,n')] - \sum_{n'=n}^{\infty} K(n,n')\phi_+^0(z,n')[1-S(z)]. \tag{21}$$

Suppose now Eq. (21) is multiplied by $z^{m-1}/2\pi i$ for $m \geq n \geq 1$ and integrated around the unit circle.

Then

$$I(m,n) = \frac{1}{2\pi i}\oint (z-z^{-1})z^{m-1}\frac{\phi(\lambda,n)dz}{\phi^+(z,0)} = \sum_{n'=n}^{\infty} K(n,n')\left(\frac{1}{2\pi i}\oint z^{m-1}\phi_+^0(z,n')dz - \frac{1}{2\pi i}\oint z^{m-1}\phi_-^0(z,n'|dz\right) + \sum_{n'=n}^{\infty} K(n,n')F_c(n',m), \qquad m \geq n \geq 1, \tag{22}$$

where

$$F_c(n',m) = \frac{1}{2\pi i}\oint z^{m-1}\phi_+^0(z,n')[1-S(z)]dz. \tag{23}$$

Now

$$\frac{1}{2\pi i}\oint z^{m-1}\phi_+^0(z,n')dz = \frac{1}{2\pi i}\oint z^{m+n'-1}dz = 0, \qquad m \geq n' \geq 1 \tag{24a}$$

and

$$\frac{1}{2\pi i}\oint z^{m-1}\phi_+^0(z,n')dz' = \frac{1}{2\pi i}\oint z^{m-n'-1}dz = \delta(m,n'). \tag{24b}$$

To evaluate $I(m,n)$ we note that the singularities within the unit circle are poles at

(i) Zeros of $\phi^+(z,0)$-bound states

and

(ii) A pole at zero.

From the bound states we obtain

$$I_b(m,n) = \sum_i (z_i)^{m-1}\frac{[z_i - z_i^{-1}]\phi_+(z_i,n)}{\phi_+(z_i,1)\dot{\phi}_+(z_i,0)} \tag{25}$$

with

$$\dot{\phi}_+(z_i,0) = \frac{d}{dz}\phi_+(z,0)|_{z_i}.$$

On using the expansion Eq. (17), this becomes

$$I_b(m,n) = \sum_i \frac{z_i - z_i^{-1}}{z_i \phi_+(z_i,1)\dot{\phi}_+(z_i,0)} \sum_{n'=n}^{\infty} K(n,n')\phi_+^0(z_i,n')z_i^m. \tag{26}$$

The contribution of the pole at $z = 0$ is found by noting that using the basic recursion relation Eq. (18) we can readily obtain the behavior of $\phi(\lambda, n)$ and $\phi_+(z, 0)$ for sufficiently small z.

Thus, for small z, we have

$$\phi(\lambda, n) \cong \frac{z^{1-n}}{\prod_{i=1}^{n} 2a(i)}$$

and

$$\phi_+(z,m) \cong \frac{z^m}{\prod_{i=m}^{\infty} 2a(i)}. \tag{27}$$

Hence

$$\frac{\phi(\lambda,n)}{\phi_+(z,0)} \approx z^{1-n} \prod_{i=n+1}^{\infty} 2a(i). \tag{28}$$

However, from Eq. (19) we see that

$$\prod_{i=n+1}^{\infty} 2a(i) = \frac{K(n+1,n+1)}{K(n,n)} \frac{K(n+2,n+2)}{K(n+1,n+1)} \cdots = \frac{K(\infty,\infty)}{K(n,n)}. \tag{29}$$

If the expression of Eq. (29) is inserted in Eq. (22) we see, since $m \geq n$, that there is at most a simple pole at the origin and this only occurs when $n = m$. Hence, the contribution to $I(m, n)$ is

$$I^0(m,n) = \frac{-K(\infty,\infty)\delta(n,m)}{K(n,n)}. \tag{30}$$

Inserting Eqs. (24a, b), (26), and (30) into Eq. (22) then yields

$$\frac{K(\infty,\infty)\delta(n,m)}{K(n,n)} = K(n,m) + \sum_{n'=n}^{\infty} K(n,n')F(n'm), \qquad m \geq n \geq 1 \tag{31}$$

with

$$F(n',m) = \frac{1}{2\pi i} \oint [1 - S(z)]\phi_+^0(z,n')z^{m-1}\,dz + \sum_i \frac{(z_i - z_i^{-1})\phi_+^0(z_i,n')z_i^m}{z_i + (z_i,1)\dot{\phi}_+(z_i,0)}. \tag{32}$$

But since $\phi_+^0(z,n') = z^{n'}$, we have

$$F(n',m) = F(n'+m) \tag{33}$$

with

$$F(m) = \frac{1}{2\pi i} \oint [1 - S(z)]z^{m-1}\,dz + \sum_i M_i^2 z_i^m, \tag{34}$$

where, as shown in the Appendix, the M_i^2 is the bound state normalization constant, i.e., $M_i^{-2} = \sum_{n=1}^{\infty} \phi_+^2(z_i, n)$.

First consider Eq. (31) for $m > n$, then if $\kappa(n,m) = K(n,m)/K(n,n)$ this becomes

$$0 = \kappa(n,m) + F(n+m) + \sum_{n'=n+1}^{\infty} \kappa(n,n')F(n'+m), \qquad m > n > 1,^6 \tag{35}$$

an inhomogeneous equation for $\kappa(n,m)$.

Then, if we put $m = n$, we obtain

$$\frac{K(\infty,\infty)}{K(n,n)^2} = 1 + F(2n) + \sum_{n'=n+1}^{\infty} \kappa(n,n')F(n'+n). \tag{36}$$

This then determines $K(n, n)$[7] after having found $\kappa(n,m)$ by solving Eq. (35). From this $a(n)$ follows from Eq. (19).

As in our previous paper,[1] all that is then determined is $\sqrt{g(n)g(n-1)}$, $n \geq 2$. (In essence this gives just the average of the potential at two consecutive points—a distinction of no importance in the continuous limit.) However, as there, there is a unique determination of $g(n)$ by the requirement that $\lim g(n) \to 1$ as $n \to \infty$. Thus, consider the now "known" quantities

$$\chi(n) = \left[\frac{1}{2a(n)}\right]^2. \tag{37}$$

We have

for n even

$$g(n)g(1) = \frac{\chi(2)\chi(4)\cdots\chi(n)}{\chi(3)\cdot\chi(5)\chi(n-1)}$$

and for n odd

$$\frac{g(1)}{g(n)} = \frac{\chi(2)\chi(4)\cdots\chi(n-1)}{\chi(3)\cdot\chi(5)\cdots\chi(n)}.$$

In either case, since $g(n) \to 1$, $\chi(n) \to 1$, we obtain

$$g(1) = \frac{\chi(2)\chi(4)\cdots}{\chi(3)\chi(5)\cdots} \tag{38}$$

Then from $a(n)$ and $g(1)$ we have all $g(n)$.

IV. AN EXAMPLE

It is instructive to consider in detail the situation when $g(n) \equiv 1$, $n > N$. Then $\phi_\pm(z, 0)$ are polynomials in $z(z^{-1})$ of order N. The integral in $F(m)$ may be done by contours. There are contributions from poles at the bound states and at the origin. The bound state contribution is

$$\sim -\frac{1}{2\pi i} \oint \frac{\phi_-(z,0)z^{m-1}}{\phi_+(z,0)}\,dz = -\frac{\phi_-(z_i,0)z_i^{m-1}}{\dot{\phi}_+(z_i,0)} = \frac{-(z_i - z_i^{-1})z_i^{m-1}}{\phi_+(z_i,1)\dot{\phi}_+(z_i,0)}, \tag{39}$$

since in this case

$$\phi_-(z_i,0) = \frac{z_i - z_i^{-1}}{\phi_i(z_i,1)}. \tag{40}$$

This just cancels the explicit bound state term in Eq. (34),

$$\therefore F(m) = \frac{1}{2\pi i} \oint' [1 - S(z)]z^{m-1}\,dz, \tag{41}$$

where $\oint'$ means to omit contributions from bound state poles.

Furthermore,

$$F(m) = \frac{1}{2\pi i} \oint' \frac{\phi_+(z,0) - \phi_-(z,0) z^{m-1} dz}{\phi_+(z,0)}$$
$$= \frac{1}{2\pi i} \oint' \frac{P_N(z) - P_N(1/z) z^{m-1} dz}{P_N(z)}, \quad (42)$$

where P_N is a polynomial of order N. The integrand we see has a pole or order at most $2N-m$ at the origin. Hence, if $m > 2N$, we conclude that

$$F(m) = 0. \quad (43)$$

For $n \geq N$, $m > n$ we have from Eq. (35) that

$$K(n,m) = 0. \quad (44)$$

From Eq. (36), we then have

$$\frac{K(\infty,\infty)}{K(n,n)^2} = 1, \quad n \geq N+1. \quad (45)$$

Thus

$$K(\infty,\infty) = 1$$

and

$$K(n,n) = 1, \quad n \geq N+1. \quad (46)$$

But then

$$a(n) = \frac{1}{2} \frac{K(n,n)}{K(n-1,n-1)} = \frac{1}{2}, \quad n \geq N+2. \quad (47)$$

As a specific example, we consider

$$S(z) = \frac{1 - (1-g)(z+1/z)z^{-N}[(z^N - z^{-N})/(z - z^{-1})]}{1 - (1-g)(z+1/z)z^{N}[(z^N - z^{-N})/(z - z^{-1})]}. \quad (48)$$

From the general argument above, we conclude that $F(m) = 0$, $m > 2N$:

$$\therefore K(\infty,\infty) = 1 = K(n,n), \quad n > N \quad (49)$$

By direct calculation,

$$F(2N) = (1-g)/g$$

Then, from Eq. (35), we see

$$K(N,m) = 0, \quad m > N.$$

Furthermore, Eq. (36) is then

$$1/K(N,N)^2 = 1 + F(2N) = 1/g, \quad K(N,N) = \sqrt{g},$$

and

$$a(N+1) = \frac{1}{2} \frac{K(N+1,N+1)}{K(N,N)} = \frac{1}{2\sqrt{g}}. \quad (50)$$

Continuing further, we find that for $n > 1$

$$F(2N-2n) = \frac{2(1-g)[2-g]^{n-1}}{g^{n+1}}. \quad (51)$$

Then, for $N-1$, we have

$$1/K^2(N-1,N-1) = 1 + F(2N-2) + K(N-1,N+1)F(2N). \quad (52)$$

But Eq. (35) gives

$$K(N-1,N+1) = -F(2N). \quad (53)$$

Thus, Eq. (52) becomes

$$1/K^2(N-1,N-1) = 1 + F(2N-2) - F(2N)^2. \quad (54)$$

Inserting the values of F, then yields

$$K(N-1,N-1) = g,$$

while

$$a(N) = \frac{1}{2} \frac{K(N,N)}{K(N-1,N-1)} = \frac{1}{2} \frac{\sqrt{g}}{g} = \frac{1}{2\sqrt{g}}. \quad (55)$$

Similarly, we find

$$K(N-n,N-n) = g, \quad n \geq 2,$$
$$\therefore a(N-2) = \frac{1}{2} \frac{K(N-1,N-1)}{K(N-2,N-2)} = \frac{1}{2} \frac{g}{g} = 1/2$$

and

$$a(N-n) = 1, \quad n \geq 2. \quad (56)$$

To compute the $g(n)$ explicitly we note that the $\chi(n)$ given by Eq. (37) is

$$\chi(n) = 1, \quad n = N, N+1$$
$$= g, \quad n \neq N, N+1. \quad (57)$$

Then

$$g(1) = \frac{\chi(2)\chi(4)\cdots}{\chi(3)\chi(5)\cdots} = 1. \quad (58)$$

Thus,

$$g(n) = 1, \quad n = 1, 2, \ldots N-1$$
$$= g, \quad n = N$$
$$= 1, \quad n > N,$$

i.e.,

$$g(n) = g\delta(n,N) \quad (59)$$

V. THE CONTINUUM LIMIT

Let us see how the classical results are obtained in the limit $\Delta \to 0$. We have

$$F(m) = F_c(m) + F_b(m), \quad (60)$$

where

$$F_c(m) = \frac{1}{2\pi i} \oint [1 - S(z)] z^{m-1} dz \quad (61)$$

and

$$F_b(m) = \sum_i M_i^2 z_i^m \quad (62)$$

In Eq. (61) we substitute $z = e^{i\theta}$ and obtain

$$F_c(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [1 - S] e^{im\theta} d\theta. \quad (63)$$

Now

$$\cos\theta = \lambda = 1 - E\Delta^2.$$

Solving for $\Delta \to 0$ gives

$$\theta = \pm k\Delta, \quad k = \sqrt{2E}. \quad (64)$$

Then

$$F_c(m) = \frac{1}{2\pi} \Delta \int_{-\infty}^{\infty} [1 - S] e^{im\Delta k} dk, \quad (65)$$

which in the limit $m \to \infty$, $\Delta \to 0$, $m\Delta = x$ gives

$$F_c(x) = \Delta F_c'(x), \quad \text{where } F_c'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[1 - S]e^{ikx}dk. \tag{66}$$

Similarly,

$$F_b(m) \to \Delta F_b'(x), \quad \text{where } F_b'(x) = \sum_i M_i'^2 e^{-kix}, \tag{67}$$

$$M_i'^2 = 1/\int_0^{\infty} \phi_+^2(-\epsilon_i, x)dx, \tag{68}$$

and

$$\kappa_i = \sqrt{2\epsilon_i}.$$

Then

$$F(m) \to \Delta F'(m).$$

Now Eq. (35) becomes

$$0 = K(n,m) + \Delta F'(n+m) + \sum_{n'=n+1}^{\infty} K(n,n')\Delta F'(n+m), \quad m > n. \tag{69}$$

Clearly, $K(n,m) = \Delta K'(n,m)$, where

$$0 = K'(n,m) + F'(n+m) + \sum_{n'=n+1}^{\infty} K'(n,n')F'(n+m)\Delta. \tag{70}$$

In the limit then

$$0 = K'(x,y) + F'(x,y) + \int_x^{\infty} K'(x,z)F'(t,y)dt, \quad y \geq x. \tag{71}$$

The Eq. (36) is then

$$\frac{K(\infty,\infty)}{K(n,n)^2} = 1 + \Delta\ F'(2n) + \sum_{n'=1}^{\infty} K'(n,n')F'(n',n)\Delta\ . \tag{72}$$

Since $F'(2n) \to 0$, we first see that $K(\infty,\infty) \to 1$, while substituting from Eq. (71)

$$1/K^2(n,n) \approx 1 - \Delta K'(n,n)$$

or

$$K(n,n) \approx 1 + \frac{\Delta}{2}K'(n,n). \tag{73}$$

Then using

$$a(n) = \frac{1}{2\sqrt{g(n)g(n-1)}} = \frac{1}{2}\frac{K(n,n)}{K(n-1,n-1)} \tag{74}$$

$$\ln g(n) = \ln K(n-1,n-1) - \ln K(n,n).$$

But $g(n) \approx 1 + q(x)\Delta^2$

$$\therefore\ q(x)\Delta^2 \cong \frac{\Delta}{2}[K'(n-1,n-1) - K'(n,n)],$$

which in the limit gives

$$q(x) = -\frac{1}{2}\frac{d}{dx}K'(x,x). \tag{75}$$

Furthermore, since $K(n,n) \to 1$, $\kappa(n,m) \to \Delta K'(n,m)$ for $m > n$ we see our basic integral representation becomes

$$\psi_+(E,x) = \psi_+^0(E,x) + \int_x^{\infty} K'(x,y)\psi_+^0(E,y)dy, \tag{76}$$

where

$$\psi_+^0(E,x) = e^{i\sqrt{2E}\,x}.$$

It is interesting that the explicit extraction of the term $\psi_+^0(E,x)$(i.e., a delta function in the general integral representation) is so obvious in passing from the discrete to the continuous case.

VI. CONCLUSION

A discretized version of the inverse scattering problem has been discussed. In particular, an analog of the Marchenko approach to the continuous problem is followed. All steps in the solution appear to be elementary and relatively obvious. The continuous limit is extraordinarily clear. While no claim is made to have made the continuous limit rigorous, it is hoped that the basic structure of the solution has been illuminated.

APPENDIX

The bound state normalization constants

In the text we have stated that the coefficients of the bound state terms are related to the normalization constants, i.e.,

$$\frac{z_i - z_i^{-1}}{z_i\phi_+(z_i,1)\dot{\phi}_+(z_i,0)} = \left[\sum_{n=1}^{\infty}\phi_+^2(z_i,n)\right]^{-1}. \tag{A1}$$

To prove this let us write out the equations for $\phi_+(z_i,n)$ and a $\phi_+(z',n)$ where z' is near z_i, i.e.,

$$a(n+1)\phi_+(z_i,n+1) + a(n)\phi_+(z_i,n-1) = \lambda_i\phi_+(z_i,n) \tag{A2}$$

and

$$a(n+1)\phi_+(z',n+1) + a(n)\phi_+(2',n-1) = \lambda'\phi_+(z',n). \tag{A3}$$

[Here $\lambda_i = \frac{1}{2}(z_i + z_i^{-1})$ and $\lambda' = \frac{1}{2}(z' + z'^{-1})$.] If we multiply Eq. (A2) by $\phi_+(z',n)$, Eq. (A3) by $\phi_+(z_i,n)$, subtract and sum over n from 1 to a, we obtain

$$\begin{aligned}(\lambda_i - \lambda')\sum_{n=1}^{N}\phi_+(z_i,n)\phi_+(z',n) &= a(N+1)[\phi_+(z_i,N+1)\phi_+(z',N) - \phi_+(z',N+1)\phi_+(z_i,N)] \\ &\quad - a(1)[\phi_+(z_i,1)\phi_+(z',0) - \phi_+(z',1)\phi_+(z_i,0)].\end{aligned} \tag{A4}$$

This can be considerably simplified. Thus, since the ϕ_+ vanish as $N \to \infty$, we can pass to this limit. Also, $\phi_+(z_i,0) = 0$ and $a(1) = \frac{1}{2}$:

$$\therefore\ \sum_{n=1}^{\infty}\phi_+(z_i,n)\phi_+(z',n) = \frac{1}{2}\phi_+(z_i,1)\frac{\phi_+(z',0)}{(\lambda' - \lambda)}. \tag{A5}$$

On passing to the limit $z' \to z_i$, we then obtain

$$\sum_{n=1}^{\infty}\phi_+(z_i,n)^2 = \frac{1}{2}\phi_+(z_i,1)\frac{d\phi_+(z',0)}{d\lambda'}\bigg|_{z'=z_i}. \tag{A6}$$

But,

$$\frac{d}{d\lambda'} = \frac{dz'}{d\lambda'}\frac{d}{dz'}$$

and

$$\frac{d\lambda'}{dz'} = \frac{1}{2}[1 - 1/z'^2] = \frac{1}{2z'}[z' - z'^{-1}]. \tag{A7}$$

Hence

$$\sum_{n=1}^{\infty}\phi_+(z_i,n)^2 = \frac{z_i\phi_+(z_i,1)\dot{\phi}_+(z_i,0)}{z_i - z_i^{-1}}. \tag{A8}$$

*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). In the future we will refer to this paper as I. Most of the proofs omitted here are sketched in the Appendix to I.

[2] R. Jost and W. Kohn, Phys. Rev. **87**, 977 (1952).

[3] I. M. Gelfand and B. M. Levitan, Izv. Akad. Nauk SSSR **15**, 309 (1951) [Am. Math. Soc. Transl. **1**, 253 (1956)].

[4] See, for example, Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).

[5] This is obvious if $g(n) = 1$, $n > N$, since $\phi_-(z,n)$ is then a finite polynomial in z^{-1}. The more general statement will be proved elsewhere.

[6] We also obtain an equation when $n = 0$, but this gives no information of relevance for our particular question.

[7] Actually, of course, this determines all the $K(n,n)$ only up to a common factor. Since, however $a(n)$ is given by a ratio of $K(i,i)$'s this is immaterial. It can indeed be shown that with our assumption $F(m) \to 0$ as $m \to \infty$, and then by passing to the limit $n \to \infty$ in Eq. (36), $K(\infty,\infty) = 1$ and can therefore be omitted.

Position operators in relativistic quantum theory: Analysis of algebraic operator relationships*

Bernice Durand

Department of Physics, University of Wisconsin, Madison, Wisconsin 53706
(Received 29 November 1972)

This work is the first of a series of three papers examining different aspects of position operators in relativistic quantum theory. In this paper the properties of the position operator **X** and the velocity operator **V** are derived for single particle matrix elements in the context of the Poincaré generator algebra. Both the physical meaning and the mathematical implications of each property are discussed. The algebraic structure of the extended set of relationships including the Poincaré generators, **X** and **V** is examined. It is found that this set defines an infinite algebra which is intractable mathematically. The Casimir operators of the Poincaré algebra are required to be Casimir operators for **X** and **V**, a new condition on **V** is formulated, and a simple solution for **K** is constructed. These conditions, together with familiar position operator properties, give the constraints and solutions for the extended algebra.

1. INTRODUCTION

The "correct" formulation of a position operator for relativistic quantum theory is a problem with a long history. Many of the problems of generalizing the classical notion of position to an acceptable relativistic quantum theoretical notion were defined in early works by Dirac,[1] Papapetrou,[2] Pryce,[3] Møller,[4] Newton and Wigner,[5] and Foldy and Wouthuysen.[6] There is a great deal of choice in what is meant by "position" as well as in the properties the position operator should have. For example, the "position" of a particle could be defined as its center of mass, center of energy, center of spin, center of charge, center of number density, that operator whose eigenstates are "most localized", that operator **X** whose eigenvalue is the coordinate **x**, or that operator whose trajectory moves with the classical velocity of the particle. All of these possibilities and more have been studied, many of them in Ref. 1 through 6. It will be shown in a paper to follow this present work that one position operator satisfying the algebra of this paper is the center of charge or number density for a system of free Dirac spin $\frac{1}{2}$ particles.

Having chosen a basic definition for the position operator, one still has the choice of constructing it as a 3-vector or 4-vector quantity. Its properties can then be studied, for example whether its time derivative gives a reasonable definition of a velocity operator. It has been found that all of the position operators listed above always lack at least one property considered necessary or reasonable. The criteria in this present series of papers for the necessary and reasonable properties are correct classical limits, correct relativistic behavior, correct quantum operator formulation, and correct field operator construction. Examples of these properties are that the position operator should be a polar vector, should have the canonical commutation relation with the momentum operator, should give the physical velocity operator, should transform under Lorentz transformations as a correctly defined field operator, and should have the correct behavior under charge conjugation and time reversal operations. These properties and others are the subjects of this paper and two companion papers.

One approach to studying the properties of operators is the algebraic approach. By this is meant the systematic use of commutation and anticommutation relations between operators to describe their relationships and derive new rules. Thus the commutator of the total angular momentum operator **J** with any operator tells how that operator behaves under spatial rotations, and the commutator with the momentum **P** tells how the operator behaves under a spatial translation. This approach leads to the construction of a set of relationships which may or may not define a useful algebra. The set of ten commutation relations between the Poincaré generators H, **P**, **J**, and **K** (the Hamiltonian or time translation operator, momentum or space translation operator, total angular momentum or spatial rotation operator, and pure Lorentz boost operator, respectively) is an example of a very familiar and useful Lie algebra. This algebra, supplemented by relationships for **X** and **V**, is the specific subject of this paper. The extended algebra is defined and its properties are studied.

This paper is the first of a series of three on different aspects of position operators. In the second paper a field operator will be constructed as a simple example of a usable position operator for field theory. This operator will be the Dirac spin-$\frac{1}{2}$ field operator

$$\mathbf{X} = \int d^3x\, \psi^\dagger(\mathbf{x})\mathbf{x}\psi(\mathbf{x}).$$

The purpose of constructing the simple field operator is to determine rigorously which of the algebraic properties in the present paper hold field theoretically. In particular, the "Lorentz covariance" property to be discussed in this paper has been a subject of controversy and has never been proven to hold for general spin. The field construction enables one to derive rigorously the conditions under which this property holds, and shows that the covariance property holds only for coordinate space operator densities and single particle momentum space matrix elements, but not for the operators themselves. Because of this, all algebraic relations in the present paper should be taken as conditions on the operators which appear in single particle momentum space matrix elements. The third paper will extend to general spin the results of the first two papers on algebraic and field theoretic properties of the position operator.

It should be noted that the approaches of this paper and Paper II, namely the algebraic and field operator approaches, are different from the very important formulation of position operators worked out by Newton and Wigner.[5] Newton and Wigner made assumptions about the states of a most localized system and then derived the position operator whose eigenstates were those most localized states. Here an operator is postulated to obey certain algebraic properties, with no emphasis on states. In Paper II a field operator will be construc-

Copyright © 1973 by the American Institute of Physics

ted which can be compared to the Newton–Wigner operator and whose algebraic properties can be derived. The three approaches are different but compatible.

The algebraic properties under consideration in this paper will now be introduced with a brief historical background before discussing them in detail in Sec. 2. The properties are all formulated as commutators or anticommutators, where the commutator of operators A and B is defined to be

$$[A,B] = AB - BA$$

and the anticommutator is defined to be

$$\{A,B\} = AB + BA.$$

The properties are first listed, and then discussed individually. Note indices run from 1 to 3 and bold face indicates 3-vectors.

$$[J_i, X_j] = i\epsilon_{ijk}X_k, \tag{1.1}$$

$$[H, \mathbf{X}] = -i\mathbf{V}, \tag{1.2}$$

$$[X_i, P_j] = i\delta_{ij}, \tag{1.3}$$

$$[X_i, X_j] = 0, \tag{1.4}$$

$$[X_i, K_j] = \tfrac{1}{2}\{X_j, [X_i, H]\}, \tag{1.5}$$

$$\tfrac{1}{2}\{H, \mathbf{V}\} = \mathbf{P}. \tag{1.6}$$

The first three properties require little justification. Equation (1.1) states the requirement that a position operator $\mathbf{X}$ be a 3-vector under rotations. While the parity operator $\mathcal{P}$ was omitted in this collection (as were the charge conjugation operator $\mathcal{C}$ and the time reversal operator $\mathcal{T}$), $\mathbf{X}$ clearly must be a polar vector, that is have odd parity. Equation (1.2) is the definition of the velocity operator as the time derivative of the position operator. In quantum theory the time derivative of any operator $A(t)$ is given by

$$\frac{dA(t)}{dt} = i[H, A(t)] + \frac{\partial A(t)}{\partial t}.$$

Thus $[H, \mathbf{X}]$ gives $-i\mathbf{V}$ for time-independent $\mathbf{X}$. The further properties of $\mathbf{V}$ need justification, but not this initial definition. About Eq. (1.3) there can be some debate, as will be discussed. However, it is generally accepted that the canonically conjugate coordinates q_i and momenta p_j of classical physics, which obey the Poisson bracket identity

$$[q_i, p_j]_{\mathrm{PB}} = \delta_{ij},$$

should become the canonically conjugate position operator $\mathbf{X}$ and momentum operator $\mathbf{P}$ in quantum theory, and hence obey eq. (1.3). In a 4-vector theory of position operators extra terms can be added to Eq. (1.3). These are discussed by Zmuidzinas[7] and Johnson[8], and will be mentioned in Sec. 2.

Equation (1.4) is a more debatable condition. Classically, the coordinates of a particle commute with each other. However, there is no *a priori* reason to expect the quantum operator $\mathbf{X}$ to commute with itself, provided the correct classical limit pertains. There are even classical definitions of positions which lead in the quantum mechanical case to noncommuting position operators. For example, the center of relativistic mass of a particle which is spinning shifts coordinates depending on the frame of reference from which it is viewed. For a spherically symmetric particle with spin $\mathbf{S}$ the position of the center of relativistic mass viewed by an observer moving with relative velocity $\mathbf{V}$ is shifted from the position viewed by an observer at rest relative to the particle by a factor proportional to $\mathbf{V} \times \mathbf{S}$.

This term $\mathbf{V} \times \mathbf{S}$ does not commute with itself quantum mechanically. Thus one might expect position operators for particles with spin to obey Eq. (1.4) only under very special construction. The position operator of Newton and Wigner[5] is constructed as the operator whose eigenstates are "most localized". This operator $\mathbf{X}_{\mathrm{NW}}$ obeys Eq. (1.4). For this reason Fleming[9] suggested calling the property "locality". A "local" position operator commutes with itself. In this series of papers, however, the terminology "self-commuting" is used, to avoid confusion with other meanings of the word "local". Self-commutation is taken as a reasonable requirement for $\mathbf{X}$ partly for the purpose of finding out when it fails or when it can be relaxed without losing other desirable properties of $\mathbf{X}$.

Equation (1.5) is usually called the "Lorentz covariance" condition and will be one subject of Paper II in this series. As can be seen, it involves a more complicated algebraic relationship than the preceding equations, and it is consequently harder to prove. Currie, Jordan, and Sudarshan[10] attempted to prove Eq. (1.5) for general spin position operators, but could not, as it involves a nonlinear equation in expectation values. $\mathbf{K}$ is the Lorentz boost operator, so $[X_i, K_j]$ should give the behavior of $\mathbf{X}$ under boosts. Pryce[3] derived Eq. (1.5) in the form

$$[X_i, K_j] = iX_jV_i$$

by assuming that the position of a particle behaved as a point, or alternatively as a classical coordinate. When the right side of this equation is symmetrized for quantum theory, one gets Eq. (1.5),

$$[X_i, K_j] = \frac{i}{2}\{X_j, V_i\}. \tag{1.5}$$

Many of the possible position operators do not behave as points. For example the center of relativistic mass of a spinning particle, which was already used to illustrate self-commutativity, is a "dynamic" position operator. That is, it is defined as an average over a dynamical property. Any position operator which involves a spatial integral will change as the frame in which the integral is performed changes. Pryce's[3] derivation breaks down for such "smeared out" average positions. This does not mean that they cannot be covariant, as will be shown in Paper II. It only means that imposing covariance does not lead to Eq. (1.5). Fleming[9] suggested calling this property "pointlike", as it depends on the position operator tracing out a worldline rather than a smeared out worldtube. Because the "covariance" property and self-commutativity are interrelated they have often been misinterpreted. The fact that the Newton–Wigner position operator $\mathbf{X}_{\mathrm{NW}}$ obeys Eq. (1.4) and not Eq. (1.5) has been accepted as a proof that a "local" or self-commuting position operator cannot be Lorentz covariant. However, a Dirac-like[1] position operator does have both properties. The Dirac operator $\mathbf{X}_{\mathrm{D}}$ incorporates operators referring to both positive and negative energy, while the Newton–Wigner operator $\mathbf{X}_{\mathrm{NW}}$ is restricted to act on only positive energy wave functions. The explicit relationship between these operators will

be shown in Paper II, along with their covariance properties. Again, Eq. (1.5) is assumed in this paper partly to see what constraints it imposes on solutions for all the operators.

Equation (1.6) is a new condition on the velocity operator **V**. It has usually been required that **V** should reduce to either the nonrelativistic velocity $\mathbf{V} = \mathbf{P}/M$ or the relativistic velocity $\mathbf{V} = \mathbf{P}/E$. These conditions lead to

$$[H,\mathbf{X}] = -i\frac{\mathbf{P}}{M} \quad \text{or} \quad -i\frac{\mathbf{P}}{E}.$$

Some position operators do not satisfy either of these conditions. For example, the Dirac operator $\mathbf{X}_D$ does not have a velocity operator corresponding to the classical particle velocity. The well-known Zitterbewegung is added to the classical velocity. Fortunately, the expectation value of the Dirac velocity $\mathbf{V}_D$ is still the classical velocity, but the operator $\mathbf{V}_D = \boldsymbol{\alpha}$ $(c = 1)$ is not a classical velocity operator. Equation (1.6) is a new statement of the velocity constraint

$$\tfrac{1}{2}\{\mathbf{V},H\} = \mathbf{P}, \tag{1.6}$$

which reduces to the nonquantum limit

$$\mathbf{V}E = \mathbf{P} \quad \text{or } \mathbf{V} = \mathbf{P}/E.$$

The Dirac velocity satisfies Eq. (1.6).

Some relations for **X** can be stated by considering the effect which a measurement of the position of a particle should have on the state of the particle. Thus the self-commutativity of **X** can be interpreted as saying that any two components of the position can be simultaneously measured, and the canonical commutation relation between **X** and **P** leads to the uncertainty statement about measurements of position and momentum. Two additional requirements are that a measurement of the position of a particle should not change its mass or spin. These are derived in Sec. 2 as

$$[M^2,\mathbf{X}] = 0 \tag{1.7}$$

and

$$[W^2,\mathbf{X}] = 0. \tag{1.8}$$

M^2 is the invariant mass operator defined as

$$M^2 = H^2 - \mathbf{P}^2$$

and W^2 is the invariant spin operator defined as

$$W^2 = W_0^2 - \mathbf{W}^2,$$

where

$$W_0 = \mathbf{J}\cdot\mathbf{P} \quad \text{and} \quad \mathbf{W} = -H\mathbf{J} - \mathbf{P}\times\mathbf{K}.$$

M^2 and W^2 commute with all generators of the Poincaré algebra, and are called Casimir operators. The requirement that they also commute with **X** and **V** gives new constraints on the solutions for the operators as well as information about the extended operator algebra.

In Sec. 2 these eight conditions are further discussed and the entire set of commutation and anticommutation relations which can be derived for **X** and **V**, together with the Poincaré algebra, is collected. Explicit representations of **J** and **K** are used to derive some of the conditions. A solution of the algebra is given for spin $\frac{1}{2}$.

In Sec. 3 this extended algebra is studied to see if it reveals any new physics. Recently Gürsey and Orfanides[11] have studied implications of **X** within the Poincaré algebra alone. It is found in the present paper that the more extensive algebra yields no new physics in the sense that the Poincaré algebra yielded new physics. Four mathematical limitations on studying the extended algebra are discussed in the context of the definitions of Lie and Jordan algebras.

In Sec. 4 the new results of this paper are summarized and suggestions for future work on the mathematics and on the solutions to the algebra of physical use are made.

The Appendix is a tabulation of many of the relations derived in Sec. 2. In particular, the velocity operator **V** and the spin operator **S** are compared in detail for the purpose of finding an explicit representation for **V**.

The references in this paper are not intended to be a complete span of position operator literature. Further references will of course be cited in the two remaining papers of this series. Only those papers with special relevance to algebraic properties were included in this bibliography.

2. DERIVATIONS OF ALGEBRAIC RELATIONSHIPS FOR THE POSITION AND VELOCITY OPERATORS

There are two possible approaches to the problem of finding the algebraic properties of the position operator. The first approach is to argue the physical meaningfulness of a property which can then be written as an algebraic relation. When all physically reasonable properties are collected, the mathematical and physical implications of the corresponding set of algebraic relations can be studied. For example, the physical requirement that a measurement of the position of a particle not change the mass of the particle can be written algebraically as

$$[M^2,\mathbf{X}] = 0.$$

As noted in the introduction, this condition is the mathematical statement that M^2 is a Casimir operator for the extended algebra which includes **X** as well as the Poincaré generators. The physical implications of M^2 being a Casimir operator for the extended set can then be studied.

The second approach is to postulate mathematical conditions on the operator **X**, along with the standard conditions on the Poincaré operators, then to derive further conditions from the ones which were postulated. The most elegant form of this approach consists of finding the fewest and simplest conditions which must be imposed on **X** and the other operators in order to derive all other conditions. As an example of this, self-commutativity of **X** [Eq. (1.4)], along with an explicit representation of **K**,

$$[X_i,X_j] = 0 \tag{1.4}$$

and

$$\mathbf{K} = \tfrac{1}{2}\{\mathbf{X},H\},$$

are sufficient to derive the condition [Eq. (1.5)]

$$[X_i,K_j] = \tfrac{1}{2}\{X_j,[X_i,H]\}, \tag{1.5}$$

which is sometimes taken as the Lorentz covariance requirement, but will be shown in Paper II to be a stronger requirement on **X**. A problem with this second

approach is that of choosing which conditions to postulate and which to derive. Different small sets of conditions are sufficient to derive all other conditions, and it is a matter of taste which set is considered to be more fundamental or more elegant.

The approach taken in this section is to first discuss the physical meaning of the most important conditions for **X** and **V**, then to switch to the mathematical approach and choose the set of conditions which most economically leads to all other conditions on these operators. In Paper II of this series of three papers, a field operator **X** will be defined from which all properties of **X** can be derived, so that none needs to be postulated.

Equations (1.1) through (1.5) are the five standard conditions on **X**. They give the commutation relations of **X** with the Poincaré generators H, **P**, **J**, and **K**, and with itself. Two features of this set of equations are striking. First, the Eq. (1.2)

$$[H,\mathbf{X}] = -i\mathbf{V} \tag{1.2}$$

defines a new operator **V** not in the set H, **P**, **J**, **K**, **X**. Likewise, the time derivative of **V** would lead to an "acceleration" operator **A**,

$$[H,\mathbf{V}] = -i\mathbf{A},$$

which would not be in the set H, **P**, **J**, **K**, **X**, **V**. Thus in any physical situation in which there is a velocity or an acceleration one expects to have an algebra which does not close. By contrast, the Poincaré generators all close upon themselves in their commutation relations. Second, the right hand side of Eq. (1.5)

$$[X_i,K_j] = \tfrac{1}{2}\{X_j,[X_i,H]\} \tag{1.5}$$

involves an anticommutator and is not linear in the operators, two properties which are also not characteristic of the Poincaré algebra. The whole set of **X** and **V** relationships consequently is not expected to be an easy algebra to analyze; its mathematical limitations are discussed in Sec. 3. Nonetheless, it is physically imperative that the velocity operator exist if the position operator exists, so relationships must be found for both.

There is no need to further justify Eqs. (1.1) and (1.2). Classically **X** must be a polar 3-vector and **V** must be the time derivative of **X**. In quantum language these become

$$[J_i,X_j] = i\epsilon_{ijk}X_k, \tag{1.1}$$

$$[H,\mathbf{X}] = -iV, \tag{1.2}$$

as discussed in Sec. 1.

Equation (1.3),

$$[X_i,P_j] = i\delta_{ij}, \tag{1.3}$$

has justification in the classical canonical commutation relation between q_i and p_j and in the uncertainty principle for measuring position and momentum simultaneously. However, it is not ruled out that such a term as iP_iP_j/E^2 could be added to the right side of Eq. (1.3). In fact (see Refs. 7 and 8), if the position operator is postulated to be a 4-vector instead of a 3-vector operator (a possibility which is rejected here), then Eq. (1.3) becomes

$$[X_\mu,P_\nu] = -ig_{\mu\nu},$$

where $g_{\mu\nu} = 0$ for $\mu \neq \nu$ and $g_{00} = -g_{ii} = 1$. For the 3-vectors **X** and **P** this reduces to Eq. (1.3). But for $\mu = i = 1, 2, 3$ and $\nu = 0$ this becomes

$$[X_i,H] = -ig_{i0} = 0,$$

where P_0 is the operator H. This result has the physical interpretation that the velocity is zero, which is not an interesting theory. Thus the term $iP_\mu P_\nu/E^2$ might be added to the right side of the 4-vector equation to give for $\mu = i$ and $\nu = 0$

$$[X_i,H] = iV_i = iP_iH/E^2.$$

This last equation gives a reasonable velocity, but it means that the canonical commutation relation of Eq. (1.3) is replaced by

$$[X_i,P_j] = i\delta_{ij} + iP_iP_j/E^2.$$

This extra term is rejected here as being an unnecessary complication in a 3-vector theory such as the one under construction. Along with the difficulty of interpreting X_0 as the time operator, the 4-vector formulation of the position operator leads to enough nonphysical results such as the problem of $[X_\mu,P_\nu]$ that it has been rejected in this paper. Zmuidzinas[7] and Johnson[8] have studied the 4-vector position operator. Johnson[8] discusses adding $iP_\mu P_\nu/M^2$ to the right side of Eq. (1.3) for 4-vectors and rejects the idea, thus leaving his algebra open to the criticism that it produces zero velocity.

Equation (1.4),

$$[X_i,X_j] = 0, \tag{1.4}$$

has already been defended from a measurement theory point of view, namely that measuring the x component of position should not disturb the y component. This treats the position operator as behaving the same as the momentum operator, for which

$$[P_i,P_j] = 0$$

holds. Drawing an analogy between **X** and **P**, based partly on this similarity, can be quite misleading, particularly in the study of covariance properties of field operators, as will be noted in Paper II. The position operator is not necessarily like either the momentum operator or a classical coordinate, both of which have commuting components. The example of a spinning spherically symmetric particle was cited in the introduction as a case for which the center of relativistic mass is not quantum mechanically self-commuting. There are many more quantum examples. This condition is one of the least justifiable. Because it is a simple requirement, and because there is no way to construct a nonzero right side for Eq. (1.4) without knowing **X** explicitly, for example as a function of the spin operator **S**, it will be left as a constraint condition, which will be shown to hold for the field operator **X** in Paper II.

Equation (1.5),

$$[X_i,K_j] = \tfrac{1}{2}\{X_j,[X_i,H]\} = \tfrac{i}{2}\{X_j,V_i\}, \tag{1.5}$$

is justified physically purely by analogy to classical physics.

Classically the Poisson bracket for the boost operator **K** acting on **X** is

$$[X_i, K_j]_{\mathrm{PB}} = X_j V_i.$$

This is derived, following Pryce,[3] by considering the trajectory of a point particle. Figure 1 illustrates the derivation. The particle traces out a world line $X^\mu(\tau)$, where τ is proper time. Any observer will measure a position $\mathbf{X}(t)$ where t is that observer's time parameter. The 4-vector $X^\mu(\tau)$ is

$$X^\mu(\tau) = (t(\tau), \mathbf{X}(\tau)).$$

Let two observers, primed and unprimed, be related by an infinitesimal Lorentz transformation

$$\Lambda^\mu{}_\nu = g^\mu{}_\nu + \epsilon^\mu{}_\nu.$$

Both observers measure the position of the particle at time t_0,

$$t = t' = t_0,$$

where $t = X^0(\tau)$ and $t' = X'^0(\tau')$. The usual coordinate covariance holds at any point τ' on the worldline,

$$X^\mu(\tau') = X'^\mu(\tau') + \epsilon^\mu{}_\nu X'^\nu(\tau').$$

But the variation of X^μ along the worldline is given by

$$X^\mu(\tau') = X^\mu(\tau) + \frac{dX^\mu(\tau)}{d\tau}\, d\tau.$$

These two equations for $X^\mu(\tau')$ give, to first order,

$$X'^\mu(\tau') - X^\mu(\tau) = \frac{dX^\mu}{d\tau} d\tau - \epsilon^\mu{}_\nu X^\nu.$$

The $\mu = 0$ component of this identity is just the relationship between t and t', both of which are measured at t_0. Thus

$$d\tau = \frac{\epsilon^0{}_\nu X^\nu}{dt/d\tau},$$

which can be substituted back into the identity for $\mu = i$,

$$X'^i(\tau') - X^i(\tau) = \epsilon^0{}_j \frac{dX^i}{dt} X^j - \epsilon^i{}_j X^j - \epsilon^i{}_0 X^0.$$

The term with $\epsilon^i{}_j$ is a rotation. For pure boosts

$$X'^i(\tau') - X^i(\tau) = \epsilon^0{}_j V^i X^j - \epsilon^i{}_0 t.$$

This relates the position of the particle measured by two observers at $t = t' = t_0$. Now using the quantum mechanical definition

$$X'^i(t) = e^{i\epsilon^0{}_j K^j} X^i(t)\, e^{-i\epsilon^0{}_j K^j}$$
$$\cong X^i(t) + i\epsilon^0{}_j [K^j, X^i(t)]$$

to first order, the time-dependent result is Pryce's commutator

$$[K^j, X^i(t)] = i\delta^{ij} t - iV^i X^j(t).$$

Removing explicit time dependence leads to

$$[X_i, K_j] = iX_j V_i.$$

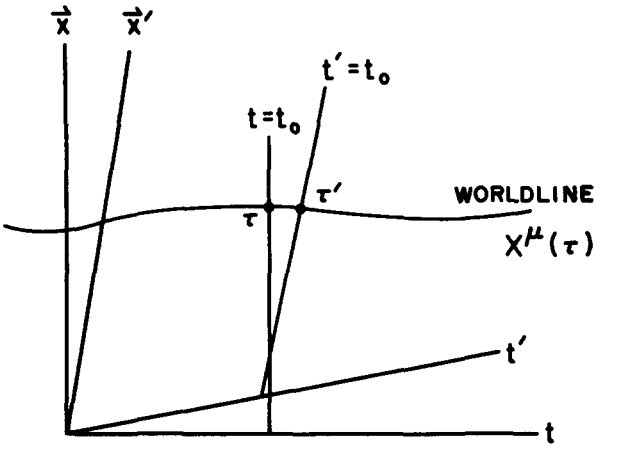

FIG. 1. The two coordinate systems $x'^\mu = (t', \mathbf{x}')$ and $x^\mu = (t, \mathbf{x})$ are related by $x^\mu = \Lambda^\mu{}_\nu x'^\nu$, where $\Lambda^\mu{}_\nu = g^\mu{}_\nu + \epsilon^\mu{}_\nu$ is an infinitesimal Lorentz transformation. $X^\mu(\tau)$ traces out the worldline of a point particle.

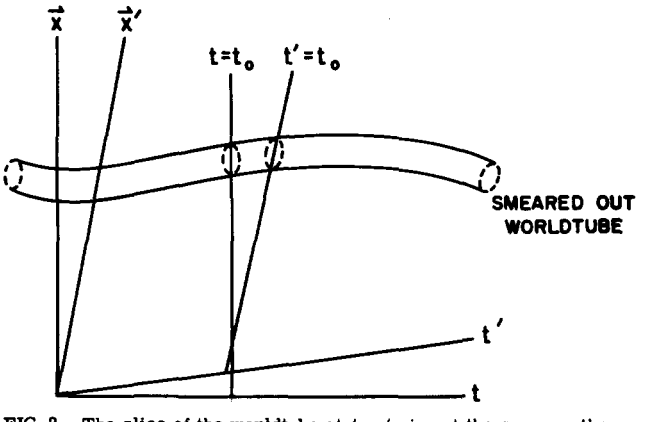

FIG. 2. The slice of the worldtube at $t = t_0$ is not the same as the slice at $t' = t_0$. Two observers will thus take constant-time averages over different slices of the distribution of the dynamical property and their average positions may not agree.

The covariance which was used in this derivation was a coordinate covariance relating two observations of a single point. Throughout the derivation it was assumed that the particle position behaved as a coordinate point.

The worldline of a point particle which obeys Eq. (1.5) is called an invariant worldline by Pryce,[3] Fleming,[9] and Currie, Jordan, and Sudarshan,[10] because two observers agree on the uniqueness of the worldline. Fleming also called the property "pointlike" rather than covariant. Any position operator defined as a spatial average over a density such as the relativistic center of mass

$$\mathbf{X}_{\mathrm{CM}} = \frac{1}{M} \int d^3q\, \mathbf{q}\rho_M(\mathbf{q})$$

will not trace out an invariant worldline. Figure 2 illustrates the different slices through the smeared out density which two observers would use in their averaging process.

In Paper II, Eq. (1.5) will be derived for a field operator density and for single particle matrix elements, but is found not to hold for the field operator **X** itself,

$$\mathbf{X} = \int d^3x \psi^\dagger(\mathbf{x}) \mathbf{x} \psi(\mathbf{x}).$$

This does not imply that the operator **X** is not Lorentz covariant, only that it does not behave as a coordinate

point. Covariance is a property which must be examined for each position operator individually. Further comments are deferred to Paper II, except to note one additional misconception about position operators, besides treating them as coordinate-like quantities. If the operator is written as a 4-vector X_μ, as in Johnson,[8] and then treated the same as the 4-momentum P_μ, one obtains the algebra

$$[X_\mu, X_\nu] = 0, \quad [P_\mu, P_\nu] = 0, \quad [X_\mu, P_\nu] = -ig_{\mu\nu},$$

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(g_{\mu\sigma}M_{\nu\rho} + g_{\nu\rho}M_{\mu\sigma} - g_{\mu\rho}M_{\nu\sigma} - g_{\nu\sigma}M_{\mu\rho}),$$

$$[M_{\mu\nu}, P_\rho] = i(g_{\nu\rho}P_\mu - g_{\mu\rho}P_\nu),$$

$$[M_{\mu\nu}, X_\rho] = i(g_{\nu\rho}X_\mu - g_{\mu\rho}X_\nu).$$

Using $K_i = M_{i0}$, the last equation yields

$$[K_i, X_j] = i\delta_{ij}X_0.$$

This is recognized as the explicitly time dependent part of Pryce's[3] identity, but the iX_iV_j term is missing. Thus even for the simple pointlike position operators, it is a mistake to form a 4-vector operator which is treated like the 4-momentum under Lorentz transformations. Equation (1.5) will be included in the operator algebra because it does hold for single particle matrix elements of some common position operators, such as the Dirac[1] operator $\mathbf{X}_D$.

Equation (1.6) is a new way of writing the condition which the relativistic velocity should satisfy,

$$\mathbf{V}E = \mathbf{P} \quad \text{(classical)},$$

$$\tfrac{1}{2}\{\mathbf{V}, H\} = \mathbf{P} \quad \text{(quantum mechanical)}. \tag{1.6}$$

Equation (1.6) has advantages over the usual condition $\mathbf{V} = \mathbf{P}/E$. First, there exist solutions to Eqs. (1.1) through (1.6), but not to Eqs. (1.1) through (1.5) plus $\mathbf{V} = \mathbf{P}/E$. Second, it is the expectation value of $\mathbf{V}$ which must be $\mathbf{P}/E$. This can be true without the operator $\mathbf{V}$ being $\mathbf{P}/E$. For example, it is true for the Dirac velocity $\mathbf{V}_D = \boldsymbol{\alpha}$. Equation (1.6) holds for the Dirac operator while $\mathbf{V}_D = \mathbf{P}/E$ does not hold.

Equations (1.7) and (1.8),

$$[M^2, \mathbf{X}] = 0 \tag{1.7}$$

and

$$[W^2, \mathbf{X}] = 0, \tag{1.8}$$

have the physical justification that a measurement of the position of a particle should not change its mass or spin. The same equations should hold when $\mathbf{V}$ replaces $\mathbf{X}$; that is, a measurement of velocity should also not alter the mass or the spin. The physically motivated properties of the position operator are summarized in Table I.

TABLE I. Physically motivated properties of the position operator $\mathbf{X}$ and velocity operator $\mathbf{V}$

Property	Equations in Text	Motivation
$[J_i, X_j] = i\epsilon_{ijk}X_k$	(1.1), (2.17)	$\mathbf{X}$ is a polar 3-vector operator.
$[H, \mathbf{X}] = -i\mathbf{V}$	(1.2), (2.21)	$\mathbf{V}$ is the time derivative of $\mathbf{X}$.
$[X_i, P_j] = i\delta_{ij}$	(1.3), (2.16)	Canonical commutation relation, uncertainty principle.
$[X_i, X_j] = 0$	(1.4), (2.18)	Simultaneous measurability of components of $\mathbf{X}$.
$[X_i, K_j] = \frac{i}{2}\{X_j, V_i\}$	(1.5), (2.23)	$\mathbf{X}$ is a coordinate- or pointlike position operator under Lorentz boosts.
$\frac{1}{2}\{\mathbf{V}, H\} = P$	(1.6), (2.38)	$\mathbf{V}$ reduces to $\mathbf{P}/E$ classically, expectation value of $\mathbf{V}$ is $\mathbf{P}/E$.
$[M^2, \mathbf{X}] = 0$	(1.7), (2.26)	Measuring $\mathbf{X}$ does not change M.
$[W^2, \mathbf{X}] = 0$	(1.8), (2.43)	Measuring $\mathbf{X}$ does not change S.

We turn now to the mathematical approach of postulating as few identities for $\mathbf{X}$ and $\mathbf{V}$ as are necessary to derive all others. This will lead to interesting constraints on the allowable solutions for all the operators. The first set of equations is the Poincaré algebra for $H, \mathbf{P}, \mathbf{J}$, and $\mathbf{K}$,

$$[H, H] = 0, \tag{2.1}$$
$$[H, \mathbf{P}] = 0, \tag{2.2}$$
$$[H, \mathbf{J}] = 0, \tag{2.3}$$
$$[H, \mathbf{K}] = -i\mathbf{P}, \tag{2.4}$$
$$[P_i, P_j] = 0, \tag{2.5}$$
$$[P_i, J_j] = i\epsilon_{ijk}P_k, \tag{2.6}$$
$$[P_i, K_j] = -i\delta_{ij}H, \tag{2.7}$$
$$[J_i, J_j] = i\epsilon_{ijk}J_k, \tag{2.8}$$
$$[J_i, K_j] = i\epsilon_{ijk}K_k, \tag{2.9}$$
$$[K_i, K_j] = -i\epsilon_{ijk}J_k. \tag{2.10}$$

The two Casimir operators of this algebra commute with all of the generators $H, \mathbf{P}, \mathbf{J}$, and $\mathbf{K}$. They are

$$M^2 = P^\mu P_\mu = H^2 - \mathbf{P}^2 \tag{2.11}$$

and

$$W^2 = W^\mu W_\mu = {W_0}^2 - \mathbf{W}^2,$$

where

$$W_0 = \mathbf{J}\cdot\mathbf{P} \tag{2.12}$$

and

$$\mathbf{W} = -H\mathbf{J} - \mathbf{P}\times\mathbf{K}. \tag{2.13}$$

Simpler forms for ${W_0}^2$ and $\mathbf{W}^2$ will be derived later.

The set of identities chosen as fundamental to deriving the $\mathbf{X}$ and $\mathbf{V}$ algebra begins with explicit representations for $\mathbf{J}$ and $\mathbf{K}$ based on both analogies to classical physics and experience with workable solutions,

$$\mathbf{J} = \mathbf{X}\times\mathbf{P} + \mathbf{S} \tag{2.14}$$

and

$$\mathbf{K} = \tfrac{1}{2}\{\mathbf{X}, H\}. \tag{2.15}$$

Two new operators have been introduced, the usual spin operator $\mathbf{S}$ and the position operator $\mathbf{X}$. One can now derive constraints on $\mathbf{X}$ and $\mathbf{S}$ such that these representations of $\mathbf{J}$ and $\mathbf{K}$ will satisfy the Poincaré algebra. A minimal set of identities for $\mathbf{X}$ and $\mathbf{S}$ (but not the only possible such set) from which all other identities are economically derived is

$$[X_i, P_j] = i\delta_{ij}, \tag{2.16}$$
$$[X_i, J_j] = i\epsilon_{ijk}X_k, \tag{2.17}$$

$$[X_i, X_j] = 0, \tag{2.18}$$

$$[S_i, P_j] = 0. \tag{2.19}$$

Again an appeal to experience (knowing $\mathbf{S}^2 = -W^2/M^2$) leads to the requirement on $\mathbf{S}^2$ that it commute with any operator A,

$$[A, \mathbf{S}^2] = 0. \tag{2.20}$$

This is easily proven for $A = \mathbf{P}, \mathbf{J}, \mathbf{K}$, or $\mathbf{X}$, but not for $A = H$. It is asserted knowing that the spin operator $\mathbf{S}$ is an $n(2s+1)$ dimensional matrix, where n is an integer, and that $\mathbf{S}^2 = s(s+1)$ is a multiple of the unit matrix and hence must commute with any other operator. The solution for $\mathbf{J}$ in Eq. (2.14) forces all operators in the set to be matrices of the same dimension as $\mathbf{S}$.

The two final equations essential for deriving the extended algebra are the definition of $\mathbf{V}$,

$$\mathbf{V} = i[H, \mathbf{X}], \tag{2.21}$$

and the Jacobi identity for commutators,

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0, \tag{2.22}$$

which holds for any three operators A, B, C.

In the Appendix these 22 equations are used to derive in a logical sequence many identities for $\mathbf{X}, \mathbf{V}$, and $\mathbf{S}$. In this section the most important identities are collected and grouped without proofs. Identities involving $\mathbf{X}$ which can be derived from the first 22 equations are

$$[X_i, K_j] = \tfrac{1}{2}\{X_j, [X_i, H]\}, \tag{2.23}$$

$$[X_i, S_j] = 0, \tag{2.24}$$

$$[X_i, V_j] = [X_j, V_i], \tag{2.25}$$

$$[M^2, \mathbf{X}] = 0, \tag{2.26}$$

$$[W^2, X_i] = [W_0{}^2 - \mathbf{W}^2, X_i] = -2iS_i \mathbf{S}\cdot\mathbf{P} + \frac{i}{2} V_i \mathbf{V}\cdot\mathbf{P} + 2iP_i \mathbf{S}^2 - \frac{i}{2} P_i \mathbf{V}^2 + \tfrac{1}{4}(\mathbf{P}^2\{V_j, [V_j, X_i]\} - P_j\{\mathbf{V}\cdot\mathbf{P}, [V_j, X_i]\}). \tag{2.27}$$

Here $W_0{}^2 - \mathbf{W}^2$ has been simplified to

$$W_0{}^2 - \mathbf{W}^2 = \mathbf{S}\cdot\mathbf{P}\ \mathbf{S}\cdot\mathbf{P} - H^2\mathbf{S}^2 - \tfrac{1}{4}[\mathbf{S}, H]\cdot[\mathbf{S}, H]. \tag{2.28}$$

Identities involving $\mathbf{V}$ and $\mathbf{S}$ are

$$[S_i, S_j] = i\epsilon_{ijk} S_k, \tag{2.29}$$

$$[V_i, V_j] = 4i\epsilon_{ijk} S_k, \tag{2.30}$$

$$[S_i, V_j] = i\epsilon_{ijk} V_k + \epsilon_{ikl} P_k [X_j, V_l], \tag{2.31}$$

$$[S_i, J_j] = i\epsilon_{ijk} S_k, \tag{2.32}$$

$$[V_i, J_j] = i\epsilon_{ijk} V_k, \tag{2.33}$$

$$[V_i, P_j] = 0, \tag{2.34}$$

$$[M^2, \mathbf{S}] = 0, \tag{2.35}$$

$$[M^2, \mathbf{V}] = 0, \tag{2.36}$$

$$[A, \mathbf{V}^2] = 0 \quad \text{for } A = H, \mathbf{P}, \mathbf{J}, \tag{2.37}$$

$$\{H, \mathbf{V}\} = 2\mathbf{P}. \tag{2.38}$$

There are many more identities for $\mathbf{S}$ and $\mathbf{V}$ collected in the Appendix, but Eqs. (2.29) through (2.37) illustrate the striking similarity in behavior between $\mathbf{V}$ and $\mathbf{S}$. This similarity, plus the condition that Eq. (2.27) reduce to

$$[W_0{}^2 - \mathbf{W}^2, \mathbf{X}] = 0,$$

suggests an explicit solution for $\mathbf{V}$ which would allow it to satisfy all of its identities and cause the right side of Eq. (2.27) to vanish. Note that Eqs. (1.1) through (1.7) have all appeared already in the postulated or derived identities. The solution to the Poincaré algebra and the constraints necessary for this to be a solution are summarized in Table II.

No identities involving $[H, \mathbf{V}]$ or $[\mathbf{X}, \mathbf{V}]$, and thus $[\mathbf{K}, \mathbf{V}]$, could be derived because the forms of $H, \mathbf{X}$, and $\mathbf{V}$ were not explicitly known. Equations (2.29) and (2.30) sug-

TABLE II. Extended algebra with constraints and solutions

Property	Equations in Text	Comments
Poincaré algebra for $H, \mathbf{P}, \mathbf{J}, \mathbf{K}$	(2.1)–(2.10)	Holds for all spin.
$M^2 = H^2 - \mathbf{P}^2$	(2.11)	Casimir operator for Poincaré algebra.
$W^2 = W_0{}^2 - \mathbf{W}^2$	(2.12), (2.13)	Casimir operator for Poincaré algebra.
$\mathbf{J} = \mathbf{X} \times \mathbf{P} + \mathbf{S}$	(2.14)	Standard solution for $\mathbf{J}$; $\mathbf{S}$ is usual spin operator, $\mathbf{X} \times \mathbf{P}$ gives orbital angular momentum operator.
$\mathbf{K} = \frac{1}{2}\{\mathbf{X}, H\}$	(2.15)	Standard classical solution for $\mathbf{K}$, gives constraints on $\mathbf{X}$.
$[X_i, P_j] = i\delta_{ij}$	(2.16), (1.3)	Constraint on $\mathbf{X}$ for $\mathbf{J}$ and $\mathbf{K}$ to satisfy Poincaré algebra.
$[X_i, J_j] = i\epsilon_{ijk} X_k$	(2.17), (1.1)	Constraint on $\mathbf{X}$ for $\mathbf{J}$ and $\mathbf{K}$ to satisfy Poincaré algebra.
$[X_i, X_j] = 0$	(2.18), (1.4)	Constraint on $\mathbf{X}$ for $\mathbf{J}$ and $\mathbf{K}$ to satisfy Poincaré algebra.
$[S_i, P_j] = 0$	(2.19)	Constraint on $\mathbf{S}$ for $\mathbf{J}$ to satisfy Poincaré algebra.
$\mathbf{V} = i[H, \mathbf{X}]$	(2.21), (1.2)	Definition needed for working with $\mathbf{X}$.
$[X_i, K_j] = \frac{1}{2}\{X_j, [X_i, H]\}$	(2.23), (1.5)	Constraint on $\mathbf{X}$ follows from (2.15) and (2.18) or vice versa.
$[M^2, \mathbf{X}] = 0$	(2.26), (1.7)	Condition on $\mathbf{X}$, derivable.
$\{H, \mathbf{V}\} = 2\mathbf{P}$	(2.38), (1.6)	Condition on $\mathbf{V}$, derivable.
$[X_i, S_j] = 0$	(2.24)	Condition on $\mathbf{X}$, derivable.
$[S_i, S_j] = i\epsilon_{ijk} S_k$	(2.29)	Derivable condition on $\mathbf{S}$, follows from (2.16) or (2.16) follows from (2.24) and (2.29).
$[V_i, V_j] = 4i\epsilon_{ijk} S_k$	(2.30)	Derivable condition on $\mathbf{V}$, suggests $\mathbf{V} = 2\rho_1 \mathbf{S}$ as one solution.
$[S_i, V_j] = i\epsilon_{ijk} V_k + \epsilon_{ikl} P_k [X_j, V_l]$	(2.31)	Derivable and suggests that $\mathbf{S}$ and $\mathbf{V}/2$ form 0(4) subalgebra if $[X_j, V_l] = 0$.
		All above equations are spin-independent.
$[W^2, \mathbf{X}] = 0$	(2.27), (2.43), (1.8)	Imposes conditions on right side of Eq. (2.27). Constraint condition which is satisfied if $\mathbf{V} = 2\rho_1 \mathbf{S}$.
$\mathbf{V} = 2\rho_1 \mathbf{S}$	(2.39)	Derived solution for half-integral spin.
$\mathbf{X} = i\nabla_p$		Derived solution for any spin.
$H = 2\rho_1 \mathbf{S}\cdot\mathbf{P} + \rho_3 M$		Holds if $\mathbf{X} = i\nabla_p$ and $\mathbf{V} = 2\rho_1 \mathbf{S}$, but gives $H^2 = E^2$ only for half-integral spin.

gest that **V** behaves as **2S**, except that **V** is a polar vector and **S** is an axial vector. There is no solution for **V** using the $2s + 1$ dimensional representation of **S**. However, a solution is easily found using a $2(2s + 1)$ dimensional representation of **S** and **V**,

$$\mathbf{V} = 2\rho_1\mathbf{S}. \tag{2.39}$$

Here ρ_1 is the $2(2s + 1)$ dimensional matrix $\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$ such that

$$\rho_1{}^2 = \mathbf{1}, \tag{2.40}$$

$$[\rho_1, \mathbf{S}] = 0, \tag{2.41}$$

$$[\rho_1, \mathbf{P}] = 0. \tag{2.42}$$

The solution $\mathbf{V} = 2\rho_1\mathbf{S}$ satisfies all **V** equations and leads to

$$[W_0{}^2 - \mathbf{W}^2, \mathbf{X}] = 0, \tag{2.43}$$

$$[W_0{}^2 - \mathbf{W}^2, \mathbf{V}] = 0, \tag{2.44}$$

and

$$[W_0{}^2 - \mathbf{W}^2, \mathbf{S}] = 0. \tag{2.45}$$

Note explicit forms for **X** and H have not yet been derived.

Careful derivations of explicit forms for **X**, **V**, and H start from the point that **S** is represented by an at least $2s + 1$ dimensional matrix, so that all other operators in the set must likewise be represented by at least $2s + 1$ dimensional matrices. There are $(2s + 1)^2$ independent $2s + 1$ dimensional matrices, and they can all be formed by taking tensor combinations of the operator **S**, up to tensors of rank $2s$. Calling these tensors $T_{kq}(\mathbf{S})$, where k indicates the rank of the tensor and q ranges from $-k$ to $+k$ [for the tensor of rank $k = 1$, such as X_q, $q = 0, \pm$ corresponds to the z and $\mp 1/\sqrt{2}$ $(x \pm iy)$ choices for the three q values], the operators **X**, **V**, and H can be expanded in terms of the $T_{kq}(\mathbf{S})$. For example, the unit matrix is a tensor of rank $k = 0$ and any vector is a tensor of rank $k = 1$. A complete representation of **X** would be

$$X_q = \sum_{k'k''} a_{k'k''} \sum_{q'q''} \begin{pmatrix} 1 & k' & k'' \\ -q & q' & q'' \end{pmatrix} T_{k'q'}(\mathbf{P})\, T_{k''q''}(\mathbf{S}).$$

In the above expression $k = 1; k' = 0, 1, 2, \ldots; k'' = 0, 1, 2, \ldots, 2s; q = z, \mp (x \pm iy); q' = -k', \ldots, +k'; q'' = -k'', \ldots, +k''$. The "3-$j$ symbol" imposes rank $k = 1$ on **X**. The tensors in P commute with the tensors in **S** and the a's are functions of scalars. The tensors $T_{kq}(\mathbf{S})$ have commutation relations with S_3 and $S^{\pm}$ given by

$$[S_3, T_{kq}(\mathbf{S})] = qT_{kq}(\mathbf{S}),$$

$$[S_\pm, T_{kq}(\mathbf{S})] = [(k \mp q)(k \pm q + 1)]^{1/2} T_{k,q\pm1}(\mathbf{S}).$$

From the above commutation relations, the identity

$$[X_i, S_j] = 0, \tag{2.24}$$

and the lack of a preferred direction one can immediately conclude that the solution for **X** is a multiple of the unit matrix $T_{00}(\mathbf{S}) = \mathbf{1}$.

Completing the solution for **X**, one uses the identities

$$[X_i, X_j] = 0 \tag{2.18}$$

and

$$[X_i, P_j] = i\delta_{ij}. \tag{2.16}$$

The most general form in the $2s + 1$ dimensional representation is found to be

$$\mathbf{X} = (i\nabla_p + i\nabla_p \ln f)\mathbf{1},$$

where f is a function of M^2, W^2, and $\mathbf{P}^2$. In a $2(2s + 1)$ dimensional space **X** can have the more general form

$$\mathbf{X} = (a\mathbf{1} + b\rho_3)(i\nabla_p + i\nabla_p \ln f).$$

In $2(2s + 1)$ dimensions three ρ matrices are introduced,

$$\rho_1 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix},$$

$$\rho_2 = \begin{pmatrix} 0 & -i\mathbf{1} \\ i\mathbf{1} & 0 \end{pmatrix},$$

$$\rho_3 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix},$$

which along with the unit matrix and the $2s + 1$ T's give $(2(2s + 1))^2$ independent matrices. The ρ_3 part of the position operator is ruled out as unphysical because it gives a different operator for positive and negative energy states, hence for particles and antiparticles. The term $i\nabla_p \ln f$ cannot be ruled out because, as an example, a change from the weight d^3p in momentum space to the weight d^3p/p_0 introduces just such a term in the position operator $i\nabla_p$, where f corresponds to $1/\sqrt{p_0}$. But as there is no physical motivation for including this extra term with a general f, the solution

$$\mathbf{X} = i\nabla_p$$

is taken as the simplest operator form, corresponding to the weight d^3p in momentum space.

In solving for **V** one uses the identities

$$[V_i, V_j] = 4i\epsilon_{ijk}S_k \tag{2.30}$$

and

$$[S_i, S_j] = i\epsilon_{ijk}S_k \tag{2.29}$$

and the fact that **V** is a polar vector while **S** is an axial vector to rule out a $2s + 1$ dimensional solution. Again there is no preferred direction so the most obvious solution for **V** is that it be proportional to $T_{1q}(\mathbf{S})$, which is just **S** itself. To obtain the correct parity either ρ_1 or ρ_2 must be introduced. Thus the most obvious solution for **V** is

$$\mathbf{V} = 2(a\rho_1 + b\rho_2)\mathbf{S},$$

where

$$a^2 + b^2 = 1.$$

A simple rotation in the ρ_1, ρ_2 space allows a or b to be eliminated, so that

$$\mathbf{V} = 2\rho_1\mathbf{S} \tag{2.39}$$

is the simplest solution. Multiplication by ρ_3 gives $\mathbf{V} = 2i\rho_2\mathbf{S}$ because of the property of the ρ matrices

$$\rho_i\rho_j = i\epsilon_{ijk}\rho_k.$$

This simple and obvious solution for **V** is unfortunately not the general solution, as will be shown.

It is now possible to further simplify the relations derived for **X**, **V**, and **S**. In particular, **X** and **V** commute,

$$[X_i, V_j] = [i\nabla_{pi}, 2\rho_1 S_j] = 0,$$

which simplifies several identities. However, it is quickly seen that one of the implications of these particular solutions is that H must be a Hamiltonian of the half-integral spin type. Because $\mathbf{X} = i\nabla_p$, the position operator picks out the coefficient of $\mathbf{P}$ in any operator with which it is commuted. Thus if the commutator of $\mathbf{X}$ with an operator is independent of $\mathbf{P}$, that operator must be linear in $\mathbf{P}$. Using the definition of $\mathbf{V}$ and this solution for $\mathbf{V}$

$$[\mathbf{X}, H] = i\mathbf{V} = 2i\rho_1\mathbf{S}$$

leads to the result that H is linear in $\mathbf{P}$,

$$H = 2\rho_1\mathbf{S}\cdot\mathbf{P} + \rho_3 M.$$

Note the solution

$$H = \rho_3(2\rho_1\mathbf{S}\cdot\mathbf{P} + \rho_3 M).$$

is rejected for H because H is the zeroth component of a 4-vector, not a scalar.

The flaw in this solution, which of course has the form of the Dirac Hamiltonian, is that

$$H^2 = 4\mathbf{S}\cdot\mathbf{P}\,\mathbf{S}\cdot\mathbf{P} + M^2$$

reduces to

$$H^2 = E^2 = \mathbf{P}^2 + M^2$$

only for spin $\frac{1}{2}$ and carefully chosen combinations of other half-integral spin tensors which reduce to a spin $\frac{1}{2}$-like term $\mathbf{S}'\cdot\mathbf{P}$ in the Hamiltonian, where $\mathbf{S}'$ is not the usual spin operator $\mathbf{S}$. This is a new approach to a result of Weinberg,[12] discussed by Nelson and Good[13] and Hammer, McDonald, and Pursey,[14] that only half-integral spin particles may have Hamiltonians linear in $\mathbf{P}$. For integral spin, $\mathbf{V}$ cannot therefore be independent of $\mathbf{P}$ and a more general solution must be found for $\mathbf{V}$ and thus the whole algebra. This will be discussed in Paper III of this series, where the general spin Foldy–Wouthuysen transformation (see Case,[15] for example) and Weinberg's[12] rules will be combined with the field theoretic results of Paper II.

Even restricting the solution to half-integral spins, the extended algebra looks interesting. For example, if $[X_i, V_j] = 0$ then equations (2.29), (2.30), and (2.31) become

$$[S_i, S_j] = i\epsilon_{ijk}S_k, \tag{2.29}$$

$$[V_i, V_j] = 4i\epsilon_{ijk}S_k, \tag{2.30}$$

$$[S_i, V_j] = i\epsilon_{ijk}V_k. \tag{2.31}$$

Then $\mathbf{S}$ and $\mathbf{V}/2$ form an 0(4) subalgebra just as $\mathbf{J}$ and $\mathbf{K}$ form an 0(3, 1) subalgebra. Also M^2 and W^2 are Casimir operators for the whole set H, $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$, $\mathbf{X}$, $\mathbf{V}$, and $\mathbf{S}$, although there might be additional Casimir operators. In Sec. 3 the mathematical structure of Eqs. (2.1) through (2.44) will be analyzed.

The Dirac position operator $\mathbf{X}_D$ and velocity operator $\mathbf{V}_D$ are a solution to Eqs. (2.1) through (2.45). The complete Dirac solution is

$$H_D = 2\rho_1\mathbf{S}\cdot\mathbf{P} + \rho_3 M, \quad \mathbf{X}_D = i\nabla_p,$$
$$\mathbf{K}_D = \tfrac{1}{2}\{H_D, \mathbf{X}_D\}, \quad \mathbf{V}_D = 2\rho_1\mathbf{S},$$
$$\mathbf{J}_D = \mathbf{X}_D \times \mathbf{P} + \mathbf{S}_D, \quad \mathbf{S}_D^2 = s(s+1) = \tfrac{3}{4}.$$

The Foldy–Wouthuysen representation of this solution is also a solution, for spin $\frac{1}{2}$.

In Paper II the Dirac position operator is constructed as the field operator center of charge or number density for a system of spin $\frac{1}{2}$ particles. The algebraic properties of this field operator can then be derived directly. It obeys Eqs. (2.1) through (2.45) in its coordinate space representation and in its momentum space single particle matrix element representation, but not as a full field operator.

3. ALGEBRAIC PROPERTIES OF THE EXTENDED SET

As mentioned earlier, it is possible to study the position operator $\mathbf{X}$ in terms of its representations within the Poincaré group. This approach has been fruitful for Gürsey and Orfanides.[11] Wightman,[16] in a discussion in depth of the localization requirements of Newton and Wigner,[5] successfully treated transformation properties of the position operator within the Poincaré covering group framework (see Sec. 5, Theorem 6, of Ref. 16). The alternative is to look for structure in Eqs. (2.1) through (2.45) which might reveal new physics as the Poincaré algebra has revealed new physics.

There are four mathematical difficulties encountered in studying the extended algebra.

A. The set does not close.

B. The commutation and anticommutation relations are not all linear in the operators.

C. There is a mixture of commutators and anticommutators.

D. Representation theory for anticommutator product algebras is much more difficult than for commutator product algebras.

These difficulties will now be discussed individually.

A. As observed early in Sec. 2 the commutator $[H, \mathbf{X}]$ introduces a new operator $\mathbf{V}$ to the set H, $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$, $\mathbf{X}$. Similarly, $[H, \mathbf{V}]$ would introduce the "acceleration" operator $\mathbf{A}$, etc. With no physical constraints on the number of nonzero time derivatives of $\mathbf{X}$, successive commutations of H with $\mathbf{X}$ introduce new operators. This can be expressed as a sequence

$$\underset{1}{[H, \mathbf{X}_0]} = -i\mathbf{X}_1 = -i\mathbf{V},$$

$$\underset{2}{[H, \underset{1}{[H, \mathbf{X}_0]}]} = (-i)^2\mathbf{X}_2 = -\mathbf{A},$$

$$\underset{n}{[H, \underset{n-1}{[\cdots \underset{1}{[H, \mathbf{X}_0]} \cdots]}]} = (-i)^n\mathbf{X}_n. \tag{3.1}$$

Technically the lack of closure of the set of operators precludes calling their relationships an algebra.[17] However, there is the possibility that a sequence as suggested above could be handled mathematically by limiting processes. It is in particular difficult to exhaust all Casimir operators for such a set, but if this were the only setback, the algebra could possibly be handled by a limiting process on the sequence (3.1).

B. Another technical difficulty with this algebra is its unusual structure constants. If a Lie algebra, such as the Poincaré algebra, has generators A_i, B_j, C_k, then relationships such as

$$[A_i, B_j] = \alpha_{ijk}C_k \tag{3.2}$$

define the algebraic structure. The coefficient α_{ijk} is called a structure constant. A Jordan algebra is defined by relationships

$$\{A_i, B_j\} = \beta_{ijk} C_k . \tag{3.3}$$

For both kinds of algebras the structure constants have only enough indices to relate the product on the left side of the equation to one operator on the right side. That is, these algebras are linear. But the extended position operator algebra has such equations as

$$[X_i, K_j] = \frac{i}{2}\{X_j, V_i\},$$

the right side of which is not linear in the operators. This corresponds to structure constants γ of the form

$$[A_i, B_j] = \gamma_{ijkl} C_k D_l. \tag{3.4}$$

There are no algebras with such structure constants classified.[17] Again, this would be a challenging mathematical problem.

C. A Lie algebra is an algebra of operators with a product defined on them which has the properties of the commutator,

$$[A_i, B_j] = \alpha_{ijk} C_k. \tag{3.2}$$

The commutator product obeys the Jacobi identity

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0. \tag{3.5}$$

A Jordan algebra is an algebra of operators with a product which has the properties of the anticommutator,

$$\{A_i, B_j\} = \beta_{ijk} C_k . \tag{3.3}$$

The anticommutator product obeys the Jordan identity

$$\{\{\{A,D\},B\},C\} + \{\{\{C,D\},B\},A\} + \{\{\{C,A\},B\},D\} = \{\{A,D\},\{B,C\}\} + \{\{C,D\},\{B,A\}\} + \{\{C,A\},\{B,D\}\}. \tag{3.6}$$

Sets of operators obeying a mixture of Eqs. (3.2) and (3.3), even those which close on themselves, are not mathematically classified.[17] This would be another interesting problem mathematically, but it does not lead to new information about the extended set of physical operators which obey both types of identities.

D. Jordan algebras were first introduced in physics by Jordan, Wigner, and Von Neumann.[18] They proved not to be useful because the ones of particular interest were "exceptional" Jordan algebras, for which there are not unique representations. It is possible that for low spin the position and velocity operators would not define an exceptional Jordan algebra if only their anticommutators were to be considered. However for spin $\frac{3}{2}$ and greater, the difficulty could probably not be circumvented. There is work in progress[17] on representations of exceptional (or nonassociative) algebras. Thus if problems A, B, and C were solved, there might be some way of getting representations for the resulting algebra.

It is unfortunate that this interesting algebra (in the physical sense) is intractable and in fact not really an algebra in the mathematical sense. Zmuidzinas[7] and Johnson[8] have derived interesting representations using the 4-vector position operator X_μ mentioned in Sec. 2.

With the usual Poincaré commutation relations between the generators $M_{\mu\nu}$ and P_ρ, X_μ is treated identically to P_μ and

$$[X_\mu, P_\nu] = -ig_{\mu\nu}.$$

However as noted in Sec. 2, the commutation relations for $\mathbf{X}$ in this case do not reduce to the correct quantum mechanical results,

$$[H, \mathbf{X}] = 0 \neq -i\mathbf{V}$$

and

$$[K_i, X_j] = i\delta_{ij}t \neq i\delta_{ij}t - iV_iX_j.$$

Thus this approach is not as physically meaningful as the 3-vector position operator construction.

Such facts as $\mathbf{S}$ and $\mathbf{V}$ forming an 0(4) subalgebra and M^2 and W^2 being Casimir operators, together with knowledge of the behavior of $\mathbf{X}$ under all Poincaré transformations, are of course useful tools in constructing representations for $\mathbf{X}$ in the Poincaré group and solutions for all of the operators for any spin.

4. SUMMARY OF RESULTS

The use of algebraic notation is a simple way to keep track of operator properties. Solutions to the operator algebra of Sec. 2 can be found for spin 0 or $\frac{1}{2}$ and should exist for higher spin, though they will in general involve spin tensors of ranks up to 2s. Some properties of the position operator might be sacrificed, such as self-commutativity or the "pointlike" condition, and the altered algebra solved for explicit representations of the operators.

The velocity condition of Eq. (2.38),

$$\{H, \mathbf{V}\} = 2\mathbf{P},$$

has not been used before, nor have all of the properties of $\mathbf{X}$ and $\mathbf{V}$ been explicitly derived from a minimal set. These derivations are in the Appendix. It has not been generally observed that Eq. (2.15) for $\mathbf{K}$ is a valid solution to the Poincaré algebra,

$$\mathbf{K} = \tfrac{1}{2}\{H, \mathbf{X}\}. \tag{2.15}$$

The solution $\mathbf{V} = 2\rho_1\mathbf{S}$ is a physically motivated solution in that one believes $[W^2, \mathbf{X}]$ should be zero. This solution leads naturally to the consideration that integral spin Hamiltonians cannot be linear in the momentum, so that velocity operators for integral spin must be more complicated than for half-integral spin. However, all equations through (2.38) should hold for any spin. The use of both Casimir operators M^2 and W^2 has not appeared before and leads to constraints on $\mathbf{X}$ and $\mathbf{V}$.

There are four interesting unsolved mathematical problems outlined in Sec. 3. The solution to any of these four could lead to interesting results, particularly when applied to a physical theory which incorporates the position operator.

Until these problems are solved, the position operator has use within the Poincaré framework (see Refs. 11 and 16) and as a field operator, as will be shown in Paper II of this series.

ACKNOWLEDGMENTS

The author is grateful to the Aspen Center for Physics for providing a place and an atmosphere in which to work and to the Atomic Energy Commission for its support during part of this research. Conversations with several people have been useful, among them D. L. Pursey, C. L. Hammer and J. M. Osborn. The most helpful advice and constructive criticisms have come from L. Durand, whose interest in this work is appreciated.

APPENDIX: TABULATION OF RELATIONS AMONG H, P, J, K, X, S, AND V

The purpose of this Appendix is to indicate how Eqs. (2.23) through (2.45) followed from Eqs. (2.1) through (2.22), as well as to list several further identities which do not belong in the basic algebra. Some of these identities are second order, such as $[H, \mathbf{X}^2]$, some are trivial but useful for gaining insight, and many have been derived exclusively for the purpose of comparing the behavior of $\mathbf{V}$ to the behavior of $\mathbf{S}$. The $\mathbf{S}$ and $\mathbf{V}$ identities are separately collected after Eq. (A75). The first 22 equations are the same as Eqs. (2.1) through (2.22). Starting with Eq. (A23), every identity can be derived solely from previous identities. This has the advantage of being logical and the disadvantage of mixing up identities in $\mathbf{X}$, $\mathbf{V}$, and $\mathbf{S}$. Thus the identities are always subgrouped according to which operator they involve and whether they are first or second order identities. The notation is to label the equations (A.1), (A.2), etc., but to refer to the equations necessary to a given proof by number only. Thus Eq. (A25) follows from Eqs. (A3), (A13), (A14), and (A15) and is written

$$\mathbf{W} = -\tfrac{1}{2}\{H, \mathbf{S}\} \quad (3, 13, 14, 15). \tag{A25}$$

It was mentioned in Sec. 2 that different identities could be assumed from those listed in the first 22 equations. Examples of this are that Eq. (A15) implies Eqs. (A26) and (A36) or vice versa, and Eq. (A16) implies (A27) and (A32) or vice versa. The general order of the equations is the most economical that could be found while avoiding any circular arguments. Identities for $\mathbf{V}$ were harder to prove than identities for $\mathbf{S}$, but even the trivial $\mathbf{S}$ identities are included to show the plausibility of the solution $\mathbf{V} = 2\rho_1\mathbf{S}$. Much of the motivation behind these proofs was to prove that W^2 is a Casimir operator for $\mathbf{X}$ and $\mathbf{V}$. The first 22 equations are grouped below.

$$[H, H] = 0, \tag{A1}$$
$$[H, \mathbf{P}] = 0, \tag{A2}$$
$$[H, \mathbf{J}] = 0, \tag{A3}$$
$$[H, \mathbf{K}] = -i\mathbf{P}, \tag{A4}$$
$$[P_i, P_j] = 0, \tag{A5}$$
$$[P_i, J_j] = i\epsilon_{ijk}P_k, \tag{A6}$$
$$[P_i, K_j] = -i\delta_{ij}H, \tag{A7}$$
$$[J_i, J_j] = i\epsilon_{ijk}J_k, \tag{A8}$$
$$[J_i, K_j] = i\epsilon_{ijk}K_k, \tag{A9}$$
$$[K_i, K_j] = -i\epsilon_{ijk}J_k, \tag{A10}$$
$$M^2 = P^\mu P_\mu = H^2 - \mathbf{P}^2, \tag{A11}$$
$$W^2 = W^\mu W_\mu = {W_0}^2 - \mathbf{W}^2,$$
$$W_0 = \mathbf{J}\cdot\mathbf{P}, \tag{A12}$$
$$\mathbf{W} = -H\mathbf{J} - \mathbf{P}\times\mathbf{K}, \tag{A13}$$
$$[H \text{ or } \mathbf{P} \text{ or } \mathbf{J} \text{ or } \mathbf{K}, M^2 \text{ or } W^2] = 0,$$
$$\mathbf{J} = \mathbf{X}\times\mathbf{P} + \mathbf{S}, \tag{A14}$$
$$\mathbf{K} = \tfrac{1}{2}\{\mathbf{X}, H\}, \tag{A15}$$
$$[X_i, P_j] = i\delta_{ij}, \tag{A16}$$
$$[X_i, J_j] = i\epsilon_{ijk}X_k, \tag{A17}$$
$$[X_i, X_j] = 0, \tag{A18}$$
$$[S_i, P_j] = 0, \tag{A19}$$
$$[A, \mathbf{S}^2] = 0 \quad \text{for any } A, \tag{A20}$$
$$\mathbf{V} = i[H, \mathbf{X}], \tag{A21}$$
$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0. \tag{A22}$$

The next 14 equations can be derived from the first 22 equations.

$$\mathbf{K} = H\mathbf{X} + \frac{i}{2}\mathbf{V} \quad \text{or}$$

$$\mathbf{K} = \mathbf{X}H - \frac{i}{2}\mathbf{V} \qquad (15, 21) \tag{A23}$$

$$W_0 = \mathbf{S}\cdot\mathbf{P} \qquad (5, 12, 14) \tag{A24}$$
$$\mathbf{W} = -\tfrac{1}{2}\{H, \mathbf{S}\}, \qquad (3, 13, 14, 15), \tag{A25}$$
$$[X_j, K_j] = \tfrac{1}{2}\{X_j, [X_i, H]\} \qquad (15, 18), \tag{A26}$$
$$[X_i, S_j] = 0 \qquad (14, 16, 17, 18), \tag{A27}$$
$$[X_i, V_j] = [X_j, V_i], \qquad (18, 21, 22), \tag{A28}$$
$$[\mathbf{P}^2, \mathbf{X}] = -2i\mathbf{P}, \qquad (16), \tag{A29}$$
$$[H, \mathbf{X}^2] = -i(\mathbf{X}\cdot\mathbf{V} + \mathbf{V}\cdot\mathbf{X}) \qquad (21), \tag{A30}$$
$$[\mathbf{S}, H] = i\mathbf{P}\times\mathbf{V} \qquad (2, 3, 14, 21), \tag{A31}$$
$$[S_i, S_j] = i\epsilon_{ijk}S_k \qquad (5, 6, 8, 14, 16, 17, 18), \tag{A32}$$
$$[\mathbf{P}, \mathbf{S}^2] = 0 \qquad (20 \text{ with } A = \mathbf{P}), \tag{A33}$$
$$[V_i, P_j] = 0 \qquad (2, 16, 21, 22) \tag{A34}$$
$$[V_i, J_j] = i\epsilon_{ijk}V_k \qquad (3, 17, 21, 22), \tag{A35}$$
$$\{H, \mathbf{V}\} = i[H^2, \mathbf{X}] = 2\mathbf{P} \qquad (1, 4, 15, 21). \tag{A36}$$

Equations (A37) through (A58) follow from the first 36 equations.

$$W_0{}^2 - \mathbf{W}^2 = \mathbf{S}\cdot\mathbf{P}\,\mathbf{S}\cdot\mathbf{P} - H^2\mathbf{S}^2 - \tfrac{1}{4}[\mathbf{S}, H]\cdot[\mathbf{S}, H] \qquad (20 \text{ with } A = H, 24, 25), \tag{A37}$$
$$[M^2, \mathbf{X}] = 0 \qquad (11, 29, 36), \tag{A38}$$
$$[W_0{}^2, \mathbf{X}] = -i\{\mathbf{S}\cdot\mathbf{P}, \mathbf{S}\} \qquad (16, 24, 27), \tag{A39}$$
$$[\mathbf{X}, \mathbf{S}^2] = 0 \qquad (20 \text{ or } 27), \tag{A40}$$
$$[\mathbf{X}, \mathbf{S}\cdot\mathbf{P}] = i\mathbf{S} \qquad (16, 27), \tag{A41}$$
$$[H^2, \mathbf{X}^2] = [\mathbf{P}^2, \mathbf{X}^2] = -2i(\mathbf{X}\cdot\mathbf{P} + \mathbf{P}\cdot\mathbf{X}) \qquad (29, 36), \tag{A42}$$
$$[J_i, S_j] = i\epsilon_{ijk}S_k \qquad (14, 19, 27, 32), \tag{A43}$$
$$[K_i, S_j] = \tfrac{i}{2}, \{X_i, (\mathbf{V}\times\mathbf{P})_j\} \qquad (15, 21, 27, 31), \tag{A44}$$
$$[S_i, V_j] = [X_j, (\mathbf{P}\times\mathbf{V})_i] = i\epsilon_{ijk}V_k + \epsilon_{ikl}P_k[X_j, V_l] \qquad (21, 22, 27, 31), \tag{A45}$$
$$[\mathbf{S}, H^2] = 0 \qquad (31, 36), \tag{A46}$$
$$[\mathbf{S}, \mathbf{S}^2] = 0 \qquad (32), \tag{A47}$$
$$i\mathbf{P}\cdot(\mathbf{V}\times\mathbf{S} - \mathbf{S}\times\mathbf{V}) = 0 \qquad (20 \text{ with } A = H, 31), \tag{A48}$$

$[H,[\mathbf{X},\mathbf{S}^2]]=0$ (20 or 27), (A49)
$[H^2,[\mathbf{X},\mathbf{S}^2]]=0$ (20 or 27), (A50)
$[\mathbf{P}^2,[\mathbf{X},\mathbf{S}^2]]=0$ (20 or 27), (A51)
$[M^2,[\mathbf{X},\mathbf{S}^2]]=0$ (20 or 27), (A52)
$[V_i,V_j]=4i\epsilon_{ijk}S_k$ (1, 10, 14, 15, 18, 21, 36), (A53)
$[\mathbf{P}^2,\mathbf{V}]=0$ (34), (A54)
$[H,\mathbf{V}^2]=0$ (21, 36), (A55)
$[\mathbf{P},\mathbf{V}^2]=0$ (34), (A56)
$[\mathbf{J},\mathbf{V}^2]=0$ (35), (A57)
$\{V_i,V_j\}+i\{H,[V_i,X_j]\}=2\delta_{ij}$ (16, 21, 36). (A58)

Equations (A59) through (A75) follow from the first 58 equations and from each other.

$$[W_0{}^2-\mathbf{W}^2,X_i]=-2iS_i\mathbf{S}\cdot\mathbf{P}+\frac{i}{2}V_i\mathbf{V}\cdot\mathbf{P}+2iP_i\mathbf{S}^2-\frac{i}{2}P_i\mathbf{V}^2+\tfrac{1}{4}(\mathbf{P}^2\{V_j,[V_j,X_i]\}-P_j\{\mathbf{V}\cdot\mathbf{P},[V_j,X_i]\})$$ (32, 37, 39), (A59)

$[M^2,\mathbf{S}]=0$ (14, 38, and $[M^2,\mathbf{J}$ or $\mathbf{P}],=0$) (A60)
$[\mathbf{P}^2,\mathbf{S}^2]=0$ (19 or 20 or 33), (A61)
$[M^2,\mathbf{V}]=0$ (22, 38, and $[M^2,H]=0$), (A62)
$[X_i,\mathbf{V}^2]=\{V_j,[X_j,V_i]\}=8i(\mathbf{X}\times\mathbf{S})_i+[\mathbf{X}\cdot\mathbf{V}+\mathbf{V}\cdot\mathbf{X},V_i]$ (18, 21, 22, 28, 53), (A63)
$[\mathbf{V},\mathbf{V}^2]=4i(\mathbf{V}\times S-S\times V)$ (53), (A64)
$[\mathbf{P}^2,\mathbf{V}^2]=0$ (54), (A65)
$[H,[\mathbf{X},\mathbf{V}^2]]=4(\mathbf{V}\times\mathbf{S}-\mathbf{S}\times\mathbf{V})$ (21, 22, 53, 55), (A66)
$[\mathbf{J},\mathbf{S}^2]=0$ (20 or 43), (A67)
$[M^2,\mathbf{S}^2]=0$ (20 or 60), (A68)
$[H^2,\mathbf{V}]=0$ (54, 62), (A69)
$[M^2,\mathbf{V}^2]=0$ (62), (A70)
$[\mathbf{P}^2,[\mathbf{X},\mathbf{V}^2]]=0$ (22, 29, 34, 65), (A71)
$[H^2,\mathbf{S}^2]=0$ (20, or 11, 61, 68), (A72)
$[H^2,\mathbf{V}^2]=0$ (11, 65, 70), (A73)
$[M^2,[\mathbf{X},\mathbf{V}^2]]=0$ (22, 38, 70), (A74)
$[H^2,[\mathbf{X},\mathbf{V}^2]]=0$ (11, 71, 74). (A75)

Before stating the explicit solution for $\mathbf{V}$ which allows W^2 to be a Casimir operator for $\mathbf{X}$ [that is, Eq. (A59) to become $[W_0{}^2-\mathbf{W}^2,X]=0$], most of the $\mathbf{S}$ and $\mathbf{V}$ identities are collected for comparison.

$[H,\mathbf{V}]$ unknown,
$[H,\mathbf{S}]=i\mathbf{V}\times\mathbf{P}$ (31),
$[P_i,V_j$ or $S_j]=0$ (19, 34),
$[J_i,V_j$ or $S_j]=i\epsilon_{ijk}(V_k$ or $S_k)$ (35, 43),
$[K_i,V_j]$ unknown
$[K_i,S_j]=\frac{i}{2}\{X_i,(\mathbf{V}\times\mathbf{P})_j\}$ (44),
$[X_i,V_j]=[X_j,V_i]$ (28),
$[X_i,S_j]=0$ (27),
$\frac{1}{4}[V_i,V_j]=[S_i,S_j]=i\epsilon_{ijk}S_k$ (32, 53),
$[S_i,V_j]=i\epsilon_{ijk}V_k+i\epsilon_{ikl}P_k[X_j,V_l]$ (45),
$[H$ or $\mathbf{P}$ or $\mathbf{J},\mathbf{V}^2$ or $\mathbf{S}^2]=0$ (20, 55, 56, 57),
$[\mathbf{K},\mathbf{V}^2]$ unknown,
$[\mathbf{K},\mathbf{S}^2]=0$ (20),
$[H^2$ or $\mathbf{P}^2$ or $M^2,[\mathbf{X},[\mathbf{V}^2$ or $\mathbf{S}^2]]=0$ (50, 51, 52, 71, 74, 75),
$[H,[\mathbf{X},\mathbf{V}^2]]=4(\mathbf{V}\times\mathbf{S}-\mathbf{S}\times\mathbf{V})$ (66),
$[H,[\mathbf{X},\mathbf{S}^2]]=0$ (49),
$[H^2$ or $\mathbf{P}^2$ or $M^2,\mathbf{V}$ or $\mathbf{S}]=0$ (46, 54, 60, 61, 62, 69),
$[\mathbf{V},\mathbf{V}^2]=4i(\mathbf{V}\times\mathbf{S}-\mathbf{S}\times\mathbf{V})$ (64),
$[\mathbf{S},\mathbf{S}^2]=0$ (47),
$-i\mathbf{P}\cdot(\mathbf{V}\times\mathbf{S}-\mathbf{S}\times\mathbf{V})=0$ (48).

It can easily be seen that $\mathbf{V}$ behaves very similarly to $\mathbf{S}$. All identities involving $[H,\mathbf{V}]$ or $[X_i,V_j]$ are unknown. The representation of $\mathbf{V}$ which was derived in Sec. 2, along with conditions on ρ_1 and the remaining provable identities, follow.

$\mathbf{V}=2\rho_1\mathbf{S}$, (A76)
$[\rho_1,\mathbf{P}]=0$, (A77)
$[\rho_1,\mathbf{S}]=0$, (A78)
$\rho_1{}^2=\mathbf{1}$, (A79)
$[W_0{}^2-\mathbf{W}^2,\mathbf{X}]=0$ (22, 27, 40, 41, 59, 76, 77, 78, 79), (A80)
$[W_0{}^2-\mathbf{W}^2,\mathbf{V}]=0$ (21, 22, 80, $[W_0{}^2-\mathbf{W}^2,H]=0$), (A81)
$[W_0{}^2-\mathbf{W}^2,\mathbf{S}]=0$ (14, $[W_0{}^2-\mathbf{W}^2$, $\mathbf{P}$ or $\mathbf{J}]=0$), (A82)
$\mathbf{V}^2=4\mathbf{S}^2$ (76, 78, 79), (A83)
$[X_i,\mathbf{V}^2]=0$ (40, 83), (A84)
$\mathbf{V}\times\mathbf{S}-\mathbf{S}\times\mathbf{V}=0$ (76, 78) (A85)
$[H,\mathbf{V}]=2[H,\rho_1]\mathbf{S}+4i\mathbf{S}\times\mathbf{P}$ (31, 76, 79), (A86)
$[S_i,V_j]=i\epsilon_{ijk}\mathbf{V}_k$ (32, 76, 78), (A87)
$\{V_j,[X_j,V_i]\}=0$ (63, 84), (A88)
$i\epsilon_{ikl}P_k[X_j,V_l]=0$ (45, 87). (A89)

Note that both M^2 and W^2 are now Casimir operators for $\mathbf{X}$, $\mathbf{V}$, and $\mathbf{S}$ (Eqs. 38, 60, 62, 80, 81, 82). The proof that $W_0{}^2-\mathbf{W}^2$ commutes with $\mathbf{X}$ could have been attacked by first showing that it commutes with $\mathbf{S}$, but this leads to the same difficulties as what has been done here.

Introducing the explicit form of $\mathbf{X}$ which was derived in Sec. 2,

$$\mathbf{X}=i\nabla_p,$$ (A90)

leads to

$$[X_i, V_j] = 0 \quad (76, 90) \tag{A91}$$

and further restrictions on H which reduce the whole algebra to essentially a spin $\frac{1}{2}$ algebra, as discussed in Sec. 2.

---

*This work was supported in part by the U.S. Atomic Energy Commission under contract AT(11−1)−881,COO−881−345.

[1]P. A. M. Dirac, *The Principles of Quantum Mechanics*, (Clarendon, Oxford, 1930).
[2]A. Papapetrou, Praktika de l'Academie d'Athènes, **14**, 540 (1939) and Praktika de l'Academie d'Athènes, **15**, 404, (1940).
[3]M. H. L. Pryce, Proc. R. Soc. **A195**, 62 (1948).
[4]C. Møller, Communications from Dublin Institute for Advanced Studies, Ser. A., No. 5, 1949 (unpublished).
[5]T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[6]L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
[7]J. S. Zmuidzinas, Jet Propulsion Laboratory Technical Report No. 32−797, 1965.
[8]J. E. Johnson, Phys. Rev. **181**, 1755 (1969).
[9]G. N. Fleming, Phys. Rev. B **137**, 188 (1965).
[10]D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).
[11]F. Gürsey and S. Orfanides, Nuovo Cimento **11A**, 225 (1972).
[12]S. Weinberg, Phys. Rev. B **133**, 1318 (1964).
[13]T. J. Nelson and R. H. Good, Jr., Rev. Mod. Phys. **40**, 508 (1968).
[14]C. L. Hammer, S. C. McDonald, and D. L. Pursey, Phys. Rev. **171**, 1349 (1968).
[15]K. M. Case, Phys. Rev. **95**, 1323 (1954).
[16]A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[17]J. M. Osborn, Department of Mathematics, University of Wisconsin at Madison, private communication.
[18]P. Jordan, J. Von Neumann, and E. P. Wigner, Ann. Math. **35**, 29 (1934). The use of Jordan algebras in physics was short-lived due to the problem of representations of exceptional Jordan algebras. There is an extensive mathematics literature on the subject. The book *Structure and Representations of Jordan Algebras* (American Mathematical Society, Providence, Rhode Island, 1968) by N. Jacobson has an excellent bibliography.

Equivalent Lagrangians in generalized mechanics

D. Anderson

Research Laboratory of Electronics, Chalmers University of Technology, Gothenburg, Sweden
(Received 23 October 1972)

A correspondence is established between differential equations and equivalence classes of Lagrangians containing higher derivatives. This suggests that a disputed problem of quantization in generalized mechanics is due to an inconvenient choice of Lagrangians characterizing the variational problems. These general concepts are then applied to the set of Lagrangians bilinear in the field variable and its derivatives.

1. INTRODUCTION

There have recently appeared several articles concerning the problem of quantization in generalized mechanics.[1-4] The aim has been to set up a Hamiltonian formalism for the case of a Lagrangian depending upon higher order derivatives of the generalized coordinate, and to develop a consistent quantization scheme. The gist of the problem is that it is necessary to find pairs of generalized canonical variables, whereas the Hamiltonian very well might be a function of an uneven number of variables. The problem has been discussed extensively by Hayes.[1,2] His treatment has been criticized by Ryan,[3] who suggests that the proper way to treat this difficulty is to pass to a new equivalent Lagrangian, which avoids the problem. Further difficulties associated with Hayes quantization procedure are discussed by Tesser.[4]

In the present article we will show, that it is possible to place the method suggested by Ryan[3] within a more general framework. A one-to-one correspondence is established between variational equations and equivalence classes of Lagrangians. We make the plausible assumption that the physical content of the variational equation is a characteristic of the corresponding equivalence class. This subsequently leads to the conclusion that problems connected with one representative of the equivalence class, as opposed to another, implies that the former Lagrangian is an inconvenient choice and should be replaced by the latter.

A specific application is made to the set of ordinary linear differential equations expressible by means of a variational principle. The corresponding equivalence classes of Lagrangians are shown to be characterized by uniquely determined Lagrangians of an especially simple form. The corresponding Hamiltonians are derived explicitly in terms of generalized canonical variables. Finally some characteristic properties of the differential equations and the Lagrangians are derived.

2. EQUIVALENT LAGRANGIANS

For simplicity we consider systems described by Lagrangians of the form

$$L = L(t, x, Dx, \ldots, D^m x), \tag{2.1}$$

where $D = d/dt$ and $x = x(t)$ is a generalized coordinate function.

The Euler–Lagrange variational equation corresponding to Eq. (2.1) is

$$\frac{\delta L}{\delta x} = \sum_{k=0}^{m} (-D)^k \left(\frac{\partial L}{\partial (D^k x)} \right) = 0 \tag{2.2}$$

which constitutes a differential equation for x. We might say that the Lagrangian is mapped on the differential equation by means of the variational procedure. However, it is well known, that this mapping is not one-to-one, i.e., the differential equation does not uniquely determine the Lagrangian.

A unique correspondence can be established as follows: Define an equivalence relation on the set of Lagrangians by saying that two Lagrangians L_1 and L_2 are equivalent (written $L_1 \equiv L_2$), if they yield the same differential equation. We then have a one-to-one correspondence between equivalence classes and differential equations admitting of a variational characterization.

But is is evident that the physical content of a differential equation is a characteristic of the corresponding equivalence class, i.e., it does not depend upon the specific Lagrangian chosen to represent the class, cf. the notion of a function in the Lesbegue theory of integration. This implies that if we encounter problems when using a certain Lagrangian and these difficulties do not occur for another Lagrangian in the same class, then these problems have no implications of physical nature. They are only due to an inconvenient representation of the class.

One question remains: Is it always possible to find in every equivalence class a Lagrangian, which avoids the difficulties mentioned above? The answer is yes, as is proved by the investigations made by Ryan.[3]

3. APPLICATIONS

In order to make the preceding discussion more explicit we shall apply it to a class of Lagrangians, which are somewhat restricted, but still sufficiently general to be of importance. Consider the set of all ordinary, linear and homogeneous differential equations with smooth coefficient-functions and which, furthermore, are expressible in terms of a variational principle. The corresponding Lagrangians must be bilinear in the generalized coordinate x and its derivatives, i.e.,

$$L = \sum_{j,k} a_{j,k} D^j x D^k x, \tag{3.1}$$

where $a_{j,k} = a_{j,k}(t)$ are smooth functions for all j and k.

According to our previous analysis, there are among the Lagrangians given by Eq. (3.1) many that will yield the same differential equations. These are grouped together into equivalence classes. We will now show, that in every equivalence class there is a uniquely determined representative which has an especially simple form.

We note first that if $j > k$, then

$$D^j x D^k x = \begin{cases} D[\tfrac{1}{2}(D^k x)^2] & \text{if } j = k+1, \\ D(D^{j-1} x D^k x) - D^{j-1} x D^{k+1} x & \text{if } j > k+1. \end{cases} \tag{3.2}$$

Copyright © 1973 by the American Institute of Physics

Repeated use of this formula yields

$$D^jxD^kx \equiv \begin{cases} 0 & \text{if } j \not\equiv k \pmod 2, \\ (-1)^{(j-k)/2}\,(D^{(j+k)/2}x)^2 & \text{if } j \equiv k \pmod 2, \end{cases} \tag{3.3}$$

where we have used the fact that a total derivative does not contribute to the variational equation.

Equation (3.3) implies that

$$L = \sum_{j,k} a_{j,k} D^jxD^kx \equiv \frac{1}{2}\sum_{k=0}^{N} a_k (D^kx)^2. \tag{3.4}$$

$a_k = a_k(t)$ are new coefficient functions determined by the $a_{j,k}$:s.

We conclude that every bilinear Lagrangian as given above has an equivalent uniquely determined "normal" form, which is quadratic in the generalized coordinate and its derivatives. Conversely, it will be shown below that the normal Lagrangian is uniquely determined by the equivalence class.

Next we will establish the Hamiltonian formalism for the normal Lagrangians and show that the disputed problem does not occur for this case. The generalized momenta, P_n, and the Hamiltonian, H, are, e.g., Hayes,[1]

$$\begin{aligned} P_n &= \sum_{j=0}^{N-n} (-D)^j \left(\frac{\partial L}{\partial (D^{n+j}x)}\right), \\ H &= \sum_{n=1}^{N} P_n D^n x - L. \end{aligned} \tag{3.5}$$

Define new variables, Q_n, as follows

$$Q_{n+1} = D^n x, \quad n = 0, 1, \ldots, N-1. \tag{3.6}$$

We want to show that $\{Q_n, P_n\}_{n=1}^{N}$ act as conjugate variables and that H can be expressed explicitly in these variables.

Note first that

$$\begin{aligned} P_N &= \frac{\partial L}{\partial (D^N x)} = a_N D^N x, \\ L &= \frac{1}{2}\sum_{j=0}^{N-1} a_j Q_{j+1}^2 + \frac{1}{2} a_N (D^N x)^2, \end{aligned} \tag{3.7}$$

Then we get for H

$$H = -\frac{1}{2} a_0 Q_1^2 + \sum_{n=1}^{N-1}\left(P_n Q_{n+1} - \frac{1}{2} a_n Q_{n+1}^2\right) + \frac{1}{2a_N} P_N^2 \tag{3.8}$$

($a_N \neq 0$, see comment below).

By means of the recurrence formulas

$$\begin{aligned} DP_{n+1} &= \frac{\partial L}{\partial (D^n x)} - P_n, \quad n = 1, 2, \ldots, N-1, \\ DP_1 &= \frac{\partial L}{\partial x}, \\ Q_{n+1} &= DQ_n, \quad n = 1, 2, \ldots, N-1 \end{aligned} \tag{3.9}$$

it is easy to show that H as given by Eq. (3.8) yields the Hamiltonian equations, viz.

$$\begin{aligned} \frac{\partial H}{\partial P_n} &= DQ_n, \\ \frac{\partial H}{\partial Q_n} &= -DP_n, \quad n = 1, 2, \ldots, N. \end{aligned} \tag{3.10}$$

Finally we will consider the example studied by Hayes.[1,2] The equation of motion of the harmonic oscillator is

$$mD^2x + kx = 0. \tag{3.11}$$

The Lagrangian used by Hayes is

$$L = -(m/2)xD^2x - (k/2)x^2. \tag{3.12}$$

This choice of Lagrangian leads to quantization difficulties, cf. 1-4. However, it is easily seen that L belongs to the equivalence class, whose normal form is the conventional Lagrangian of the problem, i.e.,

$$L = -\frac{m}{2} xD^2x - \frac{k}{2} x^2 \equiv \frac{m}{2}(Dx)^2 - \frac{k}{2} x^2 = L_2. \tag{3.13}$$

The Hamiltonian formalism based on L_2 represents no problems as is well known.

4. FURTHER COMMENTS

In this section we will point out some consequences of the normal form, Eq. (3.4): It has a unique simplicity, because the coefficient matrix $A = (a_{j,k})$ which characterizes the Lagrangian of Eq. (3.1) becomes a diagonal matrix with $a_k/2$ as the diagonal entries. Furthermore, define the order of a Lagrangian (ord L) to be the order of the highest appearing derivative, or which is equivalent, the dimension of the matrix A. Then we have the relation

$$\text{ord } L_N = \min_{L \in \{L_N\}} (\text{ord } L). \tag{4.1}$$

Here L_N denotes the normal Lagrangian and $\{L_N\}$ the equivalence class determined by L_N. In this context we note that, cf. Eq. (3.4), if $a_N = 0$ for same value of t, then ord $L_N < N$ at this point. Thus, the condition expressed in connection with Eq. (3.8) is implied by the assumption that the normal Lagrangian L_N does not degenerate at any point i.e., that ord $L_N = N$ for all t.

We can also draw some conclusions about the corresponding differential equations.

The variation equation of L_N is

$$\frac{\delta L_N}{\delta x} = \sum_{k=0}^{N} (-D)^k [a_k D^k x] = \sum_{k=0}^{N} D^k [a_k' D^k x] = 0, \tag{4.2}$$

where we have put $(-1)^k a_k = a_k'$

The last expression of Eq. (4.2) implies that the equation is selfadjoint and of even order, cf. Darboux.[5] {It can be regarded as a generalization of the Sturm–Liouville equation $D[p(t)Dx(t)] + q(t)\,x\,(t) = 0$.} Thus we have the result that every linear, ordinary differential equation obtainable from a variational principle must be of even order and selfadjoint. Furthermore, since the selfadjoint form of the differential equation is unique we conclude that the equivalence class uniquely determines the corresponding normal Lagrangian.

ACKNOWLEDGMENT

The author wishes to express his sincere thanks to Dr. J. Askne for reading the manuscript and making several fruitful suggestions.

APPENDIX

It is well known, see, e.g., Ref. 6, that if we add a total derivative to a Lagrangian this does not affect the corresponding variational equation. This result is conventionally established under the assumption that the function whose total derivative we add to the Lagrangian does not depend on derivatives of the generalized coordinate. However, in the context of generalized mechanics this restriction can be discarded since its only purpose is to ensure that the contribution to the Lagrangian does not contain derivatives of higher order than first. In this appendix we will give a simple direct proof of the invariance of the variational derivative under the so called divergence transformation of the Lagrangian without the functional restriction mentioned above, cf. Ref. 7.

Thus suppose that $F(t, x(t), \ldots, D^n x(t)) = D\Phi(t, x(t), \ldots, D^{n-1}x(t))$. Then we assert that the variational derivative of F vanishes identically, i.e.,

$$\frac{\delta F}{\delta x} \equiv 0.$$

Proof: For the total derivative of Φ we have

$$D\Phi = \frac{\partial \Phi}{\partial t} + \sum_{k=1}^{n} \frac{\partial \Phi}{\partial (D^{k-1}x)} D^k x \tag{A1}$$

The variational derivative of F is

$$\frac{\delta F}{\delta x} = \sum_{m=0}^{n} (-D)^m \left[\frac{\partial F}{\partial (D^m x)}\right]. \tag{A2}$$

With the use of (1) we obtain

$$\frac{\delta F}{\delta x} = \sum_{m=0}^{n-1} (-D)^m \left(\frac{\partial^2 \Phi}{\partial t \partial (D^m x)} + \sum_{k=1}^{n} \frac{\partial^2 \Phi}{\partial (D^{k-1}x)\partial (D^m x)} D^k x\right) + \sum_{m=1}^{n} (-D)^m \left(\frac{\partial \Phi}{\partial (D^{m-1}x)}\right). \tag{A3}$$

However, the first two terms can be grouped together into a total derivative as follows:

$$\frac{\delta F}{\delta x} = -\sum_{m=0}^{n-1} (-D)^{m+1} \left(\frac{\partial \Phi}{\partial (D^m x)}\right) + \sum_{m=1}^{n} (-D)^m \left(\frac{\partial \Phi}{\partial (D^{m-1}x)}\right) \tag{A4}$$

and with a change of summation variable in the first sum the variational derivative of F vanishes identically.

[1] C. F. Hayes, Nuovo Cimento B **58**, 494 (1968).
[2] C. F. Hayes, J. Math. Phys. **10**, 1555 (1969).
[3] C. Ryan, J. Math. Phys. **13**, 283 (1972).
[4] H. Tesser, J. Math. Phys. **13**, 796 (1972).
[5] G. Darboux, *Leçons sur la théorie générale des surfaces* (Gauthier-Villars, Paris, France, 1915), Vol. 2.
[6] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[7] E. Noether, Nachr. Kgl. Ges. Wiss. Göttingen, 235 (1918); a translation is given by M. Tavel in Transp. Theory and Stat. Phys. **1**, 183 (1971).

Thermodynamics of an anisotropic boundary of a two-dimensional Ising model

Helen Au-Yang*

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11790
(Received 15 December 1972)

We consider a semi-infinite two-dimensional Ising model with its spins on the boundary row having a different interaction energy E'_1 from the ferromagnetic bulk. We find that the boundary specific heat has two divergent terms: one of which diverges linearly at the bulk critical temperature T_c, and the other, logaritmically. The linearly divergent term is independent of E'_1, and the coefficient of the logaritmically divergent term is a decreasing function of E'_1. There is a boundary latent heat at T_c, which is identical to McCoy and Wu's result. The boundary spins, which can be either ferromagnetic or antiferromagnetic, are aligned for temperature lower than T_c. The boundary spontaneous magnetization approaches zero in the form of $A(E'_1)(1-T/T_c)^{1/2}$, and the boundary zero field magnetic susceptibility diverges at T_c in the form $-B(E'_1)\ln|1-T/T_c|$, where $A(E'_1)$ and $B(E')$ are increasing functions of E'_1.

1. INTRODUCTION

In 1967, McCoy and Wu[1] considered a semi-infinite rectangular Ising lattice. They obtained a term, in the total free energy, which is proportional to the number of lattice sites on the boundary. This term is defined to be the boundary free energy, from which they obtained the expressions for the boundary specific heat and the boundary magnetization. They found that the critical temperature of the boundary is the same as the bulk's; the specific heat has two terms, one of which diverges logarithmically at T_c and the other diverges linearly at T_c, and the spontaneous magnetization of the boundary spins approaches zero in the form of $(1-T/T_c)^{1/2}$ as $T \to T_c$.

More recently Hohenberg and collaborators[2] have considered boundary effects of a three-dimensional Ising model. They consider the situation where the interaction within the surface (E') is different from the interaction energy in the bulk. Their computation is done by means of series expansions.

In this paper we consider the two-dimensional version of the problem of Hohenberg *et al.*; i.e., we modify the half plane problem of McCoy and Wu by allowing the interaction energy in the boundary row E'_1 to be different from the interaction energy of E_1 of the other rows.

We find that the boundary specific heat has two divergent terms; one of which diverges linearly at the bulk critical temperatures T_c, and the other, logarithmically. The linearly divergent term is identical to the linearly divergent term obtained by McCoy and Wu[1] for the case $E'_1 = E_1$. The coefficient of the logarithmically divergent term is a decreasing function of E'_1. The fact that the singularity which characterizes the boundary specific heat remains the same for different values of E'_1 is indicated by Fisher and Ferdinand.[3] Moreover, we find that there is a boundary latent heat at T_c which is independent of E'_1.

We find that the spins on the boundary row are aligned at all temperature only for $E'_1 = \infty$. Otherwise, the spontaneous magnetization of the boundary spins, which can be either ferromagnetic or antiferromagnetic, is nonzero for $T < T_c$; it vanishes at the bulk critical temperature T_c in the form of $A(E'_1)(1-T/T_c)^{1/2}$. The magnetic susceptibility is divergent only when the applied field h is zero. We find it diverges logarithmically with a coefficient that depends upon E'_1. In other words, we find that the critical indices do not depend upon E'_1.

The evaluation of the boundary free energy, for the above system with the field h applied only on the boundary layer, is carried out in Sec. 2. It is given as a contour integral.

In Sec. 3, we describe in detail our method of finding the divergent terms in the specific heat, which is more systematic than that of McCoy and Wu.[1] We differentiate the integrand, and then convert the integral to complete elliptical integrals.

In Sec. 4, by letting $h \to 0$, we calculate the boundary spontaneous magnetization and the boundary magnetic susceptibility using the method described in Sec. 3.

Finally, we conclude with a qualitative discussion of the results and use our calculation to speculate about the problem of a two-dimensional layer on a three-dimensional bulk.

2. PARTITION FUNCTION AND FREE ENERGY

Consider a two-dimensional rectangular Ising lattice of $2M$ rows and $2N$ columns, whose Hamiltonian is

$$\epsilon(\Delta) = -E_1 \sum_{j=2}^{2M} \sum_{k=-N+1}^{N} \sigma_{j,k}\sigma_{j,k+1} - E'_1 \sum_{k=-N+1}^{N} \sigma_{1,k}\sigma_{1,k+1} - E_2 \sum_{j=1}^{2M-1} \sum_{k=-N+1}^{N} \sigma_{j,k}\sigma_{j+1,k} - h \sum_{k=-N+1}^{N} \sigma_{1,k}, \quad (2.1)$$

where j, and k label, respectively, the rows and the columns of the lattice. Cyclic boundary conditions are imposed in the horizontal direction; therefore, the $(N+1)$th column is identified with the $(-N+1)$th column. The spin σ on each lattice site can take two values: $+1$ or -1. $E_1(E_2)$ is the horizontal (vertical) interaction energy between spins in the bulk on neighboring sites. E'_1 is the horizontal interaction energy between the neighboring spins on the boundary row, that is, the first row. The magnetic field h is applied only on the first row, and $\Delta = E_1 - E'_1$. When $E_1 = E'_1$, $\epsilon(0) = \epsilon$ is the Hamiltonian considered by McCoy and Wu.[1]

The partition function

$$Z(\Delta) = \sum_{\sigma=\pm 1} e^{-\beta\epsilon(\Delta)} \quad (2.2)$$

can be written in terms of a Pfaffian of a $4(2M+1)(2N) \times 4(2M+1)(2N)$ antisymmetric matrix $A(\Delta)$, corresponding to the dimer problem of the oriented half plane lattice given by Fig. 1,

Copyright © 1973 by the American Institute of Physics

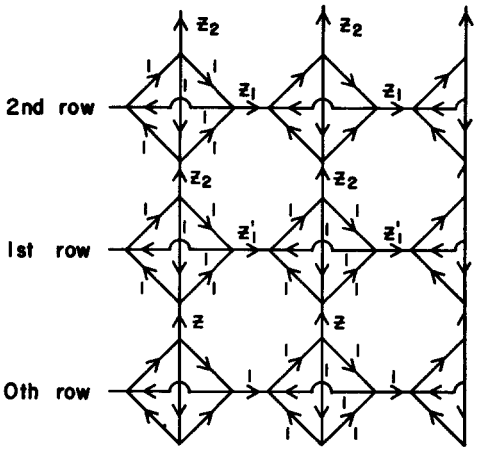

FIG. 1. Oriented half-plane lattice.

where

$$z = \tanh\beta h, \quad z_1 = \tanh\beta E_1, \quad z_1' = \tan\beta' E_1, \quad z_2 = \tanh\beta E_2. \tag{2.3}$$

That is,

$$Z(\Delta) = \tfrac{1}{2}(2\cosh\beta E_1)^{2N(2M-1)}(2\cosh\beta E_1')^{2N} \times(\cosh\beta E_2)^{2N(2M-1)}(\cosh\beta h)^{2N}\, Pf\, A(\Delta) \tag{2.4}$$

or, equivalently,

$$Z^2(\Delta) = \tfrac{1}{4}(2\cosh\beta E_1)^{4N(2M-1)}(2\cosh\beta E_1')^{4N} \times(\cosh\beta E_2)^{4N(2M-1)}(\cosh\beta h)^{4N}\det A(\Delta). \tag{2.5}$$

We have, in this paper, adapted the following notations: let Z, A, etc. be the symbols used by McCoy and Wu to describe some quantity for the case when $E_1' = E_1$, then the corresponding quantity for $E_1' \neq E_1$ is denoted in this paper by $Z(\Delta), A(\Delta)$, etc.

Notice in Fig. 1, that only the first row is different from the dimer problem considered by McCoy and Wu[1]; therefore, the antisymmetric matrix $A(\Delta)$ differs from the antisymmetrix matrix A only by matrix elements associated with the first row, which is given by

$$A(\Delta;1,k;1,k+1) = \begin{bmatrix} 0 & z_1' & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = -A(\Delta;1,k+1;1,k), \quad \text{for } -N+1 \le k \le N \tag{2.6}$$

To calculate the determinant of $A(\Delta)$, we shall use the method employed by McCoy and Wu.[1] Avoiding repetitions, we shall only outline it here.

We can write $A(\Delta)$ as a product of $2N$ determinants

$$\det A(\Delta) = \prod_{\theta} \det B(\Delta,\theta), \quad \theta = \pi(2k-1)/2N, \quad k = 1, \cdots 2N, \tag{2.7}$$

where $B(\Delta,\theta)$ is a $4(2M+1) \times 4(2M+1)$ matrix, which can be obtained from $B(\theta)$ by replacing z_1 in $B_{11}(\theta)$ by z_1'.

The determinant of $B(\Delta,\theta)$ is found to be

$$\det B(\Delta,\theta) = |1+e^{i\theta}|^2\,|1+z_1'e^{i\theta}|^2 \times|1+z_1e^{i\theta}|^{2(2M-1)}\, C_{2M}(\Delta,\theta), \tag{2.8}$$

where $C_{2M}(\Delta,\theta)$ is the determinant of a matrix $C(\Delta,\theta)$ which is different from $C(\theta)$ only in the second 2×2 diagonal block. By defining

$$a' = 2i\sin\theta z_1'/|1+z_1'e^{i\theta}|^2, \quad b' = (1-z_1'^2)/|1+z_1'e^{i\theta}|^2, \tag{2.9}$$

$C(\Delta,\theta)$ is obtained from $C(\theta)$ by putting a *prime* sign on the elements of the second 2×2 diagonal block of $C(\theta)$.

The determinant $C_n(\theta)$ of the first $2(n+1)\times 2(n+1)$ diagonal block of $C(\Delta,\theta)$, and the determinant $D_n(\theta)$ of the first $(2n+1)\times(2n+1)$ diagonal block of $C(\Delta,\theta)$, satisfy the following recursion formula:

$$\begin{bmatrix} C_n(\theta) \\ z_2 D_n(\theta) \end{bmatrix} = \begin{bmatrix} -a^2+b^2 & az_2 \\ -az_2 & z_2^2 \end{bmatrix} \begin{bmatrix} C_{n-1}(\theta) \\ z_2 D_{n-1}(\theta) \end{bmatrix} \quad \text{for } n \le 2, \tag{2.10}$$

where

$$a = 2iz_1\sin\theta\,|1+z_1e^{i\theta}|^{-2}, \quad b = (1-z_1^2)\,|1+z_1e^{i\theta}|^{-2}, \quad c = 2i\sin\theta\,|1+e^{i\theta}|^{-2}, \tag{2.11}$$

as defined in McCoy and Wu's paper.[1]

However, the boundary condition becomes

$$C_1(\Delta,\theta) = \det\begin{bmatrix} -c & 0 & 0 & 0 \\ 0 & c & z & 0 \\ 0 & -z & -a' & b' \\ 0 & 0 & -b' & a' \end{bmatrix} = -c^2(-a'^2+b'^2) - z^2 c a' \tag{2.12a}$$

and

$$D_1(\Delta,\theta) = \det\begin{bmatrix} -c & 0 & 0 \\ 0 & c & z \\ 0 & -z & -a' \end{bmatrix} = +c^2a' - cz^2. \tag{2.12b}$$

Again the hermitian matrix in (2.10) can be diagonalized by a unitary transformation,

$$\begin{bmatrix} -a^2+b^2 & az_2 \\ -az_2 & z_2^2 \end{bmatrix} = \begin{bmatrix} \nu & -i\nu' \\ -i\nu' & \nu \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \nu & i\nu' \\ i\nu' & \nu \end{bmatrix}, \tag{2.13a}$$

where λ_1 and λ_2 are the two real eigenvalues given by

$$\lambda_1 = \tfrac{1}{2}(-a^2+b^2+z_2^2) + \tfrac{1}{2}[(a^2-b^2+z_2^2)^2 - 4a^2z_2^2]^{1/2},$$

$$\lambda_2 = \tfrac{1}{2}(-a^2+b^2+z_2^2) - \tfrac{1}{2}[(a^2-b^2+z_2^2)^2 - 4a^2z_2^2]^{1/2}. \tag{2.13b}$$

Here the function $f(x) = x^{1/2}$ is defined to be positive for all real and positive x, so $\lambda_1 > \lambda_2$. Moreover, the components of the eigenvectors are given by

$$\nu = \{\tfrac{1}{2}[1+(\lambda_2-\lambda_1)^{-1}(z_2^2+a^2-b^2)]\}^{1/2} \tag{2.13c}$$

and

$$i\nu'/\nu = (z_2^2-\lambda_2)/az_2. \tag{2.13d}$$

When the number of rows $2M$ becomes large we have

$$C_{2M}(\Delta,\theta) \sim \nu^2\lambda_1^{2M-1}(-c^2)[-a'^2+b'^2+(i\nu/\nu)z_2a'] \times\left(1+\frac{z^2/c(a'-(i\nu'/\nu)z_2)}{(-a'^2+b'^2)+(i\nu'/\nu)z_2a'}\right).$$

The partition function becomes

$$Z^2(\Delta) \sim (2\cosh\beta E_1)^{4N(2M-1)}(2\cosh\beta E_1')^{4N(2M-1)} \times(\cosh\beta h)^{4N}(\cosh\beta E_2)^{4N(2M-1)}\prod_\theta |1+z_1'e^{i\theta}|^2 \times|1+z_1e^{i\theta}|^{2(2M-1)}\nu^2\lambda_1^{2M-1}[-a'^2+b'^2+(i\nu'/\nu)z_2a] \times\left(1+\frac{z^2/c(a'-(i\nu'/\nu)z_2)}{(-a'^2+b'^2)+(i\nu'/\nu)z_2a'}\right), \quad (2.14)$$

where we have used the identity

$$\prod_\theta |1+e^{i\theta}|^2 = 2\prod_\theta(-c^2) = \prod_\theta 4\sin^2\theta/|1+e^{i\theta}|^4 = 1.$$

In the thermodynamic limit, when the number of columns $2N$ also becomes large, one can convert the summation over θ into an integration over θ, that is, $\sum_\theta = (2N/2\pi)\int_{-\pi}^{\pi} d\theta$. The free energy is then given by

$$F(\Delta) = \frac{1}{\beta}\ln Z(\Delta) = 4MNF + 4NF_0(\Delta) + 2NF(\Delta,h),$$

where

$$F = \left(-\frac{1}{\beta}\right)\left(\ln(2\cosh\beta E_1)(2\cosh\beta E_2) + \frac{1}{4\pi}\int_{-\pi}^{\pi} d\theta \ln|1+z_1e^{i\theta}|^2\lambda_1(\theta)\right), \quad (2.15)$$

$$F_0(\Delta) = F_0 + F^1(\Delta) + F^1, \quad (2.16)$$

with

$$F_0 = \left(-\frac{1}{2\beta}\right)\left(-\ln\cosh\beta E_2 + \frac{1}{4\pi}\int_{-\pi}^{\pi} d\theta \ln\nu^2\right), \quad (2.16')$$

$$F^1(\Delta) = \left(-\frac{1}{2\beta}\right)\left(\ln\cosh\beta E_1' + \frac{1}{4\pi}\int_{-\pi}^{\pi} d\theta \ln|1+z_1'e^{i\theta}|^2 \times(-a'^2+b'^2+(i\nu'/\nu)z_2a)\right), \quad (2.16'')$$

$$F^1 = \left(-\frac{1}{2\beta}\right)\left(-\ln\cosh\beta E_1 - \frac{1}{4\pi}\int_{-\pi}^{\pi} d\theta \ln|1+z_1e^{i\theta}|^2\lambda_1(\theta)\right), \quad (2.16''')$$

and

$$F(\Delta,h) = \left(-\frac{1}{\beta}\right)\left[\ln\cosh\beta h + \frac{1}{4\pi}\int_{-\pi}^{\pi} d\theta \ln\left(1+\frac{z^2/c(a'-(i\nu'/\nu)z_2)}{-a'^2+b'^2+(i\nu'/\nu)z_2a'}\right)\right]. \quad (2.17)$$

Here $4MNF$ is the bulk free energy. As expected, it does not depend on E_1' or on h. The term $2N(2F_0(\Delta)+F(\Delta,h))$ is proportional to the number of lattice sites on the boundary, so it is called the boundary free energy of the system. The first part $4NF_0(\Delta)$ is the boundary free energy in the absence of the applied magnetic field and $2NF(\Delta,h)$ is then the change in free energy due to the presence of an applied field.

3. BOUNDARY SPECIFIC HEAT

Before analyzing the singularity of the boundary specific heat, let us examine $F_0(\Delta)$, which is the boundary free energy in the absence of the external field. Notice that F_0 given by (2.16') is identical to the boundary free energy McCoy and Wu[1] obtained for the case $E_1' = E_1$, which gives rise to a linearly divergent term in the specific heat and a discontinuity of the internal energy—the latent heat at T_c. When $E_1' \neq E_1$, the system gains additional free energy given by the sum of $F^1(\Delta)$ and F^1. The integral in F^1 is identical to the integral of the Onsager's solution for the bulk free energy, and the singularity of its specific heat is well known.[4] We therefore only need to find the values of z_1', z_1 and z_2 for which $F^1(\Delta)$ is singular.

It will be shown that $F^1(\Delta)$ and F^1 give rise only to the logarithmically divergent term in the specific heat. Hence the most singular behavior of the boundary specific heat is independent of E_1'.

To find the singularity of $F^1(\Delta)$, we shall first differentiate $F^1(\Delta)$ with respect to z_1, z_2, and z_1' to obtain its contribution to the boundary energy $U^1(\Delta)$. We then show these integrals are analytic except at $T = T_c$, where T_c is the bulk critical temperature. Expressions for these integrals for T near T_c are easily obtainable. By differentiating them with respect to T, we obtain the divergent terms of the specific heat.

Substituting the expression for a, b, a', b' and $i\nu'/\nu$ given by (2.11), (2.9), and (2.13) into (2.16''), we get

$$\beta F^1(\Delta) = -\tfrac{1}{2}\ln\cosh\beta E_1' - \frac{1}{8\pi}\int_{-\pi}^{\pi} d\theta \ln\{(1+z_1'^2-2z_1') \times[(1+z_1^2)(z_2^2-1)+2z_1(1+z_2^2)\cos\theta+(1-z_2^2)\chi(\theta)]\}, \quad (3.1)$$

where

$$\chi(\theta) = [(1-2\alpha_1\cos\theta+\alpha_1^2)(1-2\alpha_2^{-1}\cos\theta+\alpha_2^{-2})]^{1/2} \quad (3.2)$$

with

$$\alpha_1 = z_1\frac{1-|z_2|}{1+|z_2|}, \qquad \alpha_2^{-1} = z_1\frac{1+|z_2|}{1-|z_2|}, \quad (3.3)$$

when $\alpha_2 \leq 1$, $T \leq T_c$.

It is easy to verify that

$$F^1(z_1', z_1, z_2) = F^1(z_1', z_1, -z_2) = F^1(-z_1', -z_1, z_2).$$

Therefore, we can restrict ourself to the case of $z_1 > 0$, $z_2 > 0$ without losing generality. Differentiate $\beta F^1(\Delta)$ with respect to z_1:

$$\frac{\partial}{\partial z_1}(\beta F^1(\Delta)) = -\frac{1}{8\pi}[z_1'(1-z_1^2)/2z_1^2] \times\int_{-\pi}^{\pi}\frac{(1-z_2^2)\chi(\theta)-[(1+z_1^2)(1-z_2^2)-2z_1(1+z_2^2)\cos\theta]}{\{(1+z_1'^2-2z_1'\cos\theta)+(z_1'/2z_1)[(1+z_1^2)(z_2^2-1)+2z_1(1+z_2^2)\cos\theta+(1-z_2^2)\chi(\theta)]\}}\frac{d\theta}{\chi(\theta)}. \quad (3.4)$$

Let $w = 2i\tan\theta/2$, then define,

$$\tau_i = (1-\alpha_i)/(1+\alpha_i), \quad i = 1,2 \tag{3.5}$$

and (3.4) becomes

$$\frac{\partial}{\partial z_1}(\beta F^1(\Delta)) = -\frac{1}{4\pi i}\frac{z_1'}{2z_1^2}(1-z_1^2)(1-z_2^2) \times \int_{-i\infty}^{i\infty} \frac{[\tau_1\tau_2 + w^2 + (\tau_1^2-w^2)^{1/2}(\tau_2^2-w^2)^{1/2}]}{(1-z_1')^2 - w^2(1+z_1')^2 + (z_1'/2z_1)[(1+z_1)^2 - z_2^2(1-z_1)^2](\tau_1\tau_2+w^2) + (\tau_1^2-w^2)^{1/2}(\tau_2^2-w^2)^{1/2}]} \times \frac{dw}{[(\tau_1^2-w^2)(\tau_2^2-w^2)]^{1/2}}. \tag{3.6}$$

Denote the integrand of (3.6) by $R(w^2)$. It has four branch points at $\pm|\tau_2|$ and $\pm\tau_1$. Hence the branch cuts are the line from $-\tau_1$ to $-|\tau_2|$, and the line from $|\tau_2|$ to τ_1. Consider the contour integral

$$\tilde{I}_0 = \int_c R(w^2)\,dw,$$

where c is the contour shown in Fig. 2.

One can verify easily that

$$|R(w^2)| \to \frac{1}{|w|^2} \quad \text{as } |w| \to \infty. \tag{3.7}$$

As a result, the integration over the semicircle tends to zero as its radius tends to infinity.

The integration over the imaginary axis can be written as a sum of the residues inside c and the integration above and below the cut from $|\tau_2|$ to τ_1, that is

$$I_0 = \int_{-i\infty}^{i\infty} R(w^2)\,dw = -\int_{c_0} R(w^2)\,dw - 2\pi i \sum \text{residues of } R(w^2) \text{ inside } c, \tag{3.8}$$

where c_0 denote the two straight lines above and below the cut. The integrand $R(w^2)$ of (3.6) is finite at the origin. This means zero is not a double pole of $R(w^2)$. Therefore none of the residue of the integrand $R(w^2)$ are divergent. The divergence of the specific heat would only come from the first term in (3.8). It is obvious that the integrand $R(w^2)$ can be written in the following form:

$$R(w^2) = \frac{P(w^2)}{D(w^2)} \times \frac{1}{[(\tau_1^2-w^2)(\tau_2^2-w^2]^{1/2}} + \frac{Q(w^2)}{D(w^2)}, \tag{3.9}$$

where $D(w^2)$, $Q(w^2)$, and $P(w^2)$ are polynomials in w^2. In fact, for the integrand $R(w^2)$ in (3.6), we have

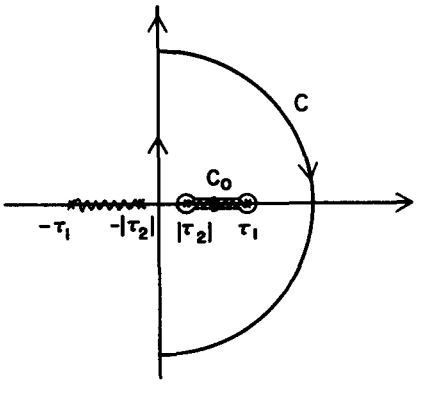

FIG. 2. The path of integration c consists of the imaginary axis, A large semicircle and c_0 which is the two straight lines just above and below the cut.

$$D(w^2) = \{(1-z_1')^2 - w^2(1+z_1')^2 + (z_1'/2z_1) \times [(1+z_1^2)(1-z_2^2) + 2z_1(1+z_2^2)] \times (\tau_1\tau_2 + w^2)\}^2 - (z_1'/2z_1)^2 \times [(1+z_1^2)(1-z_2^2) + 2z_1(1+z_2^2)]^2(\tau_1^2-w^2)(\tau_2^2-w^2), \tag{3.10}$$

$$P(w^2) = [(1-z_1')^2 - w^2(1+z_1')^2](\tau_1\tau_2+w^2) + (z_1'/2z_1) \times [(1+z_1^2) + 2z_1(1+z_2^2)][(\tau_1+\tau_2)^2w^2], \tag{3.11}$$

$$Q(w^2) = (1-z_1')^2 - w^2(1+z_1')^2. \tag{3.12}$$

From (3.9), one finds

$$\frac{\partial}{\partial z_1}(\beta F^1(\Delta)) = -\frac{1}{4\pi i}\frac{z_1'}{z_1^2}(1-z_1^2)(1-z_2^2) \times \int_{|\tau_2|}^{\tau_1} \frac{P(w^2)}{D(w^2)} \frac{dw}{[(\tau_1^2-w^2)(\tau_2^2-w^2)]^{1/2}} + \sum \text{residues}, \tag{3.13}$$

where the sum of residues is an analytic function of τ_1, τ_2 and z_1'.

The singularity and the most singular term of this integral can be determined by the following rule.

Rule: Let $P(w^2)$ and $D(w^2)$ be any polynomials in w^2, such that

(1) $P(\tau_1^2)/D(\tau_1^2)$ and $P(\tau_2^2)/D(\tau_2^2)$ are finite for some real number $0 < |\tau_2| < \tau_1 < \infty$;

(2) $D(0) \neq 0$, for $\tau_2 = 0$;

then the integral

$$I = \frac{1}{i}\int_{|\tau_2|}^{\tau_1} \frac{dw}{[(\tau_1^2-w^2)(\tau_2^2-w^2)]^{1/2}} \frac{P(w^2)}{D(w^2)} \tag{3.14}$$

is singular only for $\tau_2 = 0$. Its most singular term is given by

$$I \sim \frac{1}{\tau_1}\frac{P(\tau_2^2)}{D(\tau_2^2)} \ln\frac{4}{|\tau_2|}. \tag{3.15}$$

The derivation of this rule is given in the Appendix.

It can be verified, for the two polynomials $P(w^2)$ and $D(w^2)$ given by (3.10) and (3.11), respectively, the two conditions necessary for the above rule to hold are satisfied. Therefore, $(\partial/\partial z_1)(\beta F^1(\Delta))$ is analytic except $\tau_2 = 0$. Since $\tau_2 = (1-\alpha_2)/(1+\alpha_2)$, $\tau_2 = 0$ implies $\alpha_2 = 1$. Note α_2 is a function of E_1, E_2, and T. For given E_1 and E_2, the bulk critical temperature is defined by the equation $\alpha_2 = 1$. Hence $(\partial/\partial z_1)(\beta F^1(\Delta))$ is singular for $T = T_c$; its most singular term is given by

$$\frac{\partial}{\partial z_1}(\beta F^1(\Delta)) \sim -\frac{1}{4\pi}\left(\frac{z_1}{z_1^2}\right)(1-z_1^2)(1-z_2^2)\frac{\tau_2(\tau_1+\tau_2)}{\tau_1[(1-z_1')^2-\tau_2^2(1+z_1')^2+4z_1z_1'\tau_2]}\ln\frac{4}{|\tau_2|}. \tag{3.16}$$

We differentiate $\beta F^1(\Delta)$ with respect to z_2, and get

$$\frac{\partial}{\partial z_2}(\beta F^1(\Delta)) = -\frac{1}{8\pi}\frac{z_1'z_2}{z_1(1-z_2^2)} \times \int_{-\pi}^{\pi}\frac{d\theta\{(1-z_2^2)(1+z_1^2+2z_1\cos\theta)\chi(\theta)-(1-z_2^2)\chi^2(\theta)+4z_1[\alpha_1+\alpha_2^{-1}-(1+\alpha_1\alpha_2^{-1})\cos\theta]\}}{\chi(\theta)\{(1+z_1'^2-2z_1'\cos\theta)+(z_1'/2z_1)[(1+z_1^2)(z_2^2-1)+2z_1(1+z_2^2)\cos\theta+(1-z_2^2)\chi(\theta)]\}}. \tag{3.17}$$

The denominator of the integrand in (3.17) is the same as the denominator in (3.4), and one can verify that the integrand satisfies condition (3.7), and the conditions necessary for the above rule to hold. The most singular term can be written out without difficulties, and is

$$\frac{\partial}{\partial z_2}(\beta F^1(\Delta)) \sim \frac{1}{2\pi}\frac{z_1'z_2}{z_1(1-z_2^2)} \times \frac{-4z_1(\tau_1-\tau_2)\tau_2}{[(1-z_1')^2-\tau_2^2(1+z_1')^2+4z_1'z_2\tau_2]}\ln\frac{4}{|\tau_2|}. \tag{3.18}$$

Likewise,

$$\frac{\partial}{\partial z_1'}(\beta F^1(\Delta)) \sim 0 + \text{Taylor series in } |\tau_2|. \tag{3.19}$$

As mentioned before, the term F^1, given by (2.16″), is identical to the bulk free energy. Thus F^1 is analytic except at $T = T_c$; the asymptotic expansion for the energy per site due to this term is given by

$$\frac{\partial}{\partial\beta}(\beta F^1) \sim +E_1(1-z_1^2)\times\frac{1}{\pi}\frac{\tau_2(\tau_1-\tau_2)}{(1-z_1^2)\tau_1(1-\tau_1^2)}\ln\frac{4}{|\tau_2|} + E_2(1-z_2^2)\times\frac{1}{2\pi}\frac{(\tau_1+\tau_2)\tau_2}{z_2\tau_1(1-\tau_2^2)}\ln\frac{4}{|\tau_2|}. \tag{3.20}$$

The term F^0 given by (2.16‴) can be written as

$$\beta F_0 = \tfrac{1}{2}\ln\cosh\beta E_2 - \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}\frac{dw}{1-w^2} \times \ln\left[\tfrac{1}{2}\left(1-\frac{(\tau_1-w)(\tau_2-w)}{(\tau_1^2-w^2)^{1/2}(\tau_2^2-w^2)^{1/2}}\right)\right]. \tag{3.21}$$

[See (4.22) in McCoy and Wu's paper.[1]]

For completeness, we shall convert the above integral to elliptical integrals, and we shall also calculate the latent heat at T_c.

Differentiating the integrand $f(\tau_1\tau_2)$ in (3.21) with respect to τ_2, we get

$$\frac{\partial}{\partial\tau_2}f(\tau_1\tau_2) = \frac{1}{2(\tau_1+\tau_2)}\left(-\frac{\tau_1-w^2}{[(\tau_1^2-w^2)(\tau_2^2-w^2)]^{1/2}} - \frac{\tau_1-w}{\tau_2+w}\right)\times\frac{1}{1-w^2}. \tag{3.22}$$

The second term has only one pole, $w = 1$, on the rhs for $T < T_c(\tau_2 > 0)$; however, when $T > T_c$ $(\tau_2 < 0)$, it has two poles on the rhs, namely, $w = -\tau_2 = |\tau_2|$ and $w = 1$. Therefore, the integration of (3.31) gives for $T > T_c$

$$\frac{\partial}{\partial\tau_2}(\beta F_0) \sim \frac{1}{2\pi}\frac{\tau_1-\tau_2}{\tau_1}\ln\frac{4}{|\tau_2|} - \frac{1}{2}\times\frac{1}{1-\tau_2^2} + \frac{\tau_1-1}{4(\tau_1+\tau_2)(\tau_2+1)}. \tag{3.23a}$$

and for $T < T_c$

$$\frac{\partial}{\partial\tau_2}(\beta F_0) \sim \frac{1}{2\pi}\frac{\tau_1-\tau_2}{\tau_1}\ln\frac{4}{|\tau_2|} + \frac{\tau_1-1}{4(\tau_1+\tau_2)(\tau_2+1)}. \tag{3.23b}$$

The differentiation of βF_0 with respect to τ_1 can be obtained from (3.22) simply by interchanging τ_1 and τ_2:

$$\frac{\partial}{\partial\tau_1}f(\tau_1\tau_2) = \frac{1}{1-w^2}\left(-\frac{\tau_2^2-w^2}{[(\tau_1^2-w^2)(\tau_2^2-w^2)]^{1/2}} - \frac{\tau_2-w}{\tau_1+w}\right)\times\frac{1}{2(\tau_1+\tau_2)}. \tag{3.24}$$

It is obvious the second term in (3.25) has only one pole: $w = 1$, on the rhs for any temperature. For τ_2 small,

$$\frac{\partial}{\partial\tau_1}(\beta F_0) \sim \text{Taylor series} + 0(\tau_2^2\ln|\tau_2|) + \frac{\tau_2-1}{4(\tau_1+\tau_2)(\tau_1+1)}.$$

Combining these results, we get the following asymptotic expressions:

$$\frac{\partial}{\partial\beta}(\beta F_0) \sim \left(\frac{\partial\tau_2}{\partial\beta}\right)\left(\frac{1}{2\pi}\frac{\tau_1-\tau_2}{\tau_1}\ln\frac{4}{|\tau_2|} - \frac{1}{2(1-\tau_2^2)}\right), \quad \text{for } \tau_2 < 0 \ (T > T_c), \tag{3.25a}$$

$$\frac{\partial}{\partial\beta}(\beta F_0) \sim \left(\frac{\partial\tau_2}{\partial\beta}\right)\left(\frac{1}{2\pi}\frac{\tau_1-\tau_2}{\tau_1}\ln\frac{4}{|\tau_2|}\right), \quad \text{for } \tau_2 > 0 \ (T < T_c). \tag{3.25b}$$

As τ_2 is not a pole of the integrand given either by (3.6), (3.17), or (3.19), the discontinuity of the boundary energy comes only from (3.23), which is

$$\lim_{\delta T\to 0} U(T+\delta T) - U(T-\delta T) = -\frac{1}{2}\left(\frac{\partial\tau_2}{\partial\beta}\right)_{T=T_c} = -\frac{1}{2}S_1 \tag{3.26}$$

with $S_1 = [E_1(1-z_{1c}^2)/2z_{1c} + E_2]$.

This is identical to the latent heat obtained by McCoy and Wu.[1]

Therefore from (3.16), (3.18), (3.19), and (3.26), we find the specific heat diverges at $T = T_c$ as

$$C_v \sim -\frac{1}{2\pi}\frac{S_1}{T-T_c} - \frac{1}{2\pi T_c} S_2 \ln|1 - T/T_c| - \frac{1}{2\pi}\frac{1}{KT_c^2}\frac{4z_1'(1-z_1)}{(1-z_1')^2(1+z_1)} S_1^2 \ln|1 - T/T_c| + \frac{1}{2\pi}\frac{1}{KT_c^2}\frac{4z_1}{(1-z_1^2)} S_1^2 \ln|1 - T/T_c| \quad \text{for } z_1' \neq 1, \tag{3.27}$$

where

$$S_2 = \left[\left(\frac{\partial \tau_2}{\partial \beta}\right)^2 \Big/ \tau_1 - \frac{\partial^2 \tau_2}{\partial \beta^2}\right]_{T=T_c} KT_c. \tag{3.28}$$

It is obvious from (3.39) that if the boundary spin coupling constant E_1' becomes larger, the coefficient of the logarithmically divergent term becomes more negative. Therefore the specific heat is more divergent. It is also evident from (3.39) that for temperature slightly higher than T_c, the boundary specific heat is negative.

4. MAGNETIZATION AND SUSCEPTIBILITY

The magnetization of the boundary row is defined to be

$$M_1(h) = \lim_{N,M\to\infty} \frac{1}{2N} \left\langle \sum_{K=-N+1}^{N} \sigma_{1,K} \right\rangle = \beta^{-1} \lim_{N,M\to\infty} \frac{1}{2N} \ln Z = -\frac{\partial}{\partial h} F(\Delta, h). \tag{4.1}$$

By (2.17), this becomes

$$M_1(h) = z + \frac{1}{2\pi}(1-z^2)z \times \int_{-\pi}^{\pi} d\theta \frac{1/c[a' - (i\nu'/\nu)z_2]}{-a'^2 + b'^2 + (i\nu'/\nu)z_2 a' + z^2/c[a' - (i\nu'/\nu)z_2]}. \tag{4.2}$$

On substituting $(i\nu'/\nu)$, a', b', and c as given by (2.11), (2.9), and (2.13) into the above equation, we get

$$M_1(h) = z + \frac{1}{2\pi}(1-z^2)z \times \int_{-\pi}^{\pi} d\theta \{z^2(2+2\cos\theta) \times [(1+z_1'^2+2z_1'\cos\theta)\gamma + z_1] + 4\sin^2\theta z_1'\gamma + \tfrac{1}{2}[(1-z_2^2) \times (1+z_1^2) - 2z_1(1+z_2^2)\cos\theta] - \tfrac{1}{2}(1-z_2^2)\chi(\theta)\} \times \{z^4(2+2\cos\theta)[(1+z_1'^2+2z_1'\cos\theta)\gamma + z_1] + 8z^2\sin^2\theta z_1'\gamma + z^2[(1-z_2^2)(1+z_1^2) - 2z_1(1+z_2^2)\cos\theta] - (2-2\cos\theta)[z_1 z_2^2 - (1+z_1'^2 - 2z_1'\cos\theta)\gamma]\}^{-1}, \tag{4.3}$$

where

$$\gamma = \frac{(1-z_2^2)[z_1'(1+z_1)^2 - z_1(1+z_1')^2]}{(1-z'^2)^2}.$$

Let us first consider the special case $E_1' \to \infty$, so that $z_1 \to 1$. In that case, $\gamma \to \infty$.

The integrand in (3.3) becomes:

$$\frac{z^2(2+2\cos\theta)^2 + 4\sin^2\theta}{z^4(2+2\cos\theta)^2 + 2z^2(4\sin^2\theta) + (2-2\cos\theta)^2} = \frac{1+\cos\theta}{(1+\cos\theta) - (1-\cos\theta)}.$$

The magnetization can be calculated exactly, and is given by

$$M_1(h) = z + \frac{2}{\pi}(1-z^2)z \times \int_0^{\infty} \frac{du}{(1+u^2)(z^2+u^2)} = \begin{cases} +1 & \text{for } z > 0 \\ -1 & \text{for } z < 0 \end{cases}. \tag{4.4}$$

The spontaneous magnetization is

$$M_1(0^+) = 1. \tag{4.5}$$

This means, when $E_1' \to \infty$, the boundary spins are aligned irrespective of the bulk spins.

For the general case ($z_1' \neq 1$), the magnetization can be calculated by converting the integral in (4.3) to complete elliptical integrals; and it is for this purpose, we have written the integrand in such a way that the irrational function $\chi(\theta)$ appears only in the numerator. We shall restrict ourself to the ferromagnetic bulk spins ($z_1 > 0$).

Just as in the previous section, we change the variable of integration to $w = i\tan\theta/2$, so that the path of integration becomes the imaginary axis:

$$M_1(h) = z + \frac{1}{\pi i}(1-z^2)z \int_{-i\infty}^{i\infty} \frac{dw}{(1-w^2)} \frac{p(w^2) - \tfrac{1}{2}(1-w^2)[(1+z_1)^2 - z_2^2(1-z_1)^2][(\tau_1^2 - w^2)(\tau_2^2 - w^2)]^{1/2}}{d(w^2)}, \tag{4.6}$$

where

$$p(w^2) = 4z^2\{[(1+z_1')^2\gamma + z_1] - w^2[(1-z_1)^2\gamma + z_1]\} - 16z_1'\gamma w^2 - \tfrac{1}{2}(1-w^2) \times [(1+z_1)^2 - (1-z_1)^2 z_2^2][\tau_1\tau_2 + w^2], \tag{4.6'}$$

$$d(w^2) = 4z^4\{[(1+z_1')^2\gamma + z_1] - w^2[(1-z_1')^2\gamma + z_1]\} + 2z^2\{-16z_1'\gamma w^2 - \tfrac{1}{2}(1-w^2)[(1+z_1)^2 - z_2^2(1-z_1)^2] \times (\tau_1\tau_2 + w^2)\} + 4w^2\{[z_1 z_2^2 - (1-z_1')^2\gamma] - w^2[z_1 z_2^2 - (1+z_1')^2\gamma]\}. \tag{4.6''}$$

Again, we close the path of integration as shown in Fig. 2:

$$M_1(h) = z + \frac{1}{\pi i}(1-z^2)z \times \left(-\sum 2\pi i \text{ residues} - [(1+z_1)^2 - z_2^2(1-z_1^2)] \times \int_{|\tau_2|}^{\tau_1} \frac{dw}{[(\tau_1^2 - w)(\tau_2^2 - w^2)]^{1/2}} \frac{(\tau_1^2 - w^2)(\tau_2^2 - w^2)}{d(w^2)}\right), \tag{4.7}$$

where $d(w^2)$ is given by (4.6″).

We shall first compute the spontaneous magnetization for all temperatures, which is defined to be

$$M_1(0^+) = \lim_{h\to 0} M_1(h). \tag{4.8}$$

Away from the critical temperature ($\tau_2 \neq 0$), the integration from $|\tau_2|$ to τ_1, which is the second term in (4.7), is finite. In the limit $z \to 0$, it drops out. We only need to find the divergent residue of the integrand.

It is evident from (4.6″) that, in the limit $z \to 0$, one of the roots, say λ_1, of the polynomial $d(w^2)$ in the denominator, approaches zero. This implies, the two poles $w = \pm\sqrt{\lambda_1}$ are pinching onto the origin. The residue of this pole $w = +\sqrt{\lambda_1}$ (and only of this pole) will not be finite.

To calculate the residue at $w = \sqrt{\lambda_1}$, we first expand $\lambda_1(z^2)$ as a Taylor series in z^2,

$$\lambda_1(z^2) = \lambda_1(0) + \left(\frac{\partial}{\partial z}\,\lambda_1(z^2)\right)_{z^2=0} z^2 = \left[\frac{\partial}{\partial z}\,\lambda_1(z^2)\right]_{z^2=0} (z^2). \tag{4.9}$$

All higher terms can be neglected as we are interested in the limit $z \to 0$. The first derivative of $\lambda_1(z^2)$ is easy to calculate from the well-known formula

$$\left(\frac{\partial\lambda_1}{\partial u}\right) = -\left[(\partial d/\partial u)/(\partial d/\partial\lambda)\right]_{\lambda=\lambda_1},$$

where $d(\lambda, u)$ is a polynomial in λ, with u as a parameter, and λ_1 one of its roots. We find

$$\left(\frac{\partial\lambda_1}{\partial z^2}\right)_{z^2=0} = +\frac{[(1+z_1)^2 - z_2^2(1-z_1)^2]\tau_1\tau_2}{4[z_1z_2^2 - (1-z_1')^2\gamma]}. \tag{4.10}$$

When $z = 0$, the integrand of (4.6) becomes

$$\frac{-16w^2z_1'\gamma - \frac{1}{2}(1-w^2)[(1+z_1)^2 - z_2^2(1-z_1)^2][\tau_1\tau_2 + w^2 + (\tau_1^2-w^2)^{1/2}(\tau_2^2-w^2)^{1/2}]}{4w^2\{[z_1z_2^2 - (1-z_1')^2\gamma] - w^2[z_1z_2^2 - (1+z_1')^2\gamma]\}(1-w^2)}.$$

Therefore, the residue at $w = \sqrt{\lambda_1}$ is given by

$$\frac{-\frac{1}{2}[(1-z_1)^2 - z_2^2(1-z_2)^2][\tau_1\tau_2 + |\tau_1\tau_2|]}{z\{[(1+z_1)^2 - z_2^2(1-z_1)^2]\tau_1\tau_2/[z_1z_2^2 - (1-z_1')^2\gamma]\}^{1/2}\,4[z_1z_2^2 - (1-z_1')^2\gamma]} + 0(z). \tag{4.11}$$

One can write $\tau_1\tau_2 + |\tau_1\tau_2| = 2\tau_1\tau_2\,\theta(\tau_2)$, where

$$\theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}. \tag{4.12}$$

Hence the spontaneous magnetization is

$$M_1(0^+) = \lim_{z\to 0}\frac{1}{\pi i}\, z(-2\pi i \text{ residue at } w = \sqrt{\lambda_1}) = \frac{[(1+z_1)^2 - z_2^2(1-z_1)^2]^{1/2}}{2[z_1z_2^2 - (1-z_1')^2\gamma]^{1/2}}(\tau_1\tau_2)^{1/2}\,\theta(\tau_2). \tag{4.13}$$

The graphs and discussion of the spontaneous magnetization are included in Sec. 5.

To study the behavior of the boundary magnetization, we write

$$\frac{(\tau_1^2 - w^2)(\tau_2^2 - w^2)}{d(w^2)} = \beta_0 + \frac{\beta_1}{w^2 - \lambda_1} + \frac{\beta_2}{w^2 - \lambda_2},$$

λ_1, λ_2 are roots of $d(w^2)$.

Then the integral in (4.7) can be written as

$$-\frac{1}{i\tau_1}\left(\beta_0 K(\kappa) + \frac{\beta_1}{\tau_1^2 - \lambda_1}\,\pi(\nu_1, \kappa) + \frac{\beta_2}{\tau_1^2 - \lambda_2}\,\pi(\nu_2, \kappa)\right),$$

and again

$$\nu_i = -(\tau_1^2 - \tau_2^2)/(\tau_1^2 - \lambda_i).$$

For $|\tau_2|$ small, we get from (3.21), (A7), the magnetization

$$M_1(h) = M_1(0^+) + 0(z) - \frac{z}{\pi}[(1+z_1)^2 - z_2^2(1-z_1)^2] \times (1/\tau_1)\left(\frac{\beta_1\sqrt{\nu_1}}{\tau_2^2 - \lambda_1}\tan^{-1}\sqrt{\nu_1} + \frac{\beta_2\sqrt{\nu_2}}{\tau_2^2 - \lambda_2}\tan^{-1}\sqrt{\nu_2} + 0(\tau_2^2\ln\tau_2)\right). \tag{4.14}$$

When $z \neq 0$, the roots λ_i of $d(w^2)$ are nonvanishing. Therefore, $\nu_i \neq -1$, and the inverse tangent functions in (4.14) and its derivative are finite. This means the susceptibility is finite for $z \neq 0$. However, when $z \to 0$ there is one, and only one, root λ_1 approaching zero. Thus one writes, for small z and τ_2, $\sqrt{\nu_1} = i(1 - \tau_2^2/\tau_1^2 + \lambda_1/\tau_1^2)$. Since

$$i\tan ix = \tfrac{1}{2}\ln\left|\frac{1+x}{1-x}\right| \sim +\tfrac{1}{2}\ln|1-x|, \quad \text{for } |x| \sim 1,$$

we find the magnetization for small z and small τ_2 is

$$M_1(h) \sim M_1(0^+) + 0(z) - \frac{z}{2\pi\tau_1}[(1+z_1)^2 - z_2^2(1-z_1)^2] \times \frac{\beta_1}{\tau_2^2}\ln|\tau_2^2 - \lambda_1|. \tag{4.15}$$

The coefficient β_1 is again a function of z^2, and we expand it as a Taylor series of z^2, so

$$\beta_1 = \beta_1(0) + \beta_1'(0)z^2 = \beta_1(0) + 0(z^2).$$

$\beta_1(0)$ is given by the following formula:

$$\beta_1(0) = \lim_{w^2\to 0} w^2 \lim_{z\to 0}\frac{(\tau_1^2 - w^2)(\tau_2^2 - w^2)}{d(w^2)} = \frac{\tau_1^2\tau_2^2}{4[z_1z_2^2 - (1-z_1')^2\gamma]}. \tag{4.16}$$

Notice the first derivative of λ_1 given by (4.10) is proportional $\tau_1\tau_2$. For τ_2 small, the second derivative with respect to z^2 need to be taken into consideration. It is,

$$\frac{\partial^2\lambda_1}{\partial(z^2)^2} = -\frac{(1+z_1')^2\gamma + z_1}{4[z_1z_2^2 - (1-z_1')^2\gamma]} + 0(\tau_2).$$

Therefore,

$$\lambda_1 \sim \{z^2\tau_1\tau_2[(1+z_1)^2 - z_2^2(1-z_1)^2] - \tfrac{1}{2}z^4[(1+z_1')^2\gamma + z_1]\}/4[z_1z_2^2-(1-z_1')^2\gamma]. \quad (4.17)$$

Substituting (4.16) and (4.17) into (4.15), we find the most divergent term in the magnetization near T_c to be

$$M_1(h) \sim M(0^+) + \frac{-1}{2\pi}\frac{[(1+z_1)^2 - z_2^2(1-z_1)^2]\tau_1}{4[z_1z_2^2 - (1-z_1)^2\gamma]} \times \left(z\ln|\tau_2^2 - z^2\tau_1\tau_2\frac{[(1+z_1)^2 - z_2^2(1-z_2^2)]}{4[z_1z_2^2 - (1-z_1')^2\gamma]}\right.$$

$$\left. + \tfrac{1}{8} \times \frac{z^4[(1+z_1)^2\gamma + z_1]}{[z_1z_2^2 - (1-z_1')^2\gamma]}\right|). \quad (4.18)$$

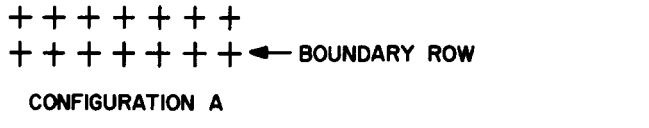

FIG. 3. Configurations.

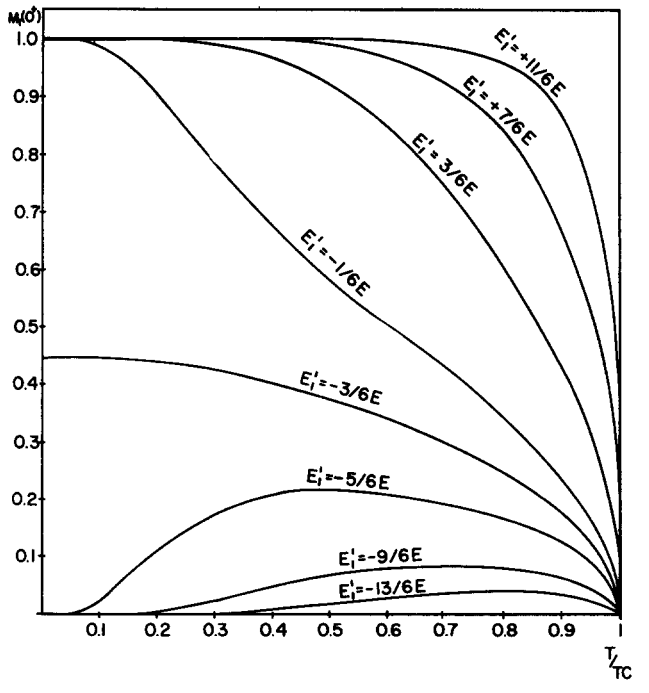

FIG. 4. The boundary spontaneous magnetization versus temperature for different values of boundary interaction energy E_1'.

Therefore, at $T = T_c$,

$$M_1(h) \sim \frac{-4}{\pi}\frac{(1-z_{1c})}{(1+z_{1c})}\frac{(1+z_{1c}')^2}{(1-z_{1c}')^2} \times z\ln z, \quad (4.19)$$

where $z_{1c} = \tanh E_1/KT_c$ and $z_{1c} = \tanh E_1'/KT_c$.

The magnetic susceptibility at zero field diverges at T_c in the following form:

$$\chi_1(0^+) = \frac{\partial}{\partial h}M_1(h)\Big|_{h=0} = -\frac{1}{KT_c} \times \frac{2}{\pi}\frac{(1-z_{1c})}{1+z_{1c}} \times \frac{(1+z_{1c}')^2}{(1-z_{1c}')^2}\ln|1 - T/T_c|. \quad (4.20)$$

Clearly when E_1' become larger, the coefficient of the susceptibility becomes more negative. Hence the susceptibility becomes more divergent.

The above results can also be obtained by the method used by McCoy and Wu.

5. DISCUSSION

In this section, we shall begin with the discussion of the boundary spontaneous magnetization, from which one can draw the conclusion of what seems to be happening between the boundary and the bulk spins. Then we shall speculate about the three-dimensional Ising case.

The terms inside the square root sign in (4.13) satisfy

$$\left.\begin{aligned}&[z_2^2(1+z_1)^2 - (1-z_1)^2] \geq 0\\ &[z_1z_2^2 - (1-z_1')^2\gamma] > 0\end{aligned}\right\} \quad \text{for } T \leq T_c.$$

Therefore, (4.13) is a valid expression for the real quantity $M_1(0^+)$. When the temperature is at absolute zero, one finds from (4.13),

$$M_1(0_+) = 1 \quad \text{at } T = 0, \quad \text{if } |E_2| > -2E_1',$$

and

$$M_1(0_+) = 0 \quad \text{at } T = 0, \quad \text{if } |E_2| < -2E', \quad (5.1)$$

and

$$M_1(0_+) = \sqrt{5/5} \quad \text{at } T = 0, \quad \text{if } |E_2| = -2E_1'.$$

To understand these results, let us consider the two configurations, given by Fig. 3, corresponding to the two possible ground states for the ferromagnetic bulk $E_1 = E_2 > 0$, and the antiferromagnetic boundary $E_1' = -|E_1'|$. Configuration A has all bulk spins pointing up, and all spins on the boundary pointing up also; and configuration B has all the bulk spins pointing up, but the spins on the boundary row are in the perfect alternating configuration.

The energy difference between the energy $E(A)$ and $E(B)$ correspond to the two configurations A and B, respectively:

$$\Delta E = E(A) - E(B) = (+|E_1'| - E_2) - (-|E_1'|) = 2|E_1'| - E_2.$$

Since at absolute zero, only the ground state is occupied, if $2|E_1'| > E_2$, configuration B is the ground state. The magnetization of this configuration is zero. On the other hand, if $2|E_1'| < E_2$, configuration A is the ground state, the magnetization of which is + 1. At the critical value $|E_1'| = E_2/2$, the ground state is highly degenerate, such that to flip one spin up will cause no change in energy.

In Fig. 4, we plot the graphs of the spontaneous magnetization $M_1(0^+)$ versus temperature corresponding to different boundary coupling constants E_1', but for fixed bulk coupling constant $E_1 = E_2 = E$.

One finds from the graph that the spontaneous magnetization decreases as E_1' decreases. [At least when $E_1' > 0$, this is a consequence of the Griffith's theorem.] When the boundary becomes antiferromagnetic, the boundary spins are still aligned for $T < T_c$. For $E_1' > -E/2$, $M_1(0^+)$ starts from 1 at $T = 0$; for $E_1' = -E/2$, the magnetization starts at $\sqrt{5/5}$ at $T = 0$; and when $E_1' < -E/2$, the spontaneous magnetization starts from 0 at the absolute zero, arises to a maximum, and then falls to zero again at $T = T_c$.

One can interpret these graphs by considering the boundary as a one-dimensional Ising chain in the presence of a magnetic field produced by the bulk spontaneous magnetization. Let us consider the case of $E_1' = -1.5$, for example. At absolute zero, the ferromagnetic *bulk* spins are all aligned to produce a magnetic field. However, the field is not strong enough to overcome the tendency of an antiferromagnetic boundary to be in a perfect alternating configuration. As the temperature rises, the *bulk* spontaneous magnetization decreases inappreciably, but the antiferromagnetic chain is less rigid. Hence, the magnetization of the boundary spins increases. When the temperature approaches the critical temperature, the bulk spontaneous magnetization falls off rapidly to zero, which causes the magnetization of the boundary to fall to zero at the bulk critical temperature.

The divergence of the boundary magnetic susceptibility is caused by the singular behavior at T_c of the field produced by the bulk spins.

One can also think of the bulk as a heat reservoir. The singularity of the bulk specific heat at T_c will give rise to the singular behavior of the boundary specific heat at T_c.

Since the idea of treating the boundary as a one-dimensional chain and the two-dimensional bulk as an external field works so well qualitatively, the same way of treating the boundary of a three-dimensional Ising as a two-dimensional Ising system and the bulk as an external field should also give a good qualitative description.

Let T_c^2 denote the critical temperature of a two-dimensional Ising model whose interaction energy is E'; and T_c^3 denotes the critical temperature of a three-dimensional Ising model whose interaction energy is E.

Consider E and E' such that $T_c^2 < T_c^3$. At the critical temperature of surface, $T = T_c^2$, the bulk spins are aligned and therefore the magnetic field produced by the bulk spins is not zero. The boundary magnetization is not zero at $T = T_c^2$. Furthermore, as the magnetic susceptibility of a two-dimensional Ising model is divergent only when the external field is zero. One expects the boundary susceptibility is not divergent at $T = T_c^2$. When the temperature is near the bulk critical temperature $T = T_c^3$, the bulk spontaneous magnetization drops to zero, with a discontinuous derivative at $T = T_c^3$. We conclude that the boundary (spontaneous) magnetization goes to zero at $T = T_c^3$ and the boundary magnetic susceptibility has a singularity at $T = T_c^3$.

Consider now the case $T_c^3 < T_c^2$. At the bulk critical temperature, $T = T_c^3$, the boundary (spontaneous) magnetization is not zero. As the bulk spontaneous magnetization goes to zero at $T = T_c^3$ with a discontinuous derivative, one expects the boundary susceptibility diverges at $T = T_c^3$. When $T > T_c^3$, the bulk spontaneous magnetization is identically zero. One can expect the boundary spontaneous magnetization to go to zero at $T = T_c^2$, and the boundary susceptibility diverges at $T = T_c^2$ with the same critical indices as the two-dimensional Ising model.

It is obvious from the above discussion that when the critical temperature of the surface is not very much higher than the critical temperature of the bulk, the boundary magnetic susceptibility cannot be determined accurately by series extrapolation method.

ACKNOWLEDGMENTS

The author would like to thank Dr. P. C. Hohenberg for suggesting this problem, and Dr. M. Blume and Dr. S. Krinsky for helpful discussions. The author is grateful to Dr. B. McCoy for his encouragement, discussions, and critical reading of the manuscript.

APPENDIX

To derive the rule given in Sec. 3, we write

$$D(w^2) = D_0(w^2 - \lambda_1) \cdots (w^2 - \lambda_n), \tag{A1}$$

where λ_i are roots of the polynomial $D(w^2)$.

By partial fractions, we get[5]

$$\frac{P(w^2)}{D(w^2)} = \frac{1}{D_0}\left(A_0 + \frac{A_1}{w^2 - \lambda_1} + \cdots + \frac{A_n}{w^2 - \lambda_n}\right). \tag{A2}$$

Substitute (A2) into (3.14), we get[5]

$$I = \int_{|\tau_2|}^{\tau_1} \frac{dw}{[(\tau_1^2 - w^2)(w^2 - \tau_2^2)]^{1/2}} \frac{P(w^2)}{D(w^2)}$$
$$= \frac{1}{D_0\tau_1}\left(A_0K(\kappa) + \frac{A_1}{\tau_1^2 - \lambda_1}\pi(\nu_1, \kappa) + \cdots + \frac{A_n}{\tau_1^2 - \lambda_n}\pi(\nu_n, \kappa)\right), \tag{A3}$$

where

$$\kappa^2 = 1 - \tau_2^2/\tau_1^2. \tag{A4}$$

Thus $\kappa = 1$, implies $\tau_2 = 0$ and

$$\nu_i = \frac{\tau_1^2 - \tau_2^2}{\tau_1^2 - \lambda_i} \quad (i = 1, \ldots, n). \tag{A5}$$

$K(\kappa)$ is the complete elliptical integral of the first kind and $\pi(\nu_i, \kappa)$ is the complete elliptical integral of the third kind. $K(\kappa)$ has one and only one singularity at $\kappa = 1$. By condition (1) of the Rule in Sec. 3, $A_i\pi(\nu_i, \kappa)/(\tau_1^2 - \lambda_i)$ can be shown to have singularity only at $\kappa = 1$ also.

When $|\tau_2|$ is small, we can write

$$K(\kappa) = \ln\frac{4}{|\tau_2|} + O(\tau_2^2 \ln|\tau_2|) \tag{A6}$$

and

$$\pi(\nu_i, k) = \frac{1}{1+\nu_i} \ln \frac{4}{|\tau_2|} + \frac{\sqrt{\nu_i}}{1+\nu_i} \tan^{-1} \sqrt{\nu_i} + 0(\tau_2^2 \ln \tau_2). \tag{A7}$$

If $\nu_i \neq 1$, the second term in (A7) is analytic. By condition (2) of the Rule, we have in the limit $\tau_2 \to 0$, $\nu_i = -\tau_1^2/\tau_1^2 - \lambda_i \neq -1$. Therefore, the singular term of the integral I is

$$I \sim \frac{1}{\tau_1} \left[\frac{1}{D_0} \left(A_0 + \frac{A_1}{\tau_1^2 - \lambda_1} \frac{1}{1+\nu_1} + \cdots + \frac{A_n}{\tau_1^2 - \lambda_n} \frac{1}{1+\nu_n} \right) \right] \ln \frac{4}{|\tau_2|}$$

$$= \frac{1}{\tau_1} \left[\frac{1}{D_0} \left(A_0 + \frac{A_1}{\tau_2^2 - \lambda_1} + \cdots + \frac{A_n}{\tau_2^2 - \lambda_n} \right) \right] \ln \frac{4}{|\tau_2|}.$$

It is evident from (A2) that the term inside the bracket is just the function $P(w^2)/D(w^2)$ evaluated at $w^2 = \tau_2^2$. This means

$$I \sim \frac{1}{\tau_1} \frac{P(\tau_2^2)}{D(\tau_2^2)} \ln \frac{4}{|\tau_2|}.$$

---

*Supported in part by the National Science Foundation Grant No. GP-32998X.

[1]Barry McCoy and T. T. Wu, Phys. Rev. **162**, 436 (1967).
[2]Private communications.
[3]M. E. Fisher and A. E. Ferdinand, Phys. Rev. Lett. **19**, 169 (1967).
[4]K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
[5]This choice sign is necessary to ensure $\chi(0) \geq 0$ [see (3.5)].

Scattering function of a two-component fully ionized gas

Chiang S. Cheng

East Stroudsburg State College, East Stroudsburg, Pennsylvania 18301
(Received 17 January 1973)

The effects of collisions on the scattering function of a two-component fully ionized gas are discussed, using an improved Brownian motion type collision model. In the collision model, the particle numbers as well as the energy and the linear momentum are conserved; and the electron–ion collisions are taken into account. The dispersion relations are discussed in detail, and explicit calculations are carried out for the scattering function and the energy-loss function in the long wavelength approximation when the wavelength is greater than the mean free path and the Debye length. The damped plasma oscillations dominate the energy-loss function; while for the scattering function thermal diffusions and sound propagations dominate the low frequency region and contribute most to the total scattering cross section. The contribution from the damped plasma oscillations to the scattering becomes important only for high frequencies near the plasma oscillation frequency.

I. INTRODUCTION

The scattering function, or autocorrelation functions in general, of a collisionless plasma has been studied extensively on the basis of the Vlasov equation.[1] Collisional effects have been discussed for a partly ionized gas by Dougherty and Farley[2] and recently by Williams and Chappell,[3] describing collisions of charged particles with neutral molecules by a Krook model. For a fully ionized gas, Coulomb force collisions among the charged particles can be properly described by the Landau collision term,[4] also called the Fokker–Planck collision term. Properties of the Fokker–Planck operator have been studied by Su,[5] Lewis,[6] and McLeod and Ong[7]; it is clear that this operator is too complex mathematically to be used to describe collisional effects on the scattering function.

A simplified model with the collision operator

$$\Omega = \gamma(\partial/\partial v_\alpha)[v_\alpha + (T/m)(\partial/\partial v_\alpha)]$$

has been used by Lenard and Bernstein[8] to discuss the effects of collision on Landau damping of an electron gas. This Brownian motion type operator is much easier to handle, yet it preserves the most important feature of small angle collisions of the long-range Coulomb interactions, namely, a diffusion in velocity space. It conserves particle number, though not momentum and energy. This collision operator has also been used by McLennan and the author[9] to describe both the electron–electron and the ion–ion collisions of a fully ionized gas, and to discuss the scattering function in the long wavelength approximation.

The modified collision operator

$$\Omega = \gamma(\partial/\partial v_\alpha)[(v_\alpha - \bar{v}_\alpha) + (\bar{T}/m)(\partial/\partial v_\alpha)],$$

in which the local velocity $\bar{v}$ and the local temperature $\bar{T}$ are functionals of the distribution function, has been used by Dougherty[10] to describe the ion–ion collisions and their effects on the gyroresonance in the radio wave scattering in the ionosphere. It is a modification of the Lenard–Bernstein operator to conserve momentum and energy as well as the particle number.

Here, this collision model is further modified to describe collisions among the charged particles of a two-component fully ionized gas (Sec. III). The collision operator is represented by a 2×2 matrix with differential and integral operators as its elements. It has a discrete spectrum including the six hydrodynamic modes which correspond to the six conservation equations of particle numbers, momentum and energy of a two-component system.

Section II is devoted to a formal solution of the coupled (electron and ion) Boltzmann equations, including the Vlasov consistent field term. The effects of collisions are clearly indicated by the appearance of the collision operator in the electrical susceptibilities and the dielectric constant. The eigenvalue problem of the collision operator plus the streaming term is solved, in Sec. IV, by a perturbation method which is valid in the long wavelength approximation when the wavelength is greater than the mean free path. These solutions provide a basis for an expansion which is required in the evaluation of the dielectric constant. In Sec. V the dispersion relations are solved in the same long wavelength approximation. The joint effect of collisions and the Vlasov term produces a pair of damped plasma oscillation poles. The sound damping constant is also changed, as compared to the case of a neutral system, by the Vlasov term. Finally, in the last section, explicit calculations are carried out for the energy-loss function and the scattering function.

II. GENERAL THEORY

The scattering function, which is proportional to the cross section of electromagnetic waves scattered by a fully ionized gas, is related to the electron density autocorrelation functions. A correlation function, such as the number density–density correlation function

$$G_{\sigma\sigma'}(x,t) = \langle (n_\sigma(x,t) - n_\sigma)(n_{\sigma'}(o,o) - n_{\sigma'}) \rangle$$

is the equilibrium ensemble average, denoted by $\langle\ \rangle$, of the fluctuations of microscopic quantities. Here, σ and σ' are particle type indices (e for electrons and i for ions), n_σ is the equilibrium number density, and $n_\sigma(x,t)$ is the microscopic number density at point x and at time t.

A correlation function can be evaluated by the kinetic method, discussed by Rostoker,[11] Weinstock,[12] and others.[13-16] We briefly summarize the connection between the fluctuation and the kinetic theory. The correlation function can be written as

$$G_{\sigma\sigma'}(x,t) = \int d^3v \int d^3v' n_{\sigma'} f_{\sigma\sigma'}(v') \psi_\sigma(v', \sigma' | v, x, t), \qquad (1)$$

where $\psi_\sigma(|v,x,t)$ behaves like a one particle distribution function and satisfies a linearized kinetic equation. In Eq. (1) $n_\sigma f_{\sigma\sigma}(v)$ is the Maxwellian

Copyright © 1973 by the American Institute of Physics

$$n_\sigma f_{o\sigma}(v) = n_\sigma (m_\sigma/2\pi T)^{3/2} \exp[-m_\sigma v^2/2T],$$

where m_σ is the mass of a σ-type particle and T is the equilibrium Kelvin temperature expressed in units such that the Boltzmann constant is set equal to one. The linear equation for ψ_σ can be obtained from the kinetic equation for the actual one-particle distribution function $F_\sigma(v,x,t)$,

$$\frac{\partial}{\partial t} F_\sigma(v,x,t) + v \cdot \frac{\partial}{\partial x} F_\sigma + \frac{e_\sigma}{m_\sigma} E_v(x,t) \cdot \frac{\partial}{\partial v} F_\sigma = \sum_{\sigma''} C_{\sigma\sigma''}(F_\sigma, F_{\sigma''}),$$

by letting $F_\sigma = n_\sigma f_{o\sigma} + \psi_\sigma$ and ignoring $o(\psi^2)$. Here $E_v(x,t)$ is the Vlasov self-consistent electric field and $C_{\sigma\sigma'}$ are the collision terms. Let us introduce the linear collision operators $\Omega_{\sigma\sigma'}$ in such a way that $\Omega_{ee}\psi_e + \Omega_{ei}\psi_i$ results from the linearized electron collision term $C_{ee} + C_{ei}$ and $\Omega_{ie}\psi_e + \Omega_{ii}\psi_i$ from the linearized ion collision term $C_{ie} + C_{ii}$. More discussion on the collision operators will be given in the next section. The coupled linear equations are then, for $\sigma = e$ or i,

$$\left(\frac{\partial}{\partial t} + v \cdot \frac{\partial}{\partial x}\right)\psi_\sigma(v'\sigma' \mid v,x,t) + \frac{n_\sigma e_\sigma}{T} f_{o\sigma}(v) v \cdot \frac{\partial}{\partial x} \int d^3x'' \frac{1}{|x - x''|} \cdot \sum_{\sigma''} \int dv'' e_{\sigma''} \psi_{\sigma''}(v',\sigma' \mid v'',x'',t) = \Omega_{\sigma e}\psi_e(v'\sigma' \mid v,x,t) + \Omega_{\sigma i}\psi_i(v'\sigma' \mid v,x,t). \quad (2)$$

The initial condition is given by

$$\psi_\sigma(v'\sigma' \mid v,x,o) = \delta_{\sigma\sigma'}\delta(v - v')\delta(x) - n_\sigma f_{o\sigma}(v)(e_\sigma e_{\sigma'}/Tx)e^{-k_0 x}, \quad (3)$$

where $k_0 = (\Sigma 4\pi n_\sigma e_\sigma^2/T)^{1/2}$ is the inverse Debye length, $\delta_{\sigma\sigma'}$ the Kronecker delta and δ ( ) is the Dirac delta function. At time $t = 0$, Eq. (1) reduces to the Debye–Hückel formula

$$G_{\sigma\sigma'}(x,o) = n_\sigma \delta_{\sigma\sigma'}\delta(x) - (n_\sigma e_\sigma n_{\sigma'} e_{\sigma'}/Tx)e^{-k_0 x}. \quad (4)$$

The light scattering function is given by the Fourier–Laplace transform of the electron density fluctuation,[17]

$$S(k,\omega) = 2\, Re\, G_{ee}(k,\omega), \quad (5)$$

where the transform of $G_{\sigma\sigma'}(x,t)$ is defined as

$$G_{\sigma\sigma'}(k,\omega) = \int_0^\infty dt e^{i\omega t} \int d^3x\, e^{ik\cdot x} G_{\sigma\sigma'}(x,t). \quad (6)$$

Another quantity of interest is the energy-loss function $E(k,\omega)$ which describes the energy loss of a charged particle passing through the plasma and is given in terms of the charge density fluctuation,[18]

$$E(k,\omega) = 2\, Re\, G(k,\omega) = 2\, Re \sum_\sigma \sum_{\sigma'} e_\sigma e_{\sigma'}(k,\omega). \quad (7)$$

The electrical conductivity of the system in response to an external longitudinal electric field is given by[19]

$$\sigma(k,\omega) = (1/T)G^{(l)}(k,\omega)[1 + (4\pi/i\omega T)G^{(l)}(k,\omega)]^{-1} \quad (8)$$

where $G^{(l)}(k,\omega)$ is the longitudinal part of the electric current fluctuation, which is related to $G(k,\omega)$ through the equation of charge conservation,

$$G^{(l)}(k,\omega) = -(i\omega/k^2)[g(k) + i\omega G(k,\omega)], \quad (9)$$

where $g(k) = \sum_\sigma \sum_{\sigma'} e_\sigma e_{\sigma'} g_{\sigma\sigma'}(k)$, and $g_{\sigma\sigma'}(k)$ is the Fourier transform of Eq. (4).

A formal solution of Eq. (2) with the initial condition (3) can be obtained in a straightforward manner using the standard Fourier–Laplace transform techniques. The use of matrix notations keep expressions compact. We have

$$\begin{pmatrix}\psi_e \\ \psi_i\end{pmatrix} = \frac{1}{R}\begin{pmatrix}\psi_e(o) \\ \psi_i(o)\end{pmatrix} - \frac{4\pi A}{Tk^2}\frac{1}{R}\begin{pmatrix}n_e f_{oe} e_e ik\cdot v \\ n_i f_{oi} e_i ik\cdot v\end{pmatrix}, \quad (10)$$

where $\psi_\sigma(o)$ is the Fourier transform of Eq. (3), R is the operator-matrix

$$R = \Omega + ik\cdot v + i\omega = \begin{pmatrix}\Omega_{ee} + ik\cdot v + i\omega, \Omega_{ei} \\ \Omega_{ie}, \Omega_{ii} + ik\cdot v + i\omega\end{pmatrix}, \quad (11)$$

and A is the Vlasov integral term

$$A(v'\sigma' \mid k\omega) = \int d^3v(e_e, e_i)\begin{pmatrix}\psi_e \\ \psi_i\end{pmatrix} = \int d^3v(e_e\psi_e + e_i\psi_i). \quad (12)$$

Equations (10) and (12) can be solved for A,

$$A = -\frac{1}{\epsilon(k,\omega)} \int d^3v(e_e, e_i)\frac{1}{R}\begin{pmatrix}\psi_e & (o) \\ \psi_i & (o)\end{pmatrix}, \quad (13)$$

in which ϵ is the longitudinal part of the dielectric tensor

$$\epsilon(k,\omega) = 1 + \frac{4\pi}{Tk^2} \int d^3v(e_e, e_i)\frac{1}{R}\begin{pmatrix}n_e f_{oe} e_e ik\cdot v \\ n_i f_{oi} e_i ik\cdot v\end{pmatrix}. \quad (14)$$

It is convenient to introduce the 1×2 matrices

$$\Delta_1(k,\omega) = \begin{pmatrix}\Delta_{1e} \\ \Delta_{1i}\end{pmatrix} = \int d^3v \frac{1}{R}\begin{pmatrix}f_{oe} & (v) \\ o\end{pmatrix},$$
$$\Delta_2(k,\omega) = \begin{pmatrix}\Delta_{2e} \\ \Delta_{2i}\end{pmatrix} = \int d^3v \frac{1}{R}\begin{pmatrix}o \\ f_{oi} & (v)\end{pmatrix}, \quad (15)$$

and the electric susceptibilities

$$\chi_e(k,\omega) = (n_e e_e^2/Tk^2)[1 - i\omega(\Delta_{1e} - \Delta_{2e})],$$
$$\chi_i(k,\omega) = (n_i e_i^2/Tk^2)[1 - i\omega(\Delta_{2i} - \Delta_{1i})]. \quad (16)$$

In terms of (16), Eq. (14) becomes

$$\epsilon(k,\omega) = 1 + 4\pi\chi_e + 4\pi\chi_i. \quad (17)$$

In the calculation we have used the fact that the collision terms vanish for Maxwellian distributions, or equivalently,

$$\Omega\begin{pmatrix}d_e & n_e & f_{oe} & (v) \\ d_i & n_i & f_{oi} & (v)\end{pmatrix} = 0,$$

for arbitrary constants d_e, d_i.

A straightforward substitution of Eqs. (10) and (13) into Eq. (1) then gives the expression

$$i\omega G_{\sigma\sigma'}(k,\omega) = -g_{\sigma\sigma'}(k) + [Tk^2\chi_\sigma/e_\sigma^2\epsilon] \times[\delta_{\sigma\sigma'}\epsilon - (4\pi e_\sigma/e_{\sigma'})\chi_{\sigma'}] + i\omega\bar{G}_{\sigma\sigma'}(k,\omega), \quad (18)$$

where the term $\bar{G}_{\sigma\sigma'}(k,\omega)$ is due to the nondiagonality of the operator Ω and is given by

$$\bar{G}_{\sigma\sigma'}(k,\omega) = -H_{\sigma\sigma'} + (4\pi\chi_\sigma/e_\sigma\epsilon)\sum_{\sigma''} e_{\sigma''}H_{\sigma''\sigma'},$$
$$H_{ee} = n_e\Delta_{2e}, \quad H_{ei} = n_i\Delta_{2e},$$
$$H_{ie} = n_e\Delta_{1i}, \quad H_{ii} = n_i\Delta_{1i}. \quad (19)$$

The charge density–density correlation function follows immediately from Eq. (18),

$$G(k,\omega) = \sum_\sigma\sum_{\sigma'} e_\sigma e_{\sigma'}G_{\sigma\sigma'}(k,\omega) = -(Tk^2/4\pi i\omega)\{\epsilon^{-1} - [k^2/(k_0^2 + k^2)]\}. \quad (20)$$

From the last equation and Eqs. (9) and (8) we obtain

$$G^{(1)}(k,\omega) = (i\omega T/4\pi)[\epsilon^{-1}(k,\omega) - 1], \quad (21)$$

$$\sigma(k,\omega) = (i\omega/4\pi)[1 - \epsilon(k,\omega)]. \quad (22)$$

The formal expressions (18), (20), (21), and (22) in terms of χ_σ and ϵ agree with the general fluctuation theory.[1] The effects of collisions are clearly indicated by the presence of the collision operator in the electrical susceptibilities and the dielectric constant.

The problem reduces to finding the quantities $\Delta_1(k,\omega)$ and $\Delta_2(k,\omega)$ introduced in Eqs. (15). This can be accomplished by solving the eigenvalue problem of the operator $\Omega + ik\cdot v$,

$$(\Omega + ik\cdot v)\Psi = \lambda\Psi. \quad (23)$$

It is convenient to factor out the Maxwellians

$$\Psi = \begin{pmatrix} n_e f_{oe}(v) & o \\ o & n_i f_{oi}(v) \end{pmatrix}\Phi = \begin{pmatrix} n_e f_{oe}(v)\ \Phi_e \\ n_i f_{oi}(v)\ \Phi_i \end{pmatrix}, \quad (24)$$

and to introduce the 2×2 operator matrix ξ whose elements $\xi_{\sigma\sigma'}$ are defined by

$$\Omega_{ee}(n_e f_{oe}(v)\Phi_e) = n_e f_{oe}(v)\xi_{ee}\Phi_e,$$
$$\Omega_{ei}(n_i f_{oi}(v)\Phi_i) = n_e f_{oe}(v)\xi_{ei}\Phi_i,$$
$$\Omega_{ie}(n_e f_{oe}(v)\Phi_e) = n_i f_{oi}(v)\xi_{ie}\Phi_e, \quad (25)$$
$$\Omega_{ii}(n_i f_{oi}(v)\Phi_i) = n_i f_{oi}(v)\xi_{ii}\Phi_i.$$

Then Eq. (23) is transformed to

$$(\xi + ik\cdot v)\Phi = \lambda\Phi, \quad (26)$$

and Eqs. (15) become

$$\Delta_1(k,\omega) = \int d^3v \begin{pmatrix} n_e f_{oe} & o \\ o & n_i f_{oi} \end{pmatrix}\frac{1}{\xi + ik\cdot v + i\omega}\begin{pmatrix} 1/n_e \\ o \end{pmatrix},$$
$$\Delta_2(k,\omega) = \int d^3v \begin{pmatrix} n_e f_{oe} & o \\ o & n_i f_{oi} \end{pmatrix}\frac{1}{\xi + ik\cdot v + i\omega}\begin{pmatrix} o \\ 1/n_i \end{pmatrix}. \quad (27)$$

Let us introduce

$$U_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (28)$$

and the scalar product

$$[A,B] = \int d^3v(A_e^*, A_i^*)\begin{pmatrix} n_e f_{oe} & o \\ o & n_i f_{oi} \end{pmatrix}\begin{pmatrix} B_e \\ B_i \end{pmatrix} = \int d^3v[n_e f_{oe}(v)A_e^*B_e + n_i f_{oi}(v)A_i^*B_i], \quad (29)$$

where * means complex conjugate. The collision operator ξ is hermitian in a Hilbert space with this scalar product, as will be discussed in the next section. Then Eqs. (27) can be written as

$$n_e\Delta_{1e} = [U_1, (\xi + ik\cdot v + i\omega)^{-1}U_1],$$
$$n_e\Delta_{1i} = [U_2, (\xi + ik\cdot v + i\omega)^{-1}U_1],$$
$$n_i\Delta_{2e} = [U_1, (\xi + ik\cdot v + i\omega)^{-1}U_2], \quad (30)$$
$$n_i\Delta_{2i} = [U_2, (\xi + ik\cdot v + i\omega)^{-1}U_2].$$

In the next section we will consider a collision model in which the spectrum of ξ is discrete. The addition of the streaming term in Eq. (26) will be treated by the perturbation method, regarding ik at first as a real quantity and then obtain the results of the correlation functions for real k by analytical continuation. Thus $\xi + ik\cdot v$ is regarded as hermitian and therefore its eigenfunctions form a complete orthonormal set. We then have the expansions

$$U_1 = \sum_n a_n\Phi_n, \quad U_2 = \sum_n b_n\Phi_n, \quad (31)$$

with

$$a_n = [\Phi_n, U_1], \quad b_n = [\Phi_n, U_2], \quad (32)$$

and the following expressions for the Δ's:

$$n_e\Delta_{1e} = \sum_n \frac{a_n a_n}{\lambda_n + i\omega}, \quad n_i\Delta_{2i} = \sum_n \frac{b_n b_n}{\lambda_n + i\omega}$$
$$n_e\Delta_{1i} = n_i\Delta_{2e} = \sum_n \frac{a_n b_n}{\lambda_n + i\omega}. \quad (33)$$

We have used the fact that all the a's and b's are real so long as ik is regarded as real. This will be shown in Sec. IV.

Because of the last equation of (33) and also the overall charge neutrality of the plasma, $n_e e_e + n_i e_i = 0$, the second term of $\bar{G}_{\sigma\sigma'}$ in Eq. (19) vanishes. The electron density–density correlation function, given by Eq. (18) thus simplifies to

$$G_{ee}(k,\omega) = -\frac{n_e}{n_i}\sum_n \frac{a_n b_n}{\lambda_n + i\omega} - \frac{g_{ee}(k)}{i\omega} + \frac{Tk^2}{e_e^2}\frac{\chi_e(1 + 4\pi\chi_i)}{i\omega\epsilon(k\omega)}. \quad (34)$$

Eqs. (33) can be substituted into Eqs. (16) to obtain

$$\chi_e(k,\omega) = (n_e e_e^2/Tk^2)\sum_n \lambda_n a_n[(a_n/n_e) - (b_n/n_i)]/(\lambda_n + i\omega),$$
$$\chi_i(k,\omega) = (n_i e_i^2/Tk^2)\sum_n \lambda_n b_n[(b_n/n_i) - (a_n/n_e)]/(\lambda_n + i\omega). \quad (35)$$

The facts that $[U_1,U_1] = n_e = \sum_n a_n a_n$, $[U_2,U_2] = n_i = \sum_n b_n b_n$, and $[U_1,U_2] = 0 = \sum_n a_n b_n$ have been used. The problem reduces to finding the coefficients a's and b's by solving the eigenvalue problem (26) with an appropriate collision model.

III. A COLLISION MODEL

We start by considering the Landau collision terms given by[4]

$$C_{\sigma\sigma'}(F_\sigma, F_{\sigma'}) = -\frac{2\pi\lambda_c e_\sigma^2 e_{\sigma'}^2}{m_\sigma}\frac{\partial}{\partial v_\alpha} \times \int d^3v' W_{\alpha\beta}\left(\frac{F_\sigma(v)}{m_{\sigma'}}\cdot\frac{\partial F_{\sigma'}(v')}{\partial v'_\beta} - \frac{F_{\sigma'}(v')}{m_\sigma}\cdot\frac{\partial F_\sigma(v)}{\partial v_\beta}\right),$$

where

$$W_{\alpha\beta} = \frac{1}{w^3}(w^2\delta_{\alpha\beta} - w_\alpha w_\beta), \qquad w_\alpha = v_\alpha - v'_\alpha,$$

and λ_c is the Coulomb logarithm. To obtain the corresponding Landau ξ operator, we let the distribution functions be $F_\sigma = n_\sigma f_{o\sigma}(1+\phi_\sigma)$ and perform the linearization procedure as described in the previous section. We have

$$\xi_{ee}\phi_e = -\frac{2\pi\lambda_c e_e^2}{m_e^2}\left(\frac{\partial}{\partial v_\alpha} - \frac{m_e}{T}v_\alpha\right) \times\left(\int d^3v' E^{(e)}_{\alpha\beta}(v,v')\frac{\partial\phi_e(v')}{\partial v'_\beta} - B_{\alpha\beta}(v)\frac{\partial\phi_e(v)}{\partial v_\beta}\right),$$

$$\xi_{ei}\phi_i = -\frac{2\pi\lambda_c e_e^2}{m_e m_i}\left(\frac{\partial}{\partial v_\alpha} - \frac{m_e}{T}v_\alpha\right)\int d^3v' E^{(i)}_{\alpha\beta}(v,v')\frac{\partial\phi_i(v')}{\partial v'_\beta},$$

$$\xi_{ie}\phi_e = -\frac{2\pi\lambda_c e_i^2}{m_e m_i}\left(\frac{\partial}{\partial v_\alpha} - \frac{m_i}{T}v_\alpha\right)\int d^3v' E^{(e)}_{\alpha\beta}(v,v')\frac{\partial\phi_e(v')}{\partial v'_\beta},$$

$$\xi_{ii}\phi_i = -\frac{2\pi\lambda_c e_i^2}{m_i^2}\left(\frac{\partial}{\partial v_\alpha} - \frac{m_i}{T}v_\alpha\right) \times\left(\int d^3v' E^{(i)}_{\alpha\beta}(v,v')\frac{\partial\phi_i(v')}{\partial v'_\beta} - B_{\alpha\beta}(v)\frac{\partial\phi_i(v)}{\partial v_\beta}\right), \tag{36}$$

where

$$E^{(\sigma)}_{\alpha\beta}(v,v') = n_\sigma e_\sigma^2 f_{oo}(v')W_{\alpha\beta},$$

$$B_{\alpha\beta}(v) = \int d^3v'[E^{(e)}_{\alpha\beta}(v,v') + E^{(i)}_{\alpha\beta}(v,v')].$$

Two important properties of the Landau ξ operator can be shown easily. First, ξ is hermitian in a Hilbert space whose scaler product is defined by Eq. (29). Specifically,

$$\begin{aligned}
&\int d^3v n_e f_{oe}\psi_e^*\xi_{ee}\phi_e = \int d^3v n_e f_{oe}\phi_e\xi_{ee}\psi_e^*,\\
&\int d^3v n_e f_{oe}\psi_e^*\xi_{ei}\phi_i = \int d^3v n_i f_{oi}\phi_i\xi_{ie}\psi_e^*,\\
&\int d^3v n_i f_{oi}\psi_i^*\xi_{ie}\phi_e = \int d^3v n_e f_{oe}\phi_e\xi_{ei}\psi_i^*,\\
&\int d^3v n_i f_{oi}\psi_i^*\xi_{ii}\phi_i = \int d^3v n_i f_{oi}\phi_i\xi_{ii}\psi_i^*.
\end{aligned} \tag{37}$$

The second property is that ξ has a six-fold degenerate eigenvalue zero:

$$\xi\begin{pmatrix}1\\1\end{pmatrix} = o, \quad \xi\begin{pmatrix}-1\\1\end{pmatrix} = o, \quad \xi\begin{pmatrix}m_e v_\alpha\\ m_i v_\alpha\end{pmatrix} = o, \quad \xi\begin{pmatrix}\frac{1}{2}m_e v^2\\ \frac{1}{2}m_i v^2\end{pmatrix} = o, \tag{38}$$

which corresponds to the conservation of the particle numbers, linear momentum, and kinetic energy.

The eigenvalue problem of the Landau ξ operator is very complex mathematically. There are indications that it may have a continuous spectrum extending down to the origin. Indeed, the operator

$$\xi_e\phi = -\frac{2\pi\lambda_c e_e^2}{m_e^2}\left(\frac{\partial}{\partial v_\alpha} - \frac{m_e}{T}v_\alpha\right) \times\int d^3v' E^{(e)}_{\alpha\beta}(v,v')\left(\frac{\partial\phi(v')}{\partial v'_\beta} - \frac{\partial\phi(v)}{\partial v_\beta}\right)$$

for an electron gas system has been studied by Su,[5] Lewis,[6] and McLeod and Ong,[7] and is shown to have a continuous spectrum for the entire negative real axis. Furthermore, the streaming term in Eq. (26) can not be easily handled by a perturbation method if ξ has a continuous spectrum extending to the origin. Because of these mathematical complexity, it is desirable to consider a collision model which retains as many important features of the Landau operator as possible, yet on the other hand is simple enough to be solved.

In a previous paper McLennan and the author[9] have used the simplified collision model,

$$\begin{aligned}
&\Omega_{ee} = \gamma_e(\partial/\partial v_\alpha)[v_\alpha + (T/m_e)(\partial/\partial v_\alpha)], \qquad \Omega_{ei} = 0,\\
&\Omega_{ii} = \gamma_i(\partial/\partial v_\alpha)[v_\alpha + (T/m_i)(\partial/\partial v_\alpha)], \qquad \Omega_{ie} = 0,
\end{aligned} \tag{39}$$

to calculate the conductivity, energy-loss function, and scattering function. It was first suggested and used by Lenard and Berstein,[8] for the case of a single component electron gas, to discuss the effects of collisions on Landau damping. This Brownian motion type collision operator describes a diffusion in velocity space, a very important character of the long-range Coulomb interactions. Yet it is so simple that even the eigenvalue problem (26) with the streaming term included can be solved analytically.[20]

The simple collision operator (39) does not conserve momentum and energy, though it does conserve particle numbers. The effects of collisions between electrons and ions are also ignored. Here we consider an improved model which includes all the six hydrodynamic modes and also takes into account the interspecies collisions.

For a one-component system, the Brownian motion type collision term

$$C(F,F) = \gamma(\partial/\partial v_\alpha)[(v_\alpha - \bar{v}_\alpha)F + (\bar{T}/m)(\partial F/\partial v_\alpha)]$$

has been used by Dougherty[10] to discuss the effect of ion–ion collisions on the gyroresonance in the ionosphere radio wave scattering experiments. Here γ is a constant and $\bar{v}_\alpha(x,t)$ and $\bar{T}(x,t)$ are functionals of the distribution function $F(v,x,t)$. In order to have conservations of momentum and energy, it requires $\bar{v}_\alpha$ and $\bar{T}$ to be the local velocity and temperature,

$$\begin{aligned}
&\bar{v}_\alpha = n(x,t)^{-1}\int d^3v v_\alpha F,\\
&\bar{T} = (2/3n(x,t))\int d^3v (m/2)(v_\alpha - \bar{v}_\alpha)^2 F,
\end{aligned}$$

where $n(x,t) = \int d^3v F(x,v,t)$ is the local number density.

The linearization process described before can be carried out to find the corresponding ξ operator. Namely,

put $F = n_o f_o(v)(1 + \phi)$, ignoring $o(\phi^2)$ and identify ξ from $C = n_o f_o \xi\phi$. We find

$$\xi\phi = \gamma\{I_o + (m/T)v_\alpha[v_\alpha|\phi] + (4/3)E[E|\phi]\},$$

where

$$I_o = -v_\alpha \partial/\partial v_\alpha + (T/m)\partial^2/\partial v_\alpha^2, \qquad (40)$$
$$E = (mv^2/2T) - (3/2),$$

and the notation $[\ |\ ]$ is defined as

$$[A|B] = \int d^3v f_o(v) A^* B. \qquad (41)$$

We now consider a two-component system of electrons and ions. Suggested by the form of the Landau ξ operator of Eq. (36) we assume the following for the ξ operator,

$$\begin{aligned}
\xi_{ee}\phi_e &= \gamma_e I_{oe}\phi_e + \alpha_{ee} v_\alpha[v_\alpha|\phi_e]_e + \beta_{ee}(4/3)E_e[E_e|\phi_e]_e,\\
\xi_{ei}\phi_i &= \alpha_{ei} v_\alpha[v_\alpha|\phi_i]_i + \beta_{ei}(4/3)E_e[E_i|\phi_i]_i,\\
\xi_{ie}\phi_e &= \alpha_{ie} v_\alpha[v_\alpha|\phi_e]_e + \beta_{ie}(4/3)E_i[E_e|\phi_e]_e,\\
\xi_{ii}\phi_i &= \gamma_i I_{oi}\phi_i + \alpha_{ii} v_\alpha[v_\alpha|\phi_i]_i + \beta_{ii}(4/3)E_i[E_i|\phi_i]_i.
\end{aligned} \qquad (42)$$

Here I_{oo}, E_o and $[\ |\ ]_o$ are defined similar to I_o, E and $[\ |\ ]$ in Eqs. (40) and (41), except that the species characteristics are included; that is, m is replaced by m_o and $f_o(v)$ by $f_{oo}(v)$. The α's, β's, γ's are constants to be adjusted to retain the two important properties of the Landau ξ, namely, the hermitian property (37) and the conservation laws (38). The former results in the relations $n_e\alpha_{ei} = n_i\alpha_{ie}$, $n_e\beta_{ei} = n_i\beta_{ie}$, and the latter in $\gamma_e = (T/m_e)(\alpha_{ee} + \alpha_{ei}) = \beta_{ei} + \beta_{ee}$ and $\gamma_i = (T/m_i)(\alpha_{ie} + \alpha_{ii}) = \beta_{ie} + \beta_{ii}$. Introducing two new constants, $\alpha = \alpha_{ei}/n_i = \alpha_{ie}/n_e$ and $\beta = \beta_{ei}/n_i = \beta_{ie}/n_e$, we can put Eqs. (42) in the form

$$\begin{aligned}
\xi_{ee}\phi_e &= \gamma_e I_{oe}\phi_e + [(m_e/T)\gamma_e - n_i\alpha] v_\alpha[v_\alpha|\phi_e]_e\\
&\quad + (4/3)(\gamma_e - n_i\beta)E_e[E_e|\phi_e]_e,\\
\xi_{ei}\phi_i &= n_i\alpha v_\alpha[v_\alpha|\phi_i]_i + (4/3)n_i\beta E_e[E_i|\phi_i]_i,\\
\xi_{ie}\phi_e &= n_e\alpha v_\alpha[v_\alpha|\phi_e]_e + (4/3)n_e\beta E_i[E_e|\phi_e]_e,\\
\xi_{ii}\phi_i &= \gamma_i I_{oi}\phi_i + [(m_i/T)\gamma_i - n_e\alpha] v_\alpha[v_\alpha|\phi_i]_i\\
&\quad + (4/3)(\gamma_i - n_e\beta)E_i[E_i|\phi_i]_i.
\end{aligned} \qquad (43)$$

There are still four free parameters α, β, γ_e, and γ_i in the collision model. It will be shown later that α is related to the electrical conductivity; γ_e and γ_i are the collisional frequencies. In the sense that those parameters are adjustable to fit experimental results, the present theory, as in the case of the Krook model or the Lenard–Bernstein model, is not a purely microscopic one, but must be rather regarded as a semiempirical theory.

A few words remarking the notation $[\ |\ ]_o$ and the scalar product $[,]$ defined in Eq. (29): They are different, and are related,

$$[A, B] = n_e[A_e|B_e]_e + n_i[A_i|B_i]_i. \qquad (44)$$

For further discussion of the collision model, it is convenient to introduce the dimensionless quantities

$$u_\alpha = (m_e/2T)^{1/2} v_\alpha, \qquad u'_\alpha = (m_i/2T)^{1/2} v_\alpha. \qquad (45)$$

Then the differential operators I_{oe} and I_{oi} become

$$\begin{aligned}
I_{oe} &= (1/2)(\partial^2/\partial u_\alpha^2) - u_\alpha(\partial/\partial u_\alpha)\\
I_{oi} &= (1/2)(\partial^2/\partial u_\alpha'^2) - u'_\alpha(\partial/\partial u'_\alpha),
\end{aligned} \qquad (46)$$

and the integral operators $[\ |\ ]_e$ and $[\ |\ ]_i$ become

$$\begin{aligned}
[A|B]_e &= \pi^{-3/2}\int d^3u e^{-u^2} A^* B,\\
[A|B]_i &= \pi^{-3/2}\int d^3u' e^{-u'^2} A^* B.
\end{aligned} \qquad (47)$$

Let us also introduce the complete orthonormal set

$$p_{nlm}(\mathbf{u}) = [2\pi^{3/2} n!/\Gamma(n + l + 3/2)^3] L_n^{l+1/2}(u^2) u^l Y_{lm}(\theta, \phi), \qquad (48)$$

where (u, θ, ϕ) is the spherical coordinates of the vector $\mathbf{u}$, Γ is the gamma function,[21] L_n^a is the Laguerre polynomial,[21] and Y_{lm} is the spherical harmonic.[21] The indices n and l can take all nonnegative integers, and the value of m is restricted to $-l, -l+1, \cdots +l$ for a given l. It can be easily shown that (48) constitutes the complete set of the eigenfunctions of the operator I_{oe}, with the corresponding discrete eigenvalues $-(2n + l)$. In other words,

$$I_{oe} p_{nlm}(\mathbf{u}) = -(2n + l) p_{nlm}(\mathbf{u}). \qquad (49)$$

Similarly,

$$I_{oi} p_{nlm}(\mathbf{u}') = -(2n + l) p_{nlm}(\mathbf{u}'). \qquad (50)$$

The orthonormality of p_{nlm} is refering to the integral products (47),

$$\begin{aligned}
[p_{nlm}(\mathbf{u})|p_{n'l'm'}(\mathbf{u})]_e &= \delta_{nn'}\delta_{ll'}\delta_{mm'},\\
[p_{nlm}(\mathbf{u}')|p_{n'l'm'}(\mathbf{u}')]_i &= \delta_{nn'}\delta_{ll'}\delta_{mm'}.
\end{aligned} \qquad (51)$$

Noticing that from Eqs. (45) and (48), the velocity vector $\mathbf{v}$ and the dimensionless energy E_o can be written as

$$\begin{aligned}
v_1 &= i(T/2m_e)^{1/2}[p_{011}(\mathbf{u}) + p_{01\text{-}1}(\mathbf{u})]\\
&= i(T/2m_i)^{1/2}[p_{011}(\mathbf{u}') + p_{01\text{-}1}(\mathbf{u}')],\\
v_2 &= -(T/2m_e)^{1/2}[p_{011}(\mathbf{u}) - p_{01\text{-}1}(\mathbf{u})]\\
&= -(T/2m_i)^{1/2}[p_{011}(\mathbf{u}') - p_{01\text{-}1}(\mathbf{u}')],\\
v_3 &= (T/m_e)^{1/2} p_{010}(\mathbf{u}) = (T/m_i) p_{010}(\mathbf{u}'),\\
E_e &= u^2 - (3/2) = -(3/2)^{1/2} p_{100}(\mathbf{u}),\\
E_i &= u'^2 - (3/2) = -(3/2)^{1/2} p_{100}(\mathbf{u}'),
\end{aligned}$$

We can put Eq. (43) in the following form

$$\begin{aligned}
\xi_{ee}\phi_e &= \gamma_e I_{oe}\phi_e + [\gamma_e - (T/m_e) n_i\alpha] p_{01m}(\mathbf{u})[p_{01m}(\mathbf{u})|\phi_e]_e\\
&\quad + 2(\gamma_e - n_i\beta) p_{100}(\mathbf{u})[p_{100}(\mathbf{u})|\phi_e]_e,\\
\xi_{ei}\phi_i &= (T/m_i) n_i\alpha p_{01m}(\mathbf{u}')[p_{01m}(\mathbf{u}')|\phi_i]_i\\
&\quad + 2n_i\beta p_{100}(\mathbf{u})[p_{100}(\mathbf{u}')|\phi_i]_i,\\
\xi_{ie}\phi_e &= (T/m_e) n_e\alpha p_{01m}(\mathbf{u})[p_{01m}(\mathbf{u})|\phi_e]_e\\
&\quad + 2n_e\beta p_{100}(\mathbf{u}')[p_{100}(\mathbf{u})|\phi_e]_e,\\
\xi_{ii}\phi_i &= \gamma_i I_{oi}\phi_i + [\gamma_i - (T/m_i) n_e\alpha] p_{01m}(\mathbf{u}')[p_{01m}(\mathbf{u}')|\phi_i]_i\\
&\quad + 2(\gamma_i - n_e\beta) p_{100}(\mathbf{u}')[p_{100}(\mathbf{u}')|\phi_i]_i.
\end{aligned} \qquad (52)$$

Here the summation over $m = -1, 0, 1$ is implied.

Now with the form of Eq. (52) for our collision model, the eigenvalue problem, $\xi\phi = \mu\phi$, is almost trivial. We list

below all the possible eigenvalues and their corresponding degenerate eigenfunctions:

$$\mu = 0,$$

$$\phi_0^{(1)} = N^{-1/2}\begin{pmatrix}1\\1\end{pmatrix}, \quad \phi_0^{(2)} = N^{-1/2}\begin{pmatrix}-1\\1\end{pmatrix},$$

$$\phi_0^{(3)} = N^{-1/2}\begin{pmatrix}p_{100}(\mathbf{u})\\p_{100}(\mathbf{u}')\end{pmatrix}, \tag{53}$$

$$\phi_0^{(4,5,6)} = \begin{pmatrix}(m_e/\rho)^{1/2}\ p_{01m}(\mathbf{u})\\(m_i/\rho)^{1/2}\ p_{01m}(\mathbf{u}')\end{pmatrix}, \quad m = 0, 1, -1,$$

$$\mu = -\mu_1,$$

$$\phi_1^{(m)} = \begin{pmatrix}(m_i n_i/\rho n_e)^{1/2}\ p_{01m}(\mathbf{u})\\-(m_e n_e/\rho n_i)^{1/2}\ p_{01m}(\mathbf{u}')\end{pmatrix}, \quad m = -1, 0, 1,$$

$$\mu = -\mu_2,$$

$$\phi_2 = \begin{pmatrix}(n_i/n_e N)^{1/2}\ p_{100}(\mathbf{u})\\-(n_e/n_i N)^{1/2}\ p_{100}(\mathbf{u}')\end{pmatrix}, \tag{54}$$

$$\mu = -\mu_{nl,\sigma}, \quad (n, l) \neq (0, 0), (0, 1), (1, 0),$$

$$\phi_{nl,\sigma}^{(m)} = \begin{pmatrix}n_\sigma^{-1/2}\ p_{nlm}(\mathbf{u})\\0\end{pmatrix}, \quad m = -l, \ldots, +l$$

Here $N = n_e + n_i$ is the total number density, $\rho = n_e m_e + n_i m_i$ is the mass density and

$$\mu_1 = (T\rho/m_e m_i)\alpha, \quad \mu_2 = 2N\beta, \quad \mu_{nl,\sigma} = (2n + l)\gamma_\sigma. \tag{55}$$

The results (53) are expected; they are the conservation laws of Eqs. (38). The degenerate eigenfunctions listed above form a complete orthonormal set, also expected, since ξ is hermitian.

Actually the only crucial point involved in the collision model (52), as compared to the Landau ξ, is the assumption that ξ has a purely discrete spectrum. It is likely that results obtained on the basis of this assumption are still approximately valid, perhaps in an asymptotic sense, if a continuous spectrum is involved.

IV. TREATING THE STREAMING TERM BY PERTURBATION METHOD

We now consider the eigenvalue problem of Eq. (26), with the eigenvalues and eigenfunctions of ξ given by Eqs. (53) and (54). Introducing the quantity

$$\kappa = ik(2T/m_e)^{1/2}, \tag{56}$$

we can write Eq. (26) as

$$(\xi + \kappa \cdot u)\Phi = \lambda\Phi. \tag{57}$$

Since our primary interest is to evaluate the correlation functions for small k, the streaming term $\kappa \cdot u$ shall be treated as the perturbation. In order to make the operator $\xi + \kappa \cdot u$ a hermitian, and therefore to be able to apply the standard perturbation method, we first consider κ as real, and later obtain results for imaginary κ (i.e., real k) by analytic continuation. In order to avoid complication of algebraic nature we assume that the ions in the plasma are monovalent: $e_i = -e_e = e =$ electronic charge, and $n_e = n_i = n = N/2$. We also choose κ to be along the direction of the polar axis of the spherical coordinates so that $\kappa \cdot u = \kappa u \cos\theta = 2^{-1/2}\kappa p_{010}(\mathbf{u})$. The condition under which the streaming term can be treated as a perturbation to the collision operator will be discussed later.

The perturbation calculation is complicated by the degeneracy but is otherwise straightforward. It is carried out to the second order of κ for the eigenvalues and the first order for the eigenfunctions. We first list the results of the calculation and then explain the notations:

$$\lambda_0^{(1)} = \sigma\kappa^2,$$

$$\Phi_0^{(1)} = \phi_0^{(2)} - K_0^{(2)} + (M/m_e)^{1/2}(\Delta m/M)\sigma\kappa\phi_0^{(5)}$$

$$\lambda_0^{(2)} = D\kappa^2, \quad \Phi_0^{(2)} = (2/5)^{1/2}[\phi_0^{(1)} - K_0^{(1)} + (3/2)^{1/2}(\phi_0^{(3)} - K_0^{(3)}) + (M/m_e)^{1/2}D\kappa\phi_0^{(5)}],$$

$$\lambda_0^{(3,4)} = \pm c\kappa + \Gamma\kappa^2,$$

$$\Phi_0^{(3,4)} = (3/10)^{1/2}(1 \mp q_2\kappa)\{\phi_0^{(1)} - K_0^{(1)} \mp q_1\kappa\phi_0^{(2)} - (2/3)^{1/2}[1 \pm c(M/m_e)D\kappa](\phi_0^{(3)} - K_0^{(3)}) + (\pm c)(M/m_e)^{1/2}[1 \pm (\Gamma - (5/3)(\Delta m/M)^2\sigma)\kappa/c](\phi_0^{(4)} - K_0^{(4)})\}, \tag{58}$$

$$\lambda_0^{(5)} = \eta\kappa^2, \quad \Phi_0^{(5)} = \phi_0^{(5)} - K_0^{(5)},$$

$$\lambda_0^{(6)} = \eta\kappa^2, \quad \Phi_0^{(6)} = \phi_0^{(6)} - K_0^{(6)},$$

$$\lambda_1^{(m)} = -\mu_1 + O(\kappa^2), \quad \Phi_1^{(m)} = \phi_1^{(m)} - K_1^{(m)},$$

$$\lambda_2 = -\mu_2 + O(\kappa^2), \quad \Phi_2 = \phi_2 - K_2,$$

$$\lambda_{nl,e}^{(m)} = -\mu_{nl,e} + O(\kappa^2), \quad \Phi_{nl,e}^{(m)} = \phi_{nl,e}^{(m)} - K_{nl,e}^{(m)},$$

$$\lambda_{nl,i}^{(m)} = -\mu_{nl,i} + O(\kappa^2), \quad \Phi_{nl,i}^{(m)} = \phi_{nl,i}^{(m)} - K_{nl,i}^{(m)}.$$

The subscripts and superscripts on the λ's and Φ's are self-explanatory. The K's are first order in κ and are solutions of the following equations,

$$\xi K_0^{(j)} = (\kappa \cdot u)\phi_0^{(j)} - \sum_{i=1}^{6}[\phi_0^{(i)}, \kappa \cdot u\phi_0^{(j)}]\phi_0^{(i)},$$

$$(\xi + \mu_1)K_1^{(m)} = (\kappa \cdot u)\phi_1^{(m)} - \sum_{m'}[\phi_1^{(m)}, \kappa \cdot u\phi_1^{(m')}]\phi_1^{(m')} = (\kappa \cdot u)\phi_1^{(m)},$$

$$(\xi + \mu_2)K_2 = (\kappa \cdot u)\phi_2 - [\phi_2, \kappa \cdot u\phi_2] = (\kappa \cdot u)\phi_2,$$

$$(\xi + \mu_{nl,e})K_{nl,e}^{(m)} = (\kappa \cdot u)\phi_{nl,e}^{(m)} - \sum_{m'}[\phi_{nl,e}^{(m)}, \kappa \cdot u\phi_{nl,e}^{(m')}]\phi_{nl,e}^{(m')} = (\kappa \cdot u)\phi_{nl,e}^{(m)},$$

$$(\xi + \mu_{nl,i})K_{nl,i}^{(m)} = (\kappa \cdot u)\phi_{nl,i}^{(m)},$$

together with the conditions

$$[\phi_0^{(i)}, K_0^{(j)}] = 0, \quad [\phi_1^{(m)}, K_1^{(m')}] = 0,$$
$$[\phi_2, K_2] = 0, \quad [\phi_{nl,e}^{(m)}, K_{nl,e}^{(m')}] = 0, \tag{59}$$
$$[\phi_{nl,i}^{(m)}, K_{nl,i}^{(m')}] = 0.$$

Explicit form of the K's can be obtained, if the properties of ξ listed in Eqs. (53) and (54) are used,

$$K_0^{(1)} = -(\Delta m/M)K_0^{(2)}, \quad K_0^{(2)} = (\kappa/2\mu_1)(M/m_i)^{1/2}\phi_1^{(0)},$$

$$K_0^{(3)} = -(2/3)^{1/2}K_0^{(1)} - (5/12)^{1/2}\kappa[(\phi_{11,e}^{(0)}/\mu_{11,e}) + (m_e/m_i)^{1/2}(\phi_{11,i}^{(0)}/\mu_{11,i})],$$

$$K_0^{(4)} = -(2m_e/3M)^{1/2}\kappa[(\phi_{02,e}^{(0)}/\mu_{02,e}) + (\phi_{02,i}^{(0)}/\mu_{02,i})],$$

$$K_0^{(5)} = -(m_e/2M)^{1/2}\kappa[(\phi_{02,e}^{(-1)}/\mu_{02,e}) + (\phi_{02,i}^{(-1)}/\mu_{02,i})],$$

$$K_0^{(6)} = -(m_e/2M)^{1/2}\kappa[(\phi_{02,e}^{(1)}/\mu_{02,e}) + (\phi_{02,i}^{(1)}/\mu_{02,i})]$$

$$\begin{aligned}K_1^{(0)} = {} & (M/m_i)^{1/2}(\kappa/2\mu_1)[\phi_0^{(1)} + (\Delta m/M)\phi_0^{(2)}] \\ & - (M/6m_i)^{1/2}\kappa[(\Delta m/M)(\phi_0^{(3)}/\mu_1) + \phi_2/(\mu_1 - \mu_2)] \\ & + (2/3)^{1/2}\kappa\{[(m_i/M)^{1/2}/(\mu_1 - \mu_{02,e})]\phi_{02,e}^{(0)} \\ & + [(m_e^2/m_iM)^{1/2}/(\mu_1 - \mu_{02,i})]\phi_{02,i}^{(0)}\}. \end{aligned} \tag{60}$$

The constants D, c, and Γ are found to be

$$\begin{aligned} D &= (3/5)[\lambda_t - (l^2/\sigma)] = (3/5)\lambda_t - (2/5)(\Delta m/M)^2\sigma, \\ c &= (5m_e/3M)^{1/2}, \\ \Gamma &= (\lambda_t/5) + (2\eta/3) + (7/10)(\Delta m/M)^2\sigma, \end{aligned} \tag{61}$$

and σ, l, λ_t, and η are defined as

$$\begin{aligned} \sigma &= -[K_0^{(2)}, \xi K_0^{(2)}]/\kappa^2 = M/4\mu_1 m_i \\ l &= -[K_0^{(2)}, \xi K_0^{(3)}]/\kappa^2 = (2/3)^{1/2}(\Delta m/M)\sigma, \\ \lambda_t &= -[K_0^{(3)}, \xi K_0^{(3)}]/\kappa^2 = (2/3)(\Delta m/M)^2\sigma \\ &\quad + (5/12)[(1/\mu_{11,e}) + (m_e/m_i\mu_{11,i})], \\ \eta &= -[K_0^{(4)}, \xi K_0^{(4)}]/\kappa^2 = (m_e/2M)[(1/\mu_{02,e}) + (1/\mu_{02,i})]. \end{aligned} \tag{62}$$

Also, $\Delta m = m_i - m_e$, and

$$\begin{aligned} q_1 &= c\sigma\Delta m/m_e, \\ q_2 &= (2c)^{-1}[(3/5)\lambda_t + (2/3)\eta - (37/30)(\Delta m/M)^2\sigma]. \end{aligned} \tag{63}$$

We are now ready for the calculation of the coefficients a's and b's given by Eqs. (32). Notice that U_1 and U_2 are orthogonal to all the ϕ's and K's except $\phi_0^{(1)}$, $\phi_0^{(2)}$, and $K_1^{(0)}$.

The results are:

$$\begin{aligned} a_0^{(1)} &= -(n/2)^{1/2} + o(\kappa^2), \\ a_0^{(2)} &= (n/5)^{1/2}(1 + \alpha_2\kappa^2), \\ a_0^{(3)} &= (3n/20)^{1/2}[1 + (q_1 - q_2)\kappa + \alpha_3\kappa^2], \\ a_0^{(4)} &= (3n/20)^{1/2}[1 - (q_1 - q_2)\kappa + \alpha_4\kappa^2], \\ a_1^{(0)} &= -(nm_i/2M)^{1/2}(\kappa/\mu_1) + o(\kappa^2), \\ b_0^{(1)} &= (n/2)^{1/2} + o(\kappa^2), \\ b_0^{(2)} &= (n/5)^{1/2}(1 + \beta_2\kappa^2), \\ b_0^{(3)} &= (3n/20)^{1/2}[1 - (q_1 + q_2)\kappa + \beta_3\kappa^2], \\ b_0^{(4)} &= (3n/20)^{1/2}[1 + (q_1 + q_2)\kappa + \beta_4\kappa^2], \\ b_1^{(0)} &= (nm_e^2/2m_iM)^{1/2}(\kappa/\mu_1) + o(\kappa^2). \end{aligned} \tag{64}$$

All the other a's and b's are at least second order in κ. The second order terms $\alpha_2\kappa^2$, etc., can be determined only by going to the next higher order perturbation. They are indicated in Eqs. (64) for the purpose of later calculation of the scattering function. The detail forms of α's and β's are not needed.

Since κ is still regarded as real, all the a's and b's are also real—a fact that has been used in the derivation of Eqs. (33).

For monovalent ions, the electrical susceptibilities of Eqs. (35) become

$$\chi_e(k,\omega) = (e^2/Tk^2)\sum_n [\lambda_n a_n(a_n - b_n)/(\lambda_n + i\omega)],$$

$$\chi_i(k,\omega) = (e^2/Tk^2)\sum_n [\lambda_n b_n(b_n - a_n)/(\lambda_n + i\omega)],$$

and the dielectric constant of Eq. (17) becomes

$$\epsilon(k,\omega) = 1 + (4\pi e^2/Tk^2)\sum_n \lambda_n(a_n - b_n)^2/(\lambda_n + i\omega).$$

After the substitution of the a's and b's,

$$\begin{aligned}\chi_e(k,\omega) = (ne^2/Tk^2)\Bigg(& \frac{\lambda_0^{(1)}[1 + o(\kappa^2)]}{\lambda_0^{(1)} + i\omega} \\ & + \frac{\lambda_1^{(0)}(\kappa^2/2\mu_1^2)}{\lambda_1^{(0)} + i\omega} + \frac{\lambda_0^{(2)}(\alpha_2 - \beta_2)\kappa^2/5}{\lambda_0^{(2)} + i\omega} \\ & + \frac{\lambda_0^{(3)}(3/10)q_1\kappa\{1 + [q_1 - q_2 + (\alpha_3 - \beta_3)/2q_1]\kappa\}}{\lambda_0^{(3)} + i\omega} \\ & + \frac{\lambda_0^{(4)}(-3/10)q_1\kappa\{1 - [q_1 - q_2 + (\alpha_4 - \beta_4)/2q_1]\kappa\}}{\lambda_0^{(4)} + i\omega} \\ & + \cdots\Bigg), \end{aligned} \tag{65}$$

$$\begin{aligned}\chi_i(k,\omega) = (ne^2/Tk^2)\Bigg(& \frac{\lambda_0^{(1)}[1 + o(\kappa^2)]}{\lambda_0^{(1)} + i\omega} \\ & + \frac{\lambda_1^{(0)}(m_e/m_i)(\kappa^2/2\mu_1^2)}{\lambda_1^{(0)} + i\omega} + \frac{\lambda_0^{(2)}(\beta_2 - \alpha_2)/5}{\lambda_0^{(2)} + i\omega} \\ & + \frac{\lambda_0^{(3)}(-3/10)q_1\kappa\{1 - [q_1 + q_2 + (\beta_3 - \alpha_3)/2q_1]\kappa\}}{\lambda_0^{(3)} + i\omega} \\ & + \frac{\lambda_0^{(4)}(3/10)q_1\kappa\{1 + [q_1 + q_2 + (\beta_4 - \alpha_4)/2q_1]\kappa\}}{\lambda_0^{(4)} + i\omega} + \cdots\Bigg), \end{aligned} \tag{66}$$

$$\begin{aligned}\epsilon(k,\omega) = 1 + (4\pi ne^2/Tk^2)\Bigg(& \frac{\lambda_0^{(1)}[2 + o(\kappa^2)]}{\lambda_0^{(1)} + i\omega} + \frac{\lambda_1^{(0)}(2\sigma\kappa^2/\mu_1)}{\lambda_1^{(0)} + i\omega} \\ & + \frac{\lambda_0^{(2)}(\alpha_2 - \beta_2)^2\kappa^4/5}{\lambda_0^{(2)} + i\omega} + \frac{\lambda_0^{(3)}(3/5)q_1^2\kappa^2[1 + \kappa(\alpha_3 - \beta_3)/q_1]}{\lambda_0^{(3)} + i\omega} \\ & + \frac{\lambda_0^{(4)}(3/5)q_1^2\kappa^2[1 - \kappa(\alpha_4 - \beta_4)/q_1]}{\lambda_0^{(4)} + i\omega} + \cdots\Bigg). \end{aligned} \tag{67}$$

From here on κ can be replaced by the imaginary value of Eq. (56). But we will keep the notation whenever it is convenient.

We now discuss the validity of the perturbation method. Let us first introduce the collision times $\tau_e = \gamma_e^{-1}$, $\tau_i = \gamma_i^{-1}$, the mean free path $\lambda_{mfp} = (T/m_e)^{1/2}\tau_e = (T/m_i)^{1/2}\tau_i$, and the conductivity relaxation time τ_c. The last quantity has the same order of magnitude as the electron collision time τ_e and is related to the constant o of Eq. (62), $\tau_c = (4m_i/M)\sigma$, as shall be shown in the next section. Let us investigate the perturbed eigenvalues and eigenfunctions of Eqs. (58). The terms $\sigma\kappa^2, D\kappa^2, \Gamma\kappa^2$, and $\eta\kappa^2$, are all the same order of magnitude. The perturba-

tion parameter is then $(\Gamma\kappa^2/c\kappa) \sim (\sigma\kappa/c) \sim (m_i/m_e)^{1/2} \cdot (\lambda_{mfp}/\lambda) \equiv z$, where $\lambda = k^{-1}$ is the wavelength. Furthermore, all the first order perturbation quantities $(M/m_e)^{1/2}(\Delta m/M)o\kappa$, $(M/m_e)^{1/2}D\kappa, q_1\kappa, q_2\kappa, c(M/m_e)D\kappa$ and all the K's given by Eqs. (60) have the same order of magnitude as the perturbation parameter z. The condition for the validity of the perturbation method is therefore

$$z = (m_i/m_e)^{1/2}(\lambda_{mfp}/\lambda) < 1; \tag{68}$$

that is, when the wavelength is much larger than the mean free path.

Perturbation expansions have been done formally and will be done again later in term of k (or κ)—instead of z—which is directly proportional to z.

V. DISPERSION RELATIONS

From Eqs. (20), (21), and (34), it is obvious that the singularities of the correlation functions in the ω plane are determined by the dispersion equation

$$\epsilon(k, \omega) = 0.$$

We first consider the long wavelength limit of the dielectric constant given by Eq. (67). For $k \to 0$,

$$\epsilon(o, \omega) = 1 + \omega_p^2/i\omega(i\omega - \mu_1), \tag{69}$$

where

$$\omega_p = (4\pi ne^2 M/m_e m_i)^{1/2} \tag{70}$$

is the plasma oscillation frequency. The static conductivity can be obtained from Eqs. (69) and (22),

$$\sigma_c = \lim_{\omega\to 0}\lim_{k\to 0} \sigma(k, \omega) = \omega_p^2/4\pi\mu_1. \tag{71}$$

The familiar formula $\sigma_c = \omega_p^2\tau_c/4\pi$ defines the conductivity relaxation time τ_c. We learn therefore the meaning of the constant μ_1 as the inverse of the conductivity relaxation time,

$$\mu_1 = 1/\tau_c. \tag{72}$$

The constant σ, which is related to μ_1 through the first equation of (62), is directly related to the static conductivity

$$\sigma = (m_e/4ne^2)\sigma_c. \tag{73}$$

To solve the dispersion equation we put $i\omega = -\lambda_n + \delta_n = \nu_n$ in Eq. (67), expand δ_n in a power series of κ (or k), and then determine the coefficients of the expansion. We find the following solutions,

$$\begin{aligned}
&\nu_0^{(2)} = -\lambda_0^{(2)} + [(\alpha_2 - \beta_2)^2 D(D - \sigma)/10\sigma]\kappa^6,\\
&\nu_0^{(3)} = -\lambda_0^{(3)} + (5/6)(\Delta m/M)^2\sigma\kappa^2(1 + \gamma_3\kappa),\\
&\nu_0^{(4)} = -\lambda_0^{(4)} + (5/6)(\Delta m/M)^2\sigma\kappa^2(1 + \gamma_4\kappa),\\
&\nu_0^{(5)} = -\lambda_0^{(5)} + o(\kappa^6), \qquad \nu_0^{(6)} = -\lambda_0^{(6)} + o(\kappa^6),\\
&\nu_1^{(1)} = -\lambda_1^{(1)} + o(\kappa^4), \qquad \nu_1^{(-1)} = -\lambda_1^{(-1)} + o(\kappa^4),\\
&\nu_2 = -\lambda_2 + o(\kappa^2), \qquad \nu_{nl,e} = -\lambda_{nl,e} + o(\kappa^2),\\
&\nu_{nl,i} = -\lambda_{nl,i} + o(\kappa^2).
\end{aligned} \tag{74}$$

The constant γ_3 is given by

$$\gamma_3 = -(m_e c/16\pi ne^2\sigma) + [2\Gamma - \sigma - (5/6)(\Delta m/M)^2\sigma]/c + (\alpha_3 - \beta_3)/q_1 + (c/\mu_1) + (3/20)(cq_1^2/\sigma), \tag{75}$$

and γ_4 is given by a similar equation.

We pay special attentions to the cases $\nu_0^{(1)} = -\lambda_0^{(1)} + \delta_0^{(1)}$ and $\nu_1^{(0)} = -\lambda_1^{(0)} + \delta_1^{(0)}$. They are not listed in Eqs. (74). Both $\delta_0^{(1)}$ and $\delta_1^{(0)}$ are nonzero for $\kappa = 0$. It is more instructive to try solutions in the form

$$i\omega = A + B\kappa + \cdots$$

for $A \neq 0, \mu_1, \mu_2, \mu_{nl,e}$ or $\mu_{nl,i}$. The dispersion equation becomes

$$0 = 1 + \frac{4\pi ne^2}{T}\left(-\frac{2T}{m_e}\frac{1}{\kappa^2}\right)\left(\frac{2\sigma\kappa^2}{A + B\kappa + \cdots} + \frac{(-\mu_1)(2\sigma\kappa^2/\mu_1)}{(A - \mu_1) + B\kappa + \cdots} + \cdots\right),$$

from which we obtain $A^2 - \mu_1 A + \omega_p^2 = 0$ and $B = 0$. The quadratic equation gives a pair of complex conjugate roots, because for almost all types of plasma the Debye length $\lambda_D = k_0^{-1} = (T/8\pi ne^2)^{1/2}$ is much less than the mean free path,

$$y = (\lambda_D/\lambda_{mfp}) \ll 1, \tag{76}$$

or, equivalently, $\omega_p\tau_c \sim \omega_p\tau_e \sim y^{-1} \gg 1$.

Denoting this pair of solutions by ζ_1 and ζ_2, we add to the list of Eq. (73) the damped plasma oscillation poles

$$\begin{aligned}
\zeta_1 &= (2\tau_c)^{-1} + i\Omega_p + o(\kappa^2),\\
\zeta_2 &= (2\tau_c)^{-1} - i\Omega_p + o(\kappa^2),
\end{aligned} \tag{77}$$

where Ω_p is the oscillation frequency

$$\Omega_p = (4\omega_p^2\tau_c^2 - 1)^{1/2}/2\tau_c \cong \omega_p. \tag{78}$$

The relations

$$\zeta_1\zeta_2 = \omega_p^2 + o(\kappa^2), \qquad \zeta_1 + \zeta_2 = \mu_1 + o(\kappa^2) \tag{79}$$

are very useful for later calculations. The damped oscillation poles can also be obtained from Eq. (69) and $\epsilon(o, \omega) = 0$.

We have demonstrated here very clearly the combined effect of the Vlasov and the collision term. The charge relaxation pole $\sigma\kappa^2$ and the collisional relaxation pole $-\mu_1$ on one hand (pure collision), and the plasma oscillation poles $\pm i\omega_p$ on the other hand (collisionless), merge into the damped plasma oscillation poles. The Vlasov term also has a very interesting effect on the damping constant of the sound poles,

$$\begin{aligned}
\nu_0^{(3)} &= -i\bar{c}k + [\bar{\Gamma} - (5/6)(\Delta m/M)^2\bar{\sigma}]k^2,\\
\nu_0^{(4)} &= +i\bar{c}k + [\bar{\Gamma} - (5/6)(\Delta m/M)^2\bar{\sigma}]k^2,
\end{aligned} \tag{80}$$

where

$$\begin{aligned}
\bar{c} &= c\kappa/ik = (10T/3M)^{1/2},\\
\bar{\sigma} &= -\sigma\kappa^2/k^2 = (2T/m_e)\sigma = (T/2ne^2)\sigma_c,\\
\bar{\Gamma} &= -\Gamma\kappa^2/k^2 = (2T/m_e)\Gamma,
\end{aligned} \tag{81}$$

and $\bar{c}$ is the sound propagation speed. The sound damping constant is decreased by an amount of $(5/6)(\Delta m/M)^2\bar{\sigma}$ because of the collective effect of the Vlasov term.

VI. THE ENERGY-LOSS FUNCTION AND THE SCATTERING FUNCTION

From the expression (67) and the dispersion relations (74) and (77), it suggests that $\epsilon^{-1}(k,\omega)$ can be put in the following forms:

$$\frac{1}{\epsilon(k,\omega)} = \frac{(i\omega+\lambda_0^{(1)})(i\omega+\lambda_1^{(0)})(i\omega+\lambda_0^{(2)})\cdots}{(i\omega-\zeta_1)(i\omega-\zeta_2)(i\omega-\nu_0^{(2)})\cdots} \tag{82}$$

$$= 1 + [s_1/(i\omega-\zeta_1)] + [s_2/(i\omega-\zeta_2)] + [r_{02}/(i\omega-\nu_0^{(2)})] + \cdots \tag{83}$$

The residues $s_1, s_2\cdots$ can be determined from Eq. (82) and

$$s_1 = [(i\omega-\zeta_1)/\epsilon(k,\omega)]_{i\omega=\zeta_1}, \text{ etc.} \tag{84}$$

The calculation is straightforward,

$$\begin{aligned}
&s_1 = \omega_p^2/(\zeta_2-\zeta_1) + o(\kappa^2), \quad s_2 = \omega_p^2/(\zeta_1-\zeta_2) + o(\kappa^2),\\
&r_{02} = [(D-\sigma)^2\mu_1 D(\alpha_2-\beta_2)^2/(10\omega_p^2\sigma)]\kappa^8,\\
&r_{03} = (\mu_1 c/\omega_p^2)(5/6)(\Delta m/M)^2\sigma\kappa^3(1+\delta_3\kappa),\\
&r_{04} = -(\mu_1 c/\omega_p^2)(5/6)(\Delta m/M)^2\sigma\kappa^3(1+\delta_4\kappa),\\
&r_{05}, r_{06} = o(\kappa^8), \quad r_{11}, r_{1,-1} = o(\kappa^6).
\end{aligned} \tag{85}$$

All the other r's are at least $o(\kappa^2)$. Here δ_3 is given by

$$\delta_3 = \gamma_3 - (5/6)(\Delta m/M)^2\sigma(2c)^{-1} + [\Gamma - \sigma - (5/6)(\Delta m/M)^2\sigma]c^{-1} + (c/\mu_1) - (c\mu_1/\omega_p^2) \tag{86}$$

and δ_4 is given by a similar equation.

We now substitute the expression (83) into Eqs. (20) and (7) to obtain the energy-loss function

$$E(k,\omega) = \frac{Tk^2}{2\pi}\left(\frac{\zeta_1\zeta_2(s_1\zeta_2/\zeta_1 + s_2\zeta_1/\zeta_2) + \omega^2(s_1+s_2)}{(\zeta_1\zeta_2-\omega^2)^2 + \omega^2(\zeta_1+\zeta_2)^2} + \frac{r_{02}}{\omega^2+(\nu_0^{(2)})^2} + \cdots\right). \tag{87}$$

To the lowest order of k,

$$E(k,\omega) = \frac{Tk^2}{2\pi}\frac{\omega_p^2/\tau_c}{(\omega^2-\omega_p^2)^2 + \omega^2/\tau_c^2}. \tag{88}$$

The damped plasma oscillations dominate the energy loss of a passing charged particle through a plasma. The domination of the plasma oscillation poles can also be demonstrated by calculating the total contribution. From Eqs. (87) and (83),

$$\begin{aligned}
\int_0^\infty E(k,\omega)d\omega &= (Tk^2/4)[(s_1/\zeta_1) + (s_2/\zeta_2) + (r_{02}/\nu_0^{(2)}) + \cdots]\\
&= (Tk^2/4)[1-\epsilon(k,0)^{-1}]\\
&= (Tk^2/4)[k_0^2/(k_0^2+k^2)].
\end{aligned}$$

Since $(s_1/\zeta_1) + (s_2/\zeta_2) = 1 + o(k^2)$, the plasma oscillation poles give dominate contribution and the contributions from all other poles are order of $(k^2/k_0^2) = (\lambda_D/\lambda)^2 = (m_e/m_i)y^2z^2 \ll 1$. In the above calculation we have used the fact that

$$\epsilon(k,0) = (k_0^2+k^2)/k^2 \tag{89}$$

which follows from Eqs. (17) and (35).

The scattering function is given by Eqs. (5) and (34). It can be easily shown that $G_{ee}(k,\omega)$ is analytic at $i\omega = 0$ or $-\lambda_n$; the singularities come only from $\epsilon(k,\omega) = 0$. Let

$$\begin{aligned}
&[4\pi\chi_e(1+4\pi\chi_i)/\epsilon] - (4\pi e^2/Tk^2)i\omega\sum_n [a_n b_n/(\lambda_n+i\omega)]\\
&\quad = [s_1'/(i\omega-\zeta_1)] + [s_2'/(i\omega-\zeta_2)] + [r_{02}'/(i\omega-\nu_0^{(2)})] + \cdots .
\end{aligned} \tag{90}$$

We then have

$$G_{ee}(k,\omega) = \frac{Tk^2}{4\pi e^2}\left(\frac{s_1'/\zeta_1}{i\omega-\zeta_1} + \frac{s_2'/\zeta_2}{i\omega-\zeta_2} + \frac{r_{02}'/\nu_0^{(2)}}{i\omega-\nu_0^{(2)}} + \cdots\right) \tag{91}$$

$$= -\frac{Tk^2}{4\pi e^2}\left(\frac{\zeta_1\zeta_2(s_1'\zeta_2/\zeta_1 + s_2'\zeta_1/\zeta_2) + \omega^2(s_1'+s_2') + i\omega(\cdots)}{(\omega^2-\zeta_1\zeta_2)^2 + \omega^2(\zeta_1+\zeta_2)^2} + \frac{r_{02}'[1+i\omega(\cdots)]}{\omega^2+\bar{D}k^2}\right. \tag{92}$$

$$\left. + \frac{-r_{03}'[(\omega+\bar{c}k)/\bar{c}k + (\bar{\Gamma}_1 k/i\bar{c})(1+(\omega+\bar{c}k)/\bar{c}k)}{(\omega+\bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2} + \frac{r_{04}'[(\omega-\bar{c}k)/\bar{c}k + (\bar{\Gamma}_1 k/i\bar{c})(1-(\omega-\bar{c}k)/\bar{c}k)]}{(\omega-\bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2} + \cdots\right).$$

The terms denoted by $(\cdots)$ will not contribute to the scattering function in the lowest orders of k. The constant

$$\bar{\Gamma}_1 = \bar{\Gamma} - (5/6)(\Delta m/M)^2\bar{\sigma} \tag{93}$$

is the sound damping constant, and

$$\bar{D} = -D\kappa^2/k^2 = (T/2m_e)[(3/5)\lambda_t - (2/5)(\Delta m/M)^2\sigma] \tag{94}$$

is the thermal diffusion coefficient. In deriving Eq. (91) we have used the relation

$$-[(s_1'/\zeta_1) + (s_2'/\zeta_2) + \ldots] = (4\pi e^2/Tk^2)g_{ee}(k) = n(2\pi e^2/Tk^2)[1+2(k^2/k_0^2)]/[1+(k^2/k_0^2)] \tag{95}$$

which follows from Eqs. (90) and (35).

The residues $s_1', s_2', r_{02}'\cdots$ can be obtained from

$$s_1' = \{[(i\omega-\zeta_1)/\epsilon(k,\omega)][4\pi\chi_i(1+4\pi\chi_e)]\}_{i\omega=\zeta_1} = s_1\bar{s}_1, \text{ etc.}$$

where $s_1, s_2\cdots$ are given by Eqs. (85), and

$$\bar{s}_1 = [4\pi\chi_i(1+4\pi\chi_e)]_{i\omega=\zeta_1}, \text{ etc.}$$

can be calculated by using Eqs. (65) and (66). The calculation is straightforward though tedious. We find

$$\begin{aligned}
&s_1' = (m_e/M)^2[\omega_p^2/(\zeta_1-\zeta_2)],\\
&s_2' = -(m_e/M)^2[\omega_p^2/(\zeta_1-\zeta_2)],\\
&r_{02}' = -(2\pi ne^2/T)(2\bar{D}/5),\\
&r_{03}' = (3\pi ne^2/5T)(i\bar{c}/k) + (2\pi ne^2/5T)\bar{D},
\end{aligned} \tag{96}$$

$$r'_{04} = r'^{*}_{03} = -(3\pi ne^2/5T)(i\bar{c}/k) + (2\pi ne^2/5T)\bar{D},$$
$$r'_{05}, r'_{06} = o(\kappa^4)$$

and all the other r's are at least $o(\kappa^2)$. In the calculation for r'_{02} and r'_{03} the unknown constants $\alpha_2, \beta_2, \alpha_3$, and β_3 introduced in Eqs. (64) do not appear in the final results. Eqs. (75) and (86) have been used in the elimination of α_3 and β_3. The fact that $r'_{04} = r'^{*}_{03}$ follows from $\nu_0^{(4)} = \nu_0^{(3)*}$.

Finally, a straightforward substitution of Eqs. (96) into Eqs. (92) and (7) gives the scattering function $S(k,\omega) = S_1(k,\omega) + S_2(k,\omega)$,

$$S_1(k,\omega) = \left(\frac{m_i}{M}\right)^2 \frac{Tk^2}{2\pi e^2} \frac{\omega_p^2/\tau_c}{(\omega^2 - \omega_p^2)^2 + \omega^2/\tau_c^2}, \tag{97}$$

$$S_2(k,\omega) = \frac{nk^2}{2}\left(\frac{(4/5)\bar{D}}{\omega^2 + (\bar{D}k^2)^2} + \frac{(3/5)[\bar{\Gamma}_1 + (\bar{\Gamma}_1 + 2\bar{D}/3)(\omega + \bar{c}k)/\bar{c}k]}{(\omega + \bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2} + \frac{(3/5)[\bar{\Gamma}_1 - (\bar{\Gamma}_1 + 2\bar{D}/3)(\omega - \bar{c}k)/\bar{c}k]}{(\omega - \bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2}\right)$$
$$= \frac{A\omega^2 + B}{[\omega^2 + (\bar{D}k^2)^2][(\omega + \bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2][(\omega - \bar{c}k)^2 + (\bar{\Gamma}_1 k^2)^2]}. \tag{98}$$

Here, $A = (3\bar{\Gamma}_1 - \bar{D})\bar{c}^2k^4[1 + o(z^2)] = 2\bar{\eta}\bar{c}^2k^4[1 + o(z^2)]$ and $B = \bar{D}\bar{c}^4k^6[1 + o(z^2)]$.

It is easy to show that $S_1(k,\omega)$ has maxima at $\omega = \pm\omega_p$ and $S_2(k,\omega)$ has maxima at $\omega = 0$ and $\omega = \pm\bar{c}k$, and minima at $\omega = \pm h\bar{c}k$, where $h = (1/2)\{[9(\bar{D}/2\bar{\eta})^2 + 8(\bar{D}/2\bar{\eta})]^{1/2} - 3(\bar{D}/2\bar{\eta})\}^{1/2} < 1$. Their order of magnitude values at these points are

$$S_2(k,\bar{c}k) < S_2(k,0) \sim [n\tau_c(m_i/m_e)]z^{-2},$$
$$S_2(k,h\bar{c}k) \sim [n\tau_c(m_i/m_e)],$$
$$S_2(k,\omega_p) \sim [n\tau_c(m_i/m_e)](m_e/m_i)^4z^4y^4,$$
$$S_1(k,0) \sim S_2(k,\bar{c}k) \sim [n\tau_c(m_i/m_e)](m_e/m_i)^2z^2y^4,$$
$$S_1(k,\omega_p) \sim [n\tau_c(m_i/m_e)](m_e/m_i)^2z^2y^2.$$

We recall that both z and y, given by Eqs. (68) and (76) respectively, are less than 1. It is clear that scatterings occur mostly in the low frequency region when $\omega \sim \bar{c}k \ll \omega_p$, and are dominated by the thermal diffusion pole and the sound poles. The damped plasma oscillation poles become relatively important, however, in the high frequency region when the frequencies are comparable to the plasma oscillation frequency.

For an unpolarized incident wave the scattering cross section per unit solid angle per unit frequency is given by[22]

$$d^2\Sigma/d\Omega d\omega = (1/4\pi)(e^2/m_ec^2)^2(1 + \cos^2\theta)S(k,\omega), \tag{99}$$

if the frequencies of both the incident and scattered waves, ω_i and ω_s, are much larger than the plasma frequency—true in most experimental situations. In the last equation c is the speed of light, θ is the scattered angle, $\omega = \omega_s - \omega_i$, and $k = |k_s - k_i| = 2(\omega_i/c)\sin(\theta/2)$, where k_i and k_s are, respectively, the incident and the scattered wavenumber vectors.

The total scattering cross section is related to $\int_{-\infty}^{\infty} S(k,\omega)d\omega$ which can be obtained from Eqs. (91) and (95):

$$\int_{-\infty}^{\infty} S(k,\omega)d\omega = 2\pi g_{ee}(k) = n\pi[1 + 2(k^2/k_0^2)]/[1 + (k^2/k_0^2)]. \tag{100}$$

This last result is independent of any collision model including collisionless model.[23] As a check to the scattering function we have obtained, let us intregrate Eqs. (97) and (98). We find

$$\int_{-\infty}^{\infty} S_2(k,\omega)d\omega = n\pi + o(k^2/k_0^2),$$
$$\int_{-\infty}^{\infty} S_1(k,\omega)d\omega = 4n\pi(m_i/M)^2(k^2/k_0^2), \tag{101}$$

which agree with Eq. (100) at least to the first order of $(k/k_0) = (m_e/m_i)^{1/2}zy$. It is clear from Eqs. (101) that the hydrodynamic poles (thermal diffusion and sound poles) contribute most to the total scattering cross section.

The total scattering cross section can be obtained by integrating Eq. (99) over the frequency and the solid angle. For $(k^2/k_0^2) \ll 1$,

$$\Sigma = (1/2)n[(8\pi/3)(e^2/m_ec^2)^2], \tag{102}$$

which is one-half of the Thomson free electron scattering cross section, in agreement with an earlier result.[23]

ACKNOWLEDGMENTS

The author wishes to thank Dr. James A. McLennan of Lehigh University, Bethlehem, Pennsylvania for many helpful suggestions and discussions. The hospitality offered by Lehigh University during the summers of 1971 and 1972 is gratefully acknowledged.

This work was supported by the National Science Foundation through a grant to Lehigh University.

[1] A. G. Sitenko, *Electromagnetic Fluctuations in Plasma* (Academic, New York, 1965).
[2] J. P. Dougherty and D. T. Farley, Jr., J. Geophys. Res. **68**, 5473 (1963).
[3] R. H. Williams and W. R. Chappell, Phys. Fluids **14**, 591 (1971).
[4] B. A. Trubnikov, in *Review of Plasma Physics*, edited by M. A. Leontovich (Consultants Bureau, New York, 1965), Vol. 1, p. 105.
[5] C. H. Su, J. Math. Phys. **8**, 248 (1967).
[6] J. D. Lewis, J. Math. Phys. **8**, 791 (1967).
[7] J. E. MeLeod and R. S. B. Ong, J. Math. Phys. **8**, 240 (1967).
[8] A. Lenard and I. B. Bernstein, Phys. Rev. **112**, 1456 (1958).
[9] C. S. Cheng and J. A. McLennan, Phys. Fluids **15**, 1285 (1972).
[10] J. P. Dougherty, Phys. Fluids **7**, 1788 (1964).
[11] N. Rostoker, Nucl. Fusion **1**, 101 (1961); Phys. Fluids **7**, 479 (1964).
[12] J. Weinstock, Phys. Rev. A **139**, 388 (1965).
[13] J. A. McLennan, Phys. Fluids **9**, 1581 (1966).
[14] J. Van Leeuwen and S. Yip, Phys. Rev. A **139**, 1138 (1965).
[15] M. H. Ernst, Physica (Utr.) **31**, 493 (1965); Physica (Utr.) **32**, 209 (1966).
[16] D. Montgomery, Phys. Fluids **12**, 804 (1969).
[17] L. Van Hove, Phys. Rev. **95**, 249 (1959); D. C. Montgomery and D. A. Tidman, *Plasma Kinetic Theory* (McGraw-Hill, New York, 1964).
[18] David Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1964).
[19] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
[20] P. Resibois, *Electrolyte Theory* (Harper and Row, New York, 1968).
[21] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[22] See Eq. (10.20) of Ref. 1.
[23] J. P. Dougherty and D. T. Farley, Proc. R. Soc. A **259**, 79 (1960).

Multichannel stationary scattering theory in two-Hilbert space formulation*

E. Prugovečki

Department of Mathematics, University of Toronto, Toronto 181, Canada
(Received 30 November 1972)

Two-Hilbert space versions of the basic integral-representation formulas for wave and transition operators are derived within the framework of time-independent scattering theory. Multichannel scattering theory for state vectors as well as for statistical operators is presented in a two-Hilbert space formulation to which these formulas become applicable. This formulation stays valid when long-range interactions, for which renormalized wave operators exist, are present.

1. INTRODUCTION

The customary[1,2] formulations of multichannel scattering theory do not bear a close formal resemblance to two-body scattering theory. This is primarily due to the absence[3] of a single S matrix for N-body scattering if $N \geq 3$. It is instead necessary to introduce an array[3] of S matrices, which can be grouped together in single operators $S_{\alpha\beta}$ only for each given pair of arrangement[1,2] channels, corresponding to a given incoming clustering α and given outgoing clustering β of the basic constituent particles into fragments. Hence, if more than one arrangement channel is nonempty then we are dealing with more than one scattering operator $S_{\alpha\beta}$. These operators are partial isometries with, in general, overlapping initial and final domains. Of course, despite the fact that these operators cannot be replaced by a single S operator in the Hilbert space $\mathcal{H}$ in which the N-body Hamiltonian H acts, it is possible to construct[4] a new Hilbert space $\tilde{\mathcal{H}}$ as a direct sum of the initial or the final domains of the operators $S_{\alpha\beta}$, and combine these operators into a single operator S from $\tilde{\mathcal{H}}$ to $\tilde{\mathcal{H}}$.

One unsatisfactory feature of this ad hoc procedure is that it does not embody any of the other important objects of scattering theory, such as the wave operators, the Lippmann-Schwinger equations for the distorted waves, etc. However, this objection can be circumvented by replacing[5] the free Hamiltonian H_0 with a physically equivalent Hamiltonian $\tilde{H}$ acting in $\tilde{\mathcal{H}}$, then relating $\mathcal{H}$ to $\tilde{\mathcal{H}}$ by means of an identication operator[6,7] J, and after that resorting to two-Hilbert space scattering theory.[6,7]

In Sec. 2 of this paper we derive those formulas of stationary two-Hilbert space scattering theory which are easily identifiable as generalizations of corresponding single-Hilbert space scattering theory formulas. All these formulas become automatically valid for multichannel scattering theory recast in the two-Hilbert space formalism, as shown in Sec. 3. Moreover, these formulas stay valid when long-range interactions (such as asymptotically Coulombic interactions) are present.

As first shown by F. Coester[7] such a two-Hilbert space approach has the additional advantage of being equally well applicable to relativistic scattering and to scattering involving identical particles.

As a matter of fact, the two-Hilbert space formulation is convenient[8] even in two-body scattering in long-range potentials, when the "strong" wave-operators

$$W_\pm = \underset{t\to\pm\infty}{\text{s-lim}} \exp(iHt)\exp(-iH_0t) \tag{1.1}$$

do not exist. It has been recently observed[9] that for spherically symmetric and twice continuously differentiable potentials $q(r)$ for which there is a constant C such that

$$\left|\frac{d^\kappa}{dr^\kappa} q(r)\right| \leq C(1+r)^{-\rho-\kappa}, \quad \kappa = 0, 1, 2 \tag{1.2}$$

for some fixed $\rho > 0$, there is a densely defined operator Z in terms of which the "renormalized"[10,11] wave operators $W(\frac{r}{\pm})$ for such potentials satisfy (throughout this paper $\mathfrak{D}(A)$ denotes the domain of definition of an operator A)

$$W(\tfrac{r}{\pm})\Psi = \underset{t\to\pm\infty}{\text{s-lim}} \exp(iHt)Z\exp(-iH_0t)\Psi, \quad \Psi \in \mathfrak{D}(Z). \tag{1.3}$$

A similar result is valid[8] for the case of asymptotically Coulombic potentials (which are not necessarily spherically symmetric), and probably extends to other classes of long-range potentials. By restricting Z to any subspace $\mathcal{H}^\Delta$ of $\mathcal{H}$ corresponding to a finite range Δ of positive energies, we get a bounded operator $Z^\Delta = E^{(0)}(\Delta)Z$, where $E^{(0)}(\Delta)$ is the projector onto $\mathcal{H}^\Delta$. Then we can take $\mathcal{H}^\Delta$ to be $\tilde{\mathcal{H}}$ and Z^Δ to be the identification operator from $\tilde{\mathcal{H}}$ to $\mathcal{H}$. Thus, we obtain a two-Hilbert space formulation of two-body scattering theory for long-range potentials. Since the procedure is valid for an arbitrarily large finite range of energies of the reduced free particle (as long as it does not include the zero energy), this formulation is suitable for the description of any realistic scattering experiment. The same idea can be used with slight modifications[8] in constructing the identification operator of multichannel scattering theory for long-range interactions. In that case, we have in general a different operator Z_α for each arrangement channel α, which can be used (in the manner indicated in Sec. 3) in the construction of the identification operator J.

In the last section we extend the formalism to the Liouville spaces (i.e., Hilbert-Schmidt classes) $\mathfrak{B}_2(\mathcal{H})$ and $\mathfrak{B}_2(\tilde{\mathcal{H}})$ of $\mathcal{H}$ and $\tilde{\mathcal{H}}$, respectively. We show how a scattering theory for the statistical operator can be developed directly in terms of transformers[12] on these spaces.

2. STATIONARY SCATTERING IN THE TWO-HILBERT SPACE APPROACH

In this section we demonstrate how the basic formulas of single-Hilbert space stationary scattering theory can be generalized to the two-Hilbert space theory. The more important features of the time-dependent counterpart of such two-Hilbert space theories have been derived in Ref. 6, while some of the properties of the stationary two-Hilbert space theory for trace-class interactions have been studied in Ref. 13. In this section we restrict ourselves to deriving two-Hilbert space counterparts to the Lippmann-Schwinger equations for wave and transition operators. These results can be then easily specialized to the case of nonrelativistic multichannel scattering. This is done in the next section.

Copyright © 1973 by the American Institute of Physics

Let $\bar{H}$ and H be self-adjoint operators in the Hilbert spaces $\bar{\mathcal{H}}$ and $\mathcal{H}$, respectively. For any given *bounded* linear operator J (called the identification operator)[6] from $\bar{\mathcal{H}}$ to $\mathcal{H}$ we define the *generalized wave operators* corresponding to J by

$$W_{\pm}^{J} = \text{s-}\lim_{t\to\pm\infty} \exp(iHt)J\exp(-i\bar{H}t) \tag{2.1}$$

if the above strong limits exist. The case when the above limits exist only on some closed subspace $\bar{\mathcal{H}}'$ of $\bar{\mathcal{H}}$ which is left invariant by $\bar{H}$ can be easily incorporated in the present framework by simply restricting $\bar{H}$ to $\bar{\mathcal{H}}'$ and then substituting in all the results $\bar{\mathcal{H}}'$ instead of $\bar{\mathcal{H}}$.

In the following we use the notation in which $\langle\cdot\,|\,\cdot\rangle$ and $(\cdot\,|\,\cdot)$ denote the inner products in $\mathcal{H}$ and $\bar{\mathcal{H}}$, respectively; similarly, $\|\cdot\|$ and $|\cdot|$ denote the vector norm in $\mathcal{H}$ and operator bound in $\mathfrak{B}(\mathcal{H})$, as opposed to the vector norm $\|\cdot\|_0$ and operator bound $|\cdot|_0$ related to $\bar{\mathcal{H}}$. The spectral functions of H and $\bar{H}$ are E_λ and $\bar{E}_\lambda$, respectively.

Theorem 2.1: For any self-adjoint operators H and $\bar{H}$ and any identification operator J from $\bar{\mathcal{H}}$ to $\mathcal{H}$ the following weak[2,14] *Riemann-Stieltjes integrals*

$$\begin{aligned}\Omega_{\pm\epsilon}^{J} &= \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H-\lambda\pm i\epsilon}\, J\, d_\lambda \bar{E}_\lambda \\ &= \int_{-\infty}^{+\infty} d_\lambda E_\lambda J \frac{\pm i\epsilon}{\lambda-\bar{H}\pm i\epsilon}\end{aligned} \tag{2.2}$$

exist. Furthermore, $\Omega_{\pm\epsilon}^{J}$ can be represented by means of the following Bochner[2] *integrals:*

$$\Omega_{\pm\epsilon}^{J} = \pm\epsilon\int_{0}^{\pm\infty} \exp(\mp\epsilon t)\exp(iHt)J\exp(-i\bar{H}t)dt. \tag{2.3}$$

Proof: The existence of the Bochner integrals in (2.3) follows[2] from the fact that $\langle\Phi\,|\,\exp(\mp\epsilon t + iHt)J\exp(-i\bar{H}t)\Psi\rangle$ is continuous in t for any $\Phi\in\mathcal{H}$, and that

$$\|\exp(\mp\epsilon t + iHt)J\exp(-i\bar{H}t)\Phi\| \le \exp(\mp\epsilon t)\,|J|\,\|\Psi\|_0,$$

where the above majorizing functions are Lebesgue integrable on the respective domains $[0,+\infty)$ and $(-\infty,0]$. Since the integrand is weakly continuous, this integral exists also in the weak Riemann sense. By applying the spectral theorem twice and using theorems on cross-iterated Riemann-Stieltjes integrals,[2,14] we can justify (cf. Ref. 2, pp. 440-48) each interchange of the order of integration in the following derivation:

$$\begin{aligned}\langle\Phi\,|\,\Omega_{\pm\epsilon}^{J}\Psi\rangle &= \pm\epsilon\int_{0}^{\pm\infty} dt\exp(\mp\epsilon t)\int_{-\infty}^{+\infty}\exp(-i\lambda t) \\ &\quad\times d_\lambda\langle\exp(-iHt)\Phi\,|\,J\bar{E}_\lambda\Psi\rangle \\ &= \pm\epsilon\int_{-\infty}^{+\infty} d_\lambda\int_{0}^{\pm\infty} dt\exp(\mp\epsilon t - i\lambda t) \\ &\quad\times\int_{-\infty}^{+\infty}\exp(i\lambda' t)d_{\lambda'}\langle\Phi\,|\,E_{\lambda'}J\bar{E}_\lambda\Psi\rangle \\ &= \pm\epsilon\int_{-\infty}^{+\infty} d_\lambda\int_{-\infty}^{+\infty} \\ &\quad\times\left(\int_{0}^{\pm\infty}\exp(\mp\,\epsilon t - i\lambda t + i\lambda' t)dt\right) \\ &\quad\times d_{\lambda'}\langle\Phi\,|\,E_{\lambda'}J\bar{E}_\lambda\Psi\rangle \\ &= \int_{-\infty}^{+\infty} d_\lambda\langle\Phi\,|\,\frac{\mp\epsilon}{\mp\epsilon - i\lambda + iH}J\bar{E}_\lambda\Psi\rangle.\end{aligned}$$

The equality of the first and last expressions in the above chain of equations represents a statement equivalent to the first relation in (2.2). Since the derivation of the second relation in (2.2) is quite analogous, the theorem is completely established.

We note that from the definition of weak Riemann-Stieltjes integrals we can immediately infer that

$$\begin{aligned}(\Omega_{\pm\epsilon}^{J})^* &= \int_{-\infty}^{+\infty} d_\lambda\bar{E}_\lambda J^*\frac{\mp i\epsilon}{H-\lambda\mp i\epsilon} \\ &= \int_{-\infty}^{+\infty}\frac{\mp i\epsilon}{\lambda-\bar{H}\mp i\epsilon}J^* d_\lambda E_\lambda\end{aligned} \tag{2.4}$$

where the adjoint J^* of J maps $\mathcal{H}$ into $\bar{\mathcal{H}}$.

Theorem 2.2: If the limits

$$\Omega_{\pm}^{J} = \text{s}-\lim_{\epsilon\to+0}\Omega_{\pm\epsilon}^{J} \tag{2.5}$$

exist, then they satisfy the following intertwining relationships:

$$\exp(iHs)\Omega_{\pm}^{J} = \Omega_{\pm}^{J}\exp(i\bar{H}s), \qquad s\in\mathbb{R}^1, \tag{2.6}$$

$$E_\lambda\Omega_{\pm}^{J} = \Omega_{\pm}^{J}\bar{E}_\lambda, \qquad \lambda\in\mathbb{R}^1, \tag{2.7}$$

$$H\Omega_{\pm}^{J} = \Omega_{\pm}^{J}\bar{H}\Psi, \qquad \Psi\in\mathfrak{D}(\bar{H}). \tag{2.8}$$

Moreover, the limits $\Omega_{\pm}^{J}$ exist if the limits $W_{\pm}^{J}$ in (2.1) exist, and in that case $W_{\pm}^{J} = \Omega_{\pm}^{J}$.

Remark: Since we do not require that J is a partial isometry from $\bar{\mathcal{H}}$ into $\mathcal{H}$, the operators $\Omega_{\pm}^{J}$ are not[6] in general partial isometries, as is the case in single-Hilbert space scattering theory.

Proof: The relation (2.6) follows from the equation

$$\begin{aligned}\exp(iHs)\Omega_{\pm\epsilon}^{J} &= \pm\epsilon\exp(\pm\epsilon s)\int_{0}^{\pm\infty}\exp[\mp\epsilon(t+s)] \\ &\quad\times\exp[iH(t+s)]J\exp(-i\bar{H}t)dt \\ &= \mp\epsilon\exp(\pm\epsilon s)\int_{0}^{s}\exp(\mp\epsilon\tau + iH\tau)J \\ &\quad\times\exp(-i\bar{H}\tau)d\tau + \exp(\pm\epsilon s)\Omega_{\pm\epsilon}^{J}\exp(i\bar{H}s)\end{aligned}$$

by letting $\epsilon\to+0$ on both sides (while keeping $s\in\mathbb{R}^1$ fixed). Then (2.7) follows from (2.6) by means of one of the standard constructions (cf. Ref. 2, pp. 241-53) of the spectral function of the operator $\bar{H}$ from the family $e^{i\bar{H}s}$, $s\in\mathbb{R}^1$. Finally, (2.8) can be deduced from (2.7) by means of a standard corollary to Stone's theorem, involving formal differentiation of both sides of (2.8) (cf. Ref. 2, p. 288).

We note that (2.6)-(2.8) are consequences of the time-independent definitions (2.3) [or their equivalent forms (2.2)] of $\Omega_{\pm\epsilon}$ and do not rely on the time-dependent derivation[6] of these relations. However, if the operators $W_{\pm}^{J}$ in (2.1) exist, then we conclude that $\Omega_{\pm}^{J}$ exist and that $W_{\pm}^{J} = \Omega_{\pm}^{J}$ by noting that for any given $\delta > 0$

$$\begin{aligned}\|W_{\pm}^{J}\Psi - \Omega_{\pm\epsilon}^{J}\Psi\| &\le \pm\int_{0}^{\pm\infty}\exp(\mp u)\,\|W_{\pm}^{J}\Psi - \exp\left(iH\frac{u}{\epsilon}\right) \\ &\quad\times J\exp\left(-i\bar{H}\frac{u}{\epsilon}\right)\Psi\|\,du \le \pm\int_{0}^{a_\pm(\delta)} \pm\int_{a_\pm(\delta)}^{\pm\infty} \\ &< \frac{\delta}{2}+\frac{\delta}{2} = \delta,\end{aligned}$$

if $a_\pm(\delta) = \pm(\delta/2)\{\|W_{\pm}^{J}\Psi\| + |J|\,\|\Psi\|_0\}^{-1}$ and ϵ is chosen so that

$$\| W_{\pm}^{J}\Psi - \exp\left[iH\frac{a_{\pm}(\delta)}{\epsilon}\right] J \exp\left[-i\bar{H}\frac{a_{\pm}(\delta)}{\epsilon}\right]\Psi\| < \frac{\delta}{2}.$$

We can recast the integrals in (2.2) and (2.4) in forms which lead to relations which are easily recognizable as being generalizations of the single-Hilbert space versions[2] of the Lippmann-Schwinger equations. For example, to get a generalization of

$$\Psi_{\pm} = \Psi + \text{s-}\lim_{\epsilon\to+0}\int_{-\infty}^{+\infty}\frac{1}{\lambda - \bar{H} \pm i\epsilon} V d_{\lambda}E_{\lambda}\Psi_{\pm}, \tag{2.9}$$

where $\Psi_{\pm} = \Omega_{\pm}\Psi$ and $\Omega_{\pm}$ are the wave-operators (2.5) for the conventional case when $\mathfrak{K} = \bar{\mathfrak{K}}$, $J = \mathbb{1}$, and $V = H - \bar{H}$, we make use of the identity

$$\frac{\pm i\epsilon}{\lambda - \bar{H} \pm i\epsilon} = \mathbb{1} - \frac{\lambda - \bar{H}}{\lambda - \bar{H} \pm i\epsilon}$$

in order to derive from (2.4) that

$$(\Omega_{\pm\epsilon}^{J})^* = J^* + \int_{-\infty}^{+\infty}\frac{1}{\lambda - \bar{H} \mp i\epsilon}(\lambda - \bar{H})J^* d_{\lambda}E_{\lambda}. \tag{2.10}$$

Now, we observe that

$$\int_{-\infty}^{+\infty}\frac{1}{\lambda - \bar{H} \mp i\epsilon}\lambda J^* d_{\lambda}E_{\lambda} = \int_{-\infty}^{+\infty}\frac{1}{\lambda - \bar{H} \mp i\epsilon} J^* H d_{\lambda}E_{\lambda} \tag{2.11}$$

as a consequence of the following general lemma.

Lemma 2.1: Suppose E_{λ} is the spectral function of an operator H in $\mathfrak{K}$ and $A(\lambda)$ is a family of operators in $\bar{\mathfrak{K}}$ such that $I^*A^*(\lambda)\Phi \in \mathfrak{D}(H)$, $\lambda \in \mathbb{R}^1$, for a dense set $\mathfrak{D}_0$ of vectors $\Phi \in \mathfrak{K}$. If there is a constant C such that $|A(\lambda)| \leq C$ for all $\lambda \in \mathbf{R}^1$ and if I is a bounded linear mapping of $\mathfrak{K}$ into $\bar{\mathfrak{K}}$ then

$$\int_{-\infty}^{+\infty} A(\lambda)I\lambda d_{\lambda}E_{\lambda}\Psi = \int_{-\infty}^{+\infty} A(\lambda)IH d_{\lambda}E_{\lambda}\Psi, \qquad \Psi \in \mathfrak{D}(H). \tag{2.12}$$

Furthermore, if one of the two integrals in (2.12) is known to exist and if either (a) $\mathfrak{D}_0 = \mathfrak{K}$ or (b) the vector norms of the Riemann-Stieltjes sums for the other integral are uniformly bounded, then the other integral also exists.

Remark: Strictly speaking, the Riemann-Stieltjes sums which in the limit converge to the integral on the right-hand side of (2.11) are operators with domain equal to $\mathfrak{D}(H)$. However, since this integral is a bounded operator, we can extend it in a unique manner[2] to a bounded operator defined on the entire $\mathfrak{K}$.

Proof: We show first that for arbitrary $\Phi \in \mathfrak{D}_0$, $\Psi \in \mathfrak{D}(H)$ and $a < b$

$$\langle\Phi|\int_a^b A(\lambda)I(\lambda - H)d_{\lambda}E_{\lambda}\Psi\rangle = 0$$

by proving that the corresponding Riemann-Stieltjes sums converge weakly to zero in the limit of finer and finer subdivisions of $[a, b]$. In fact, for any subdivision $\lambda_0 = a < \lambda_1 < \lambda_2 < \cdots < \lambda_n = b$ of $[a, b]$ and any $\lambda_k' \in [\lambda_{k-1}, \lambda_k]$, $k = 1, \ldots, n$, we get the following estimate by using the Schwarz-Cauchy inequalities in $\mathfrak{K}$ and in $l^2(n)$:

$$\left|\left(\Phi\middle|\sum_{k=1}^{n} A(\lambda_k')I(\lambda_k' - H)(E_{\lambda_k} - E_{\lambda_{k-1}})\Psi\right)\right|$$

$$\leq \left(\sum_{k=1}^{n}\|(E_{\lambda_k} - E_{\lambda_{k-1}})(\lambda_k' - H)I^*A(\lambda_k')\Phi\|_0^2\right)^{1/2} \times\left(\sum_{k=1}^{n}\|(E_{\lambda_k} - E_{\lambda_{k-1}})\Psi\|_0^2\right)^{1/2} \leq \|\Psi\|_0 \times\left(\sum_{k=1}^{n}|(E_{\lambda_k} - E_{\lambda_{k-1}})(\lambda_k' - H)|^2|I^*A^*(\lambda_k')|^2\right)^{1/2}\|\Phi\|. \tag{2.13}$$

Since $|A^*(\lambda_k')| \leq C$ and

$$|(E_{\lambda_k} - E_{\lambda_{k-1}})(\lambda_k' - H)| \leq \lambda_k - \lambda_{k-1}$$

we conclude that the expression on the right-hand side of (2.12) can be made arbitrarily small by choosing $\max|\lambda_k - \lambda_{k-1}|$ sufficiently close to zero. Since this conclusion is true for any Φ from the dense set $\mathfrak{D}_0$, we infer that

$$\int_a^b A(\lambda)I\lambda d_{\lambda}E_{\lambda}\Psi = \int_a^b A(\lambda)IH d_{\lambda}E_{\lambda}\Psi. \tag{2.14}$$

Moreover, we see that when $\mathfrak{D}_0 = \mathfrak{K}$ the existence of one of these integrals implies the existence of the other. In fact, weak convergence of one set of Riemann-Stieltjes sums implies, according to the above result, the convergence of $\langle\Phi|\Psi^{(1)}\rangle, \langle\Phi|\Psi^{(2)}\rangle, \cdots$ for the corresponding vectors $\Psi^{(1)}, \Psi^{(2)}, \cdots$ of the Riemann-Stieltjes sums for the other integral. Since this is true for all $\Phi \in \mathfrak{K}$ if $\mathfrak{D}_0 = \mathfrak{K}$, the desired weak convergence of $\Psi^{(1)}, \Psi^{(2)}, \cdots$ is established. On the other hand, if $\mathfrak{D}_0 \neq \mathfrak{K}$ but condition (b) stipulated in the statement of the lemma is satisfied, i.e., $\|\Psi^{(n)}\|_0 \leq C$ for all $n = 1, 2, \ldots$, then convergence of $\langle\Phi|\Psi^{(n)}\rangle$ for any Φ in the dense set $\mathfrak{D}_0$ implies convergence for arbitrary $\Phi \in \mathfrak{K}$.

Finally, (2.13) follows from (2.14) and the definition of weak Stieltjes integrals, which implies that

$$\int_{-\infty}^{+\infty} A(\lambda)I(\lambda - H)d_{\lambda}E_{\lambda}\Psi = \text{w-}\lim_{m\to\infty}\sum_{n=-m}^{+m}\int_n^{n+1} A(\lambda)I(\lambda - H)d_{\lambda}E_{\lambda}\Psi.$$

If we have that $J\mathfrak{D}(\bar{H}) \subset \mathfrak{D}(H)$ then $\mathfrak{D}_0 = \mathfrak{K}$ for the integrals appearing in the relation (2.11), and the above lemma can be applied to derive (2.11), since $|(\lambda - \bar{H} \pm i\epsilon)^{-1}| \leq \epsilon^{-1}$. By combining (2.10) with (2.11), and then submitting the rest of the relations in (2.2) and (2.4) to the same type of procedure, we arrive at the following auxiliary result.

Lemma 2.2: Suppose that the operators $H, \bar{H}$ and J are such that $J\mathfrak{D}(\bar{H}) \subset \mathfrak{D}(H)$ and $J^*\mathfrak{D}(H) \subset \mathfrak{D}(\bar{H})$. Then the operators $\Omega_{\pm}^{J}$ defined by (2.2) or (2.3) and their adjoints have the following representations:

$$\Omega_{\pm\epsilon}^{J} = J - \int_{-\infty}^{+\infty}(H - \lambda \pm i\epsilon)^{-1}(HJ - J\bar{H})d_{\lambda}\bar{E}_{\lambda} = J - \int_{-\infty}^{+\infty} d_{\lambda}E_{\lambda}(HJ - J\bar{H})(\lambda - \bar{H} \pm i\epsilon)^{-1}, \tag{2.15}$$

$$(\Omega_{\pm\epsilon}^{J})^* = J^* - \int_{-\infty}^{+\infty} d_{\lambda}\bar{E}_{\lambda}(J^*H - \bar{H}J^*)(H - \lambda \mp i\epsilon)^{-1} = J^* - \int_{-\infty}^{+\infty}(\lambda - \bar{H} \mp i\epsilon)^{-1}(J^*H - \bar{H}J^*)d_{\lambda}E_{\lambda}. \tag{2.16}$$

Remark: The two preconditions appearing in the preceding Lemma 2.2 (which therefore make their appearance also in Theorem 2.3 and Theorem 2.4) are the result of using condition (a) in Lemma 2.1 in deriving Lemma 2.2. Consequently, a different set of pre-

conditions could be obtained by using condition (b) in Lemma 2.1. However, in practice H is relatively bounded with respect to $J\bar{H}$, and both (a) and (b) are satisfied.

As an immediate consequence of (2.15) and (2.16) we have

Theorem 2.3: If $\Psi_\pm = \mathrm{s}-\lim \Omega^J_{\pm\epsilon}\Psi$ *exist for some* $\Psi \in \mathfrak{K}$ *and if* $J\mathfrak{D}(\bar{H}) \subset \mathfrak{D}(H)$, $J^*\mathfrak{D}(H) \subset \mathfrak{D}(\bar{H})$, *then*

$$\Psi_\pm = J\Psi - \mathrm{s}\text{-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{H-\lambda\pm i\epsilon}\,(HJ - J\bar{H})d_\lambda \bar{E}_\lambda \Psi, \tag{2.17}$$

$$J^*\Psi_\pm = \Psi + \mathrm{w}\text{-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} \frac{1}{\lambda - \bar{H} \mp i\epsilon} \times (J^*H - \bar{H}J^*)d_\lambda E_\lambda \Psi_\pm. \tag{2.18}$$

If in addition $(\Omega^J_+)^*\Psi_\pm = \mathrm{s}\text{-}\lim(\Omega^J_{\pm\epsilon})^*\Psi_\pm$, *then the weak limit in (2.18) is also a strong limit.*

The above eqns. (2.17) and (2.18) are the two-Hilbert space theory counterparts of the Lippmann-Schwinger equations relating distorted waves to the free waves. In fact, when $\mathfrak{K} = \bar{\mathfrak{K}}$ and $J = \mathbf{1}$ Eq. (2.18) assumes the form (2.9), which in turn leads[2,15] to the well-known equations (we denote by the same letter A the extension to bra-space[15] of any given closed Hilbert space operator A)

$$\phi^\pm_\lambda = \phi_\lambda + \lim_{\epsilon\to+0} \frac{1}{\lambda - \bar{H} \mp i\epsilon} V\phi^\pm_\lambda \tag{2.19}$$

relating the eigenfunctions ϕ_λ of $\bar{H}$ to the eigenfunctions $\phi^\pm_\lambda$ of H. We note that in the present case we can expect that (2.19) will be replaced by[15]

$$J^\dagger\phi^\pm_\lambda = \phi_\lambda + \lim_{\epsilon\to+0} \frac{1}{\bar{H} - \lambda \mp i\epsilon}\,(J^\dagger H - \bar{H}J^\dagger)\phi^\pm_\lambda, \tag{2.20}$$

where $J^\dagger$ denotes the bra-adjoint[15] of J.

The results of Lemma 2.2 can be also used to derive a two-Hilbert space generalization of the single-Hilbert space ($\mathfrak{K} = \bar{\mathfrak{K}}$, $J = \mathbf{1}$) representation[2] of the T operator:

$$T = \frac{S-\mathbf{1}}{2\pi i} = \frac{1}{\pi}\,\mathrm{s}\text{-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} d\bar{E}_\lambda V \frac{\epsilon}{(H-\lambda)^2 + \epsilon^2}\,\Omega_-. \tag{2.21}$$

The above formula leads[2,15] to the well-known formulas for the T matrix $\langle\phi_\lambda|T|\phi_{\lambda'}\rangle$ on the energy-shell,

$$\langle\phi_\lambda|T|\phi_\lambda\rangle = \langle\phi_\lambda|V\Omega_-|\phi_\lambda\rangle = \langle\phi_\lambda|V + V\frac{1}{\lambda - \bar{H} + i0}T|\phi_\lambda\rangle, \tag{2.22}$$

and in case that $\mathfrak{K} \neq \bar{\mathfrak{K}}$ and/or $J \neq \mathbf{1}$, it assumes the form presented in the following theorem.

Theorem 2.4: If the limits $\Omega^J_\pm$ *in (2.5) exist and* $J^*\mathfrak{D}(H) \subset \mathfrak{D}(\bar{H})$ *then*

$$T^J = \frac{1}{2\pi i}(\Omega^J_+ - \Omega^J_-)^*\Omega^J_- = \frac{1}{\pi}\,\mathrm{w}\text{-}\lim_{\epsilon\to+0} \int_{-\infty}^{+\infty} \times d_\lambda \bar{E}_\lambda (J^*H - \bar{H}J^*) \frac{\epsilon}{(H-\lambda)^2 + \epsilon^2}\,\Omega^J_-. \tag{2.23}$$

The proof of the above statement is obtained by noting that the first set of Eqs. (2.16) imply that

$$(\Omega^J_{+\epsilon} - \Omega^J_{-\epsilon})^* = \int_{-\infty}^{+\infty} d_\lambda \bar{E}_\lambda (J^*H - \bar{H}J^*) \frac{2i\epsilon}{(H-\lambda)^2 + \epsilon^2}. \tag{2.24}$$

From this (2.23) immediately follows.

3. MULTICHANNEL SCATTERING THEORY FOR WAVE-PACKETS

Let $\mathfrak{K}$ be the Hilbert space in which a Hamiltonian H describing an N-particle system acts, and let H_α be the Hamiltonian of the arrangement channel[1,2] corresponding to a clustering α of these N particles into fragments. We allow the possibility that the interactions among the particles have long ranges, but require that for each arrangement channel (or clustering) α there is a bounded operator Z_α such that the wave operators

$$\Omega^{(\alpha)}_\pm = \mathrm{s}\text{-}\lim_{\epsilon\to+0} \Omega^{(\alpha)}_{\pm\epsilon} J_\alpha, \tag{3.1}$$

$$\Omega^{(\alpha)}_{\pm\epsilon} = \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H - \lambda \pm i\epsilon} Z_\alpha d_\lambda E_\lambda^{(\alpha)} = \int_{-\infty}^{+\infty} d_\lambda E_\lambda Z_\alpha \frac{\pm i\epsilon}{\lambda - H_\alpha \pm i\epsilon} \tag{3.2}$$

exist as partial isometries with initial domains $\mathfrak{K}_\alpha = J_\alpha\mathfrak{K}$, where J_α is an orthogonal projection operator in $\mathfrak{K}$ which commutes with H_α and $E_\lambda^{(\alpha)}$ is the spectral function of H_α. The only additional assumption which we make is that

$$E^{(\alpha)}_\pm E^{(\beta)}_\pm = \delta_{\alpha\beta}E^{(\beta)}_\pm, \quad E^{(\alpha)}_\pm = \Omega^{(\alpha)}_\pm \Omega^{(\alpha)}_\pm{}^*. \tag{3.3}$$

Usually these assumptions can be shown to be satisfied by resorting to time-dependent methods.[16]

Remark: A weaker version of these assumptions might be desirable when dealing with certain types of long-range interactions. Thus we could replace the requirement that $\Omega^{(\alpha)}_\pm$ are partial isometries from $\mathfrak{K}_\alpha$ to $R^\pm_\alpha = E^{(\alpha)}_\pm$ by the weaker conditions that they are invertible bounded operators from $\mathfrak{K}_\alpha$ onto $R^\pm_\alpha$. This would affect only Theorems 3.1 and Theorem 4.2 to the extent that the partial isometry property derived in these theorems would not be retained under these weaker assumptions. Since we do not expect that $\Omega^J_\pm$ are partial isometries for all two-Hilbert space scattering theories (cf. the remark following Theorem 2.2), this would not limit to any extent the applicability of the results in Sec. 2 to the cases treated in Sec. 3 and Sec. 4 under these weaker conditions in $\Omega^{(\alpha)}_\pm$.

We can integrate the customary piecemeal description of each clustering and each rearrangement collision as a separate occurrence into a unified description of the scattering process by introducing

$$\bar{H} = \bigoplus_\alpha J_\alpha H_\alpha J_\alpha = \bigoplus_\alpha H_\alpha J_\alpha \tag{3.4}$$

as a self-adjoint[5] operator acting on the direct sum

$$\bar{\mathfrak{K}} = \bigoplus_\alpha H_\alpha, \qquad \mathfrak{K}_\alpha = J_\alpha\mathfrak{K}. \tag{3.5}$$

We denote by $\bar{E}_\lambda$ the spectral function of $\bar{\mathfrak{K}}$. The spaces $\bar{\mathfrak{K}}$ and $\mathfrak{K}$ can be related by means of the identification operator J, where

$$J = \sum_\alpha Z_\alpha \mathcal{J}_\alpha, \qquad \mathcal{J}_\alpha\{\bigoplus_\beta \Psi_\beta\} = \Psi_\alpha \in \mathfrak{K}_\alpha. \tag{3.6}$$

As a matter of fact, J is a bounded operator from $\bar{\mathfrak{K}}$ to $\mathfrak{K}$ with adjoint

$$J^* = \bigoplus_\alpha J_\alpha^* Z_\alpha^*. \tag{3.7}$$

This is due to the fact that there is only a finite number of clusterings of N particles, and for each clustering α the corresponding Z_α is bounded.

Theorem 3.1: The operators

$$\Omega_\pm^J = \sum_\alpha \Omega_\pm^{(\alpha)} J_\alpha \tag{3.8}$$

are partial isometries with initial domain $\tilde{\mathcal{K}}$ and final domains

$$R_\pm = \sum_\alpha E_\pm^{(\alpha)} \mathcal{K} \tag{3.9}$$

in $\mathcal{K}$. Both these operators satisfy (2.5), with $\Omega_{\pm\epsilon}^J$ given by (2.2), (3.4), and (3.6).

Proof: The partial isometry properties of $\Omega_\pm^J$ are immediate consequences of the following two relations, obtained by using the partial isometry properties of $\Omega_\pm^{(\alpha)}$ and (3.3):

$$(\Omega_\pm^J)^* \Omega_\pm^J = \sum_{\alpha,\beta} J_\alpha^* \Omega_\pm^{(\alpha)*} \Omega_\pm^{(\beta)} J_\beta = \sum_{\alpha,\beta} \delta_{\alpha\beta} J_\alpha^* E_\pm^{(\alpha)} J_\alpha = \sum_\alpha J_\alpha^* J_\alpha, \tag{3.10}$$

$$\Omega_\pm^J (\Omega_\pm^J)^* = \sum_{\beta,\alpha} \Omega_\pm^{(\beta)} J_\beta J_\alpha^* \Omega_\pm^{(\alpha)*} = \sum_\alpha \Omega_\pm^{(\alpha)} \Omega_\pm^{(\alpha)*}. \tag{3.11}$$

According to (3.1)

$$\Omega_\pm^J = \text{s-}\lim_{\epsilon \to +0} \sum_\alpha \Omega_{\pm\epsilon}^J J_\alpha. \tag{3.12}$$

Since $[J_\alpha, E_\lambda^{(\alpha)}] = 0$ we have $E_\lambda^{(\alpha)} J_\alpha = J_\alpha E_\lambda^{(\alpha)} J_\alpha$. Hence, by (3.2),

$$\sum_\alpha \Omega_{\pm\epsilon}^J J_\alpha = \sum_\alpha \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H - \lambda \pm i\epsilon} Z_\alpha J_\alpha d_\lambda \tilde{E}_\lambda = \Omega_{\pm\epsilon}^J, \tag{3.13}$$

where, in view of $\tilde{J}_\alpha \tilde{J}_\beta^* = \delta_{\alpha\beta} J_\alpha$,

$$\tilde{E}_\lambda = \bigoplus_\alpha \tilde{J}_\alpha^* E_\lambda^{(\alpha)} \tilde{J}_\alpha \tag{3.14}$$

is the spectral function of $\tilde{H}$ on the Hilbert space $\tilde{\mathcal{K}}$.

Thus, (3.12) and (3.13) have the appearance of (2.5) and (2.2), respectively. Consequently, the last statement of Theorem 3.1 has been also established to be true.

It easily follows from the definitions of J and $\Omega_\pm^J$ that the following theorem is also true.

Theorem 3.2: The operators $\Omega_\pm^J$ have the intertwining properties (2.6)-(2.8). Moreover, if

$$W_\pm^{(\alpha)} = \text{s-}\lim_{t \to \pm\infty} \exp(iHt) Z_\alpha \exp(-iH_\alpha t) J_\alpha \tag{3.15}$$

exist, then

$$\Omega_\pm^J = \sum_\alpha W_\pm^{(\alpha)} J_\alpha = \text{s-}\lim_{t \to \pm\infty} \exp(iHt) J \exp(-i\tilde{H}t). \tag{3.16}$$

All the other results of Sec. 2 can be immediately applied to the present case. In particular, (2.17), (2.18), and (2.23) hold true. We note, however, that the weak limit in (2.18) cannot be replaced by a strong limit even in the case when long-range forces are absent—except when the free channel is the only nonempty arrangement channel (cf. Ref. 5, Proposition 3.10).

The scattering operator

$$S = (\Omega_+^J)^* \Omega_-^J = \sum_{\alpha,\beta} J_\alpha^* \Omega_+^{(\alpha)*} \Omega_-^{(\beta)} J_\beta \tag{3.17}$$

contains all the scattering operators $S_{\alpha\beta} = \Omega_+^{(\alpha)*} \Omega_+^{(\beta)}$ for any rearrangement collision leading from an initial clustering β to the final clustering α. Since $\Omega_\pm^J$ are partial isometries, we get as a consequence of a standard theorem (cf. Ref. 2, p. 420) on partial isometries that S is a unitary operator on $\tilde{\mathcal{K}}$ if an only if $R_+ = R_-$. Thus, we see that the entire theory of multichannel scattering assumes on $\tilde{\mathcal{K}}$ and $\mathcal{K}$ a remarkable formal similarity to the short-range two-body scattering theory even when long-range forces [for which (3.1) and (3.2) hold true] are present.

4. MULTICHANNEL SCATTERING THEORY FOR STATISTICAL OPERATORS

The above formal analogy with ordinary two-body scattering theory can be pushed even further to include any statistical operator ρ by considering ρ as an object in Liouville space. Mathematically, the Liouville spaces of $\mathcal{K}$ and $\tilde{\mathcal{K}}$ are the Hilbert-Schmidt classes $\mathfrak{B}_2(\mathcal{K})$ and $\mathfrak{B}_2(\tilde{\mathcal{K}})$ of operators on $\mathcal{K}$ and $\tilde{\mathcal{K}}$, respectively. Two-Liouville space scattering theory can be formulated directly in terms of transformers[12,17] acting from $\mathfrak{B}_2(\tilde{\mathcal{K}})$ to $\mathfrak{B}_2(\mathcal{K})$. Thus, the arrangement channel wave operators $\Omega_\pm^{(\alpha)}$ induce the wave transformers $\underline{\Omega}_\pm^{(\alpha)}$, where

$$\underline{\Omega}_\pm^{(\alpha)} A = \Omega_\pm^{(\alpha)} A \Omega_\pm^{(\alpha)*}, \quad A \in \mathfrak{B}_2(\mathcal{K}). \tag{4.1}$$

Naturally, (4.1) as well as all the later relations holding on $\mathfrak{B}_2(\mathcal{K})$ or $\mathfrak{B}_2(\tilde{\mathcal{K}})$ can be immediately applied to any statistical operator ρ, since the family of all such operators belongs to the respective trace-classes $\mathfrak{B}_1(\mathcal{K})$ or $\mathfrak{B}_1(\tilde{\mathcal{K}})$ and $\mathfrak{B}_1(\mathcal{K}) \subset \mathfrak{B}_2(\mathcal{K})$, $\mathfrak{B}_1(\tilde{\mathcal{K}}) \subset \mathfrak{B}_2(\tilde{\mathcal{K}})$.

Instead of (3.1), we expect to have

$$\underline{\Omega}_\pm^{(\alpha)} \rho = \text{h-}\lim_{\epsilon \to +0} \underline{\Omega}_{\pm\epsilon}^{(\alpha)} \underline{J}_\alpha \rho, \tag{4.2}$$

where "h-lim" denotes the limit in $\mathfrak{B}_2(\mathcal{K})$ taken in the Hilbert-Schmidt norm, and for any $A \in \mathfrak{B}_2(\mathcal{K})$

$$\underline{\Omega}_{\pm\epsilon}^{(\alpha)} A = \Omega_{\pm\epsilon}^{(\alpha)} A \Omega_{\pm\epsilon}^{(\alpha)*}, \tag{4.3}$$

$$\underline{J}_\alpha A = J_\alpha A J_\alpha. \tag{4.4}$$

We note that $\mathfrak{B}_2(\mathcal{K})$ and $\mathfrak{B}_2(\tilde{\mathcal{K}})$ are Hilbert spaces with the respective inner products

$$\langle A | B \rangle_2 = \text{Tr}[A^* B], \quad (\tilde{A} | \tilde{B})_2 = \text{Tr}[\tilde{A}^* \tilde{B}].$$

It is then to be expected that formulas analogous to (3.2) would hold for the transformer $\underline{\Omega}_{\pm\epsilon}^{(\alpha)}$. These formulas would enable the direct computation of $\underline{\Omega}_{\pm\epsilon}^{(\alpha)}$ without recourse to (3.2).

In addition to the expected usefulness of such relations in problems related to quantum statistical mechanics,[18,19] these formulas for wave transformers have the advantage over the corresponding formulas for wave-operators of being completely unaffected by shifts of the energy spectrum. This makes them attractive in any renormalization program involving removals of cut-offs in which the ground state energy is cut-off dependent.[12]

Theorem 4.1: The families $\underline{U}(t)$ and $\underline{\tilde{U}}(t)$, $t \in \mathbf{R}^1$,

$$\underline{U}(t) A = \exp(-iHt) A \exp(iHt),$$
$$\underline{\tilde{U}}(t) \tilde{A} = \exp(-i\tilde{H}t) \tilde{A} \exp(i\tilde{H}t), \tag{4.5}$$

are additive strongly continuous groups of unitary transformers on the Hilbert spaces $\mathfrak{B}_2(\mathcal{K})$ and $\mathfrak{B}_2(\tilde{\mathcal{K}})$, res-

pectively. Their respective infinitesimal generators $-i\underline{H}$ and $-i\underline{\tilde{H}}$ *are such that*

$$\underline{H}A = (HA)^{**} - (AH)^{**}, \quad \underline{\tilde{H}}\tilde{A} = (\tilde{H}\tilde{A})^{**} - (\tilde{A}\tilde{H})^{**} \tag{4.6}$$

for any $A \in \mathfrak{D}(\underline{H})$ and $\tilde{A} \in \mathfrak{D}(\underline{\tilde{H}})$, *with both* $\underline{H}$ and $\underline{\tilde{H}}$ *being self-adjoint transformers. The relationship (3.16) is valid if and only if*

$$\underline{\Omega}_{\pm}^{J}\rho = \sum_{\alpha} \underline{W}_{\pm}^{(\alpha)} J_{\alpha}\rho = \text{h-}\lim_{t\to\pm\infty} \underline{U}(t)\underline{J}\,\underline{\tilde{U}}(-t)\rho \tag{4.7}$$

for any statistical operator $\rho \in \mathfrak{B}_1(\tilde{\mathfrak{H}})$.

Proof: The relations (4.6) are obtained by specializing Theorem 3.2 and Theorem 4.2 in Ref. 18 to $\mathfrak{B}_2(\mathfrak{H})$ and $\mathfrak{B}_2(\tilde{\mathfrak{H}})$, respectively. The rest of the first two statements in Theorem 4.1 are equivalent to Theorem A1 in Ref. 12.

Let us set $U(t) = \exp(-iHt)$ and $\tilde{U}(t) = \exp(-i\tilde{H}t)$. From the relation

$$\langle [\underline{\Omega}_{\pm}^{J} - \underline{U}(t)\underline{J}\,\underline{\tilde{U}}(-t)]|\Psi)(\Psi|\,|\,[\underline{\Omega}_{\pm}^{J} - \underline{U}(t)\underline{J}\,\underline{\tilde{U}}(-t)]|\Psi)(\Psi|\rangle_2 = |\langle [\Omega_{\pm}^{J} - U(t)J\tilde{U}(-t)]\Psi|[\Omega_{\pm}^{J} - U(t)J\tilde{U}(-t)]\Psi\rangle|^2$$

we infer that (3.16) holds true if (4.7) holds true. Conversely, assume that (3.16) is true. Take any $\rho \in \mathfrak{B}_1(\tilde{\mathfrak{H}})$ with the canonical decomposition

$$\tilde{A} = \sum_{n=1}^{\infty} |\Psi_n\rangle\lambda_n\langle\Psi_n|, \quad \sum_{n=1}^{\infty} |\lambda_n| < \infty,$$

where the set $\{\Psi_n\}$ is an orthonormal system of vectors in $\tilde{\mathfrak{H}}$. Since

$$\|U(t)JU(-t)\Psi\| \leq |J|\,\|\Psi\|_0$$

for all $t \in \mathbb{R}^1$ and all $\Psi \in \tilde{\mathfrak{H}}$, we have the following estimate

$$\begin{aligned}\langle [\Omega_{\pm}^{J} - \underline{U}(t)\underline{J}\,\underline{\tilde{U}}(-t)]\rho\,|\,[\Omega_{\pm}^{J} - \underline{U}(t)\underline{J}\,\underline{\tilde{U}}(-t)]\rho\rangle_2 &\leq \sum_{m,n=1}^{r} \lambda_m\lambda_n |\langle [\Omega_{\pm}^{J} - U(t)J\tilde{U}(-t)]\Psi_m \\ &\times |[\Omega_{\pm}^{J} - \tilde{U}(t)JU(-t)\Psi_n]\rangle|^2 \\ &+ \sum_{m+n=r+1}^{\infty} (|\Omega_{\pm}| + |J|)^4\lambda_m\lambda_n.\end{aligned}$$

Hence, for any fixed $\delta > 0$ we can make the second of the above sums smaller than δ by choosing r sufficiently large. After that, if (3.16) holds, the first sum can also be made smaller than δ by taking $|t| > t_0(\delta)$ for some appropriate choice of $t_0(\delta)$. Thus, all the statements in Theorem 4.1 have been verified.

Theorem 4.2: The transformers

$$\underline{\Omega}_{\pm}^{J} = \sum_{\alpha} \underline{\Omega}_{\pm}^{(\alpha)}\underline{J}_{\alpha}, \quad \underline{\Omega}_{\pm}^{(\alpha)}A = \Omega_{\pm}^{(\alpha)}A\Omega_{\pm}^{(\alpha)*}, \tag{4.8}$$

are partial isometries with initial domain $\mathfrak{B}_2(\tilde{\mathfrak{H}})$ *and final domains*

$$\underline{R}_{\pm} = \sum_{\alpha} \underline{E}_{\pm}^{(\alpha)}\mathfrak{B}_2(\mathfrak{H}), \quad \underline{E}_{\pm}^{(\alpha)}A = E_{\pm}^{(\alpha)}AE_{\pm}^{(\alpha)}. \tag{4.9}$$

For any density operator $\rho \in \mathfrak{B}_1(\tilde{\mathfrak{H}})$.

$$\underline{\Omega}_{\pm}^{J}\rho = \text{h-}\lim_{\epsilon\to+0} \underline{\Omega}_{\pm\epsilon}^{J}\rho\ , \tag{4.10}$$

if and only if (2.5) is true, where

$$\underline{\Omega}_{\pm\epsilon}^{J}A = \Omega_{\pm\epsilon}^{J}A(\Omega_{\pm\epsilon}^{J})^*, \quad A \in \mathfrak{B}_2(\mathfrak{H}). \tag{4.11}$$

Proof: The partial isometry properties of $\underline{\Omega}_{\pm}^{J}$ are immediate consequences of the corresponding properties of $\Omega_{\pm}^{J}$. Furthermore, taking into consideration that

$$\|\underline{\Omega}_{\pm\epsilon}^{J}A\|_2 \leq |\Omega_{\pm\epsilon}^{J}|^2\|A\|_2 \leq |J|^2\|A\|_2$$

we can derive that the last statement of Theorem 4.2 is true by using the method employed in establishing the last statement of Theorem 4.1.

By stating and proving the preceding two theorems we have set the stage for applying the results of Sec. 2 to any statistical operator ρ in the same manner in which these results were applied to a state vector Ψ. The only distinction between the present case and the one treated in the preceding section is that we are now working on the Hilbert spaces $\mathfrak{B}_2(\tilde{\mathfrak{H}})$ and $\mathfrak{B}_2(\mathfrak{H})$ instead of the Hilbert spaces $\tilde{\mathfrak{H}}$ and $\mathfrak{H}$. Moreover, the present theory for the ρ matrix can be formulated without direct recourse to the theory for wave-packets by working directly with the spectral functions[12] $\underline{E}_{\lambda}$ and $\underline{\tilde{E}}_{\lambda}$ of the transformers $\underline{H}$ and $\underline{\tilde{H}}$, respectively.

---

*Supported in part by the National Research Council of Canada.

[1] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[2] E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).
[3] H. Ekstein, Phys. Rev. **101**, 880 (1956).
[4] C. van Winter, Mat. Fys. Skr. Dan Vid. Selsk. **2** (10), ,(1965).
[5] C. Chandler and A. G. Gibson, "Transition from time-dependent to time-independent multichannel quantum scattering theory" (University of New Mexico, preprint).
[6] T. Kato, J. Funct. Anal. **1**, 342 (1967).
[7] F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[8] E. Prugovečki and J. Zorbas, "Modified Lippmann-Schwinger equations for two-body scattering theory with long-range interactions" (University of Toronto, preprint).
[9] V. B. Matveev and M. M. Skriganov, Teor. Mat. Fiz. **10**, 238 (1972).
[10] E. Prugovečki, Nuovo Cimento B **4**, 105 (1971).
[11] V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. **2**, 367 (1970).
[12] E. Prugovečki, J. Math. Phys. **13**, 969 (1972).
[13] A. L. Belopolskiĭ and M. Sh. Birman, Izv. Akad. Nauk SSSR **32**, 1162 (1968).
[14] E. Prugovečki, SIAM J. Math Anal. **3**, 183 (1972).
[15] E. Prugovečki, "The bra and ket formalism in extended Hilbert space", J. Math. Phys. (to appear).
[16] J. D. Dollard, Rocky Mountain J. Math. **1**, 5 (1971).
[17] E. Prugovečki and A. Tip, "Semigroups of rank-preserving transformers on minimal norm ideals in $\mathfrak{B}(\mathfrak{H})$" (preprint).
[18] A. Tip, Physica (Utr.) **52**, 493 (1971).
[19] E. Prugovečki and A. Tip, "On ergodic limits of normal states in quantum statistical mechanics" (preprint).

Calculation of reflection and transmission coefficients for a class of one-dimensional wave propagation problems in inhomogeneous media

Alfredo Baños Jr.

Department of Physics, University of California, Los Angeles, California 90024.
(Received 24 October 1972)

Making use of an extension of Langer's method, we have solved the one-dimensional Helmholtz equation $w'' + k_0^2 g(z)w(z) = 0$, $k_0 = 2\pi/\lambda_0$, in which λ_0 is the free space wavelength and in which $g(z)$ represents any one of the three characteristic symmetric profiles. The main objective of this paper is the computation of the reflection and transmission coefficients for such a family of profiles making use of one and the same method throughout. First, we introduce our modification of Langer's method that reduces the problem to the solution of a comparison differential equation of the parabolic cylinder type, which we choose to solve in terms of confluent hypergeometric functions. Then, we set up a solution $w_+(z)$ valid for $0 \leq z$ which, as $z \to \infty$, becomes the transmitted plane wave. At this juncture we apply the method of analytic continuation, as z is changed into $-z$, and we obtain a linear combination $w_-(z) = w_r(z) + w_i(z)$ valid for $z \leq 0$ which, as $z \to -\infty$, yields the superposition of the incident and reflected waves. The reflection and transmission coefficients then ensue readily upon division by the coefficient of $\exp(i k_0 z)$ in the incident wave. The body of the work is first carried out for symmetric profiles, which we then proceed to extend to nonsymmetric profiles. Because we are using Langer's method in lowest approximation, we examine in detail the conditions underlying the applicability of the method and compare our results with the exact results derived from symmetric Epstein profiles. Although this paper is concerned mainly with the solution of the reflection–transmission problems, we have in fact set up all the necessary mathematical machinery for the computation of the actual wavefunctions in the vicinity of the origin. We have undertaken to do so, for specific parameter values, in a companion paper to appear shortly.

1. INTRODUCTION

We are concerned here with the solution of the one-dimensional Helmholtz equation

$$w'' + k_0^2 g(z)w(z) = 0 \tag{1.1}$$

where $k_0 = 2\pi/\lambda_0$ is the free space wavenumber and where the function $g(z)$ is characterized by the symmetric curves shown in Fig. 1, all of which have the common properties that $g'(0) = 0$ and $g(z) \to 1$ as $|z| \to \infty$, but are otherwise arbitrary. The method discussed here, as shown below, is readily extended to nonsymmetric profiles. The objective is to obtain (approximate) solutions of (1.1) which are uniformly valid in the whole interval $-\infty < z < \infty$, whether we are dealing with curves of type (1), which exhibit no turning points, or curves of type (3), which exhibit two turning points that may or may not be far apart, and maybe in fact coalesce at $z = 0$ to yield curve (2) with a quadratic turning point. It is clear that curve (2), for which $g(0) = 0$, is the boundary curve that separates the curves of class (1) and class (3), for which $g(0)$ is positive and negative respectively.

In Sec. 2 we give a detailed account of our own modification of Langer's method,[1] which leads to a comparison differential equation of the parabolic cylinder type. In Sec. 3 we carry out the reduction of this equation to the canonical form of the confluent hypergeometric equation, on which our entire analysis is based. As shown below, the application of Langer's method introduces a new independent variable $x = x(z)$ and, in Sec. 4, we examine in detail this functional relation, which in general is implicit and quite complicated, but which asymptotically leads to straightforward formulas. In Sec. 5 we take up the process of analytic continuation, as it applies to the problem at hand, followed in Sec. 6 by a study of the asymptotic limit of the wavefunctions, transmitted, reflected and incident. In Sec. 7 we solve completely the reflection-transmission problem, and we examine useful limiting cases depending on the magnitude of $g(0)$. Section 8 deals with the modifications that have to be introduced when the profiles of Fig. 1 are no longer symmetric. In *Appendix* A we discuss the conditions underlying the applicability of Langer's method to the problem at hand, and we show that the method provides essentially exact results for the transmission and reflection coefficients when the parameter ratio $\lambda/\lambda_0 \geq 1$, where λ is a proper characteristic scale length of the profiles of Fig. 1 and λ_0 is the free space wavelength. In *Appendix* B we confirm the above statements through a comparison with exact results derived from symmetric Epstein contours.

2. LANGER'S TRANSFORMATION

Following Langer,[1] the first step consists of resolving the dependent variable $w(z)$ in (1.1) into the product

$$w(z) = u(z)v(x(z)), \tag{2.1}$$

where $u(z)$ is a factor to be determined presently, $v(x)$ is a new dependent variable, and $x(z)$ a new independent

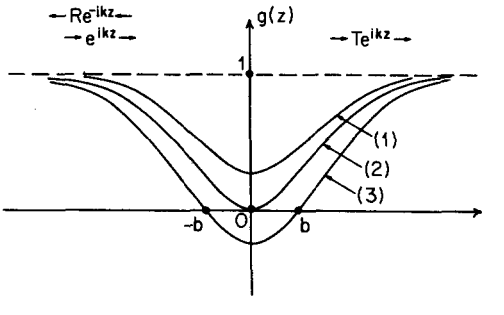

FIG. 1. Characteristics curves of $g(z)$ arising from symmetric profiles, which give rise to (1) no turning points, (2) a quadratic turning point, or (3) two turning points.

Copyright © 1973 by the American Institute of Physics

variable that, henceforth, we shall refer to as Langer's variable. Introducing (2.1) into (1.1), we obtain the differential equation

$$ux'^2\frac{d^2v}{dx^2} + (ux'' + 2u'x')\frac{dv}{dx} + (u'' + k_0^2gu)v(x) = 0. \tag{2.2}$$

Equating to zero the coefficient of dv/dx, we obtain at once

$$ux'' + 2u'x' = 0,$$

which, upon integration, yields the desired factor

$$u(z) = Nk_0^{1/2}x'^{-1/2}, \tag{2.3}$$

where N is a dimensionless normalizing factor to be determined *a posteriori*. Next, we are to divide the remaining terms of (2.2) by ux'^2, the coefficient of d^2v/dx^2 and, in the spirit of Langer's method, we put

$$k_0^2g/x'^2 = x^2 + 4\eta, \tag{2.4}$$

where η is a convenient parameter characterizing the curves of Fig. 1 as we shall see, with η being positive, zero, or negative depending on whether $g(0)$ is positive, zero, or negative. We now rewrite (2.4) as a differential equation

$$(x^2 + 4\eta)x'^2 = k_0^2g(z), \tag{2.5}$$

from which, upon extracting the square root and integrating, we obtain the functional relation $x = x(z)$ as the equality of two definite integrals,

$$\int_0^{|x|}(\xi^2 + 4\eta)^{1/2}d\xi = k_0\int_0^{|z|}\sqrt{g(s)}\,ds, \tag{2.6}$$

in which we have arranged matters in such a way that $x = 0$ when $z = 0, x > 0$ when $z > 0$, and $x < 0$ when $z < 0$ and in which it is understood that $g(s)$ is an even function of its argument, $g(-s) = g(s)$.

Combining (2.3) and (2.5) to eliminate x' and substituting into (2.1), we obtain Langer's solution

$$w(z) = N[(x^2 + 4\eta)/g(z)]^{1/4}v(x(z)), \tag{2.7}$$

which is uniformly valid for $-\infty < z < \infty$, and in which $v(x)$ satisfies the differential equation

$$\frac{d^2v}{dx^2} + \left(\frac{u''}{ux'^2} + x^2 + 4\eta\right)v(x) = 0, \tag{2.8}$$

which is intractable as it stands. However, we show in *Appendix* A that the unwanted term u''/ux'^2 is "small and slowly varying". Thus, in lowest order approximation, we neglect the unwanted term outright, to obtain for $v(x)$ the so-called comparison differential equation

$$v'' + (x^2 + 4\eta)v(x) = 0, \tag{2.9}$$

which we know how to solve in terms of parabolic cylinder functions or better still, in the present instance, in terms of solutions of the confluent hypergeometric equation, as based on Ref. 2, *Higher Transcendental Functions*, hereinafter to be referred to as HTF for brevity.

To conclude, it should be pointed out at this juncture that our proposed transformation (2.4) is, in fact, a special case (with $m = 2$) of the more general expression.

$$k_0^2g/x'^2 = x^m + 4\eta, \tag{2.10}$$

with m unrestricted. In this case we would have, instead of (2.9), the comparison differential equation

$$v'' + (x^m + 4\eta)v(x) = 0, \tag{2.11}$$

which, for $\eta = 0$, is still solvable in terms of Bessel functions. However, in the present instance, we had very much in mind the possibility of comparing our approximate results with the exact results derived from symmetric Epstein profiles, as explained in *Appendix* B, which automatically restricts us to the case $m = 2$ in (2.10), because the Epstein profiles belong to this class.

3. REDUCTION TO THE CONFLUENT HYPERGEOMETRIC EQUATION

We see at once from (2.9) that, for finite values of the parameter η, the WKB solutions for $x \to \infty$ can be written as $v(x) \sim x^{-1/2}\exp(\pm\frac{1}{2}ix^2)$. Hence, putting

$$v(x) = e^{ix^2/2}\phi(x) \tag{3.1}$$

into (2.9), we ascertain that the new dependent variable $\phi(x)$ satisfies the differential equation

$$\phi'' + 2ix\phi' + (i + 4\eta)\phi(x) = 0. \tag{3.2}$$

Next, introducing the new independent variable

$$w = -ix^2, \tag{3.3}$$

which, of course, is not to be confused with the dependent variable $w = w(z)$ in (1.1), we obtain at once from (3.2) the canonical form of the confluent hypergeometric equation

$$w\phi'' + (c - w)\phi' - a\phi(w) = 0, \tag{3.4}$$

with parameters

$$c = \tfrac{1}{2} \quad \text{and} \quad a = \tfrac{1}{4} - i\eta. \tag{3.5}$$

We know that the confluent hypergeometric equation (3.4) always possesses the fundamental set of solutions $F_5(w)$ and $F_7(w)$, which, according to HTF, Section 6.1, exhibit for $|w| \to \infty$ the asymptotic leading terms

$$F_5(w) \sim w^{-a}, \qquad F_7(w) \sim w^{a-c}e^{w - i\pi\epsilon(a-c)}, \tag{3.6}$$

wherein $\epsilon = \text{sgn}\{\text{Im } w\}$. In the present instance $w = -ix^2$ $\text{Im}w < 0$, and $\epsilon = -1$, provided $x > 0$ and $x \to \infty$. Hence, in this case, making use of (3.1), the most general solution of (2.9) can be written as the linear combination

$$v(x) = e^{ix^2/2}[AF_5(-ix^2) + BF_7(-ix^2)], \tag{3.7}$$

with arbitrary coefficients A and B.

As an example, in the case of curve (2) of Fig. 1, $\eta = 0$ and the parameters become $c = 1/2$ and $a = 1/4$. In this case, making use of (3.6), we obtain from (3.7) the asymptotic form

$$v(x) \sim (-ix^2)^{-1/4}(Ae^{ix^2/2} + Be^{-ix^2/2 - i\pi/4}), \tag{3.8}$$

which we identify, for $x \to \infty$, with the superposition of a progressive wave associated with $F_5(-ix^2)$ and a regressive wave associated with $F_7(-ix^2)$. For $x < 0$ and $x \to -\infty$, we should need to use in (3.7) the analytic continuation of the functions involved, subject to which we return in Sec. 5. Here, we merely wish to point out that the exponentials in (3.8), as pointed out by Weibel,[3] ex-

change roles as we change x into $-x$. This fact is readily established by first writing the corresponding phases of the exponentials as $\phi = \pm \frac{1}{2}x^2 - \omega t$ and then putting $d\phi = 0 = \pm x dx - \omega dt$, from which the phase velocities emerge as $v = dx/dt = \omega/(\pm x)$, which clearly depend on $\mathrm{sgn}\{x\}$.

4. FUNCTIONAL RELATION $x = x(z)$

Let us consider first curves of class (1) for which $0 \leq g(0) \leq 1$ and $\eta \geq 0$ but finite. In the sequel we shall endeavor to determine η so as to characterize the curves of class (1), and we shall find out that only when $g(z)$ in (1.1) can be allowed to be expanded to second order (parabolic profile) are we able to give a precise prescription for the computation $\eta \geq 0$. The functional relation $x = x(z)$ is given implicitly by the equality of the two definite integrals in (2.6). First, let us examine the asymptotic limit of the integral on the right-hand side of (2.6), namely,

$$\text{rhs} = k_0 \int_0^{|z|} \sqrt{g(s)}\, ds \xrightarrow[|z|\to\infty]{} k_0(|z| - z_0), \tag{4.1}$$

where $k_0 z_0$ is a phase shift that we compute as follows. Consider $z > 0$ and $z \to \infty$. Figure 2 shows a graph of

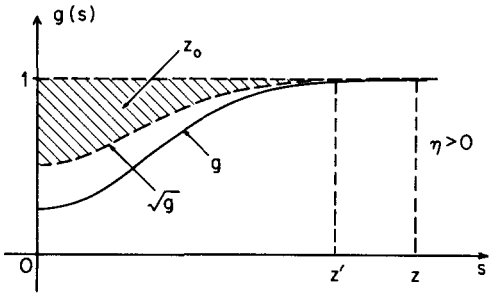

FIG. 2. Graphical computation of the phase shift z_0, for $\eta > 0$, as given by (4.2).

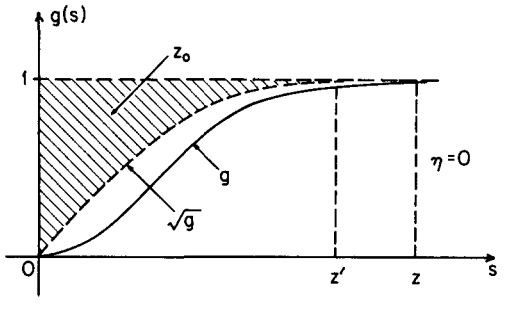

FIG. 3. Graphical computation of the phase shift z_0, for $\eta = 0$, as given by (4.2).

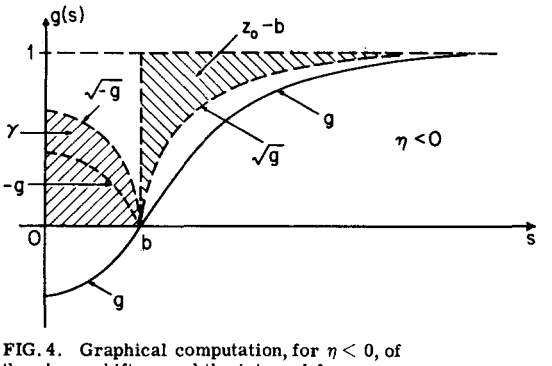

FIG. 4. Graphical computation, for $\eta < 0$, of the phase shift z_0 and the integral for γ as given by (4.6).

$g(s)$ and $\sqrt{g(s)}$ for $\eta > 0$. Then, choosing z' sufficiently large, we have

$$\int_0^z \sqrt{g(s)}\, ds \xrightarrow[z\to\infty]{} \int_0^{z'} \sqrt{g(s)}\, ds + \int_{z'}^z 1 \cdot ds = z - z' + \int_0^{z'} \sqrt{g(s)}\, ds = z - z_0,$$

where

$$z_0 = z' - \int_0^{z'} \sqrt{g(s)}\, ds = \int_0^{z'} [1 - \sqrt{g(s)}]ds \simeq \int_0^{\infty} [1 - \sqrt{g(s)}]ds, \tag{4.2}$$

which is the value for z_0 to be inserted into (4.1) and which is represented by the shaded area in Fig. 2. For $\eta = 0$, corresponding to curve (2) of Fig. 1, the value of z_0 is still given by (4.2) corresponding to the shaded area illustrated in Fig. 3.

Next we examine the asymptotic limit of the integral on the left-hand side of (2.6), again confining our remarks to curves of class (1), $\eta \geq 0$. The fact that the upper limit is written as $|x|$ tells us that the integral is an even function of x. Thus, confining our remarks to $x \geq 0$, we first have the elementary integral

$$\text{lhs} = \int_0^x (\xi^2 + 4\eta)^{1/2} d\xi = \tfrac{1}{2}x(x^2 + 4\eta)^{1/2} + 2\eta \ln \frac{x + (x^2 + 4\eta)^{1/2}}{\sqrt{4\eta}}, \tag{4.3}$$

which remains unchanged, for $x < 0$, when we replace x by $-x$, indicating that the integral above, regarded as a function of x in the interval $-\infty < x < \infty$, exhibits a branch point at $x = 0$. Proceeding to the asymptotic limit we have

$$\int_0^x (\xi^2 + 4\eta)^{1/2} d\xi \xrightarrow[|x|\to\infty]{} \tfrac{1}{2}x^2 + \eta + \cdots + \eta \ln \frac{x^2}{\eta} = \tfrac{1}{2}x^2 + \eta \ln x^2 + \eta(1 - \ln\eta) + O(x^{-2}), \tag{4.4}$$

where, it is recalled, $\eta \geq 0$ but finite.

We now proceed similarly for curves of class (3), for which $g(0) \leq 0$ and $\eta \leq 0$ but finite. First we examine the asymptotic limit of the integral on the right-hand side of (2.6) for $z > 0$ and $z \to \infty$. Thus we have (Fig. 1):

$$\text{rhs} = k_0 \int_0^b \sqrt{g(s)}\, ds + k_0 \int_b^z \sqrt{g(s)}\, ds \xrightarrow[|z|\to\infty]{} ik_0\gamma + k_0(z - z_0), \tag{4.5}$$

which should be contrasted with (4.1) and in which

$$\gamma = \int_0^b \sqrt{-g(s)}\, ds \quad \text{and} \quad z_0 = b + \int_b^{\infty} [1 - \sqrt{g(s)}]ds, \tag{4.6}$$

which we establish as follows with reference to Fig. 4. Thus, the first integral on the right of (4.5) may be written

$$\int_0^b \sqrt{g(s)}\, ds = i \int_0^b \sqrt{-g(s)}\, ds = i\gamma,$$

where γ is defined by the first of Eqs. (4.6) and is illustrated graphically by the left hand cross-hatched area in Fig. 4. To establish the second integral in (4.6), we proceed in a manner completely analogous to (4.1) and (4.2). Thus, for sufficiently large z', we write

$$\int_b^z \sqrt{g(s)}\, ds \xrightarrow[z\to\infty]{} \int_b^{z'} \sqrt{g(s)}\, ds + \int_{z'}^z 1 \cdot ds$$

$$= z - z' + \int_b^{z'} \sqrt{g(s)}\, ds = z - z_0,$$

where, this time

$$z_0 = z' - \int_b^{z'} \sqrt{g(s)}\, ds = b + \int_b^{z'} [1 - \sqrt{g(s)}]ds \simeq b + \int_b^{\infty} [1 - \sqrt{g(s)}]ds,$$

which should be contrasted with (4.2) and wherein the last integral is represented graphically by the shaded area on the right of Fig. 4.

Finally, we examine the asymptotic limit of the left-hand side of (2.6) for $\eta \leq 0$. First, we observe that the integrand in question vanishes when $\xi^2 + 4\eta = 0$ or $\xi = \pm\sqrt{-4\eta}$, and we note that, when $\eta < 0$, we can write $\eta = e^{i\pi}(-\eta)$. Hence, making use of (4.3), we compute the integral

$$\int_0^{\sqrt{-4\eta}} (\xi^2 + 4\eta)^{1/2} d\xi = 2\eta \ln \frac{\sqrt{-4\eta}}{\sqrt{4\eta}} = 2\eta \ln e^{-i\pi/2} = -i\pi\eta,$$

and equating this result to the first integral in (4.5), that is, equating imaginary parts, we obtain finally

$$-\pi\eta = k_0\gamma, \tag{4.7}$$

an important equality which determines $\eta \leq 0$ in terms of the original profile $g(z)$ in Fig. 1. More importantly, we see that (4.7) guarantees that, in Langer's solution (2.7), the zeros in the numerator of the fourth root bracket coincide precisely with the zeros of $g(z)$ in the denominator, thus insuring the uniform validity of the solution for $-\infty < z < \infty$. To complete the asymptotic analysis of the left-hand side of (2.6) for $\eta \leq 0$, we return to (4.4), and write, with $\eta = e^{i\pi}(-\eta)$,

$$\int_0^x (\xi^2 + 4\eta)^{1/2} d\eta \xrightarrow[|x|\to\infty]{} \tfrac{1}{2}x^2 + \eta + \cdots + \eta \ln \frac{x^2 e^{i\pi}}{-\eta} = \tfrac{1}{2}x^2 + \eta[1 - \ln(-\eta)] - i\pi\eta + O(x^{-2}), \tag{4.8}$$

which should be contrasted with (4.4). Equating the imaginary parts of (4.5) and (4.8) merely recovers (4.7), but in addition it tells us that the imaginary parts that accrue asymptotically on both sides of (2.6), when $\eta < 0$, simply cancel each other out.

Thus, finally, whether $\eta \geq 0$ or $\eta \leq 0$, upon equating the corresponding forms for the lhs and rhs of (2.6), and then solving for $\frac{1}{2}x^2$ in each case, we obtain for the exponential $e^{ix^2/2}$ the asymptotic limit

$$e^{ix^2/2} \xrightarrow[|x|\to\infty]{} |x|^{-2i\eta} e^{-i\eta(1-\ln|\eta|)} e^{ik_0(|z|-z_0)}, \tag{4.9}$$

wherein η is given by (4.7) for $\eta \leq 0$ but remains still unspecified for $\eta \geq 0$ (except as noted at the beginning of the present section), and in which z_0 is given by (4.2) for $\eta \geq 0$ and by the second of (4.6) for $\eta \leq 0$. We note from (4.9) that, to obtain the asymptotic form of the negative exponential $e^{-ix^2/2}$, we need only take the complex conjugate of (4.9).

5. THE PROCESS OF ANALYTIC CONTINUATION

As shown in (3.8), the function $F_5(w) = F_5(-ix^2)$, for $x > 0$ and $x \to \infty$, is associated with a progressive traveling wave, which we shall eventually identify as the transmitted wave in our reflection-transmission problem. Thus, we are led to examine the analytic continuation of $F_5(-ix^2)$ as x is changed into $-x$. For the purpose we note that, for $x < 0$, we can write $x = e^{i\pi}(e^{-i\pi}x) = e^{i\pi}(-x)$, whence $x^2 = e^{2\pi i}(-x)^2 = x^2 e^{2\pi i}$. Thus, we must investigate the properties of the function $F_5(we^{2\pi i})$ recalling that the functions $F_5(w)$ and $F_7(w)$ exhibit a branch point at $w = 0$. Invoking the formulas HTF 6.8 (17) with $m = 1$, combined with 6.7(10), we can write $F_5(we^{2\pi i})$ as the linear combination

$$F_5(we^{2\pi i}) = AF_5(w) + BF_7(w), \tag{5.1}$$

in which the coefficients A and B are given explicitly in terms of the parameters c and a, which characterise the confluent hypergeometric equation (3.4).

Thus, putting $c = \frac{1}{2}$ and $a = \frac{1}{4} - i\eta$, in accordance with (3.5), and putting $\epsilon = \mathrm{Im}w = -1$, in accordance with (3.3), we obtain the coefficients

$$A = -ie^{-2\pi\eta} \xrightarrow[\eta=0]{} -i, \tag{5.2}$$

$$B = (i + e^{-2\pi\eta})\Gamma[\tfrac{1}{4} + i\eta]/\Gamma[\tfrac{1}{4} - i\eta] \xrightarrow[\eta=0]{} 1 + i. \tag{5.3}$$

In the sequel, we will have occasion to examine the limiting form of the ratio of gamma functions in (5.3) for $4|\eta| \gg 1$. Thus, writing for brevity

$$\phi = \eta(1 - \ln|\eta|) \tag{5.4}$$

and making use of the leading term of Stirling's series, we obtain

$$\Gamma[\tfrac{1}{4} + i\eta]/\Gamma[\tfrac{1}{4} - i\eta] \xrightarrow[\eta\to\pm\infty]{} e^{-2i\phi \mp i\pi/4}, \tag{5.5}$$

which we have found very useful even when $|\eta| = 1$, when the error is less than 10%.

6. ASYMPTOTIC FORM OF THE WAVEFUNCTIONS

We now proceed to set up the wavefunctions, in the manner of (2.7), which represent asymptotically the incident, reflected, and transmitted waves for the problem at hand (Fig. 1). For the purpose, we first set up, for $z > 0$ and $z \to \infty$, a solution of the form (2.7) which we identify as the transmitted wave. Recalling from (4.9) that the exponential $e^{ix^2/2}$ is associated, for $x \to \infty$, with such a progressive wave, we write down at once, for $z \geq 0$, the wavefunction

$$w_+(z) = N\left(\frac{x^2 + 4\eta}{g(z)}\right)^{1/4} F_5(-ix^2)e^{ix^2/2}, \tag{6.1}$$

where $x = x(z)$ in accordance with (2.6) and Sec. 4 and where N is a normalizing factor still to be determined. For $x < 0$ and $x \to -\infty$, we must replace $F_5(-ix^2)$ in (6.1) by its analytic continuation (5.1), to obtain, for $z \leq 0$, the wavefunction

$$w_-(z) = N\left(\frac{(x^2 + 4\eta)}{g(z)}\right)^{1/4} [AF_5(-ix^2) + BF_7(-ix^2)]e^{ix^2/2}, \tag{6.2}$$

in which the coefficients A and B, in the present instance are given by (5.2) and (5.3), respectively. It can be shown that the postulated wavefunctions (6.1) and (6.2) satisfy, at $z = 0$, the continuity conditions

$$w_+(0) = w_-(0) \quad \text{and} \quad w_+'(0) = w_-'(0). \tag{6.3}$$

First we examine the asymptotic behavior of (6.1) as $z \to \infty$ and $g(z) \to 1$. To this end, making use of the first of (3.6), with $a = \frac{1}{4} - i\eta$, we deduce the asymptotic form

$$e^{ix^2/2}F_5(-ix^2) \sim (-ix^2)^{-a}e^{ix^2/2} = (-i)^{-a}x^{-1/2+2i\eta}e^{ix^2/2}, \quad (6.4)$$

which is valid as $x \to \infty$. Next, replacing $e^{ix^2/2}$ in (6.4) by its asymptotic limit (4.9) and substituting the combination into (6.1), we obtain the asymptotic form

$$w_+(z) \xrightarrow[z\to\infty]{} N(-i)^{-a}e^{-i\phi}\cdot e^{ik_0(z-z_0)}, \quad (6.5)$$

where ϕ has been defined by (5.4) and z_0 is given in Sec. 4. If we now choose the normalizing factor N in such a way that

$$N(-i)^{-a}e^{-i\phi} \equiv 1, \quad (6.6)$$

then, from (6.5), we obtain finally the asymptotic limit

$$w_+(z) \xrightarrow[z\to\infty]{} e^{ik_0(z-z_0)}, \quad (6.7)$$

which represents a progressive wave of unit amplitude that we identify as the *transmitted* wave.

Proceeding similarly, we now examine the asymptotic behavior of the two terms contained in (6.2) as $z \to -\infty$. We note that the first term, involving $F_5(-ix^2)$ still satisfies the first form in (6.4) but with $x < 0$; hence, the factor $e^{ix^2/2}$ must now be replaced by the appropriate limiting form of (4.9) as $z \to -\infty$. Thus, we obtain

$$e^{ix^2/2}F_5(-ix^2) \sim (-i)^{-a}(-x)^{-1/2}e^{-i\phi}\cdot e^{-ik_0(z+z_0)}. \quad (6.8)$$

Next, making use of the second of (3.6), with $c = \frac{1}{2}$, $a = \frac{1}{4} - i\eta$, and $\epsilon = -1$, we deduce from the second term in (6.2) the asymptotic form

$$e^{ix^2/2}F_7(-ix^2) \sim e^{-ix^2/2}(-ix^2)^{a-c}e^{i\pi(a-c)} = (-i)^{-a}(-x)^{-(1/2)-2i\eta}e^{-ix^2/2-i\pi/4}, \quad (6.9)$$

which is valid for $x \to -\infty$ and which should be contrasted with (6.4). Then, replacing $e^{ix^2/2}$ by its asymptotic form, which is the complex conjugate of (4.9) for $z \to -\infty$, we obtain from (6.9)

$$e^{ix^2/2}F_7(-ix^2) \sim (-i)^{-a}(-x)^{-1/2}e^{i\phi}\cdot e^{ik_0(z+z_0)-i\pi/4}, \quad (6.10)$$

which should be contrasted with (6.8).

Thus, finally, inserting (6.8) and (6.10) into (6.2) and making use of the normalization condition (6.6), we obtain, for $z \to -\infty$ and $g(z) \to 1$, the asymptotic form

$$w_-(z) \xrightarrow[z\to-\infty]{} Ae^{-ik_0(z+z_0)} + Be^{2i\phi-i\pi/4}e^{ik_0(z+z_0)}, \quad (6.11)$$

which we readily interpret as the superposition of a regressive (reflected) wave and a progressive (incident) wave and in which the coefficients A and B are given respectively by (5.2) and (5.3).

7. COMPUTATION OF REFLECTION AND TRANSMISSION COEFFICIENTS

The computation of the reflection and transmission coefficients R and T (see Fig. 1) now follows immediately from the asymptotic leading terms for $w_+(z)$ and $w_-(z)$ given above by (6.7) and (6.11), respectively. All we need to do is to reduce the incident wave in (6.11) to a plane wave of unit amplitude, and this is accomplished by dividing the transmitted and incident waves by the full coefficient of e^{ik_0z} in (6.11), namely, $B\exp(ik_0z_0 + 2i\phi - i\pi/4)$. Below we consider in turn several cases: (1) the general formulas for arbitrary $|\eta|$ finite; (2) the important practical case of a quadratic turning point corresponding to $\eta = 0$, curve (2) of Fig. 1; (3) the case in which $|\eta| \geq 1$, which leads to a great simplification of the general formulas; (4) the case corresponding to a parabolic profile, which arises when conditions are such that $g(z)$ in Fig. 1 can be expanded in Taylor's series retaining only second order terms in the analysis.

A. General formulas

Proceeding as indicated above, and making use of (5.2) and (5.3), we have at once

$$T = (i + e^{-2\pi\eta})^{-1}[\Gamma(\tfrac{1}{4} - i\eta)/\Gamma(\tfrac{1}{4} + i\eta)] \times \exp(-2ik_0z_0 - 2i\phi + i\pi/4), \quad (7.1)$$

and

$$R = (-ie^{-2\pi\eta})T, \quad (7.2)$$

in which $|\eta|$ is finite, k_0z_0 is a phase shift characteristic of the profile shape (Fig. 1), with z_0 given by (4.2) for $\eta \geq 0$, as illustrated in Fig. 2 for $\eta > 0$ and in Fig. 3 for $\eta = 0$, and in which z_0 is given by the second of (4.6) as illustrated in Fig. 4. The phase shift $\phi(\eta)$ appearing in (7.1) is defined by (5.4). We notice from (7.1) and (7.2) that

$$|T|^2 = 1/(1 + e^{-4\pi\eta}) \quad \text{and} \quad |R|^2 = e^{-4\pi\eta}/(1 + e^{-4\pi\eta}), \quad (7.3)$$

from which we deduce quite generally the principle of conservation of energy,

$$|T|^2 + |R|^2 = 1. \quad (7.4)$$

B. The case $\eta = 0$

When $\eta = 0$, we are dealing with curve (2) of Fig. 1, which exhibits a quadratic turning point. This is a case of considerable practical importance and great simplicity. Thus, putting $\eta = 0$ in (7.1) and (7.2), we obtain the transmission and reflection coefficients

$$T = [1/(1 + i)]e^{-2ik_0z_0+i\pi/4} = (1/\sqrt{2})e^{-2ik_0z_0}, \quad (7.5)$$

$$R = [-i/(1 + i)]e^{-2ik_0z_0+i\pi/4} = -(i/\sqrt{2})e^{-2ik_0z_0}, \quad (7.6)$$

in which z_0 is given by (4.2) and illustrated in Fig. 3 and in which we readily have

$$|T|^2 = |R|^2 = \tfrac{1}{2}, \quad (7.7)$$

indicating that in this case half of the incident energy is reflected and half is transmitted, a well-known fact that had previously been verified by Weibel,[3] Phinney,[4] and many other authors too numerous to mention. The novel feature of our results (7.5) and (7.6) resides in the fact that we have provided the means of computing the phase shifts $2k_0z_0$ for arbitrary contours in Fig. 1.

C. The case $|\eta| \geq 1$

A glance at the general formulas (7.1) and (7.2) discloses the fact that the exponential $e^{-2\pi\eta}$ becomes very large or very small depending on the sign of η, which

leads to considerable simplification. Because of the double sign in the exponent of (5.5), we must treat separately the cases $\eta \geq 1$ and $-\eta \geq 1$. For curves of class (1), with no turning points and $\eta \geq 1$, we obtain from (7.1) and (7.2), making use of (5.5), the simpler formulas

$$T \simeq ie^{-2ik_0z_0}/(i + e^{-2\pi\eta}), \qquad |T|^2 = 1/(1 + e^{-4\pi\eta}) \simeq 1, \tag{7.8}$$

$$R \simeq (-ie^{-2\pi\eta})T, \qquad |R|^2 = e^{-4\pi\eta}/(1 + e^{-4\pi\eta}) \simeq 0, \tag{7.9}$$

which state simply that, when $\eta \geq 1$, the disturbance caused by the plasma slab is essentially nil, meaning that the incident wave goes through without any measurable reflection. This is not to say that the wavefunction itself is not modified in the vicinity of $z = 0$, but in this paper we are not concerned with the waveforms.

The case corresponding to curves of class (3) in Fig. 1, with two turning points and $-\eta \geq 1$, where η this time is given by (4.7) under all circumstances, likewise by making use of (5.5) inserted into (7.1) and (7.2), yields the results

$$T \simeq e^{-2ik_0z_0}/(i + e^{-2\pi\eta}), \qquad |T|^2 = 1/(1 + e^{-4\pi\eta}) \simeq 0, \tag{7.10}$$

$$R \simeq (-ie^{-2\pi\eta})T, \qquad |R|^2 = e^{-4\pi\eta}/(1 + e^{-4\pi\eta}) \simeq 1, \tag{7.11}$$

which should be contrasted with (7.8) and (7.9) and which state that, for $-\eta \geq 1$, corresponding to a very wide evanescent portion of the plasma profile, the transmission is nil and all the incident power is reflected. Again, we remind the reader that the above statements do not imply that the incident and reflected waves are not modified in the vicinity of the first turning points at $z = -b$ in Fig. 1. In fact, we expect power amplification in this vicinity and some measure of penetration into the evanescent region; but these are matters which we defer to another paper.[5]

D. The case of parabolic profiles

In *Appendix* A we show that the conditions underlying the applicability of Langer's method demand that the parameter ratio

$$\lambda/\lambda_0 \geq 1, \tag{7.12}$$

where λ_0 is the free space wavelength and λ is some suitable scale length characterizing the curves of Fig. 1; for example, we might choose λ to be the value of z corresponding to the halfway mark in the profiles in Fig. 1. When (7.12) is satisfied and $|g(0)|$ is sufficiently small, we may represent $g(z)$ in (1.1) by the parabolic profile

$$g(z) = g(0) + \tfrac{1}{2}g''(0)z^2 + \cdots, \tag{7.13}$$

upon retaining only second order terms. We see at once from (7.13) upon putting $g(z) = 0$, that the turning points in Fig. 1 are given by

$$b^2 = -2g(0)/g''(0), \tag{7.14}$$

which are real and occur at $z = \pm b$ when $g(0) \leq 0$. We note, however, that when $g(0) > 0$, the roots of $g(z) = 0$ in (7.13) now constitute a pair of complex conjugate pure imaginary roots, still given correctly by (7.14).

Making use of (7.13) to evaluate the first integral in (4.6), we find, in conjunction with (4.7), that $\eta \leq 0$ may be computed from the formula

$$4\eta = k_0 g(0)/[\tfrac{1}{2}g''(0)]^{1/2}, \tag{7.15}$$

which was derived for $g(0) \leq 0$ and the quadratic form (7.12) for $g(z)$. However, when we apply the same technique for the computation of $\eta \geq 0$, that is, when we demand that the pair of complex conjugate imaginary roots of $g(z)$, as given by (7.14) for $g(0) \geq 0$, be cancelled by the pair of imaginary roots in the numerator $(x^2 + 4\eta)$ of the fourth root factor in Langer's solution (2.7), we find that the condition is exactly satisfied by (7.15). Thus, this formula is valid for $\eta \leq 0$ or $\eta \geq 0$, provided only that we are dealing with the parabolic profile (7.13). In summary, we have in (4.7) a universally valid formula to compute $\eta \leq 0$, which reduces to (7.15) for parabolic profiles. On the other hand, for $\eta \geq 0$, we have no universally valid formula and can only resort to (7.15) when applicable. Fortunately, the cases which occur in practice with $\eta > 0$ are generally of little physical interest.

To make the above considerations quantitative and to illustrate the domain of applicability of (7.14) and (7.15), we now adopt for $g(z)$ the Epstein profile[6]

$$g(z) = -h + (1 + h)\tanh^2(z/2\lambda) = -h + (1 + h)(z/2\lambda)^2 + \cdots, \tag{7.16}$$

in which λ is the scale length and from which, according to (7.13), we have

$$g(0) = -h, \qquad \tfrac{1}{2}g''(0) = (1 + h)/4\lambda^2, \tag{7.17}$$

in which we shall assume that $h \ll 1$. Substituting the above values into (7.15), we obtain

$$4\eta = -2\lambda k_0 h/\sqrt{1 + h}) = -(4\pi h/\sqrt{1 + h})\lambda/\lambda_0 \tag{7.18}$$

in terms of the parameter ratio $\lambda/\lambda_0 \geq 1$. We see at once that, even if $h \ll 1$, it does not necessarily follow that $|\eta|$ will also be small; in fact, it very much depends on the parameter ratio λ/λ_0, which in some problems (e.g. ionosphere) can be very large. If we put $\lambda/\lambda_0 = 1$, the minimum tolerable value, we see from (7.18) that $-\eta$ is $O(\pi h)$ for $h \ll 1$. Finally, substituting (7.17) into (7.14), we obtain for the location of the two turning points

$$\frac{b}{\lambda} = \pm\,[4h/(1 + h)]^{1/2} \quad \text{or} \quad b/\lambda_0 = \pm\,[4h/(1 + h)]^{1/2}\lambda/\lambda_0, \tag{7.19}$$

the first one of which clearly indicates, for $h \ll 1$, that b/λ is well within the parabolic region, and the latter states that the half-width of the evanescent region in Fig. 1 very much depends on the magnitude of the parameter ratio $\lambda/\lambda_0 \geq 1$.

In conclusion, we have shown how to deal with parabolic profiles of the form (7.13) when $|g(0)| \ll 1$. We have shown how to compute η and the location of the turning points, whether real ($\eta < 0$) or imaginary ($\eta > 0$). We have shown that $|\eta|$ can be large or small depending on the parameter ratio $\lambda/\lambda_0 \geq 1$. When $|\eta| \geq 1$, we can use the simpler formulas for the transmission and reflection coefficients given in Sec. 7C. For other values of η we can always resort to the general formulas derived in Sec. 7A.

8. NONSYMMETRIC PROFILES

The extension of the preceding results to nonsymmetric profiles is quite straightforward. Consider first the case $\eta \geq 0$ and $0 \leq g(0) < 1$ and a nonsymmetric profile characterized by the curve of Fig. 5, which should be contrasted with curves of class (1) in Fig. 1 and which has been drawn in the manner of Fig. 2. We note that, for nonsymmetric profiles, $g(-s) \neq g(s)$ and we can no longer use the right-hand side of the functional relation (2.6) as written. For $z > 0$ and $z \to \infty$, we have, instead of (4.1)

$$\text{rhs} = k_0 \int_0^z \sqrt{g(s)}\, ds \xrightarrow[z\to\infty]{} k_0(z - z_1), \tag{8.1}$$

where

$$z_1 = \int_0^\infty [1 - \sqrt{g(s)}]ds \tag{8.2}$$

which is formally identical to (4.2) and which is represented graphically by the shaded area on the right-hand side of Fig. 5, $0 \leq s \leq \infty$. However, for $z < 0$ and $z \to -\infty$, we must now write, instead of (4.1),

$$\text{rhs} = -k_0 \int_0^z \sqrt{g(s)}\, ds = k_0 \int_0^{-z} \sqrt{g(-s)}\, ds \xrightarrow[z\to-\infty]{} -k_0(z + z_2), \tag{8.3}$$

where now

$$z_2 = \int_0^\infty [1 - \sqrt{g(-s)}]ds \tag{8.4}$$

and is represented by the shaded area on the left-hand side of Fig. 5.

For $\eta = 0$, a nonsymmetric profile corresponding to curve (2) of Fig. 1 can still be handled, in a form completely analogous to the above treatment, by making use of a nonsymmetric Fig. 3. Thus, in brief, the formulas (8.2) and (8.4) are generally applicable to nonsymmetric profiles for which $\eta \geq 0$ and $0 \leq g(0) \leq 1$.

The case for $\eta \leq 0$ and $g(0) \leq 0$, when the profiles are nonsymmetric, is illustrated in Fig. 6 and requires a somewhat more complicated treatment. First we address ourselves to the question of computing the value of $\eta < 0$ which, in the case of symmetric profiles is given by (4.7). Proceeding in complete analogy with the earlier treatment, we now write

$$-2\pi i\eta = ik_0 \int_{-b_2}^{-b_1} \sqrt{-g(s)}\, ds = 2ik_0\gamma, \tag{8.5}$$

or

$$-\pi\eta = k_0\gamma, \tag{8.6}$$

which is formally identical to (4.7), but we now have that 2γ represents the double cross-hatched area in the center of Fig. 6. Next, we determine the abscissa c which divides the area 2γ into two equal parts, that is,

$$\gamma = \int_{-b_2}^{c} \sqrt{-g(s)}\, ds = \int_c^{b_1} \sqrt{-g(s)}\, ds. \tag{8.7}$$

Then, we rewrite the functional relation (2.6) in the form

$$\int_0^{|x|} [\xi^2 + 4\eta]^{1/2} d\xi = \pm k_0 \int_0^z \sqrt{g(s)}\, ds, \tag{8.8}$$

which implies that $x = 0$ when $z = c$ as determined from (8.7) and in which we have $x > 0$ when $z > c$ and $x < 0$ when $z < c$.

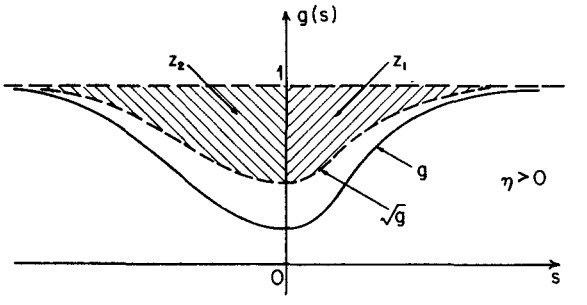

FIG. 5. Graphical computation of the phase shifts z_1 and z_2, for nonsymmetric profiles and $\eta > 0$, as given respectively by (8.2) and (8.4).

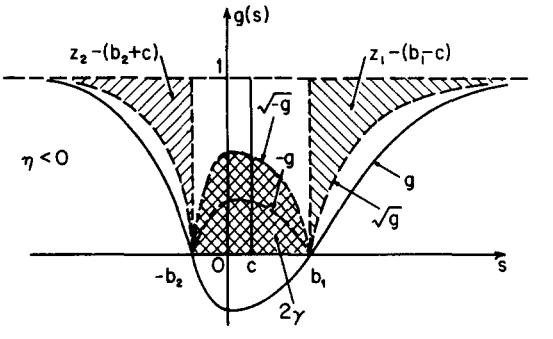

FIG. 6. Graphical computation of the phase shifts z_1 and z_2, as well as for the integral for γ, for nonsymmetric profiles and $\eta < 0$, as given respectively by (8.10), (8.12), and (8.7).

Thus, proceeding as in Sec. 4, we now write for $z > c$ and $z \to \infty$, instead of (4.5),

$$\text{rhs} = k_0 \int_c^{b_1} \sqrt{g(s)}\, ds + k_0 \int_{b_1}^\infty \sqrt{g(s)}\, ds \xrightarrow[z\to\infty]{} ik_0\gamma + k_0(z - z_1), \tag{8.9}$$

where γ is given by (8.7) and where

$$z_1 = (b_1 - c) + \int_{b_1}^\infty [1 - \sqrt{g(s)}]ds, \tag{8.10}$$

where the integral is represented by the shaded area on the right-hand side of Fig. 6. Proceeding similarly for $z < c$ and $z \to -\infty$, we now write

$$\text{rhs} = -k_0 \int_c^{-b_2} \sqrt{g(s)}\, ds - k_0 \int_{-b_2}^z \sqrt{g(s)}\, ds = k_0 \int_c^{b_2} \sqrt{g(-s)}\, ds + \int_{b_2}^{-z} \sqrt{g(-s)}\, ds \xrightarrow[z\to-\infty]{} ik_0\gamma - k_0(z + z_2), \tag{8.11}$$

where

$$z_2 = (b_2 + c) + \int_{b_2}^\infty [1 - \sqrt{g(-s)}]ds, \tag{8.12}$$

in which the integral is represented graphically by the shaded area on the left-hand side of Fig. 6.

The above analysis shows that, in the case of nonsymmetric profiles, we now have two phase shifts, $k_0 z_1$ and $k_0 z_2$, instead of the single phase shift $k_0 z_0$. Here, z_1 and z_2 are given respectively by (8.2) and (8.4) for $\eta \geq 0$ and by (8.10) and (8.12) for $\eta \leq 0$. The value of $\eta \leq 0$ is determined by (8.6), as explained above, in terms of the geometry of the profile between the turning points at $s = -b_2$ and $s = b_1$. The value of c, $-b_2 <$

$c < b_1$, exactly partitions the area 2γ as indicated by (8.6) and Fig. 6.

With these antecedents, our general formulas (7.1) and (7.2) for the transmission and reflection coefficients, as well as all the special cases discussed above, can now be readily extended to nonsymmetrical profiles by means of the following simple arguments. We merely observe that, for nonsymmetric profiles, the exponential corresponding to the *transmitted* wave (6.7) must now exhibit the phase shift $-k_0z_1$ instead of the previous $-k_0z_0$. On the other hand, in (6.11), the exponential of the *incident* wave must have the phase shift k_0z_0 replaced by k_0z_2, and the exponential of the reflected wave must have the phase shift $-k_0z_0$ replaced by $-k_0z_2$. Taking these replacements into account, we see that, for nonsymmetric profiles, we can use the formulas (7.1) and (7.2) as written, provided we replace $2z_0$ in (7.1), for the transmission coefficient T, by putting

$$2z_0 = z_1 + z_2, \tag{8.13}$$

and provided that in (7.2), for the reflection coefficient R, we merely write

$$z_0 = z_2. \tag{8.14}$$

We conclude by remarking that if one is only interested in the absolute values of the reflection and transmission coefficients, then the formulas (7.3) for $|T|^2$ and $|R|^2$ are universally valid, whether the profiles of Fig. 1 be symmetric or not; that is, the absolute values $|T|^2$ and $|R|^2$ take no cognizance of the geometric details of the profiles, except insofar as the value of η which characterizes a given curve reflects the actual geometry of the situation. To conclude, we remind the reader that, for $\eta \leq 0$, we have provided a universally valid prescription for its computation in terms of the geometry of the profile. On the other hand, for $\eta \geq 0$, we are able to provide a formula only for parabolic profiles, but, as stated before, the case $\eta > 0$ is of considerable less physical interest than the case $\eta \leq 0$.

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor Burton D. Fried, of the University of California, Los Angeles, and Professor Harold Levine, of Stanford University, for many illuminating discussions and critical remarks which proved invaluable in the course of these studies. The body of the work was performed while the author was on a tour of duty, during 1970-71, with the Office of Naval Research, London Branch Office, and served as the basis for a number of lectures that the author gave on the subject in England and on the Continent. He is particularly grateful to Dr. K. G. Budden, of the University of Cambridge, and to Dr. Andrzej Skorupski, of the University of Warsaw, for equally illuminating discussions and critical remarks. The preparation of this paper in manuscript form was supported in part by the Office of Naval Research, under contract N00014-69-A-0200-4023, and by the Atomic Energy Commission, under contract AT(04-3)-34, Project 157.

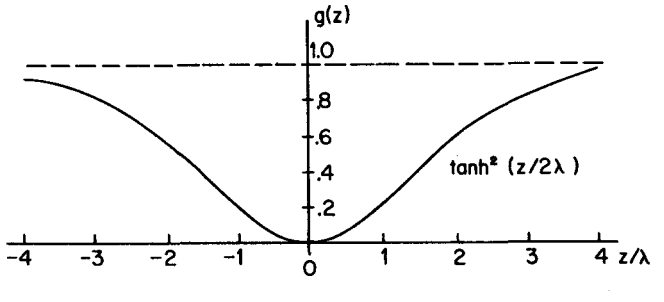

FIG. 7. The Epstein profile for $g(0) = 0$ and $\eta = 0$, illustrating the role of the scale length λ.

APPENDIX A: CONDITIONS UNDERLYING THE APPLICABILITY OF LANGER'S METHOD

In Sec. 2 we showed that the application of Langer's method, in the present instance, leads to the differential equation (2.8), which is intractable as written because of the presence of the complicated term u''/ux'^2, which we alleged was small and slowly varying and, therefore, could altogether be ignored in lowest order approximation, thus leading to the comparison differential equation (2.9), which we know how to solve. We shall now endeavour to justify the omission of the unwanted term.

From (2.3) we see that $u(z) \propto x'^{-1/2}$ whence, quite generally, we have

$$u''/ux'^2 = \tfrac{3}{4}x'^{-4}x''^2 - \tfrac{1}{2}x'^{-3}x''', \tag{A1}$$

which we now proceed to examine asymptotically. From (2.5) we have, for finite η and the fact that $g(z) \to 1$ as $z \to \infty$, the asymptotic limit

$$x' \xrightarrow[x\to\infty]{} k_0/x, \tag{A2}$$

which, when substituted into (A1), leads to the result

$$u''/ux'^2 \xrightarrow[x\to\infty]{} -3/4x^2. \tag{A3}$$

This important result is quite general, that is, independent of η and of the geometric details of $g(z)$ in (1.1), and states simply that, asymptotically, the neglected term in (2.8) behaves as x^{-2} which, side by side with $x^2 + 4\eta$, soon enough becomes completely negligible. The consequence of this result is quite clear: The neglect of the unwanted term u''/ux'^2 in (2.8) can have no effect in our asymptotic results, and hence on the computation of transmission and reflection coefficients.

The behavior of the neglected term (A1) in the vicinity of the origin is quite a different story in the sense that it becomes very cumbersome to obtain quite generally a power series expansion valid in the vicinity of $x = 0$. To simplify the presentation, we confine our attention here to the case $\eta = 0$, a quadratic turning point, and we adopt for $g(z)$ in Fig. 1 the Epstein profile (7.16) with $h = 0$, that is,

$$g(z) = \tanh^2(z/2\lambda), \tag{A4}$$

which we illustrate in Fig. 7 and which forms the basis of our analysis in *Appendix* B. Making use of (2.6), we see that the functional relation $x = x(z)$ now becomes explicit and may be written, for $x \geq 0$, as

$$\tfrac{1}{2}x^2 = k_0 \int_0^z \tanh\left(\frac{s}{2\lambda}\right) ds = 2\lambda k_0 \ln \cosh\left(\frac{z}{2\lambda}\right), \tag{A5}$$

from which we obtain, to lowest order,

$$x^2 = \frac{\pi z^2}{\lambda\lambda_0} + \cdots, \tag{A6}$$

where, it is recalled, λ_0 is the free space wavelength

and λ the scale length of the Epstein profile (A4).

Differentiating both sides of the first equality in (A5) with respect to z yields

$$xx' = k_0 \tanh(z/2\lambda) = Az + Bz^3 + Cz^5 + \cdots, \quad \text{(A7)}$$

where the expansion coefficients are

$$A = \pi/\lambda\lambda_0, \quad B = -\pi/12\lambda_0\lambda^3, \quad C = \pi/120\lambda_0\lambda^5.$$

By successive differentiation with respect to z of both sides of (A7), followed at each step by evaluation at $x = 0$ and $z = 0$, we obtain the derivatives of $x = x(z)$ evaluated at the origin:

$$\begin{aligned} &x'(0) = (\pi/\lambda\lambda_0)^{1/2}, \quad x''(0) = 0, \\ &x''' = -(1/8\lambda^2)(\pi/\lambda\lambda_0)^{1/2}, \quad x^{iv}(0) = 0, \\ &x^{v}(0) = (9/64\lambda^4)(\pi/\lambda\lambda_0)^{1/2}, \end{aligned} \quad \text{(A8)}$$

in which derivation we have made use of the coefficients (A7).

We now return to (A1) and expand each quantity therein contained, making use of (A8), to second order in z. Finally, expressing z^2 in terms of x^2 through (A6), we obtain the two-term expansion for the neglected term (A1) in the form

$$\frac{u''}{ux'^2} = \frac{1}{16\pi}\left(\frac{\lambda_0}{\lambda}\right) - \frac{3}{128\pi^2}\left(\frac{\lambda_0}{\lambda}\right)^2 x^2 + \cdots, \quad \text{(A9)}$$

which is valid for x^2 sufficiently small and which we now proceed to analyze in detail. To simplify the discussion, we rewrite (A9) in the form

$$u''/ux'^2 = \alpha - \beta x^2 + \cdots, \quad \text{(A10)}$$

with coefficients

$$\alpha = (1/16\pi)(\lambda_0/\lambda), \quad \beta = (3/128\pi^2)(\lambda_0/\lambda)^2. \quad \text{(A11)}$$

We notice from (A11) that if we put $\lambda/\lambda_0 = 1$, the numerical coefficients become $\alpha = 2 \times 10^{-2}$ and $\beta = 2.37 \times 10^{-3}$; that, is, with this parameter ratio, we should add (A10), with the numerical coefficients (A11) to x^2 in (2.8) with $\eta = 0$ to obtain, at least in the vicinity of $x = 0$, the differential equation that we really ought to solve. But we see that this addition of the neglected term is tantamount to adding a constant term $4\eta = 2 \times 10^{-2}$ (or $\eta = 0.005$) and modifying the coefficient of x^2 in (2.8), which is unity, to read $1 - 2.37 \times 10^{-3} = 0.997\,63 \simeq 1$. We therefore conclude that a sufficient and necessary criterion for the applicability of Langer's method is to impose the condition

$$\lambda/\lambda_0 \geq 1. \quad \text{(A12)}$$

At the lowest permissible value of the parameter ratio (A12), namely $\lambda/\lambda_0 = 1$, we see from Fig. 7 that this condition means that we should have at least one free space wavelength λ_0 fit within one scale length λ, a condition not difficult to satisfy in the laboratory. For $\lambda/\lambda_0 \gg 1$, say $\lambda/\lambda_0 = 100$, the numerical coefficients in (A11) become $\alpha = 2 \times 10^{-4}$ and $\beta = 2.37 \times 10^{-7}$, which state simply that the unwanted term is altogether negligible. The above analysis, albeit carried only for $\eta = 0$ and for the Epstein contour (A4), gives us confidence that the computation of the actual wave forms, which we undertake in our next paper,[5] can be regarded as essentially exact so long as condition (A12) is satisfied.

APPENDIX B: COMPARISON WITH EXACT RESULTS FROM SYMMETRIC EPSTEIN CONTOURS

To simplify the presentation, once again we confine our attention to the case $\eta = 0$ and we adopt the symmetric Epstein contour (A4). In (7.5) and (7.6) we have already given the transmission and reflection coefficients for $\eta = 0$ and an arbitrary symmetric profile. To complete the computation, in the present instance, we need only compute z_0, making use of (4.5). Thus we obtain

$$z_0 = \int_0^\infty [1 + \tanh(s/2\lambda)]ds = 2\lambda \ln 2, \quad \text{(B1)}$$

which, when inserted into the exponent of (7.5) and (7.6), yields

$$2ik_0z_0 = 4ik_0\lambda \ln 2 = 2i\alpha \ln 2, \quad \text{(B2)}$$

where we define, for convenience,

$$\alpha = 2k_0\lambda = 4\pi(\lambda/\lambda_0), \quad \text{(B3)}$$

noting at once that $\alpha \gg 1$ even when $\lambda/\lambda_0 = 1$, in accordance with (A12). By making use of (B2) and (7.5) and (7.6), the transmission and reflection coefficients become, for the problem at hand,

$$T = (1/\sqrt{2})\,e^{-2i\alpha \ln 2} \quad \text{and} \quad R = -(i/\sqrt{2})\,e^{-2i\alpha \ln 2}. \quad \text{(B4)}$$

It is well known that when the profiles of Fig. 1 can be represented as Epstein profiles, the solution of (1.1) becomes exact and can be represented in terms of hypergeometric functions. It is not our intention here to give a detailed account of the Epstein theory. For brevity we rely here on the extremely readable account of the theory given by Phinney,[4] who shows how to compute the parameters ϵ_1, ϵ_2, and ϵ_3 which characterize the three-parameter family of Epstein contours and who gives the necessary formulas for the transmission and reflection coefficients. Thus, following Phinney, we ascertain that the parameters ϵ_1, ϵ_2, and ϵ_3, which characterize the symmetric Epstein contour (A4), as depicted in Fig. 7, are given by

$$\begin{aligned} \epsilon_1 &= 1 = -(c-1)^2/\alpha^2, \\ \epsilon_2 &= 1 = -(a-b)^2/\alpha^2, \\ \epsilon_3 &= -4 = (a+b-c+1)(a+b-c-1)/\alpha^2, \end{aligned} \quad \text{(B5)}$$

in which a, b, and c are familiar parameters of the hypergeometric functions and in which α is our parameter as defined by (B3). Solving for the parameters a, b, and c from (B5), taking into account $\alpha^2 \gg 1$, we obtain

$$a = \tfrac{1}{2} + 2i\alpha, \quad b = \tfrac{1}{2} + i\alpha, \quad c = 1 + i\alpha. \quad \text{(B6)}$$

Again following Phinney, making use of his equation (42), we have for the transmission coefficient

$$\begin{aligned} T &= \frac{\Gamma(1-a)\Gamma(1+b-c)}{\Gamma(b-a+1)\Gamma(1-c)} = \frac{\Gamma(\frac{1}{2}-2i\alpha)\Gamma(\frac{1}{2})}{\Gamma(1-i\alpha)\Gamma(-i\alpha)} \\ &= \frac{\Gamma(i\alpha)}{\Gamma(-i\alpha)}\,\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}-2i\alpha)}{\Gamma(i\alpha)\Gamma(1-i\alpha)}, \end{aligned} \quad \text{(B7)}$$

into which we have substituted our parameters (B6). Making use of the well-known formula to compute $\Gamma(i\alpha) \times \Gamma(1 - i\alpha)$, and of Stirling's leading asymptotic term to compute the ratio $\Gamma(i\alpha)/\Gamma(-i\alpha)$ as well as the leading term of $\Gamma(\frac{1}{2} - 2i\alpha)$, which appear in the last form of (B7), we obtain finally

$$T = (1/\sqrt{2})\, e^{-2i\alpha \ln 2}, \tag{B8}$$

in complete accord with the first of (B4). The value of R follows at once from (7.2) with $\eta = 0$ and, thus, again in complete agreement with the second of (B4). Hence, we have shown that, provided $\alpha \gg 1$ in (B3), which conforms with our applicability criterion (A12), our results for the transmission and reflection coefficients agree completely with the *exact* results derived from the Epstein theory. This perfect agreement of course was to be expected, since we already showed in Appendix A that our asymptotic results could not possibly be affected by the neglect of the unwanted term in (2.8), and we have also shown that, provided the applicability criterion (A12) is satisfied, the computation of actual wave forms in the vicinity of the origin can also be regarded as *exact*.

---

[1]R. E. Langer, Trans. Am. Math. Soc. **40**, 545 (1934); Phys. Rev. **51**, 669 (1937). For additional references and a most readable account of Langer's method, see D. S. Jones, *Theory of Electromagnetism*, (Pergamon, New York, 1964), Sec. 6.22, p. 354.

[2]A. Erdélyi, W. Magnus, F. Obberhettinger, and F. Tricomi, *Higher Transcendental Functions* (Bateman Manuscript Project) (McGraw-Hill, New York, 1953), Vol. 1, Chap. 6.

[3]E. S. Weibel, "A Note on the Differential Equation $y'' + x^2 y = 0$," UCLA Plasma Physics Group Report R–72 (April, 1970).

[4]R. A. Phinney, Rev. Geophys. Space Phys. **8**, 517 (1970).

[5]Alfredo Baños, Jr. and Daniel L. Kelly, "Electromagnetic Wave Functions for Parabolic Density Profiles," UCLA Plasma Physics Group Report PPG–125 (October, 1972).

[6]P. S. Epstein, Proc. Natl. Acad. Sci. USA **16**, 627 (1930). See also L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960), Sec. 14, p. 172, as well as K. G. Budden, *Radio Waves in the Ionosphere*, (Cambridge U. P., Cambridge, 1966), Sec. 17.4, p. 378. A more recent and readable account of the Epstein theory is given by Phinney in Ref. 4.

Covariant disturbances and exceptional waves

Guy Boillat

Département de Mathématiques, Université de Clermont, Clermont-Ferrand, France
(Received 30 November 1971)

The disturbances of a field are determined for each kind of wave by the hyperbolic field equations. In a space-time formulation one may consider tensors which also depend on the 4-normal to the wavefront. It is shown that the disturbance of such a tensor has a definite covariant form only when the wave is exceptional. For the sake of illustration a perfect fluid is considered. It appears that the sound waves are exceptional not only when the fluid is incompressible but also when a special equation of state is used. This relation between pressure and density corresponds in the nonrelativistic limit to a usual approximation for subsonic flow.

1. INTRODUCTION

Consider a field $\mathbf{u}(x^\alpha)$ (a column vector) that satisfies a hyperbolic quasilinear system of N first-order partial differential equations.[1] Let $\mathbf{u}_0(x^\alpha)$ be a solution of class C^k ($k \geq 1$) that is a solution with continuous derivatives of order k. Then in the neighborhood of the wavefront S: $\varphi(x^\alpha) = 0$, the perturbed state can be written as

$$\mathbf{u} = \mathbf{u}_0 + \varphi\,\delta\mathbf{u} + \mathbf{O}(\varphi^2).$$

Introduce the normal velocity through

$$\lambda = -\varphi_t/|\nabla\varphi|, \qquad \mathbf{n} = \nabla\varphi/|\nabla\varphi|, \qquad \mathbf{n}^2 = \sum n_i^2 = 1, \qquad \alpha = 0, 1, 2, 3, \quad i = 1, 2, 3 \tag{1}$$

($x^0 = t$ is a time variable; the x^i are space variables). The normal speeds $\lambda^{(i)}$, $i = 1, 2, 3, \ldots, \leq N$, are determined by the field equations[2] so that if one put[3]

$$\psi^{(i)}(\mathbf{u}, \varphi_\alpha) = \varphi_t + |\nabla\varphi|\,\lambda^{(i)}(\mathbf{u}, \mathbf{n}), \tag{2}$$

the wave surface propagating with the speed $\lambda = \lambda_0^{(i)}$ into the unperturbed state $\mathbf{u}_0$ satisfies the differential equation

$$\psi^{(i)}(\mathbf{u}_0, \varphi_\alpha) = 0. \tag{3}$$

Now if f is some function continuously differentiable of the field,

$$\delta f(\mathbf{u}, \mathbf{n}) = (\nabla f)_0\,\delta\mathbf{u} = \nabla_0 f_0\,\delta\mathbf{u} \quad (\delta\mathbf{n} = 0). \tag{4}$$

Here ∇ note the gradient with respect to the components of $\mathbf{u}$. Obviously $\delta\mathbf{n} = 0$ since $\varphi \in C^2$ [being solution of (3)]. This formula allows us to compute the disturbances of quantities f from their value on the wavefront. We shall often simply write, for short,

$$\delta f = \nabla f \cdot \delta\mathbf{u}.$$

2. POSITION OF THE PROBLEM: THE PERFECT FLUID

When the field equations are covariant the components of $\mathbf{u}$ are tensors, and one encounters quantities that depend on the 4-gradient φ_α of φ (a tensor). If $T(\mathbf{u})$ is some tensor, then

$$\delta T(\mathbf{u}) = \nabla T \cdot \delta\mathbf{u}. \tag{5}$$

In analogy with (4) take, more generally, a tensor $T(\mathbf{u}, \varphi_\alpha)$; what is δT then? *A priori* it seems that (5) still holds, for, in virtue of the above considerations, φ has continuous second-order derivatives

$$\delta\varphi_\alpha = 0. \tag{6}$$

However, we shall see on an example that this problem deserves a closer look. First we make an important remark.

Remark: It is well known that at a given point (x^α) it is always possible, either in special or in general relativity, to introduce a pseudo-Cartesian frame for which the metric tensor is

$$g_{\alpha\beta} = \text{diag}\,(1, -1, -1, -1).$$

Further, in special relativity, $g_{\alpha\beta}$ is assumed to have continuous derivatives (a property of the curvilinear coordinates which are used) while in general relativity discontinuities of second-order derivatives only occur across a null surface[4]: $\mathcal{G} = g^{\alpha\beta}\varphi_\alpha\varphi_\beta = 0$. Thus, in any case, $\delta g_{\alpha\beta} = 0$ (and also $\delta^2 g_{\alpha\beta} = 0$ whenever $\mathcal{G} \neq 0$) and we can compute the disturbances of T in any frame we like.

As a simple example we consider a perfect fluid[5]:

$$\nabla_\alpha T^{\alpha\beta} = 0, \qquad \nabla_\alpha(ru^\alpha) = 0, \qquad T^{\alpha\beta} = rfu^\alpha u^\beta - pg^{\alpha\beta},$$

where $u^\alpha(u_\alpha u^\alpha = 1)$ is the four-dimensional velocity, r the matter density, S the specific entropy, $p(r, S)$ the pressure, and f the index or relativistic enthalpy of the fluid,

$$df = r^{-1}dp + \theta dS$$

(θ is the temperature). Moreover, $rf = \rho + p$ and ρ is the energy density ($T_{\alpha\beta}u^\alpha u^\beta = \rho$),

$$d\rho = f dr + r\theta dS. \tag{7}$$

The field equations become

$$\nabla_\alpha(ru^\alpha) = 0, \quad rfu^\alpha\nabla_\alpha u^\beta - \gamma^{\alpha\beta}\partial_\alpha p = 0, \quad u^\alpha\partial_\alpha S = 0, \quad \gamma^{\alpha\beta} = g^{\alpha\beta} - u^\alpha u^\beta.$$

Making the replacement

$$\partial_\alpha, \nabla_\alpha \to \varphi_\alpha\delta,$$

we get

$$r\varphi_\alpha\delta u^\alpha + \mathfrak{U}\delta r = 0, \qquad \mathfrak{U}\delta S = 0, \qquad rf\mathfrak{U}\delta u^\beta - \gamma^{\alpha\beta}\varphi_\alpha\delta p = 0. \tag{8}$$

We put

$$\mathfrak{U} = u^\alpha\varphi_\alpha, \qquad \mathcal{G} = g^{\alpha\beta}\varphi_\alpha\varphi_\beta, \qquad \mathfrak{C} = \gamma^{\alpha\beta}\varphi_\alpha\varphi_\beta = \mathcal{G} - \mathfrak{U}^2.$$

As is well known, we have two types of waves:

Copyright © 1973 by the American Institute of Physics

(a) a contact surface (or entropy wave)

$$\mathcal{U} = 0,$$

with the associated perturbations

$$\delta p = 0, \varphi_\alpha \delta u^\alpha = u_\alpha \delta u^\alpha = 0,$$

δS and two components of δu^α being arbitrary;

(b) the sound waves [easily obtained by contraction of (8) by φ_β],

$$\mathcal{U}^2 + p'\mathcal{C} = 0, \tag{9}$$

with [see (7)]

$$p' = \left(\frac{\partial p}{\partial \rho}\right)_S = f^{-1}\left(\frac{\partial p}{\partial \tau}\right)_S$$

and the disturbances

$$\tau f \mathcal{U} \delta u^\beta = \gamma^{\alpha\beta}\varphi_\alpha \delta p, \qquad \delta S = 0, \qquad \delta p = p' \delta \rho. \tag{10}$$

For the sake of clarity we denote now by $\bar{\delta}$ the operator applied to any quantity $T(\mathbf{u}, l_\alpha)$ with the result

$$\bar{\delta} T(\mathbf{u}, l_\alpha) = \nabla T(\mathbf{u}, l_\alpha)_0 \cdot \delta \mathbf{u} \qquad (\bar{\delta} l_\alpha = 0, \qquad \bar{\delta}\mathbf{u} = \delta \mathbf{u}).$$

Consider the invariant

$$T = -\mathcal{S}/\mathcal{U}^2. \tag{11}$$

Taking account of the remark and of (6), (10), we immediately obtain

$$\delta T(\mathbf{u}, \varphi_\alpha) = \bar{\delta} T = 2\mathcal{S}\mathcal{U}^{-3} \varphi_\alpha \delta u^\alpha = (2T/\tau f)\delta \rho. \tag{12}$$

On the other hand, by (9),

$$T = -1 + 1/p', \tag{13}$$

which yields

$$\delta T = -(p''/p'^2)\delta \rho. \tag{14}$$

This result coincides with the above one only if

$$(\rho + p)p'' + 2p'(1 - p') = 0, \tag{15}$$

an equality never satisfied for a real fluid (see below).

This situation is somewhat annoying. How can we, for instance, define the disturbance of the ray velocity since this velocity can be given different expressions all equivalent on the wave front? The next section is devoted to this problem.

3. COVARIANT DISTURBANCES: EXCEPTIONAL WAVES

The characteristic equation of S has the covariant form[1]

$$\psi(\mathbf{u}_0, \varphi_\alpha) = 0, \qquad \psi(\mathbf{u}, \varphi_\alpha) = p^{-1} G^{\alpha_1 \alpha_2 \cdots \alpha_p} \varphi_{\alpha_1} \varphi_{\alpha_2} \cdots \varphi_{\alpha_p}, \qquad \alpha_1, \alpha_2, \ldots, \alpha_p = 0, 1, 2, 3, \tag{16}$$

where $G(\mathbf{u})$ is a completely symmetric tensor. Inserting (1) into (16) yields

$$\psi(\mathbf{u}, \varphi_0, \nabla\varphi) = |\nabla\varphi|^p \psi(\mathbf{u}, -\lambda, \mathbf{n}),$$

and the normal velocities $\lambda^{(i)}$, $i = 1, 2, \ldots, p$, satisfy

$$\psi(\mathbf{u}_0, -\lambda, \mathbf{n}) = 0. \tag{17}$$

Equation (17) becomes an identity when λ is replaced by any root of the polynomial. Therefore, making use of (4), one gets[1]

$$|\nabla\varphi| \frac{\partial \psi}{\partial \varphi_0} \delta\lambda(\mathbf{u}, \mathbf{n}) = \bar{\delta}\psi, \tag{18}$$

where λ (for short) stands for any $\lambda^{(i)}$.

When

$$\delta\lambda(\mathbf{u}, \mathbf{n}) \equiv \nabla\lambda \cdot \delta\mathbf{u} \equiv 0,$$

the corresponding wave is called *exceptional*.[6-8] The covariant form of this criterion is, thus,

$$\bar{\delta}\psi = p^{-1}\varphi_{\alpha_1}\varphi_{\alpha_2} \cdots \varphi_{\alpha_p} \delta G^{\alpha_1 \alpha_2 \cdots \alpha_p} = 0.$$

Let T_1, T_2 be two tensors equal on S. Since one derives from the other by means of (16), we can write

$$T_1(\mathbf{u}, \varphi_\alpha) = T(\mathbf{u}, \varphi_\alpha; \psi), \qquad T_2(\mathbf{u}, \varphi_\alpha) = T(\mathbf{u}, \varphi_\alpha; 0).$$

For instance, (11) and (13) are related by

$$-\mathcal{S}/\mathcal{U}^2 = -1 + 1/p' - 2\psi/(p'\mathcal{U}^2), \quad \psi = \tfrac{1}{2}(\mathcal{U}^2 + p'\mathcal{C}). \tag{19}$$

Further $(\partial T/\partial\psi)(\mathbf{u}_0, \varphi_\alpha; 0)$ exists if we assume that T_1, T_2 are continuously differentiable with respect to φ_α,

$$\frac{\partial T_1}{\partial \varphi_\alpha} = \frac{\partial T_2}{\partial \varphi_\alpha} + \frac{\partial T}{\partial \psi}(\mathbf{u}_0, \varphi_\alpha; 0) \frac{\partial \psi_0}{\partial \varphi_\alpha}. \tag{20}$$

Consequently, we have

$$\bar{\delta} T_1 = \bar{\delta} T_2 + \frac{\partial T}{\partial \psi}(\mathbf{u}_0, \varphi_\alpha; 0)\bar{\delta}\psi \tag{21}$$

and $\bar{\delta} T_1 \neq \bar{\delta} T_2$ (though $T_1 = T_2$ on S) unless the wave is exceptional. [For the sound wave of the preceding section this will be the case if (15) is true.]

In the same way we distinguished $\mathbf{u}$ and $\mathbf{u}_0$ we introduce the 4-vector l_α such that

$$l_\alpha = \varphi_\alpha$$

on S. (In contradistinction to φ_α, l_α may have discontinuities δl_α.)

The problem is thus: Given $T(\mathbf{u}, l_\alpha)$, is it possible to determine (in a unique way) its disturbance? To this aim we first note that (because the equation of the wavefront $\varphi = 0$ is invariant through the change $\varphi \to k\varphi$, $k = \text{const}$) all meaningful quantities are homogeneous functions of the φ_α. The degree of homogeneity can always be assumed (multiplying by some power of $-\mathcal{S}$ or $-\mathcal{C}$) to be zero.[9] Hence,

$$T(\mathbf{u}, l_\alpha)_0 = T(\mathbf{u}_0, \varphi_\alpha) = T(\mathbf{u}_0, \varphi_0, \nabla\varphi) = T(\mathbf{u}_0, -\lambda(\mathbf{u}_0, \mathbf{n}), \mathbf{n}) = T(\mathbf{u}_0, \mathbf{n}),$$

where λ is a root of (17). Then we can apply (4) to get

$$\delta T(\mathbf{u}, l_\alpha) = \delta T(\mathbf{u}, \mathbf{n}) = -|\nabla\varphi| \frac{\partial T}{\partial \varphi_0} \delta\lambda + \bar{\delta} T(\mathbf{u}, \varphi_\alpha),$$

or, by virtue of (18),

$$\frac{\partial \psi_0}{\partial \varphi_0}(\delta T - \bar{\delta} T) + \frac{\partial T_0}{\partial \varphi_0}\bar{\delta}\psi = 0. \tag{22}$$

If we consider this formula to be still valid for any degree of homogeneity, we obtain, taking $T(\mathbf{u}, l_\alpha) = \psi(\mathbf{u}, l_\alpha)$,

$$\delta\psi(\mathbf{u}, l_\alpha) = 0$$

and, $T = l_\alpha$,

$$(\partial\psi/\partial\varphi_0)\delta l_\alpha + \delta^0_\alpha \bar{\delta}\psi = 0 \Rightarrow \delta l_i = 0.$$

Clearly from the very way it has been obtained δT is perfectly determined. With this formula two tensors equal on S ($T_1 = T_2$) have the same disturbances ($\delta T_1 = \delta T_2$), a result which can also easily be checked by (20), (21). In particular the perturbation of (11) is given by (14),

$$\delta(\mathcal{G}/\mathcal{U}^2) = (p''/p'^2)\delta\rho.$$

where it is understood that $\mathcal{U} = \mathcal{U}(l) = u^\alpha l_\alpha$; $\mathcal{U}(l) = \mathcal{U}(\varphi)$ on S but—in general—$\delta\mathcal{U}(l) \neq \delta\mathcal{U}(\varphi) = \bar{\delta}\mathcal{U}(\varphi) = \varphi_\alpha \delta u^\alpha$ and similarly for $\mathcal{G}$.

However it is obvious that $\delta T(\mathbf{u}, l_\alpha)$ so defined is not a tensor under general space–time transformations unless (evidently) T does not depend on S on the φ_α [as (11) equals to (13)] or unless the wave is exceptional. In this important case,

$$\delta T = \bar{\delta} T, \qquad \bar{\delta}\psi = 0.$$

When applied to the ray velocity[1],

$$v^\alpha = \frac{1}{\mathfrak{N}}\frac{\partial \psi}{\partial \varphi_\alpha} \qquad (v_\alpha v^\alpha = g_{\alpha\beta}v^\alpha v^\beta = 1), \tag{23}$$

(22) yields[10]

$$\delta v^\alpha = -\frac{1}{\mathfrak{N}}\left(\frac{\partial^2\psi}{\partial\varphi_0\partial\varphi_\beta}\Big/\frac{\partial\psi}{\partial\varphi_0}\right)(\delta^\alpha_\beta - v^\alpha v_\beta)\bar{\delta}\psi + \bar{\delta}v^\alpha,$$
$$\bar{\delta}v^\alpha = \mathfrak{N}^{-1}(\delta^\alpha_\beta - v^\alpha v_\beta)\varphi_{\alpha_2}\varphi_{\alpha_3}\cdots\varphi_{\alpha_p}\delta G^{\beta\alpha_2\alpha_3\cdots\alpha_p}. \tag{24}$$

Although (24) has no tensorial character, it is not difficult to show that $\varphi_\alpha \delta v^\alpha$ indeed is an invariant. Differentiating the Euler identity,

$$\varphi_\alpha \frac{\partial\psi}{\partial\varphi_\alpha} = p\psi,$$

gives

$$\varphi_\alpha \frac{\partial^2\psi}{\partial\varphi_0\partial\varphi_\alpha}\Big/\frac{\partial\psi}{\partial\varphi_0} = p - 1.$$

On the other hand,

$$\varphi_\alpha \bar{\delta}v^\alpha = \bar{\delta}(\varphi_\alpha v^\alpha) = (p/\mathfrak{N})\bar{\delta}\psi,$$

so that

$$\varphi_\alpha \delta v^\alpha = \mathfrak{N}^{-1}\bar{\delta}\psi. \tag{25}$$

When the wave is exceptional, the disturbance of the ray velocity is tangential. The same is true in three-dimensional formulation for

$$\mathbf{n}\cdot\delta\Lambda^{(i)} = \delta(\Lambda^{(i)}\cdot\mathbf{n}) = \delta\lambda^{(i)},$$

where $\Lambda^{(i)}$ is the ray velocity of components $\Lambda^{(i),j} = \partial\psi^{(i)}/\partial\varphi_j$, i.e. (see eq. 3)[3],

$$\Lambda^{(i)}(\mathbf{u}, \mathbf{n}) = \lambda^{(i)}\mathbf{n} + \frac{\partial\lambda^{(i)}}{\partial\mathbf{n}} - \left(\mathbf{n}\cdot\frac{\partial\lambda^{(i)}}{\partial\mathbf{n}}\right)\mathbf{n}. \tag{26}$$

Let us transcribe (25) with these notations. Consider the wave propagating with the speed $\lambda^{(i)}$ [a root of (17)]. Then (16) can be written

$$\psi = \omega\psi^{(i)}, \qquad \omega = p^{-1}G^{00\cdots0}\prod_{(j)\neq(i)}\psi^{(j)}. \tag{27}$$

By (23) and (26) one has, on S,

$$v^0 = \omega\mathfrak{N}^{-1}, \qquad v^j = v^0\Lambda^{(i),j} \qquad (\psi^{(i)} = 0),$$

and the first member of (25) becomes

$$|\nabla\varphi|\{-\lambda^{(i)}\delta v^0 + \mathbf{n}\cdot\delta(v^0\Lambda^{(i)})\} = |\nabla\varphi|v^0\delta\lambda^{(i)},$$

which is identical to the right member [see (27) or (18)].

As a consequency the jump (across S) of the velocity divergence vanishes for an exceptional wave,

$$[\nabla_\alpha v^\alpha] = 0, \qquad [\mathrm{div}\Lambda] = 0,$$

and a similar result holds for a simple wave.[11] Also, if we have some (generalized) conservation law,

$$\nabla_\alpha T^{\alpha\beta\cdots}(\mathbf{u}) = Q^{\alpha\beta\cdots}(\mathbf{u}),$$

where Q is continuous, then

$$\varphi_\alpha \delta T^{\alpha\beta\cdots} = 0$$

and in the exceptional case (where $\delta = \bar{\delta}$) the quantity $l_\alpha T^{\alpha\beta\cdots}$ is not disturbed,

$$\delta(l_\alpha T^{\alpha\beta\cdots}) = 0, \qquad \bar{\delta}\psi = 0.$$

4. INCOMPRESSIBLE FLUID: EQUATIONS OF STATE

Here we shall only be concerned with the sound waves since the contact surface $u^\alpha\varphi_\alpha = 0$ is always exceptional: $\varphi_\alpha\delta u^\alpha = 0$.[12] We have seen that (12) and (14) are equal only if (15) holds. From (21) we may therefore conclude that the sound waves are exceptional provided that

$$(\rho + p)p'' + 2p'(1 - p') = 0. \tag{28}$$

This equation has already been found.[13] We have [see (9), (20)]

$$\frac{\partial\psi}{\partial\varphi_\alpha} = \mathcal{U}(1 - p')u^\alpha + p'g^{\alpha\beta}\varphi_\beta, \qquad \mathfrak{N}^2 = \mathcal{U}^2(1 - p') \geq 0$$

(the relativistic limitation on magnitude of velocities implies that this vector must not be spacelike) so that[14]

$$0 < p' \leq 1.$$

The pressure of a real fluid satisfies

$$p'' \geq 0, \tag{29}$$

and the only solution of (28) compatible with these conditions is

$$p' = 1, \qquad p = \rho + \mathrm{funct}(S).$$

This is the equation of state of the incompressible relativistic perfect fluid.[13,15] Then the ray velocity is a null vector colinear to the normal of the wavefront.

However, if we drop (29), the general solution of (28) reads

$$p = b - a^2/(\rho + b), \qquad a = a(S), \qquad b = b(S). \tag{30}$$

In this case $p' \neq 1$ and the normalized velocity can be written

$$v_\alpha = (1 - p')^{1/2} u_\alpha \pm \sqrt{p'}\,\varphi_\alpha/\sqrt{-\mathcal{G}}.$$

Hence,

$$\delta v_\alpha = u_\alpha \delta(1 - p')^{1/2} + (1 - p')^{1/2}\delta u_\alpha \pm (\varphi_\alpha/\sqrt{-\mathcal{G}})\delta\sqrt{p'}.$$

Taking account of (10) and

$$\mathfrak{U}(1 - p')^{1/2} = \pm\,(-p'\mathcal{G})^{1/2},$$

results in

$$\delta v_\alpha = 0.$$

The ray velocity is not disturbed.[16]

It is not difficult to see that for a nonrelativistic fluid all the waves are exceptional if[17]

$$p = b - a^2/r. \tag{31}$$

If we reintroduce the velocity of light (that had been chosen as unity), (30) takes the form

$$p = b - a^2/(\rho + bc^{-2}), \tag{32}$$

since ρ has the same dimension as pc^{-2}, i.e., also bc^{-2}. In the limit of an infinite light velocity, ρ goes over to r and (32) to (31).

Frequently, in particular in the theory of subsonic flow, the adiabatic equation $p = p(r)$ is approximated by (31). Aside from an additive constant, this relation corresponds to a polytropic gas with $\gamma = -1$.[18] It does not seem that (32) has ever been used in relativity.

5. CONCLUSION

Given a tensor $T(\mathbf{u}, \varphi_\alpha)$ on S, we have seen that its disturbance can be defined in different ways:

(a) $\delta T(\mathbf{u}, \varphi_\alpha) = \bar{\delta} T(\mathbf{u}, \varphi_\alpha)$, a tensor, with the counterpart that, generally, $T_1 = T_2 \not\Rightarrow \bar{\delta}T_1 = \bar{\delta}T_2$;

(b) By $\delta T(\mathbf{u}, l_\alpha)$ [Eq. (22)]. Though no longer a tensor this expression has the great advantage to coincide with the result obtained in a pseudo-Cartesian frame when a three-dimensional formalism is used. Moreover, it solves the previous indeterminacy. From Sec. 3 one has

$$\delta T_1(\mathbf{u}, l_\alpha) = \delta T_2(\mathbf{u}, l_\alpha) + \frac{\partial T_0}{\partial \psi}\,\delta\psi(\mathbf{u}, l_\alpha)$$

and it suffices to choose l_α such that,

$$\delta\psi(\mathbf{u}, l_\alpha) = \bar{\delta}\psi + \frac{\partial \psi}{\partial \varphi_\alpha}\,\delta l_\alpha = 0, \tag{33}$$

i.e.,

$$v^\alpha \delta l_\alpha + \mathfrak{N}^{-1}\bar{\delta}\psi = 0. \tag{34}$$

In passing, remark that (34) is nothing but (25) since (33) and $v^\alpha l_\alpha = (p/\mathfrak{N})\psi$ imply $l_\alpha \delta v^\alpha + v^\alpha \delta l_\alpha = 0$.

When the wave is exceptional the definitions (a) and (b) are one and the same thing and we may conclude that the disturbance of a tensor $T(\mathbf{u}, \varphi_\alpha)$ which depends on the 4-normal to the wavefront has a definite covariant form only when the wave is exceptional.

---

[1] G. Boillat, J. Math. Phys. **10**, 452 (1969).
[2] They are eigenvalues of A_n (cf. Ref. 1) while $\delta\mathbf{u}$ are eigenvectors.
[3] G. Boillat, *La propagation des ondes* (Gauthier-Villars, Paris, 1965).
[4] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955). Incidentally, this is also true for the discontinuities of the first-order derivatives (shocks); cf. A. Lichnerowicz, Compt. Rend. **273A**, 528 (1971); G. Boillat, Compt. Rend. **274A**, 1018 (1972); **275A**, 1255 (1972).
[5] See for instance, *Relativistic Fluid Dynamics*, 1st C.I.M.E. session 1970, edited C. Cattaneo (Cremonese, Rome, 1971).
[6] P. D. Lax, Ann. Math. Studies **33**, 211 (1954); Commun. Pure Appl. Math. **10**, 537 (1957).
[7] A. Jeffrey and T. Taniuti, *Non-linear Wave Propagation* (Academic, New York, 1964).
[8] G. Boillat, Ref. 3; Compt. Rend. **270A**, 217, 1134 (1970).
[9] $\mathcal{G} \leqslant 0$, $\mathcal{C} < 0$: in a pseudo-Cartesian frame $\mathcal{G} = |\nabla\varphi|^2(\lambda^2 - 1) \leqslant 0$ since the normal velocity must not exceed the speed of light (taken as unity); besides if this frame is a rest frame ($u^0 = 1$, $u^i = 0$), $\mathcal{C} = -|\nabla\varphi|^2 < 0$.
[10] The formula $\bar{\delta}v^\alpha = \delta v^\alpha$ we gave in Ref. 1 is valid for an exceptional wave.
[11] G. Boillat, J. Math. Phys. **11**, 1482 (1970).
[12] G. Boillat, J. Math. Phys. **11**, 941 (1970) Sec. 1.
[13] Y. Choquet-Bruhat, J. Math. Pures Appl. **48**, 117 (1969). See p. 149.
[14] From the thermodynamics it is known that $p' < 0$. Note that $v_\alpha v^\alpha \geqslant 0$ and $v^\alpha\varphi_\alpha = 0 \Rightarrow \mathcal{G} \leqslant 0$, $\mathcal{C} < 0$; cf. Ref. 9. In a pseudo-Cartesian rest frame $\lambda = \pm\sqrt{p'}$.
[15] G. Boillat, "Sur la propagation de la chaleur en relativité," in Ref. 5. The incompressible fluid can be considered polytropic with a constant $\gamma = C_p/C_v = 2$.
[16] A similar result holds in nonlinear electrodynamics enabling us to show that whenever an exceptional Lagrangian (e.g., Born-Infeld's) is used only the magnetic field is perturbed in the rest frame. Cf. Ref. 12, Sec. 2.I.
[17] Write $\delta\lambda = 0$. In the case of magnetohydrodynamics this gives[3]
$r\,\frac{\partial^2 p}{\partial r^2} + \frac{5\partial p}{\partial r} = 0$, i.e., $p = -a^2(S)r^{-4} + b(S)$.
[18] We quote: R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience, New York, 1948), footnote on p. 10.

Group structure of the relativistic equivalent oscillator*

Eugene F. Chaffin

Department of Physics, Oklahoma State University, Stillwater, Oklahoma 74074
(Received 20 September 1971)

The Dirac Hamiltonian with an equivalent oscillator potential introduced by Swamy is discussed from the point of view of group theory. It is shown to possess an $SO(4) \times SU(2)$ invariance group and an $SO(4,1) \times SU(2)$ noninvariance group.

1. INTRODUCTION

There has been a great deal of interest in recent years in the symmetries and dynamical groups of simple quantum mechanical Hamiltonians which possess exact solutions.[1] Study of these happens to be useful in the context of symmetries occurring in elementary particle physics. Several papers have appeared on the invariance and noninvariance groups of the nonrelativistic hydrogen atom, the approximately relativistic Symmetric Coulomb Hamiltonian,[2] the exact Dirac Coulomb problem, and the nonrelativistic harmonic oscillator.[3-5] Of particular interest is the group $SO(4,1)$. It has been shown that the $V_{0,\sigma}$ and the $V_{1/2,\sigma}$ representations of the continuous class of irreducible representations of this group are realized by the states of the nonrelativistic hydrogen atom and the exact Dirac Coulomb problem. It is the purpose of this paper to point out that the $\pi^{+}_{1/2,1/2}$ representation of the discrete class can be realized by the solutions of the relativistic equivalent oscillator Hamiltonian proposed by Swamy.[6] This latter Dirac Hamiltonian possesses exact solutions and reduces in the nonrelativistic limit to an isotropic harmonic oscillator with spin-orbit coupling. It is, therefore, likely to provide an interesting alternative to the hydrogen atom for the study of symmetry groups. With this motivation, it is shown in this paper that this Hamiltonian belongs to the $SO(4,1) \times SU(2)$ dynamical symmetry group and that the solutions with fixed energy realize irreducible representations of the group $SO(4) \times SU(2)$.

2. INVARIANCE GROUP

In the notation of Ref. 6 and with the help of the "oscillator helicity operator"[7] $\sigma\cdot\mathbf{b}$, the Hamiltonian can be written as

$$H = \rho_1 \sigma\cdot\mathbf{b} + \rho_3 m, \tag{1}$$

where

$$\sigma\cdot\mathbf{b} = \sigma\cdot p + i\lambda^2 \sigma\cdot\mathbf{r}(\sigma\cdot\mathbf{L} + 1)/|\sigma\cdot\mathbf{L} + 1|. \tag{2}$$

The normalized solutions are

$$\Phi_{v\kappa M} = \begin{bmatrix} \sqrt{(E + m)}/2E\,|v\kappa\mu\rangle \\ \\ S(\kappa)\sqrt{(E - m)/2E}\,|v - \kappa\mu\rangle \end{bmatrix}, \tag{3}$$

where

$$E = \sqrt{m^2 + 4\lambda^2(v + |K| + \tfrac{1}{2})}, \tag{4}$$

$$|v\kappa\mu\rangle = F_{vl}\chi^{\mu}_{\kappa}, \qquad |v - \kappa\mu\rangle = iF_{vl(\kappa)}\chi^{\mu}_{-\kappa} \tag{5}$$

and F_{vl} is the usual radial wavefunction for the nonrelativistic harmonic oscillator.

The number of states possessing a given energy is easily shown to be

$$d = 2(n + 1)(n + 2), \qquad \text{where } n = 0, 1, 2, 3, \cdots. \tag{6}$$

By comparing the above with the well-known formulas for the dimensionalities of the irreducible representations of $SO(4)$ and $SU(3)$, one is led to suspect that the group of this Hamiltonian is $SO(4) \times SU(2)$ or $SU(3) \times SU(2) \times SU(2)$. The irreducible representations of the $SU(2)$ groups occurring in the direct products are restricted to the two-dimensional ones. We will now show that the first of these two is applicable.

The angular momentum operator

$$\mathbf{J} = \mathbf{L} + \tfrac{1}{2}\sigma \tag{7}$$

commutes with the Hamiltonian. There is thus an $SU(2)$ subgroup which accounts for $2j + 1$ degenerate states.

In addition, an operator $\sigma\cdot\mathbf{b}$ analogous to the "helicity" operator of the free particle Dirac equation commutes with the Hamiltonian. This operator generates a Lie algebra isomorphic to that of the Pauli spin matrices [hence to the Lie algebra of $SU(2)$]. If we consider

$$X_1 = (\sigma\cdot\mathbf{b})/\sqrt{H^2 - m^2}, \qquad X_2 = i\rho_3 X_k \underset{\sim}{\kappa}/|\underset{\sim}{\kappa}|,$$
$$X_3 = \rho_3 \underset{\sim}{\kappa}/|\underset{\sim}{\kappa}|, \tag{8}$$

where $\underset{\sim}{\kappa} = \sigma\cdot\mathbf{L} + 1$, these satisfy

$$[X_k, X_1] = 2i\epsilon_{klm}X_m, \qquad \mathbf{X}\cdot\mathbf{X} = 3. \tag{9}$$

The operator X_1 transforms $\Phi_{v\kappa\mu}$ into $\Phi_{v-\kappa\mu}$. This accounts for the degeneracy with respect to the sign of κ.

If we now consider all the states of the invariance group, we need ladder operators which will change v and κ, but keep the energy fixed. From the Lie algebra of the generators of the $SO(4)$ group,

$$\begin{aligned} [J_k, J_l] &= i\epsilon_{klm}J_m, \\ [J_k, M_l] &= i\epsilon_{klm}M_m, \\ [M_k, M_l] &= i\epsilon_{klm}J_m, \end{aligned} \tag{10}$$

we know that the operator which we seek should be a vector operator. The operator

$$\sigma \times \mathbf{L}$$

is a pseudovector operator which changes the value of kappa by one. However, it only changes the angular part of the wavefunction, and needs to be multiplied by an operator that changes the radial part. With the help of the recurrence relations

$$\left(\frac{d}{dr} + \frac{l + 1}{r} - \lambda^2 r\right)F_{vl} = 2\lambda\sqrt{v + 1}\,F_{v+1,l-1}, \tag{11}$$

$$\left(\frac{d}{dr} + \frac{l}{r} - \lambda^2 r\right)F_{vl} = -2\lambda\sqrt{v}\,F_{v-1,l+1}, \tag{12}$$

Copyright © 1973 by the American Institute of Physics

we construct the operators

$$\Omega^+ = (\sigma \times \mathbf{L})\sigma\cdot\mathbf{a}, \quad \Omega = \sigma\cdot\mathbf{a}(\sigma \times \mathbf{L}), \tag{13}$$

where

$$\sigma\cdot\mathbf{a} = \sigma\cdot\mathbf{p} - i\lambda^2\,\sigma\cdot\mathbf{r}(\sigma\cdot\mathbf{L} + 1)/|\sigma\cdot\mathbf{L} + 1|. \tag{14}$$

These operators change the value of kappa by one, but do not form a closed Lie algebra. Making them Hermitian does not help. The difficulty can be traced to the fact that they do not have the same matrix elements as the **M** operators of the $SO(4)$ group. In the basis of the physical angular momentum $|j\mu\rangle$, we have[8]

$$\begin{aligned} M_\pm|j\mu\rangle = {} & \pm\sqrt{(j\mp\mu)(j\mp\mu-1)}\,C_j|j-1,\mu\pm1\rangle \\ & -\sqrt{(j\mp\mu)(j\pm\mu-1)}A_j|j,\mu\pm1\rangle \\ & \pm\sqrt{(j\pm\mu)(j\pm\mu+2)}C_{j+1}|j+1,\mu\pm1\rangle, \end{aligned} \tag{15}$$

$$\begin{aligned} M_z|j,\mu\rangle = {} & \sqrt{(j+\mu)(j-\mu)}\,C_j|j-1,\mu\rangle - \mu A_j|j,\mu\rangle \\ & +\sqrt{(j+\mu+1)(j-\mu+1)}\,C_{j+1}|j+1,\mu\rangle, \end{aligned} \tag{16}$$

where

$$A_j = -\frac{j_0 j_1}{j(j+1)}, \quad C_j = \frac{i}{j}\sqrt{\frac{(j^2-j_0^2)(j_1^2-j^2)}{4j^2-1}},$$

$$M_\pm = M_x \pm iM_y. \tag{17}$$

Here j_0 is the lowest value of j within an irreducible representation, and j_1 is one plus the highest value of j.

By comparing these matrix elements with those of the omega operators, it is possible to see how to construct the operator **M**. For kappa greater than zero, we get

$$\begin{aligned} \mathbf{M} = {} & \frac{-i\sqrt{[(H^2-m^2)/4\lambda^2]+|\kappa|-\frac{1}{2}}}{2\lambda(2|\kappa|-1)}\,\Omega \\ & + i\,\Omega^+\frac{\sqrt{[(H^2-m^2)/4\lambda^2]+|\kappa|-\frac{1}{2}}}{2\lambda(2|\kappa|-1)} \\ & - \tfrac{1}{2}\frac{H^2-m^2}{4\lambda^2(|\kappa|-\frac{1}{2})(|\kappa|+\frac{1}{2})}\,\mathbf{J}. \end{aligned} \tag{18}$$

The operator for the other sign of kappa can be similarly obtained and with the help of the projection operators

$$\tfrac{1}{2}(1+\kappa/|\kappa|) \quad \text{and} \quad \tfrac{1}{2}(1-\kappa/|\kappa|), \tag{19}$$

we get **M** in the general form

$$\begin{aligned} \mathbf{M} = {} & -i\frac{\sqrt{[(H^2-m^2)/4\lambda^2]+|\kappa|-\frac{1}{2}}}{2\lambda(2|\kappa|-1)}^{-1/2} \\ & \times[\Omega+\Omega^+)+(\kappa/|\kappa|)(\Omega^+-\Omega)] \\ & + i\tfrac{1}{2}[\Omega-\Omega^+)+(\kappa/|\kappa|)(\Omega-\Omega^+)] \\ & \times\frac{\sqrt{[(H^2-m^2)/4\lambda^2]+\kappa-\frac{1}{2}}}{2\lambda(2|\kappa|-1)} \\ & - \tfrac{1}{2}\frac{H^2-m^2}{4\lambda^2(|\kappa|-\frac{1}{2})(|\kappa|-\frac{1}{2})}\,\mathbf{J}. \end{aligned} \tag{20}$$

It is a straightforward task to verify

$$\begin{aligned} & [M_k, M_l] = i\epsilon_{klm}J_m, \\ & [J_k, M_l] = i\epsilon_{klm}M_m, \\ & [X_k, M_l] = 0, \end{aligned} \tag{21}$$

and **M** is Hermitian. We notice thus that the Lie Algebra of $SO(4) \times SU(2)$ is realized and between the invariants **M**, **J**, and **X** the degeneracy is exhausted.

3. NONINVARIANCE GROUP

It now remains to establish the noninvariance group for this problem. Similar problems have been solved by Malkin and Man'ko[4] who discussed the symmetric Coulomb problem of Biedenharn and Swamy and by Kiefer and Fradkin,[5] who treated the exact Dirac Coulomb problem. The approach of the latter is applicable here.

We will first show that the states with fixed sign of κ realize an irreducible representation of the group $SO(4,1)$ and then give the noninvariance operators appropriate to the Lie algebra of this group. Strom[9] has given the irreducible representations of $SO(4,1)$ in the basis applicable here. The states with fixed sign of kappa form a basis for the infinite-dimensional irreducible representation $\pi^+_{1/2,1/2}$, in the notation of Strom. This representation belongs to the discrete class, contrary to the cases discussed by Fradkin Kiefer. To determine which irreducible representation the solutions of this Hamiltonian belong to, we proceed as follows. The Casimir operators of the $SO(4)$ group are

$$\begin{aligned} & C(1) = \tfrac{1}{2}(\mathbf{M}\cdot\mathbf{M} + \mathbf{J}\cdot\mathbf{J}) = \tfrac{1}{2}\{[(H^2-m^2)^2/16\lambda^4]-\tfrac{3}{4}]\}, \\ & C(2) = \mathbf{J}\cdot\mathbf{M} = \tfrac{1}{2}[(H^2-m^2)/4\lambda^2]. \end{aligned} \tag{22}$$

In the appropriate basis we get the eigenvalues of the above invariant operators as

$$\begin{aligned} & C(1) = \tfrac{1}{2}[(v+|\kappa|+\tfrac{1}{2})^2-\tfrac{3}{4}], \\ & C(2) = \tfrac{1}{2}(v+|\kappa|+\tfrac{1}{2}). \end{aligned} \tag{23}$$

Following Pauli,[8] we have

$$\begin{aligned} & K^2 = \tfrac{1}{2}[C(1)+C(2)] = j_2(j_2+1), \\ & L^2 = \tfrac{1}{2}[C(1)-C(2)] = j_1(j_1+1). \end{aligned} \tag{24}$$

Hence, we find that

$$j_1 = (v+|\kappa|-1)/2, \quad j_2 = (v+|\kappa|)/2 \tag{25}$$

and the dimensionalities of the $SO(4)$ irreducible representations realized are

$$\begin{aligned} & d' = (2j_1+1)(2j_2+1) = (n+1)(n+2), \\ & \text{where } n = v+|\kappa|-1, \quad n = 0,1,2,3,\cdots. \end{aligned} \tag{26}$$

Comparing Eqs. (6) and (26) it is clear that the doubling of states in d is due to the degeneracy in the sign of kappa. The equivalence of this $(j_1 j_2)$ parametrization to the irreducible representation $\pi^+_{r,q}$ of $SO(4,1)$ of Strom is seen as follows. According to Strom, we have

$$\begin{aligned} & r = \min(j_1+j_2), \quad q = r, r-1, \cdots, \tfrac{1}{2} \\ & n = j_2 - j_1, \quad l = j_1 + j_2 + 1. \end{aligned} \tag{27}$$

We, therefore, get

$$r = n = q = \tfrac{1}{2},$$

$$l = v + |\kappa| + \tfrac{1}{2}. \tag{28}$$

The allowed values of j_1 and j_2 are shown in Fig. 1. This shows that the states (with sign of κ fixed) span an irreducible representation of $SO(4,1)$.

We now proceed with the explicit construction of the relevant operators. In addition to the $SO(4)$ generators, there exist also a scalar operation T and a vector operation $\mathbf{N}$. The Lie algebra of $SO(4,1)$ is

$$\begin{aligned}
&[J_k, J_l] = i\epsilon_{klm}J_m, \qquad [J_k, T] = 0\\
&[J_k, N_l] = i\epsilon_{klm}N_m, \qquad [J_k, M_l] = i\epsilon_{klm}M_m,\\
&[M_k, M_l] = i\epsilon_{klm}J_m, \qquad [N_k, N_l] = -i\epsilon_{klm}J_m,\\
&[N_k, M_l] = i\delta_{kl}T, \qquad [N_k, T] = iM_k,\\
&[M_k, T] = iN_k,
\end{aligned} \tag{29}$$

For the irreducible representations of interest, the scalar operator T has the matrix elements

$$T|v,j,\mu\rangle = \tfrac{1}{2}\sqrt{(v+2j+2)(v+1)}\,|v+1,j,\mu\rangle + \tfrac{1}{2}\sqrt{(v+2j+1)v}\,|v-1,j,\mu\rangle.$$

This operator is represented for positive kappa by

$$T = \frac{1}{4\lambda}\left(\sqrt{\frac{H^2-m^2}{4\lambda^2}} + |\kappa| - \tfrac{1}{2}\,\frac{\sigma\cdot\mathbf{b}}{\sqrt{H^2-m^2}}\,\sigma\cdot\mathbf{a} + \sigma\cdot\mathbf{a}\,\frac{\sigma\cdot\mathbf{b}}{\sqrt{H^2-m^2}}\,\sqrt{\frac{H^2-m^2}{4\lambda^2}} + |\kappa| - \tfrac{1}{2}\right). \tag{30}$$

The operator for the other sign of kappa is given by interchanging the operators $\sigma\cdot\mathbf{b}/(H^2-m^2)$ and $\sigma\cdot\mathbf{a}$. The projection operators

$$\tfrac{1}{2}(1 + \kappa/|\kappa|) \quad \text{and} \quad \tfrac{1}{2}(1 - \kappa/|\kappa|), \tag{19}$$

then enable us to construct T in general. The vector operator is then constructed by means of the relation

$$\mathbf{N} = -i[\mathbf{M}, T]. \tag{31}$$

This completes the set of $SO(4,1)$ operators. The $SU(2)$ operators $\mathbf{X}$ in Eq. (8) are easily seen to commute with T, and hence with $\mathbf{N}$. Hence all the operators of the $SO(4,1) \times SU(2)$ group have been realized.

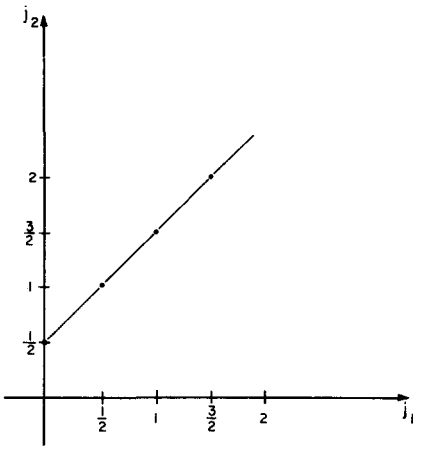

FIG. 1. The allowed values of j_1 and j_2 within the irreducible representation $\pi^+_{\frac{1}{2},\frac{1}{2}}$ of $SO(4,1)$.

ACKNOWLEDGMENTS

I would like to thank Dr. N. V. V. J. Swamy for suggesting the problem and for patient guidance during the work. I am grateful to the National Science Foundation for a fellowship which enabled me to complete this study.

*Work supported by the Research Foundation of Oklahoma State University.

[1] I. A. Malkin and V. I. Man'ko, Yad. Fiz. **3**, 372 (1966) [Sov. J. Nucl. Phys. **3**, 267 (1966)]; A. Bohm, Nuovo Cimento **43**, 667 (1966); N. Mukunda, L. O'Raiferteaigh, and E. C. G. Sudarshan, Phys. Rev. Lett. **15**, 1041 (1965).
[2] L. C. Biedenharn, Bull. Am. Math. Soc. **7**, 364 (1962); L. C. Biedenharn and N. V. V. J. Swamy, Phys. Rev. **133**, B1353 (1964).
[3] I. A. Malkin and V. I. Man'ko, P. N. Lebedev Physical Institute, Moscow, USSR, Report No. N 9, 1968 (unpublished); Zh. Eksp. Teor. Fiz. **55**, 287 (1968) [Sov. Phys.-JETP **28**, 151 (1969)].
[4] D. M. Fradkin and H. M. Kiefer, Phys. Rev. **180**, 1282 (1969).
[5] R. H. Pratt and T. H. Jordan, Phys. Rev. **148**, 1276 (1966); R. Musto, Phys. Rev. **148**, 1274 (1966); H. Bacry, *Lectures in Theoretical Physics* (Colorado U. P., Boulder, Colorado, 1966), Vol. 9A.
[6] N. V. V. J. Swamy, Phys. Rev. **180**, 1225 (1969).
[7] J. Braun and N. V. V. J. Swamy, Nucl. Phys. **A160**, 641 (1971).
[8] W. Pauli, Ergeb. Exakten Naturwiss. **37**, 85, (1964).
[9] S. Ström, Ark. Fys. **30**, 455 (1965).

Quantization by Feynman's path integration method*

Kuo-Shung Cheng

Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11790
(Received 26 January 1973)

The uniqueness of the quantization of a quantum mechanical system by Feynman's path integration method is discussed.

The problem of the uniqueness of quantization by Feynman's path integration method was discussed by some authors.[1,2] They arrived at the conclusion that different approximations of the short time action A

$$A = \int_{t,q(t)}^{t+\epsilon,q(t+\epsilon)} \mathcal{L}\,dt' \tag{1}$$

give different kinds of quantum Hamiltonians. They concluded that, in order to get the supposedly well-known Schrödinger equation, one has to add an effective potential to the action. This effective potential is different for the different approximations of the short time action A.

In a previous paper,[3] we said that Feynman's path integration method gives a unique Schrödinger equation. That is a correct statement, but in that paper the argument was not explicitly and fully presented. We shall present it here.

(a) In the lattice definition of the functional integration, Feynman[4] made it clear that we have to approximate the short time action A by the action $A_{\text{classical}}$, while

$$A_{\text{classical}} = \text{stationary value of } \int_{t,q(t)}^{t+\epsilon,q(t+\epsilon)} \mathcal{L}\,dt'. \tag{2}$$

Besides the intuitive reason that the biggest contribution to the functional integral is the classical path, we have another mathematical reason to explain why the action along the classical path is necessary. Consider the functional integration B,

$$B = N\int\left[\frac{\delta}{\delta q(t)}\exp\left(i\int_{t_a}^{t_b}\mathcal{L}\,dt'\right)\right]\prod_t Dq(t). \tag{3}$$

By any reasonable definition B should be zero.[5] Now turn to the lattice definition; let us devide the time interval as $t_1 = t_a, t_2, \ldots, t_n, t_{n+1}, \ldots, t_{n+m} = t_b$, such that

$$t_n < t < t_{n+1}.$$

Then

$$B = \lim_{(\max \Delta t_k)\to 0} N\int\cdots\int \exp\left(i\sum_k S(q(t_k), q(t_{k+1}))\right) \times\left[\frac{\partial\mathcal{L}}{\partial q(t)} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}(t)}\right)\right]\prod_i Dq(t_i) \tag{4}$$

where $S(q(t_k), q(t_{k+1})) = \int_{t_k}^{t_{k+1}}\mathcal{L}\,dt'$.

$$\left[\frac{\partial\mathcal{L}}{\partial q(t)} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}(t)}\right)\right]$$

is a function of $q(t_n)$ and $q(t_{n+1})$, and depends on which path we approximate the action A.

If we approximate the short time action as the action along the classical path, then

$$\left[\frac{\partial\mathcal{L}}{\partial q(t)} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}(t)}\right)\right] = 0,$$

B is equal to zero. If we do not, then

$$\left[\frac{\partial\mathcal{L}}{\partial q(t)} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}(t)}\right)\right]$$

is not necessary equal to zero and B would not neccessarily vanish.

(b) When we replace the short time action A by the action along the classical path, we have to count *all* possible classical paths. For example, consider a unit mass particle moving in a unit cicle; the Lagrangian for this system is $L = \frac{1}{2}\dot{\theta}^2$, the classical paths satisfying the initial conditions

$$\theta(t')|_{t'=t} = \theta(t), \qquad \theta(t')|_{t'=t+\epsilon} = \theta(t+\epsilon) \tag{5}$$

are

$$\theta(t') = \{[2n\pi + \theta(t+\epsilon) - \theta(t)]/\epsilon\}t' + \theta(t) \tag{6}$$

with n an integer. The actions along these classical paths are

$$A^{(n)}_{\text{classical}} = \tfrac{1}{2}\{[2n\pi + \theta(t+\epsilon) - \theta(t)]/\epsilon\}^2\epsilon\,. \tag{7}$$

The wave function $\psi(\theta(t+\epsilon), t+\epsilon)$ at time $t+\epsilon$ is related to the wavefunction $\psi(\theta(t), t)$ at time t by the relation

$$\psi(\theta(t+\epsilon), t+\epsilon) = \frac{1}{N}\int_0^{2\pi}\sum_{n=-\infty}^{\infty} \exp\left(\frac{i}{2\hbar\epsilon}(2n\pi + \theta(t+\epsilon) - \theta(\epsilon))^2\right)\psi(\theta(t), t)d\theta(t). \tag{8}$$

[Note that the summation over *all* classical paths is essential in Eq. (8).] It is evident from Eq. (8) that

$$\psi(\theta(t+\epsilon), t+\epsilon) = \psi(\theta(t+\epsilon) + m2\pi, t+\epsilon).$$

Suppose we do not take the summation over *all* classical paths; then we would not have the periodic condition. This fact strongly shows us that the replacement of A by $A_{\text{classical}}$ is necessary, since no other approximation of the short time action A will give the same results. [Note that without summing over all paths, one cannot obtain the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}\frac{\partial^2\psi}{\partial\theta^2} \tag{9}$$

from Eq. (8). To see this, one remembers that if one includes only one path, the integration over intermediate x would not extend from $-\infty$ to ∞. Now

Copyright © 1973 by the American Institute of Physics

$$\frac{1}{\sqrt{i\pi\epsilon}}\int_{-a}^{b} e^{ix^2/\epsilon}\,dx \underset{\epsilon\to 0}{\cong} 1+\sqrt{\epsilon}\,f(a,b)+\cdots.$$

The term $\sqrt{\epsilon}\,f$ prevents one from obtaining the simple Eq. (9).]

(c) From the above two sections, it is clear that we *must* insert $A_{\text{classical}}$ for A. For the sake of completeness, we like to write down the equation of motion of the wavefunction corresponding to different kinds of approximations of the short time action A. If we replace $\dot{q}$ by $[q(t+\epsilon)-q(t)]/\epsilon$ in the calculation of A, we get the following equation of motion:

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2}g^{ij}\frac{\partial^2}{\partial q^i\partial q^j}(\sqrt{g}\,\psi). \tag{10}$$

If we replace $\dot{q}$ by

$$\dot{q}^i(t+\epsilon)=\frac{\Delta q^i}{\epsilon}-\frac{1}{2\epsilon}\begin{Bmatrix} i \\ mn\end{Bmatrix}\Delta q^m\Delta q^n \tag{11}$$

we get[3]

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^m}\left(\sqrt{g}\,g^{mn}\frac{\partial\psi}{\partial q^n}\right)+\frac{\hbar^2(R-A)}{6}\phi, \tag{12}$$

$$A=g_{i\delta}\left[\frac{\partial}{\partial q^\gamma}\begin{Bmatrix} i \\ \alpha\beta\end{Bmatrix}+\begin{Bmatrix} i \\ m\alpha\end{Bmatrix}\begin{Bmatrix} m \\ \beta\gamma\end{Bmatrix}\right]\quad(\alpha\beta\gamma\delta). \tag{13}$$

Finally if we calculate A along the classical path, as we did in Ref. 3, we get the following equation of motion:

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^m}\left(\sqrt{g}\,g^{mn}\frac{\partial\psi}{\partial q^n}\right)+\frac{\hbar^2R}{6}\psi. \tag{14}$$

Equations (10) and (12) are not covariant under general coordinate transformation and hence are not correct equations of motion.

(d) It seems to us from the above that the only possible meaningful definition of the path integral is to take the approximation of inserting $A_{\text{classical}}$ of Eq. (2) for A. But one may still raise the additional question: Agreeing with that, why should one not add a term $\hbar^2R/6$ to the Lagrangian and obtain an equation of motion like (14) without the last term?

The answer to this question, we believe, is yes, one could. Nevertheless, we are inclined to believe that the action principle is the more fundamental dynamical principle, since it is applicable with or without constraints and since the action integral *is* the phase of a dynamical system times $\hbar$ in quantum mechanics. We are inclined to the view that adding the term $\hbar^2R/6$ to the integrand in the action integral is contradictory to the spirit of Feynman's formulation of quantum dynamics in terms of interfering contributions from different paths, each with its own phase. We are also inclined to the view that adding the term $\hbar^2R/6$ to the integrand in the action integral is contradictory to the original spirit of Schrödinger's formulation of quantum mechanics in itself. We see no advantage in forcing into an equation of motion like (14) but without the last term.

ACKNOWLEDGMENTS

The author would like to express his thanks to Professor C. N. Yang for his guidance and discussions with the author. He would like to thank Professor J. S. Dowker for his useful comments.

---

*Supported in part by National Science Foundation Grant No. GP-32998X.

[1]I. W. Mayes and J. S. Dowker, J. Math. Phys. **14**, 434 (1973).
[2]L. Cohen, J. Math. Phys. **11**, 3296 (1970).
[3]K. S. Cheng, J. Math. Phys. **13**, 1723 (1972).
[4]R. P. Feynman, Rev. Mod. Phys. **20**, 327 (1948).
[5]G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).

Analytic structure of the Laplace transforms of distributions with restricted support

G. R. Screaton and A. Truman*

Mathematical Institute, University of Oxford, Oxford, England
(Received 12 June 1972; final revised manuscript received 3 November 1972)

A relationship between the support property of a distribution and the analytic structure of its Laplace transform is obtained. This relationship is more general than the usual tube theorems.

I. INTRODUCTION

In this paper we prove two theorems concerning the relationship between the support property of a tempered distribution F and the analytic structure of its Laplace transform, $\mathcal{L}F(\xi + i\eta) = \mathcal{F}[e^{-\eta\cdot x}F](\xi)$, where $\mathcal{F}$ denotes Fourier transform.[1] These theorems are required to obtain certain results in another paper by the same authors.[2] Because of their general nature we decided to publish them separately. It will be seen that our theorems include the tube theorems of quantum field theory[3] as a special case. The tube theorems deal with distributions whose supports are contained in the light cone. Our theorems are applicable to distributions whose supports are more general.

Before stating Theorem 1 we define the set $\hat{K}$ associated with a set K in R^N (real N space). $\hat{K}$ is the portion of the unit sphere given by

$$\hat{K} = \{n : x \cdot n \geq \alpha(n) > -\infty, \forall x \in K\}, \tag{1}$$

where $x \cdot n = \sum_{j=1}^{N} x_j n_j$ and $|n|^2 = n \cdot n = 1$. The geometrical significance of $\alpha(n)$ and the relationship between K and $\hat{K}$ is illustrated in Figs. 1, 2, and 3. In Fig. 2 we see that $\hat{K}$ is the whole of the unit sphere if and only if K is bounded. Fig. 3 depicts the situation in which K is contained in a cone C and is asymptotic to C. In this case, $\hat{K}$ is the portion of the unit sphere bounded by the cone $\hat{C}$. A change of origin 0 changes the value of $\alpha(n)$ but does not alter $\hat{K}$.

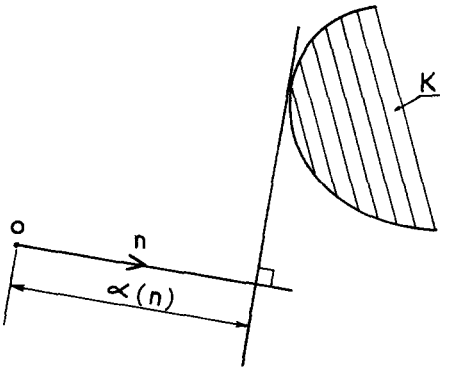

FIG. 1. Geometrical significance of $\alpha(n)$.

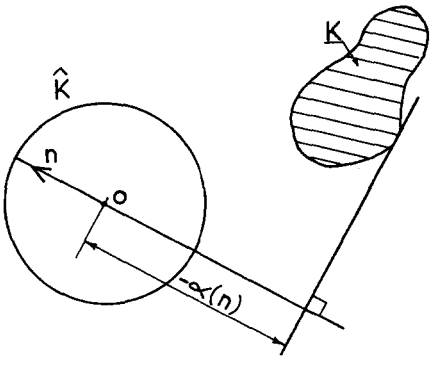

FIG. 2. $\hat{K}$, when K is bounded.

Theorem 1: If $F \in \mathcal{S}'$ has support K and if n is an interior point of $\hat{K}$, then the Laplace transform of F, $\mathcal{L}F(\omega)$, is analytic $\omega = \xi + i|\eta|n$, $|\eta| > 0 (\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ real) with an $\mathcal{S}'$ boundary value. If further, $n \in B \subset A \subset \hat{K}$, where A is open and B is compact,

$$|\mathcal{L}F(\omega)| < Ce^{-|\eta|\alpha(n)}(|\omega|^R + 1)(|\eta|^{-S} + 1), \tag{2}$$

for some nonnegative integers R and S and some constant C.

To establish this result we require no continuity properties of $\alpha(n)$ and assume only $\alpha(n)$ is finite, $n \in \hat{K}$. The integers R and S do not depend on B. When $\alpha(n) > 0$ the bound is better than a polynomial, $|\eta| > a > 0$, and when $\alpha(n) < 0$ worse. However, in view of the converse theorem (Theorem 2) this bound is the best possible.

For the case illustrated in Fig. 2, $\mathcal{L}F(\omega)$ is entire and the bound (2) holds for all n. For the case in Fig. 3, $\mathcal{L}F(\omega)$ is analytic for η interior to $\hat{C}$ and the bound holds in any interior cone.

Theorem 2: If $\tilde{F}(\xi + i\eta)$ is analytic, $\eta = |\eta|n (|n| = 1)$ and, either

$$\text{(i)} \quad |\tilde{F}(\xi + i|\eta|n)| < Ce^{-|\eta|\alpha(n)}(|\omega|^R + 1), \quad |\eta| > a \tag{3}$$

and $\tilde{F}(\xi + i|\eta|n)$ has an $\mathcal{S}'$ boundary value $|\eta| \to 0$; or

$$\text{(ii)} \quad |\tilde{F}(\xi + i|\eta|n)| < Ce^{-|\eta|\alpha(n)}(|\omega|^R + 1)(|\eta|^{-S} + 1), \tag{4}$$

where R and S are nonnegative integers and C a constant then $\tilde{F}$ is the Laplace transform of a tempered distribution with support in the region $x \cdot n \geq \alpha(n)$.

If (x, n) is the Minkowski scalar product, $(x, n) = x \cdot Pn$, where P denotes space inversion. Thus, if the Laplace transform is taken with the Minkowski scalar product, then $\hat{K}$ is replaced by $P\hat{K}$. In particular taking, $K = V_+$, the forward light cone, $\hat{K}$ is given by the intersection of the unit sphere with V_+, $\alpha(n) = 0$, $n \in \hat{K}$. Theorem 1 gives the usual tube theorem and Theorem 2 its converse.

II. PROOF OF THEOREM 1

To prove Theorem 1 we show that

$$\mathcal{L}F(\omega) = \langle F, e^{i\omega\cdot x}f\rangle, \tag{5}$$

when $\omega = \xi + i|\eta|n$, $n \in B \subset A \subset \hat{K}$, where $f \in C^\infty$ with

$$\begin{aligned} f(x) &= 1, \quad x \in K, \\ f(x) &= 0, \quad x \notin K_\delta, \end{aligned} \tag{6}$$

Copyright © 1973 by the American Institute of Physics

and

$$K_\delta = \{x : |x - y| \leq \delta, \quad \text{for some } y \in K\}. \tag{7}$$

Then it will be deduced that $\langle F, e^{i\omega\cdot x} f\rangle$ is analytic in ω and hence $\mathcal{L}F(\omega)$ is analytic. Notice, if n is an interior point of $\hat{K}$, the open set A and compact set B, with $n \in B \subset A \subset \hat{K}$, exist. Finally, f is chosen to yield the bound (2).

To see that $e^{i\omega\cdot x} f(x)$ is a test function we obtain two preliminary results—Lemmas 1 and 2.

Lemma 1: There exists a constant α_B, such that

$$x\cdot n \geq \alpha_B, \quad n \in B, x \in K. \tag{8}$$

Proof: If $n^j \in A, j = 1, 2, \ldots, N$, are linearly independent then

$$T(n^1, n^2, \ldots, n^N) = \{n : n = \sum_{j=1}^{N} \lambda^j n^j / |\sum_{j=1}^{N} \lambda^j n^j|, \lambda^j > 0\} \tag{9}$$

is an open set in $\hat{K}$. For $n \in T$, $x \in K$ we have

$$x\cdot n \geq \operatorname*{Inf}_{\lambda^j>0} \sum_{j=1}^{N} \lambda^j \alpha(n^j) / \left|\sum_{j=1}^{N} \lambda^j n^j\right| \geq \operatorname*{Inf}_{\lambda^j>0} \alpha_0 \sum_{j=1}^{N} \lambda_j / \left|\sum_{j=1}^{N} \lambda^j n^j\right|, \tag{10}$$

where $\alpha_0 = \min(\alpha(n^1), \alpha(n^2), \ldots, \alpha(n^N))$.

A T can be found for each $n \in B$ such that $n \in T$. Thus, as B is compact, it can be covered by a finite number of such T's. From eq. (10) we deduce the existence of α_B.

Lemma 2: There exist constants α and k such that

$$|x - \alpha n| < k(x\cdot n - \alpha); \quad n \in B, x \in K. \tag{11}$$

Proof: An open set C and a compact set D exist with $B \subset C \subset D \subset A$. As B is compact we can find $\theta (0 < \theta < \pi/2)$ so that, whenever $n \in B$ and m, $|m| = 1$, satisfies $m\cdot n \geq \cos\theta$, then $m \in C$. Taking $\alpha = \alpha_D$ we have from Lemma 1,

$$(x - \alpha n)\cdot m \geq \alpha(1 - n\cdot m) \geq 0, \quad n \in B, m \in D \text{ and } x \in K. \tag{12}$$

For each $x \in K$ write

$$x - \alpha n = |x - \alpha n|(\sin\phi n + \cos\phi n_\perp), n\cdot n_\perp = 0 \tag{13}$$

with $0 \leq \phi \leq \pi/2$. Then, defining m by

$$m = \cos\theta n - \sin\theta n_\perp, \tag{14}$$

$m \in C \subset D$. Therefore, referring back to eq. (12),

$$(x - \alpha n)\cdot m = |x - \alpha n| \sin(\phi - \theta) \geq 0. \tag{15}$$

Hence, $\phi \geq \theta$ and from eqn. (13)

$$(x - \alpha n)\cdot n = |x - \alpha n| \sin\phi \geq |x - \alpha n| \sin\theta, \tag{16}$$

which is the required result with $k > \csc\theta$.

We are now in a position to prove that $e^{i\omega\cdot x} f(x) \in \mathcal{S}$. For $x \in K_\delta$, choose $y \in K$, such that $|x - y| < \delta$ and apply Lemma 2 to give

$$|x_j| \leq |x| = |y - \alpha n + x - y + \alpha n| \leq k(y\cdot n - \alpha) + \delta + |\alpha|. \tag{17}$$

Then, as $|x\cdot n - y\cdot n| < \delta$, we have

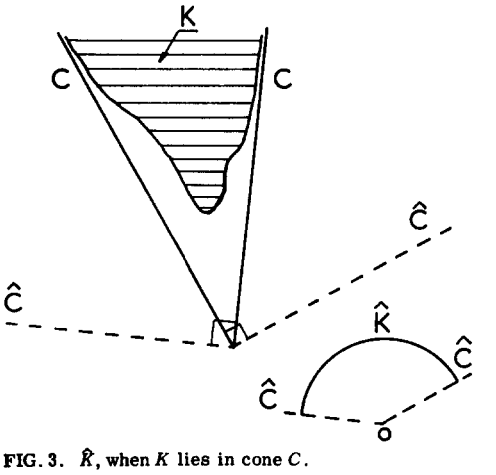

FIG. 3. $\hat{K}$, when K lies in cone C.

$$|x_j| \leq k(x\cdot n - \alpha(n)) + k(\alpha(n) - \alpha + \delta) + \delta + |\alpha|. \tag{18}$$

If x is in the support of $e^{i\omega\cdot x} f$, then $x \in K_\delta$ and so $x\cdot n - \alpha(n) \geq -\delta$. Thus, on observing

$$|e^{i\omega\cdot x}| = e^{-|\eta|\alpha(n)} e^{-|\eta|(x\cdot n - \alpha(n))} \tag{19}$$

and using the bound (18) we deduce

$$\|e^{i\omega\cdot x} f\|_{R,S} < \infty, \quad R, S = 0, 1, 2, \ldots, \tag{20}$$

i.e., $e^{i\omega\cdot x} f(x) \in \mathcal{S}$, where

$$\|h\|_{R,S} = \sum_{s=0}^{S} \sum_{r=0}^{R} \sup |x^\sigma D^\rho h(x)|, \tag{21}$$

$$x^\sigma = \prod_{i=1}^{N} x_i^{\sigma_i}, \quad D^\rho = \prod_{i=1}^{N} \left(\frac{\partial}{\partial x_i}\right)^{\rho_i} \tag{22}$$

and σ_i, ρ_i are nonnegative integers with $s = \sum_{i=1}^{N} \sigma_i$, $r = \sum_{i=1}^{N} \rho_i$.

We now show that $\langle F, e^{i\omega\cdot x} f\rangle$ is the Laplace transform of F. Since the support of F is K, $\langle F, e^{i\omega\cdot x} f\rangle$ is a function of ω which is independent of the particular choice of f and δ. As F is continuous on $\mathcal{S}$, there exists a constant C and nonnegative integers R, S such that

$$|\langle F, h\rangle| < C\|h\|_{R,S}, \quad h \in \mathcal{S}. \tag{23}$$

Hence, $\langle F, e^{i\omega\cdot x} f\rangle$ is polynomially bounded in ξ for fixed η. Furthermore $\mathcal{S}$ is dense in $\mathcal{S}'$,[4] and so there exists a sequence, $\{F_r \in \mathcal{S} : r = 1, 2, \cdots\}$, which converges in $\mathcal{S}'$ to F. Thus, for $h \in \mathcal{S}$

$$\begin{aligned}\langle\langle F, e^{i\omega\cdot x} f\rangle, h(\xi)\rangle &= \lim\langle\langle F_r, e^{i\omega\cdot x} f\rangle, h\rangle = \lim\langle \mathfrak{F}(e^{-\eta\cdot x} f F_r), h\rangle \\ &= \lim\langle e^{-\eta\cdot x} f F_r, \mathfrak{F}h\rangle = \langle e^{-\eta\cdot x} f F, \mathfrak{F}h\rangle \\ &= \langle e^{-\eta\cdot x} F, \mathfrak{F}h\rangle = \langle \mathfrak{F}(e^{-\eta\cdot x} F), h\rangle. \end{aligned} \tag{24}$$

This proves that $\langle F, e^{i\omega\cdot x} f\rangle$ is the Laplace transform of F. It also follows from eq. (24) that $\langle F, e^{i\omega\cdot x} f\rangle$ has an $\mathcal{S}'$ boundary value, $\mathfrak{F}F$, as $|\eta| \to 0$.

In the topology of $\mathcal{S}$,

$$e^{i\omega\cdot x} f(x)\,(e^{i\zeta x_j} - 1)/\zeta \to \frac{\partial}{\partial\omega_j} e^{i\omega\cdot x} f(x)$$

as $\zeta \to 0$. It follows that $\langle F, e^{i\omega\cdot x} f\rangle$ and, hence, $\mathcal{L}F(\omega)$ is

a differentiable function of ω_j, $j = 1, 2, \ldots, N$, and is, therefore, analytic in ω, $\omega = \xi + i|\eta|n$, $|\eta| > 0$, $n \in B$.

We now choose f to establish the bounds in eqn. (2). Let $g \in C_0^\infty$ be a function with

$$\int g(x)d^Nx = 1 \tag{25}$$

and

$$g(x) = 0, \quad |x| > 1. \tag{26}$$

Define $g_{|\eta|}$ and $u_{|\eta|}$ by

$$g_{|\eta|}(x) = |\eta|^N g(|\eta|x) \tag{27}$$

and

$$\begin{aligned} u_{|\eta|}(x) &= 1, \quad x \in K_{|\eta|^{-1}}, \\ &= 0, \quad x \notin K_{|\eta|^{-1}}. \end{aligned} \tag{28}$$

Then

$$f_{|\eta|} = g_{|\eta|} * u_{|\eta|} \tag{29}$$

is an infinitely differentiable function with

$$\begin{aligned} f_{|\eta|}(x) &= 1, \quad x \in K, \\ &= 0, \quad x \notin K_{2|\eta|^{-1}}. \end{aligned} \tag{30}$$

Consider the partial derivatives of $f_{|\eta|}$.

$$D^\rho f_{|\eta|} = (D^\rho g_{|\eta|}) * u_{|\eta|}. \tag{31}$$

As

$$|D^\rho g_{|\eta|}| \leq |\eta|^{r+N} \sup|D^\rho g(x)| \tag{32}$$

and as the support of $g_{|\eta|}$ is contained in the closed ball $\{x: |x| \leq |\eta|^{-1}\}$, there exists a constant C independent of ω such that

$$|D^\rho f_{|\eta|}(x)| < C|\eta|^r \leq C|\omega|^r. \tag{33}$$

We also have

$$\begin{aligned} |(x\cdot n - \alpha(n))^s e^{i\omega\cdot x}| &= e^{-|\eta|\alpha(n)}|\eta|^{-s} \\ &\times \| \eta|^s((x\cdot n) - \alpha(n))^s e^{-|\eta|(x\cdot n - \alpha(n))}| \end{aligned} \tag{34}$$

and, for $x \in K_{2|\eta|^{-1}}$, $n \in B$, $|\eta|(x\cdot n - \alpha(n)) \geq -2$.

Therefore, there exists a constant C such that

$$|(x\cdot n - \alpha(n))^s e^{i\omega\cdot x}| < C|\eta|^{-s} e^{-|\eta|\alpha(n)}, \quad x \in K_{2|\eta|^{-1}}. \tag{35}$$

Putting $\delta = 2|\eta|^{-1}$ in Eq. (18) and noticing that $\alpha(n)$ is bounded above (unless K is empty), we have with the use of the bounds (18), (33), and (35) the result

$$\| e^{i\omega\cdot x} f_{|\eta|}(x) \|_{R,S} < e^{-|\eta|\alpha(n)} P(|\eta|^{-1}, |\omega|), \tag{36}$$

where P is a polynomial of degree R in $|\omega|$ and S in $|\eta|^{-1}$.

A constant C can be found such that

$$P(|\eta|^{-1}, |\omega|) < C(|\omega|^R + 1)(|\eta|^{-S} + 1) \tag{37}$$

and Eqs. (5), (23), (36), and (37) together imply when $n \in B$ there exist constants C, R and S such that

$$|\mathcal{L}F(\omega)| < Ce^{-|\eta|\alpha(n)}(|\omega|^R + 1)(|\eta|^{-S} + 1). \tag{38}$$

This concludes the proof of Theorem 1.

III. PROOF OF THEOREM 2

It can be shown that the Condition (ii) in Theorem 2 implies condition (i).[5] We prove Theorem 2 under the assumption that condition (i) is satisfied. As $\tilde{F}$ is polynomially bounded in ξ, the inverse Fourier transform with respect to the variable ξ of $\tilde{F}(\xi + i|\eta|n)$ is a tempered distribution which will depend on the variable $|\eta|$,

$$F(x, |\eta|) = \bar{\mathcal{F}}_\xi[\tilde{F}(\xi + i|\eta|n)]. \tag{39}$$

As $\tilde{F}$ is analytic and satisfies the bound (3), its derivative is analytic and satisfies a similar bound.[6] Hence, $\tilde{F}$ considered as a distribution in ξ is differentiable with respect to the parameter $|\eta|$ and

$$\frac{\partial}{\partial|\eta|}\tilde{F}(\xi + i|\eta|n) = i\sum_{j=1}^N n_j \frac{\partial}{\partial\xi_j}\tilde{F}(\xi + i|\eta|n). \tag{40}$$

This last equation can be rewritten in terms of $F(x, |\eta|)$ as

$$\frac{\partial}{\partial|\eta|}F(x, |\eta|) = -n\cdot xF(x, |\eta|). \tag{41}$$

However, $e^{|\eta|n\cdot x}F(x, |\eta|) \in \mathcal{D}'$ and from Eq. (41)

$$\frac{\partial}{\partial|\eta|}e^{|\eta|n\cdot x}F(x, |\eta|) = 0. \tag{42}$$

Hence, there exists $F(x) \in \mathcal{D}'$ independent of $|\eta|$ such that

$$e^{|\eta|n\cdot x}F(x, |\eta|) = F(x). \tag{43}$$

In fact, $F(x)$ is a tempered distribution because $\tilde{F}$ has an $\mathcal{S}'$ boundary value as $|\eta| \to 0$. Thus,

$$\tilde{F}(\xi + i|\eta|n) = \mathcal{L}F(\xi + i|\eta|n), \tag{44}$$

where $F \in \mathcal{S}'$.

It only remains for us to show that the support of F is contained in the set, $\{x: x\cdot n \geq \alpha(n)\}$.

Suppose $g \in \mathcal{D}$, then

$$\begin{aligned} \langle F, g\rangle &= \langle e^{|\eta|n\cdot x}F(x, |\eta|), g\rangle = \langle \mathcal{F}(F(x, |\eta|)), \bar{\mathcal{F}}(e^{|\eta|n\cdot x}g)\rangle \\ &= \langle \tilde{F}(\xi + i|\eta|n), \bar{\mathcal{F}}(e^{|\eta|n\cdot x}g)\rangle. \end{aligned} \tag{45}$$

Using the bound (3), we have

$$\begin{aligned} |\langle F, g\rangle| &< C\int d^N\xi(|\xi|^R + 1)|\bar{\mathcal{F}}h| < C'\|\bar{\mathcal{F}}h\|_{N+1+R,0} \\ &< C''\|h\|_{N+1,N+1+R}, \end{aligned} \tag{46}$$

where

$$h(x) = (|\eta|^R + 1)e^{|\eta|(n\cdot x - \alpha(n))}g(x), \tag{47}$$

and the C are constants.

If the support of g is contained in the set, $\{x: x\cdot n < \alpha(n)\}$,

then there exists $\epsilon > 0$ such that $x \cdot n - \alpha(n) \leq -\epsilon$, whenever $x \in \text{supp} g$. Therefore, $\|h\|_{N+1, N+1+R} \to 0$ as $|\eta| \to \infty$. Hence from Eq. (46) we have

$$\langle F, g \rangle = 0. \tag{48}$$

This proves Theorem 2.

With only a change of origin in η the theorems are applicable to Schwarz distributions. Suppose $F \in \mathcal{D}'$ with support K and there exists λ such that $e^{-\lambda \cdot x} F \in \mathcal{S}'$. We have, on applying Theorem 1 to $e^{-\lambda \cdot x} F$, $\mathcal{L}F(\xi + i\eta)$ is analytic, $\omega = \xi + i\eta$, for all ξ, $\eta = \lambda + |\eta - \lambda| n$, n an interior point of $\hat{K}$ and, when $n \in B$,

$$|\mathcal{L}F(\omega)| < Ce^{-|\eta| \alpha(n)} (|\omega|^R + 1) (|\eta - \lambda|^{-S} + 1). \tag{49}$$

Conversely, arguing as in Theorem 2, if $\tilde{F}(\omega)$ is analytic for all ξ, $\eta = \lambda + |\eta - \lambda| n$ ($|n| = 1$) and

$$|\tilde{F}(\xi + i\lambda + i|\eta - \lambda| n)| < Ce^{-|\eta| \alpha(n)} (|\omega|^R + 1), \quad |\eta - \lambda| > a, \tag{50}$$

then $\tilde{F}$ is the Laplace transform of a distribution $F \in \mathcal{D}'$ whose support is in the region $x \cdot n \geq \alpha(n)$. To see this, notice that in the proof of Theorem 2 the only use that is made of the condition that $\tilde{F}$ has an $\mathcal{S}'$ boundary value is to establish $F \in \mathcal{S}'$. Otherwise the proof can be taken over and gives in this case $e^{-\lambda \cdot x} F \in \mathcal{D}'$ with support $x \cdot n \geq \alpha(n)$. Hence, $F \in \mathcal{D}'$ with the same support property.

---

*Present address: Department of Mathematics, Heriot-Watt University, Edinburgh, Scotland.

[1]We are using the convention $\mathcal{F}f(\xi) = \int f(x) e^{i\xi \cdot x} d^N x$, $f \in \mathcal{S}$. A discussion of the Fourier and Laplace transforms of distributions can be found in R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964); E. J. Beltrami and M. R. Wohlers, *Distribution and the Boundary Values of Analytic Functions* (Academic, New York, 1966) and H. Bremermann, *Distributions, Complex Variables and Fourier Transforms* (Addison-Wesley, Reading, Mass., 1965).

[2]"Asymptotic causality in quantum field theory" (to be published).

[3]For a brief account of tube theorems, see H. Bremermann in Ref. 1, Sect. 15-2. For a more comprehensive account, see R. F. Streater and A. S. Wightman in Ref. 1, Sec. 2-3.

[4]See, for example, E. J. Beltrami and M. R. Wohlers in Ref. 1, Theorem 1-31.

[5]R. F. Streater and A. S. Wightman, Ref. 1, Theorem 2-10.

[6]See, for example, the discussion in R. F. Streater and A. S. Wightman, Ref. 1, p. 57.

Uniqueness of the time coordinate for certain singular space-times in general relativity

Peter Goorjian

Department of Mathematics, University of Delaware, Newark, Delaware, 19711
(Received 20 November 1972)

Many singular space-times are described locally in geodesic normal coordinates with the singularity described by the time coordinate equal to zero. Under appropriate conditions the time coordinate in this description is uniquely determined. Furthermore, every timelike curve in the coordinate patch is inextendible and has the same limit point as exactly one time coordinate line. The proof of uniqueness of the time coordinate employs methods from the theory of first order partial differential equations in a new way.

1. INTRODUCTION

Many singular solutions to the Einstein equations are described locally in geodesic normal coordinates with the singularity described by the time coordinate equal to zero.[1,2] We show that under appropriate conditions the time coordinate in this description is uniquely determined. Furthermore, every timelike curve in the coordinate patch is inextendible and has the same limit point as exatly one time coordinate line.

2. UNIQUENESS OF THE TIME COORDINATE

Consider a coordinate patch of a singular space-time in which the metric has the form

$$ds^2 = dt^2 - g_{ab}(t, x^c)dx^a dx^b, \quad a,b = 1,2,3, \tag{1}$$

$t > 0$ and the singularity is given by $t = 0$. If

$$t^2 |g^{ab}| = 0(t^{2\epsilon}) \tag{2}$$

for some $\epsilon > 0$ and for each component g^{ab} of the spatial metric, then the time coordinate t has been uniquely determined. Note that if g^{ab} were C^2 at $t = 0$, then this result would be immediate.

To prove uniqueness, we consider another coordinatization of the patch. The metric has the form

$$ds^2 = du^2 - \hat{g}_{ab}(u, \hat{x}^c)d\hat{x}^a d\hat{x}^b \tag{3}$$

and the singularity is described by $u = 0$. We shall show that the coordinate u considered as a function of t and x^a satisfies $u(t, x^a) = t$. First we show that each u coordinate line (i.e., $\hat{x}^a = \text{const}$) approaches a point of the singularity that is also approached by exactly one t coordinate line (i.e., $x^a = \text{const}$). Hence, let x^μ be a timelike or lightlike curve which passes through a point $p = (t_0, x_0^a)$ of the coordinate patch described by (1). Furthermore, assume t_0 is so small that the region R given by

$$r = \left(\sum_{a=1}^{3} |x_0^a - x^a|^2\right)^{1/2} \leq 2M\epsilon^{-1}(t_0^\epsilon - t^\epsilon), \tag{4}$$

$$0 < t \leq t_0,$$

is covered by that patch and that in R Eq. (2) is satisfied so that

$$t^2 |g^{ab}| \leq M^2 t^{2\epsilon}. \tag{5}$$

Notice that the surface $r = 2M\epsilon^{-1}(t_0^\epsilon - t^\epsilon)$, $0 < t \leq t_0$ is the light cone from p toward $t = 0$ for the "bounding" diagonal metric $g_0^{aa} = 4M^2 t^{2(\epsilon-1)}$.

Now we shall show that as x^μ continues past the point p toward the singularity (i.e., its t coordinate decreases), it remains within the region R. The tangent vector to x^μ is given by

$$T^\mu(q) = \frac{dx^\mu}{dq} = (N^t, N^\perp),$$

where the parameter q measures proper time from p for a timelike curve and q measures the decrease in the time coordinate t from p for a lightlike geodesic. N^t is the component of T^μ in the t direction and $N^\perp$ represents the components of T^μ in the space directions. For a timelike curve and a likelike geodesic, respectively,

$$T^\mu T_\mu = \left(\frac{dt}{dq}\right)^2 - g_{ab}\frac{dx^a}{dq}\frac{dx^b}{dq} = 1,$$
$$T^\mu T_\mu = \left(\frac{dt}{dq}\right)^2 - g_{ab}\left(\frac{dx^a}{dq}\right)\left(\frac{dx^b}{dq}\right) = 0. \tag{6}$$

Hence

$$\|N^t\|^2 \equiv \left(\frac{dt}{dq}\right)^2 \geq 1; \text{ also, } \|N^\perp\|^2 \equiv \sum_{a=1}^{3}\left(\frac{dx^a}{dq}\right)^2.$$

By the use of (5) and the positive definiteness of g_{ab} it follows (e.g., use an orthogonal transformation of space coordinates $x^a \to \tilde{x}^a$ such that $\tilde{g}_{ab}$ is diagonal at p) that

$$\|N^\perp\|^2 \leq 3M^2 t^{2(\epsilon-1)}\|N^t\|^2. \tag{7}$$

Therefore, at p if we follow x^μ in the decreasing t direction, we see by comparison of (4) and (7) that x^μ enters the interior of the region R. Furthermore, the preceeding argument is applicable at each point of R so that (7) holds in all of R and hence x^μ is contained in R.

Next we show that x^μ approaches a definite point of the $t = 0$ boundary of R. First, since $|dt/dq| \geq 1$ by increasing q by no more than t_0, x^μ approaches the boundary. Finally, from (7) the total variation in r is bounded by $2M\epsilon^{-1}t_0^\epsilon$ and hence x^μ approaches exactly one limit point on $t = 0$.

Now consider the part of the singularity reached by all timelike curves originating in R when they are followed in the direction for which their time coordinate decreases. We can conclude that that part of the singularity is completely specified as the limit set of the time coordinate lines in R.

Now we are ready to show that the coordinate u in (3) satisfies $u(t, x^a) = t$. Since the coordinate systems (t, x^a) and $(u, \hat{x}^a)$ are related by a coordinate transformation, it follows that u satisfies the equation

Copyright © 1973 by the American Institute of Physics

$$w_t^2 - g^{ab}(t, x^c) w_a w_b = 1, \tag{8}$$

$w_t = \partial w/\partial t, w_a = \partial w/\partial x^a, a = 1,2,3$. Certainly the function $w(t, x^a) = t$ satisfies (8) and since $u(t, x^a)$ also satisfies (8), their difference $v = t - u$ satisfies the related equation

$$(1 + u_t) v_t - g^{ab} u_a v_b = 0. \tag{9}$$

From the theory of first order linear partial differential equations,[3] the characteristic curve C of (9) starting from the point (p, v_0) [where $p = (t_0, x_0^a)$ of the region R as defined previously and $v_0 = v(t_0, x_0^a)$] is determined by the system

$$\frac{dt}{dq} = 1 + u_t, \quad \frac{dx^a}{dq} = -g^{ab} u_b, \quad \frac{dv}{dq} = 0. \tag{10}$$

Note from (10) that v is constant along C, say $v = c$. If $u_t > 0$ at p is assumed [therefore $u_t > 0$ on C by (8)], then the characteristic base curve B in R is determined by

$$\frac{dx_v^a}{dt} = -\frac{g^{ab} u_b}{1 + u_t}. \tag{11}$$

Now the u coordinate line through p is determined by

$$\frac{dx_u^a}{dt} = -\frac{g^{ab} u_b}{u_t} \tag{12}$$

and is a timelike geodesic. Hence, as we have shown, it is trapped in R and approaches a definite point of the boundary $t = 0$. By comparison of (11) and (12),

$$\left|\frac{dx_v^a}{dt}\right| \leq \left|\frac{dx_u^a}{dt}\right|$$

and therefore the same proof applies to show that B also approaches a definite point of the boundary $t = 0$.

Now let us follow B as it develops from p. First notice that at an arbitrary point $p_1 = (t_1, x_1^a)$ on B the value of u can be bounded by t_1. This bound is obtained by following the u coordinate line U originating at p_1 toward the singularity. From (3) we have $ds^2 = du^2$ for U and from (1) $du^2 = dt^2 - g_{ab} dx^a dx^b \leq dt^2$. Since U was shown to approach a definite point of the $t = 0$ boundary of R which is part of the singularity and u goes to zero at the singularity, we have

$$u(t, x_1^a) = \int_{0U}^{u} du \leq \int_{0U}^{t_1} dt = t_1.$$

Therefore, $|v| = |t - u| \leq t$ on B. As we follow B toward the singularity, t approaches zero and since v is constant on B, v must be identically zero on B and therefore $u(t_0, x_0^a) = t_0$. The preceeding argument applied to an arbitrary point in R and therefore $u(t, x^a) = t$ in R and (3) is equal to

$$ds^2 = dt^2 - \hat{g}_{ab}(t, \hat{x}^c) d\hat{x}^a d\hat{x}^b. \tag{13}$$

When the condition in Eq. (2) is not satisfied, then a counterexample is provided by the Kasner metric

$$ds^2 = dt^2 - t^2 dx^2 - dy^2 - dz^2$$

for which the Minkowski metric

$$ds^2 = du^2 - d\hat{x}^2 - dy^2 - dz^2,$$

$u = t \cosh x$, $\hat{x} = t \sinh x$ is a nonequivalent coordinatization. For other possibilities of the Kasner metric, as well as for the Friedman like metrics,[1] to mention a few examples, Eq. (2) is satisfied. Of particular interest is the application to the Schwarzschild metric

$$ds^2 = (1 - 2m/r)\, dt^2 - \frac{1}{(1 - 2m/r)} \times dr^2 - r^2 (d\theta^2 + \sin^2\theta d\phi^2). \tag{14}$$

Near the $r = 0$ singularity, (14) is approximated by

$$ds^2 = du^2 - \left(\frac{3u}{4m}\right)^{-2/3} dt^2 - [\tfrac{3}{2}(2m)^{1/2} u]^{4/3} \times (d\theta^2 + \sin^2\theta d\phi^2) \tag{15}$$

with $u = (2r^3)^{1/2}/3m^{1/2}$. Hence condition (2) is satisfied and u is the unique time coordinate.

3. FURTHER REMARKS

1: Although condition (2) is sufficient, it is not necessary. In the proof we needed to show that any timelike goedesic approached a definite point of the $t = 0$ boundary of R. Given a singular manifold, it is sufficient to show that for any timelike curve the total variation in r from (t_0, x_0^a) to the singularity is bounded.

2 More generally,[2] we may have a singular spacetime described in geodesic normal coordinates for which the singularity is given by $t = \phi(x^a)$. For such cases, it is sufficient to prove that any timelike geodesic approaches a definite point of the singularity $(t, x^a) = (\phi(x^a), x^a)$. Then uniqueness of the time coordinate holds within the class of possible time coordinates $u(t, x^a)$ such that $u(\phi(x^a), x^a) = \phi(x^a)$.

3 Finally, if the time coordinate u in Eq. (3) is not required to describe the singularity $t = 0$ by $u = 0$, then it is possible to find time coordinates $u(t, x^a)$ different from t. First prescribe a function $u_0(x^a)$ on the plane $t = 0$. Next employ a change of coordinates $\bar{t} = t^{2\epsilon/(2+k)}$ with k a positive integer. For certain smooth metrics, $g^{ab} \in C^\infty$ for $t > 0$, e.g., the Kasner metrics, the resulting Eq. (8) has C^k coefficients in the $(\bar{t}, x^a)$ coordinate system for $\bar{t} \geq 0$. Hence we may use the equations for the characteristic strip from the theory of first order partial differential equations[3] to extend the function $u_0(x^a)$ to a local function $\bar{u}(\bar{t}, x^a)$. The function $u(t, x^a) = \bar{u}(\bar{t}(t), x^a)$ is the sought after local function. [Note if $u_0(x^a) \equiv 0$ then the characteristic curve can degenerate into the initial point and this method will not lift u_0 off $t = 0$.]

ACKNOWLEDGMENT

The author is grateful to Rainer Sachs for suggesting this problem and for many helpful discussions.

[1] E. M. Lifshitz and I. M. Khalatnikov, Usp. Fiz. Nauk **80**, 391 (1963) [Sov. Phys.-Usp. **6**, 495 (1964)].
[2] D. Eardley, E. Liang, and R. Sachs, J. Math. Phys. **13**, 99 (1972).
[3] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

Product-vector basis and occupation-number basis in Fock space for bosons and fermions*

Boris Leaf

State University of New York, College at Cortland, Cortland, New York 13045
(Received 25 August 1972; revised manuscript received 15 January 1973)

A basis of products of one-particle vectors (symmetric or antisymmetric for bosons or fermions) is defined, and related to the occupation-number basis in an abstract Fock space, a generalization of the wave-mechanical configuration representation of the space of second quantization. Creation and destruction operators are simply defined in the product-vector basis and shown to have their usual properties in occupation-number representation. Algebraic properties of product vectors are developed and illustrated. The representation of operators in the product-vector basis is described, together with a brief discussion of the effects on these operators of transformations from one set of one-particle vectors to another. A simple treatment of density and reduced density operators for systems of bosons or fermions is given in the product-vector basis. The relationship between degeneracy in the Fock-space spectrum of an additive operator NA_1 and degeneracy in the spectrum of one-particle operator A_1 is obtained.

1. INTRODUCTION

Since the paper by Fock[1] in 1932, descriptions of the second-quantization formalism for systems of bosons and fermions have included the Fock configuration-space formulation.[2,3] As presented, the formulation is expressed in wave mechanical, coordinate representation. The formulation can be generalized (and in the process, greatly simplified) if the wave-mechanical representation is abandoned and the Fock space is considered as an abstract vector space. For this purpose we use the product-vector basis.

The Fock space (grand space) is a direct sum over $N = 0, 1, 2, \ldots, \infty$, of petit spaces for systems of N particles. The product-vector basis in the petit space of N particles is the set of direct products $|\mathbf{r}_1\rangle|\mathbf{r}_2\rangle\cdots|\mathbf{r}_N\rangle$, where each $|\mathbf{r}\rangle$ is a one-particle vector from the complete, orthonormal set of simultaneous eigenvectors for a complete set of commuting one-particle observables.[4] The requisite symmetry (bosons) or antisymmetry (fermions) of the product vector $|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$ is ensured by an appropriate transposition relation for each case. The set of product vectors is complete in the petit space. The scalar product of two vectors is defined by a Slater determinant (fermions) or permanent (bosons).

In the product-vector basis the creation and destruction operators have very simple forms: Apart from combinatorial factors, the eigenket (bra) for a given one-particle state is the creation (destruction) operator for that state. The commutation relations for pairs of creation operators are the same as for transposition of the eigenkets, and traces of products of creation and destruction operators are replaced by scalar products of the product vectors. For many purposes the resulting algebraic manipulations are simpler than for creation and destruction operators; the formalism is similar to that of ordinary, "first-quantization" quantum mechanics, where the identical nature of particles is ignored.

In Sec. 2, the product-vector basis is introduced and its properties developed. Comparison is made with Fock's equations for the matrix elements of creation and destruction operators. The definition of these operators in the product-vector basis is shown to yield the fundamental commutation relations for these operators. In Sec. 3, the occupation-number basis is introduced as the set of simultaneous eigenvectors of the complete commuting set of operators $\hat{N}(\mathbf{r})$ for the one-particle states $|\mathbf{r}\rangle$. The occupation-number basis and the product-vector basis are alternative bases in the same abstract Fock space, the space of second quantization. The vectors of the occupation-number basis are given in terms of product vectors (taken in "natural" order, in the case of fermions), and it is shown that the creation and destruction operators take their usual forms in the occupation-number basis.

The power and simplicity of methods utilizing the product-vector basis in the abstract Fock space are illustrated in representing the operators of Fock space in Sec. 4; by a treatment of density and reduced density operators, greatly simplified compared to that of Bogoliubov,[5] in Sec. 5; and in consideration of the effects of degeneracy in the spectrum of a one-particle operator A_1 upon the spectrum of the additive, N-particle operator, $A_N(1) = NA_1$, in Sec. 6. Also included in Sec. 4 is a discussion of transformation theory for a change from one one-particle basis to another.

2. THE PRODUCT-VECTOR BASIS

To construct basis vectors in the space of N-particle state vectors, we start with an orthonormal set of one-particle kets $|\mathbf{r}\rangle$. Each ket is a simultaneous eigenvector of a complete set of commuting observables in the vector space of a single particle; the set of kets $|\mathbf{r}\rangle$ is nondegenerate.[4] A ket $|\mathbf{r}\rangle$ is labeled by the set of eigenvalues $\mathbf{r} = (r_1, r_2, \ldots)$ to which it belongs.

The space of N-particle state vectors is spanned by the direct product of N one-particle states, $|\mathbf{r}_1\rangle|\mathbf{r}_2\rangle\cdots|\mathbf{r}_N\rangle$; the product must be symmetrized in the case of bosons, antisymmetrized for fermions. Let $\mathcal{S}$ represent the symmetrizing projector $\mathcal{S}_+$ for bosons, or the antisymmetrizing projector $\mathcal{S}_-$ for fermions; if P is a permutation operator acting on the one-particle state labels $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$, then

$$\mathcal{S}_\pm(\mathbf{r}_1\cdots\mathbf{r}_N) = (N!)^{-1}\sum_P (\pm 1)^{\pi_P} P(\mathbf{r}_1\cdots\mathbf{r}_N), \tag{2.1}$$

where π_P is the parity of the permutation and the sum is over all $N!$ permutations. A set of basis vectors in the N-particle space is accordingly given by $\mathcal{S}(\mathbf{r}_1\cdots\mathbf{r}_N)$ $|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$. This is the product-vector basis. We require that the one-particle states obey the transposition relations

$$|\mathbf{r}_i\rangle|\mathbf{r}_j\rangle = \pm\,|\mathbf{r}_j\rangle|\mathbf{r}_i\rangle \tag{2.2}$$

(upper sign, bosons; lower sign, fermions), so that

$$\mathcal{S}(\mathbf{r}_1\cdots\mathbf{r}_N)|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle = |\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle. \tag{2.3}$$

Copyright © 1973 by the American Institute of Physics

The space of N-particle state vectors is the Fock space. Following the terminology of statistical mechanics, we shall refer to the space for fixed N as the petit Fock space. The grand Fock space is the direct sum of petit spaces for $N = 0, 1, 2, \ldots, \infty$. Keeping in mind the appropriate relation (2.2) for bosons or fermions, we can write the kets of the product-vector basis in the petit space, according to (2.3), simply as $|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$. These are complete, so that

$$\sum_{\mathbf{r}_1} \cdots \sum_{\mathbf{r}_N} |\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle\langle\mathbf{r}_N| \cdots \langle\mathbf{r}_1| = \mathbf{1}_N, \tag{2.4}$$

where $\mathbf{1}_N$ is the unit operator of the petit space. The scalar product of two vectors $|\mathbf{x}_1\rangle \cdots |\mathbf{x}_N\rangle$ and $|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$ can be defined as

$$\begin{aligned}
&\langle\mathbf{x}_N| \cdots \langle\mathbf{x}_1|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \\
&\quad = \langle\mathbf{x}_N| \cdots \langle\mathbf{x}_1|\,\mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_N)|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \\
&\quad = \mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_N)\langle\mathbf{x}_1|\mathbf{r}_1\rangle\langle\mathbf{x}_2|\mathbf{r}_2\rangle \cdots \langle\mathbf{x}_N|\mathbf{r}_N\rangle \\
&\quad = \frac{1}{N!}\begin{vmatrix} \langle\mathbf{x}_1|\mathbf{r}_1\rangle, \langle\mathbf{x}_1|\mathbf{r}_2\rangle, \ldots, \langle\mathbf{x}_1|\mathbf{r}_N\rangle \\ \cdots \\ \langle\mathbf{x}_N|\mathbf{r}_1\rangle, \langle\mathbf{x}_N|\mathbf{r}_2\rangle, \ldots, \langle\mathbf{x}_N|\mathbf{r}_N\rangle \end{vmatrix},
\end{aligned} \tag{2.5}$$

where $|\cdots|$ represents a Slater determinant (fermions) or permanent (bosons). In particular, the scalar product of two vectors $|\mathbf{r}_1'\rangle \cdots |\mathbf{r}_N'\rangle$ and $|\mathbf{r}_1''\rangle \cdots |\mathbf{r}_N''\rangle$ from the same basis set is given by

$$\begin{aligned}
&\langle\mathbf{r}_N''| \cdots \langle\mathbf{r}_1''|\mathbf{r}_1'\rangle \cdots |\mathbf{r}_N'\rangle \\
&\qquad = \mathcal{S}(\mathbf{r}_1' \cdots \mathbf{r}_N')\delta(\mathbf{r}_1'', \mathbf{r}_1') \cdots \delta(\mathbf{r}_N'', \mathbf{r}_N').
\end{aligned} \tag{2.6}$$

From (2.6) it follows that

$$\begin{aligned}
&\sum_{\mathbf{r}_1} \cdots \sum_{\mathbf{r}_N} \langle\mathbf{r}_N'| \cdots \langle\mathbf{r}_1'|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \psi_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \\
&\qquad = \psi_N(\mathbf{r}_1' \cdots \mathbf{r}_N'),
\end{aligned} \tag{2.7}$$

for any quantity $\psi_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ having the symmetry of the petit space:

$$\mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_N)\psi_N(\mathbf{r}_1 \cdots \mathbf{r}_N) = \psi_N(\mathbf{r}_1 \cdots \mathbf{r}_N). \tag{2.8}$$

Using the product-vector basis $|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$ we may readily discuss creation and destruction operators. The creation operator $a^\dagger(\mathbf{r})$ for the one-particle state $|\mathbf{r}\rangle$, is defined by[6]

$$a^\dagger(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle = (N+1)^{1/2}|\mathbf{r}\rangle\,|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle; \tag{2.9}$$

the destruction operator $a(\mathbf{r})$, by

$$a(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle = N^{1/2}\langle\mathbf{r}|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle. \tag{2.10}$$

From (2.1) and (2.3),

$$\begin{aligned}
\langle\mathbf{r}|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle &= \langle\mathbf{r}|\,\mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_N)|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \\
&= N^{-1}[\delta(\mathbf{r}, \mathbf{r}_1)|\mathbf{r}_2\rangle \cdots |\mathbf{r}_N\rangle \pm \delta(\mathbf{r}, \mathbf{r}_2)|\mathbf{r}_1\rangle|\mathbf{r}_3\rangle \cdots |\mathbf{r}_N\rangle \\
&\quad + \cdots (\pm)^{N-1}\delta(\mathbf{r}, \mathbf{r}_N)|\mathbf{r}_1\rangle\,|\mathbf{r}_2\rangle \cdots |\mathbf{r}_{N-1}\rangle],
\end{aligned}$$

so that

$$\begin{aligned}
a(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle = N^{-1/2}\sum_{j=1}^{N} (\pm 1)^{j-1}\delta(\mathbf{r}, \mathbf{r}_j) \\
\times |\mathbf{r}_1\rangle \cdots |\mathbf{r}_{j-1}\rangle|\mathbf{r}_{j+1}\rangle \cdots |\mathbf{r}_N\rangle.
\end{aligned} \tag{2.11}$$

The matrix element obtained from (2.11) and (2.6), namely,

$$\begin{aligned}
&\langle\mathbf{x}_{N-1}| \cdots \langle\mathbf{x}_1|\,a(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \\
&\quad = N^{-1/2}\sum_{j=1}^{N} \delta(\mathbf{r}, \mathbf{r}_j)\mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_{j-1}, \mathbf{r}_{j+1} \cdots \mathbf{r}_N) \\
&\qquad \times \delta(\mathbf{x}_1, \mathbf{r}_1) \cdots \delta(\mathbf{x}_{j-1}, \mathbf{r}_{j-1}) \\
&\qquad \times \delta(\mathbf{x}_j, \mathbf{r}_{j+1}) \cdots \delta(\mathbf{x}_{N-1}, \mathbf{r}_N),
\end{aligned} \tag{2.12}$$

can be compared with that obtained by Fock from an *ansatz*; Fock's equation[7] omits the projector $\mathcal{S}(\mathbf{r}_1 \cdots \mathbf{r}_{j-1}, \mathbf{r}_{j-1} \cdots \mathbf{r}_N)$.

A modification of (2.11) will be useful for describing creation and destruction operators in the occupation-number basis. Let $N(\mathbf{r}_\sigma)$ designate the occupation number for the one-particle state $|\mathbf{r}_\sigma\rangle$ in the ket $|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$, i.e., the number of times $|\mathbf{r}_\sigma\rangle$ appears in the product vector. We start from the relation

$$\begin{aligned}
\sum_P (\pm 1)^{\pi_P} P(\mathbf{r}_1 \cdots \mathbf{r}_N)|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle &= \sum_{\mathbf{r}_\sigma} (\pm 1)^{\sigma-1} N(\mathbf{r}_\sigma)|\mathbf{r}_\sigma\rangle \\
&\quad \times \sum_P (\pm 1)^{\pi_P} P(\mathbf{r}_1 \cdots \mathbf{r}_N)'(|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle)',
\end{aligned} \tag{2.13}$$

where $(\cdots)'$ indicates that the state $|\mathbf{r}_\sigma\rangle$ appears only $N(\mathbf{r}_\sigma) - 1$ times among the $N - 1$ one-particle states remaining when one $|\mathbf{r}_\sigma\rangle$ has been factored on the left. The summation on $\mathbf{r}_\sigma$ runs over all *different* values of $\mathbf{r}_\sigma$ in the original product $|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$, i.e., over those one-particle states for which $N(\mathbf{r}_\sigma) \neq 0$. In the case of fermions, the transposition relation (2.2) ensures that $N(\mathbf{r}_\sigma)$ can assume only the values 0 or 1. In this case $N(\mathbf{r}_\sigma) = 1$ for occupied states, all of which are different, so that σ can be identified with $j = 1 \cdots N$ as in (2.11). Equation (2.13) corresponds to an expansion in elements of a row[8] of the Slater determinant or permanent in (2.5). From (2.13), (2.1), and (2.3),

$$|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle = N^{-1}\sum_{\mathbf{r}_\sigma} (\pm 1)^{\sigma-1} N(\mathbf{r}_\sigma)|\mathbf{r}_\sigma\rangle(|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle)', \tag{2.14}$$

so that from the definition (2.10), as an alternative form of (2.11), we obtain

$$\begin{aligned}
&a(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle \\
&\quad = N^{-1/2}\sum_{\mathbf{r}_\sigma} (\pm 1)^{\sigma-1} N(\mathbf{r}_\sigma)\delta(\mathbf{r}, \mathbf{r}_\sigma)(|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle)'.
\end{aligned} \tag{2.15}$$

To verify the suitability of the definitions (2.9), (2.10) for creation and destruction operators, we now show that they imply the fundamental commutation relations for these operators. From (2.4) and (2.2),

$$\begin{aligned}
a(\mathbf{r})a(\mathbf{r}')\mathbf{1}_N &= [N(N-1)]^{1/2}\langle\mathbf{r}|\langle\mathbf{r}'|\mathbf{1}_N \\
&= \pm [N(N-1)]^{1/2}\langle\mathbf{r}'|\langle\mathbf{r}|\mathbf{1}_N = \pm\, a(\mathbf{r}')a(\mathbf{r})\mathbf{1}_N.
\end{aligned} \tag{2.16}$$

Accordingly, in the petit space,

$$\begin{aligned}
&[a(\mathbf{r})a(\mathbf{r}') \mp a(\mathbf{r}')a(\mathbf{r})]\mathbf{1}_N = 0, \\
&[a^\dagger(\mathbf{r}')a^\dagger(\mathbf{r}) \mp a^\dagger(\mathbf{r})a^\dagger(\mathbf{r}')]\mathbf{1}_N = 0,
\end{aligned} \tag{2.17}$$

independently of the choice of N. Also, from (2.11) and (2.9),

$$\begin{aligned}
a^\dagger(\mathbf{r}')a(\mathbf{r})|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle &= a^\dagger(\mathbf{r}')N^{-1/2}\sum_{j=1}^{N} \delta(\mathbf{r}, \mathbf{r}_i) \\
&\quad \times (\pm 1)^{j-1}|\mathbf{r}_1\rangle \cdots |\mathbf{r}_{j-1}\rangle|\mathbf{r}_{j+1}\rangle \cdots |\mathbf{r}_N\rangle \\
&= \sum_{j=1}^{N} \delta(\mathbf{r}, \mathbf{r}_j)(\mathbf{r}_1\rangle \cdots |\mathbf{r}_{j-1}\rangle\,|\mathbf{r}'\rangle|\mathbf{r}_{j+1}\rangle \cdots |\mathbf{r}_N\rangle
\end{aligned} \tag{2.18}$$

and

$$a(\mathbf{r})a^\dagger(\mathbf{r}')|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle = a(\mathbf{r})(N+1)^{1/2}|\mathbf{r}'\rangle|\mathbf{r}_1\rangle \cdots |\mathbf{r}_N\rangle$$

$$= (N+1)\langle \mathbf{r}|\mathbf{r}'\rangle|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$$
$$= \delta(\mathbf{r},\mathbf{r}')|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle \pm \sum_{j=1}\delta(\mathbf{r},\mathbf{r}_j) \times|\mathbf{r}_1\rangle\cdots|\mathbf{r}_{j-1}\rangle|\mathbf{r}'\rangle|\mathbf{r}_{j+1}\rangle\cdots|\mathbf{r}_N\rangle. \quad (2.19)$$

From (2.18) and (2.19), in the petit space,

$$[a(\mathbf{r})a^\dagger(\mathbf{r}') \mp a^\dagger(\mathbf{r}')a(\mathbf{r})]\mathbf{1}_N = \delta(\mathbf{r},\mathbf{r}')\mathbf{1}_N, \quad (2.20)$$

independently of the choice of N. Since the grand Fock space is the direct sum of the petit spaces for $N \geq 0$, the unit operator $\mathbf{1}$ in the grand space is therefore the direct sum

$$\mathbf{1} = \sum_{N=0}^{\infty}\mathbf{1}_N, \quad (2.21)$$

so that the commutation relations (2.17) and (2.20) hold also in the grand space with $\mathbf{1}_N$ replaced by $\mathbf{1}$.

According to (2.4), we may write the definitions (2.9), (2.10) as

$$a(\mathbf{r})\mathbf{1}_N = N^{1/2}\langle \mathbf{r}|\mathbf{1}_N, \quad a^\dagger(\mathbf{r})\mathbf{1}_{N-1} = N^{1/2}|\mathbf{r}\rangle\mathbf{1}_{N-1}, \quad (2.22)$$

so that for $1 \leq n \leq N$,

$$a(\mathbf{r}_1)a(\mathbf{r}_2)\cdots a(\mathbf{r}_n)\mathbf{1}_N = [N!/(N-n)!]^{1/2}\langle \mathbf{r}_1|\cdots\langle \mathbf{r}_n|\mathbf{1}_N,$$
$$\mathbf{1}_N a^\dagger(\mathbf{r}_n)\cdots a^\dagger(\mathbf{r}_1) = [N!/(N-n)!]^{1/2}\mathbf{1}_N|\mathbf{r}_n\rangle\cdots|\mathbf{r}_1\rangle. \quad (2.23)$$

In particular, for $n = N$,

$$\mathbf{1}_N|\mathbf{r}_N\rangle\cdots|\mathbf{r}_1\rangle = (N!)^{-1/2}\mathbf{1}_N a^\dagger(\mathbf{r}_N)\cdots a^\dagger(\mathbf{r}_1)$$
$$= (N!)^{-1/2}a^\dagger(\mathbf{r}_N)\cdots a^\dagger(\mathbf{r}_1)|0\rangle\langle 0|, \quad (2.24)$$

which expresses the petit-space basis vector $|\mathbf{r}_N\rangle\cdots|\mathbf{r}_1\rangle$ as resulting from the action of creation operators on the vacuum state $|0\rangle$. The unit operator $\mathbf{1}_N$ to the left of a product of N creation operators, like the vacuum state $|0\rangle$ to the right, is not redundant; one implies the other. The petit space for $N = 0$ contains only one vector $|0\rangle$, such that

$$\mathbf{1}_0 = |0\rangle\langle 0|, \quad \text{trace } \mathbf{1}_0 = \langle 0|0\rangle = 1,$$
$$a(\mathbf{r})\mathbf{1}_0 = 0, \quad \langle \mathbf{r}|\mathbf{1}_0 = 0, \quad (2.25)$$
$$a^\dagger(\mathbf{r})\mathbf{1}_0 = |\mathbf{r}\rangle\mathbf{1}_0 = \mathbf{1}_1|\mathbf{r}\rangle.$$

3. OCCUPATION-NUMBER BASIS

In the special case $\mathbf{r}' = \mathbf{r}$ in (2.18),

$$a^\dagger(\mathbf{r})a(\mathbf{r})|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle = \sum_{j=1}^{N}\delta(\mathbf{r},\mathbf{r}_j)|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle. \quad (3.1)$$

Accordingly, $|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$ is an eigenvector of the number operator $\hat{N}(\mathbf{r})$ for state $|\mathbf{r}\rangle$, defined as the Hermitian operator

$$\hat{N}(\mathbf{r}) = a^\dagger(\mathbf{r})a(\mathbf{r}), \quad (3.2)$$

belonging to eigenvalue

$$N(\mathbf{r}) = \sum_{j=1}^{N}\delta(\mathbf{r},\mathbf{r}_j). \quad (3.3)$$

$N(\mathbf{r})$ is the occupation number for the state $|\mathbf{r}\rangle$ in the product $|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$, as introduced in (2.13). The total number operator $\hat{N}$ is defined as

$$\hat{N} = \sum_{\mathbf{r}}\hat{N}(\mathbf{r}) \quad (3.4)$$

so that, for any vector of the petit space,

$$\hat{N}|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle = \sum_{\mathbf{r}}\sum_{j=1}^{N}\delta(\mathbf{r},\mathbf{r}_j)(\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$$
$$= N|\mathbf{r}_1\rangle\ldots|\mathbf{r}_N\rangle, \quad (3.5)$$

or

$$\hat{N} = N\mathbf{1}_N. \quad (3.6)$$

The set of number operators $\{\hat{N}(\mathbf{r})\}$ for all of the one-particle states $|\mathbf{r}\rangle$ is a complete commuting set of observables, whose simultaneous eigenvectors form a nondegenerate, orthonormal basis in Fock space. (The case of degeneracy is considered in Sec. 6.) The vectors in this basis can be labeled by the set of eigenvalues $\{N(\mathbf{r})\}$ to which they belong. In the petit Fock space the total number of particles N must also be specified since the condition

$$N = \sum_{\mathbf{r}}N(\mathbf{r}) \quad (3.7)$$

imposes a constraint on the allowed values of $\{N(\mathbf{r})\}$; in the grand Fock space, N is variable so that (3.7) is not a constraint, but merely determines the value of N in terms of $\{N(\mathbf{r})\}$. The basis $|N\{N(\mathbf{r})\}\rangle$ is the occupation-number basis for the petit space.

A correspondence exists between vectors in the two bases. As seen in (3.1) and (3.5) every ket of the product-vector basis is an eigenket of $\{\hat{N}(\mathbf{r})\}$ and $\hat{N}$. But the correspondence is not one-to-one. We may write, according to (2.24), for a product-vector ket in the petit space,

$$|\mathbf{r}'_1\rangle\cdots|\mathbf{r}'_N\rangle$$
$$= (\pm 1)^{\pi'}(|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\alpha\rangle)(|\mathbf{r}_\beta\rangle\cdots|\mathbf{r}_\beta\rangle)\cdots(|\mathbf{r}_\sigma\rangle\cdots|\mathbf{r}_\sigma\rangle)\cdots$$
$$= (N!)^{-1/2}(\pm 1)^{\pi'}[a^\dagger(\mathbf{r}_\alpha)]^{N'(\mathbf{r}_\alpha)}\cdots[a^\dagger(\mathbf{r}_\sigma)]^{N'(\mathbf{r}_\sigma)}\cdots\mathbf{1}_0 \quad (3.8)$$

where $\mathbf{r}_\alpha, \mathbf{r}_\beta, \ldots, \mathbf{r}_\sigma, \ldots$ are values (all different) of the single-particle eigenvalue spectrum with occupation numbers $N'(\mathbf{r}_\alpha)\cdots N'(\mathbf{r}_\sigma)\cdots$ satisfying the petit-space constraint (3.7), and π' is the parity of the permutation required to write $|\mathbf{r}'_1\rangle\ldots|\mathbf{r}'_N\rangle$ in this form. It is assumed that $|\mathbf{r}_\alpha\rangle$ lies to the left of $|\mathbf{r}_\beta\rangle$, $|\mathbf{r}_\beta\rangle$ to the left of $|\mathbf{r}_\gamma\rangle$, etc., in accordance with some *a priori*, "natural" ordering of the single-particle states $|\mathbf{r}\rangle$. Similarly, if $|\mathbf{r}''_1\rangle\cdots|\mathbf{r}''_N\rangle$ is another ket of the same product-vector basis, it can be written in the same ordered way, but with parity π'' and occupation numbers $N''(\mathbf{r}_\alpha)\cdots N''(\mathbf{r}_\sigma)\cdots$. According to (2.6),

$$\langle \mathbf{r}''_N|\cdots\langle \mathbf{r}''_1|\mathbf{r}'_1\rangle\cdots|\mathbf{r}'_N\rangle$$
$$= (\pm 1)^{\pi'+\pi''}\{(N'(\mathbf{r}_\alpha)!\cdots N'(\mathbf{r}_\sigma)!\cdots]/N!\}\,\delta(\{N''(\mathbf{r})\},\{N'(\mathbf{r})\}). \quad (3.9)$$

The occupation-number basis vector $|N\{N(\mathbf{r})\}\rangle$ can be identified as

$$|N\{N(\mathbf{r})\}\rangle = (\pm 1)^{\pi}(N!/\prod_{\mathbf{r}_\sigma}N(\mathbf{r}_\sigma)!)^{1/2}|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$$
$$= (N!/\prod_{\mathbf{r}_\sigma}N(\mathbf{r}_\sigma)!)^{1/2}(|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\alpha\rangle) \times(|\mathbf{r}_\beta\rangle\cdots|\mathbf{r}_\beta\rangle)\cdots(|\mathbf{r}_\sigma\rangle\cdots|\mathbf{r}_\sigma\rangle)\cdots$$
$$= (\prod_{\mathbf{r}_\sigma}N(\mathbf{r}_\sigma)!)^{-1/2}[a^\dagger(\mathbf{r}_\alpha)]^{N(\mathbf{r}_\alpha)}\cdots[a^\dagger(\mathbf{r}_\sigma)]^{N(\mathbf{r}_\sigma)}\cdots\mathbf{1}_0, \quad (3.10)$$

where π is the parity of the permutation required to put $|\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle$ in "natural" order. According to (3.9),

$$\langle N\{N''(\mathbf{r})\}|N\{N'(\mathbf{r})\}\rangle = \delta(\{N''(\mathbf{r})\},\{N'(\mathbf{r})\}), \tag{3.11}$$

and

$$\sum_{\{N(\mathbf{r})\}(N)} |N\{N(\mathbf{r})\}\rangle\langle N\{N(\mathbf{r})\}| = 1_N. \tag{3.12}$$

The summation over sets of values of the occupation numbers $\{N(\mathbf{r})\}(N)$ is restricted by the constraint (3.7). If this constraint is removed by a further summation over N, we obtain

$$\sum_{\{N(\mathbf{r})\}} |\{N(\mathbf{r})\}\rangle\langle\{N(\mathbf{r})\}| = 1, \tag{3.13}$$

the completeness relation for the grand Fock space.

From (3.10) and the definition of the creation operator (2.9),

$$\begin{aligned} a^\dagger(\mathbf{r}_\mu)|N\{N(\mathbf{r})\}\rangle &= (N!/\prod_{\mathbf{r}_\sigma} N(\mathbf{r}_\sigma)!)^{1/2} \\ &\times (N+1)^{1/2}|\mathbf{r}_\mu\rangle|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\sigma\rangle\cdots|\mathbf{r}_\sigma\rangle \\ &= [N(\mathbf{r}_\mu)+1]^{1/2}(\pm 1)^{\mu-1}|N+1, N(\mathbf{r}_\mu)+1, \{N(\mathbf{r}\neq\mathbf{r}_\mu)\}\rangle. \end{aligned} \tag{3.14}$$

And, from (3.10) and (2.15),

$$\begin{aligned} a(\mathbf{r}_\mu)|N\{N(\mathbf{r})\}\rangle &= (N!/\prod_{\mathbf{r}_\sigma} N(\mathbf{r}_\sigma)!)N^{-1/2}\sum_{\mathbf{r}_\sigma}(\pm 1)^{\sigma-1} \\ &\times N(\mathbf{r}_\sigma)\delta(\mathbf{r}_\mu,\mathbf{r}_\sigma)(|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\alpha\rangle\cdots|\mathbf{r}_\sigma\rangle\cdots|\mathbf{r}_\sigma\rangle\cdots)' \\ &= [N(\mathbf{r}_\mu)]^{1/2}(\pm 1)^{\mu-1}|N-1, N(\mathbf{r}_\mu)-1, \{N(\mathbf{r}\neq\mathbf{r}_\mu)\}\rangle. \end{aligned} \tag{3.15}$$

In the fermion case,

$$\mu - 1 = \sum_{\sigma<\mu} N(\mathbf{r}_\sigma),$$

the number of occupied one-particle states up to the μ^{th}, arranged in the "natural" order as in (3.10). Equations (3.14) and (3.15) are the usual expressions for creation and destruction operators in the occupation-number basis.[9]

4. REPRESENTATION OF OPERATORS

An operator A_N in the petit space of N particles can be written in the occupation-number r-representation, according to (3.12), as

$$A_N = \sum_{\{N(\mathbf{r})\}(N)}\sum_{\{N'(\mathbf{r})\}(N)} |N\{N(\mathbf{r})\}\rangle A_N(\{N(\mathbf{r})\},\{N'(\mathbf{r})\}) < N\{N'(\mathbf{r})\}| \tag{4.1}$$

with

$$A_N(\{N(\mathbf{r})\},\{N'(\mathbf{r})\}) = \langle N\{N(\mathbf{r})\}|A_N|N\{N'(\mathbf{r})\}\rangle.$$

In the product-vector r-representation, according to (2.4),

$$A_N = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_N} |\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle \times A_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_N)|\langle\mathbf{r}'_N|\cdots\langle\mathbf{r}'_1|, \tag{4.2}$$

with

$$\begin{aligned} &A_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_N) \\ &= \text{trace}\,(A_N|\mathbf{r}'_1\rangle\cdots|\mathbf{r}'_N\rangle\langle\mathbf{r}_N|\cdots\langle\mathbf{r}_1|) \\ &= \langle\mathbf{r}_N|\cdots\langle\mathbf{r}_1|A_N|\mathbf{r}'_1\rangle\cdots|\mathbf{r}'_N\rangle. \end{aligned} \tag{4.3}$$

According to (2.24),

$$A_N = (N!)^{-1}\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_N} a^\dagger(\mathbf{r}_1)\cdots a^\dagger(\mathbf{r}_N) \times A_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_N)a(\mathbf{r}'_N)\cdots a(\mathbf{r}'_1). \tag{4.4}$$

For example, the unit operator 1_N of the petit space in (2.4) is expressed in the form (4.2) or (4.4) with

$$\begin{aligned} 1_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_N) &= \langle\mathbf{r}_N|\cdots\langle\mathbf{r}_1|\mathbf{r}'_1\rangle\cdots|\mathbf{r}'_N\rangle \\ &= \delta(\mathbf{r}_1\cdots\mathbf{r}_N)\prod_{j=1}^{N}\delta(\mathbf{r}_j,\mathbf{r}'_j), \end{aligned} \tag{4.5}$$

according to (2.6) and (2.7).

From (4.2) and (4.3), for two operators A_N and B_N in the petit space,

$$\text{trace}\,(A_N B_N) = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_N} A_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_N) \times B_N(\mathbf{r}'_N\cdots\mathbf{r}'_1,\mathbf{r}_1\cdots\mathbf{r}_N). \tag{4.6}$$

In a system of N particles the physical properties may depend on groups of one, two, or more particles. The operators describing these properties can be classified as one-particle, two-particle, etc., and designated by $A_N(1), A_N(2)$, etc. An n-particle property has the form

$$A_N(n) = \sum_{i_n<\cdots<i_2}\cdots\sum\sum_{i_1=1}^{N} A_n(i_1,i_2\cdots i_n)1_N. \tag{4.7}$$

But since the particles are indistinguishable, the n-particle operator $A_n(i_1\cdots i_n)$ is independent of the particle labels $i_1\cdots i_n$, so that

$$\begin{aligned} A_n(n) &= \sum_{i_n<\cdots<i_1}\cdots\sum A_n 1_N = [N!/(N-n)!n!]A_n 1_N \\ &= [N!/(N-n)!n!]\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_n} 1_N|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle \\ &\times A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_n)\langle\mathbf{r}'_n|\cdots\langle\mathbf{r}'_1|1_N \end{aligned} \tag{4.8}$$

$$\begin{aligned} &= (1/n!)\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_n} a^\dagger(\mathbf{r}_1)\cdots a^\dagger(\mathbf{r}_n) \\ &\times A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_n)a(\mathbf{r}'_n)\cdots a(\mathbf{r}'_1)1_N, \end{aligned} \tag{4.9}$$

according to (2.23). As an operator in the petit space of N particles $A_N(n)$ appears in (4.9) as the product of an operator expressing a property of n particles,

$$A_n^{gr} = \frac{1}{n!}\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}'_1}\cdots\sum_{\mathbf{r}'_n} a^\dagger(\mathbf{r}_1)\cdots a^\dagger(\mathbf{r}_n) \times A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}'_1\cdots\mathbf{r}'_n)a(\mathbf{r}'_n)\cdots a(\mathbf{r}'_1), \tag{4.10}$$

independent of $N \geq n$, and the unit operator 1_N of the petit space. Accordingly, A_n^{gr} is the operator representing the n-particle property in the grand Fock space. For $n = N$ we recover (4.4). According to (4.10), for example, the kinetic energy of a system of particles, which is a one-particle or additive property ($n = 1$), has the form

$$T = \sum_{\mathbf{r}}\sum_{\mathbf{r}'} T(\mathbf{r},\mathbf{r}')a^\dagger(\mathbf{r})a(\mathbf{r}'),$$

while the two-particle interaction potential energy ($n = 2$) becomes

$$V = \tfrac{1}{2}\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2}\sum_{\mathbf{r}'_1}\sum_{\mathbf{r}'_2} V(\mathbf{r}_2\mathbf{r}_1,\mathbf{r}'_1\mathbf{r}'_2)a^\dagger(\mathbf{r}_1)a^\dagger(\mathbf{r}_2)a(\mathbf{r}'_2)a(\mathbf{r}'_1).$$

A change in basis from the product-vector r-basis to a product-vector p-basis, where $|\mathbf{p}\rangle$ is a vector of an

orthonormal set of one-particle states different from the set $|\mathbf{r}\rangle$ used up to this point, is readily performed. In the new basis, with obvious notation,

$$A_N(n) = \frac{1}{n!}\sum_{\mathbf{p}_1}\cdots\sum_{\mathbf{p}_n}\sum_{\mathbf{p}_1'}\cdots\sum_{\mathbf{p}_n'} a^\dagger(\mathbf{p}_1)\cdots a^\dagger(\mathbf{p}_n) \times A_n(\mathbf{p}_n\cdots\mathbf{p}_1,\mathbf{p}_1'\cdots\mathbf{p}_n')a(\mathbf{p}_n')\cdots a(\mathbf{p}_1')\mathbf{1}_N \quad (4.11)$$

$$= \frac{N!}{(N-n)!n!}\sum_{\mathbf{p}_1}\cdots\sum_{\mathbf{p}_n}\sum_{\mathbf{p}_1'}\cdots\sum_{\mathbf{p}_n'} |\mathbf{p}_1\rangle\cdots|\mathbf{p}_n\rangle \times A_n(\mathbf{p}_n\cdots\mathbf{p}_1,\mathbf{p}_1'\cdots\mathbf{p}_n')\langle\mathbf{p}_n'|\cdots\langle\mathbf{p}_1'|1_N. \quad (4.12)$$

Comparison with the form (4.8) gives

$$A_n(\mathbf{p}_n\cdots\mathbf{p}_1,\mathbf{p}_1'\cdots\mathbf{p}_n') = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_n'} \langle\mathbf{p}_n|\cdots\langle\mathbf{p}_1|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_n') \times\langle\mathbf{r}_n'|\cdots\langle\mathbf{r}_1'|\mathbf{p}_1'\rangle\cdots\mathbf{p}_n'\rangle = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_n'} A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_n')\prod_{j=1}^{n}\langle\mathbf{p}_j|\mathbf{r}_j\rangle\langle\mathbf{r}_j'|\mathbf{p}_j'\rangle. \quad (4.13)$$

The last step of (4.13) uses (2.5) and the transposition symmetry of $A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_n')$ as defined in (4.3).

The effect of a change in basis on the creation and destruction operators follows from (2.23). We have the invariant,

$$\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\mathbf{1}_N a^\dagger(\mathbf{r}_n)\cdots a^\dagger(\mathbf{r}_1)\langle\mathbf{r}_1|\cdots\langle\mathbf{r}_n| = [N!/(N-n)!]^{1/2}\mathbf{1}_N = \sum_{\mathbf{p}_1}\cdots\sum_{\mathbf{p}_n}\mathbf{1}_N a^\dagger(\mathbf{p}_n)\cdots a^\dagger(\mathbf{p}_1)\langle\mathbf{p}_1|\cdots\langle\mathbf{p}_n|. \quad (4.14)$$

In particular, for $n = 1$,

$$\sum_{\mathbf{r}} a^\dagger(\mathbf{r})\langle\mathbf{r}| = \sum_{\mathbf{p}} a^\dagger(\mathbf{p})\langle\mathbf{p}|, \quad (4.15)$$

for any value of N, so that

$$a^\dagger(\mathbf{p}) = \sum_{\mathbf{r}} a^\dagger(\mathbf{r})\langle\mathbf{r}|\mathbf{p}\rangle, \qquad a(\mathbf{p}) = \sum_{\mathbf{r}} a(\mathbf{r})\langle\mathbf{p}|\mathbf{r}\rangle. \quad (4.16)$$

If the p vectors are taken to be coordinate-spin eigenvectors, (4.16) agrees with the definitions of field or quantized wavefunctions.[10]

5. DENSITY AND REDUCED DENSITY OPERATORS

These operators have been discussed by Bogoliubov.[5] His treatment utilizes "double action" operators acting on the second-quantized wavefunction on the one hand, and on the ordinary wavefunction of the molecular complex on the other. A far simpler treatment can be given, utilizing the product-vector basis described here in Sec. 2.

The density operator in the petit space of N particles has the form given in (4.2):

$$\rho_N = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_N'} |\mathbf{r}_1\rangle\cdots|\mathbf{r}_N\rangle \times\rho_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_N')\langle\mathbf{r}_N'|\cdots\langle\mathbf{r}_1'|, \quad (5.1)$$

with

$$\text{trace}\,\rho_N = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\rho_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}_1\cdots\mathbf{r}_N) = 1. \quad (5.2)$$

The expectation of an operator A_N in the state specified by ρ_N is, according to (4.6),

$$\langle A_N\rangle_N = \text{trace}(A_N\rho_N) = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_N}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_N'} A_N(\mathbf{r}_N\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_N')\rho_N(\mathbf{r}_N'\cdots\mathbf{r}_1',\mathbf{r}_1\cdots\mathbf{r}_N). \quad (5.3)$$

If A_N^{gr} is a collective property depending on n particles (molecular complex) as in (4.10), then, for $1 \le n \le N$, the expectation of A_N^{gr} in state ρ_N is

$$\langle A_n^{gr}\rangle_N = \text{trace}(A_n^{gr}\rho_N) = \frac{N!}{(N-n)!n!}\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_n'}A_n(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}_1'\cdots\mathbf{r}_n') \times\text{trace}(1_N|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle\langle\mathbf{r}_n'|\cdots\langle\mathbf{r}_1'|\rho_N). \quad (5.4)$$

Define the reduced density operator for n particles,

$$\rho_N(n) = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\sum_{\mathbf{r}_1'}\cdots\sum_{\mathbf{r}_n'}|\mathbf{r}_1'\rangle\cdots|\mathbf{r}_n'\rangle \times\rho_N(n)(\mathbf{r}_n'\cdots\mathbf{r}_1',\mathbf{r}_1\cdots\mathbf{r}_n)\langle\mathbf{r}_n|\cdots\langle\mathbf{r}_1|1_n \quad (5.5)$$

acting in the space of n particles, where

$$\rho_N(n)(\mathbf{r}_n'\cdots\mathbf{r}_1',\mathbf{r}_1\cdots\mathbf{r}_n) = \text{trace}(1_N|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle\langle\mathbf{r}_n'|\cdots\langle\mathbf{r}_1'|\rho_N) \quad (5.6)$$

by definition, and also, from (5.5),

$$\rho_N(n)(\mathbf{r}_n'\cdots\mathbf{r}_1',\mathbf{r}_1\cdots\mathbf{r}_n) = \text{trace}[1_n|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle\langle\mathbf{r}_n'|\cdots\langle\mathbf{r}_1'|\rho_N(n)]. \quad (5.7)$$

According to the definitions (5.5), (5.6), the reduced density $\rho_N(n)$ has the same normalization as ρ_N, since

$$\text{trace}[\rho_N(n)] = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\rho_N(n)(\mathbf{r}_n\cdots\mathbf{r}_1,\mathbf{r}_1\cdots\mathbf{r}_n) = \sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_n}\text{trace}[1_N|\mathbf{r}_1\rangle\cdots|\mathbf{r}_n\rangle\langle\mathbf{r}_n|\cdots\langle\mathbf{r}_1|\rho_N] = \text{trace}\rho_N = 1. \quad (5.8)$$

From (5.4),

$$\langle A_n^{gr}\rangle_N = [N!/(N-n)!n!]\,\text{trace}[A_n\rho_N(n)]. \quad (5.9)$$

We may also define reduced distribution functions

$$f_N(n) = [N!/(N-n)!n!]\rho_N(n), \qquad \text{trace}\,[f_N(n)] = N!/(N-n)!n!, \quad (5.10)$$

which are normalized to the number of groups of n particles in the system of N particles. Then from (5.9),

$$\langle A_n^{gr}\rangle_N = \text{trace}[A_n f_N(n)]. \quad (5.11)$$

For example, the number operator for the state $|\mathbf{r}\rangle$ in the petit space is

$$\hat{N}(\mathbf{r}) = a^\dagger(\mathbf{r})a(\mathbf{r})1_N = N|\mathbf{r}\rangle\langle\mathbf{r}|1_N = N\delta(\hat{\mathbf{r}},\mathbf{r})1_N, \quad (5.12)$$

where $\hat{\mathbf{r}}$ represents the complete commuting set of one-particle operators with simultaneous eigenvectors $|\mathbf{r}\rangle$, and

$$\delta(\hat{\mathbf{r}},\mathbf{r}) = \int d\mathbf{u}\,\exp[2\pi i\mathbf{u}\cdot(\hat{\mathbf{r}}-\mathbf{r})].$$

In a continuous spectrum of eigenvalues $\mathbf{r}$, the Dirac delta $\delta(\hat{\mathbf{r}} - \mathbf{r})$ replaces the Kronecker delta, so that $\hat{N}(\mathbf{r})$ becomes the number density operator. The expectation of $\hat{N}(\mathbf{r})$ in the state ρ_N is

$$\langle \hat{N}(\mathbf{r}) \rangle_N = N \operatorname{trace}(\rho_N |\mathbf{r}\rangle\langle\mathbf{r}|) = N\rho_N(1)(\mathbf{r},\mathbf{r}) = f_N(1)(\mathbf{r},\mathbf{r}). \tag{5.13}$$

Likewise a pair number operator $\hat{N}(\mathbf{r}_1,\mathbf{r}_2)$ can be defined as

$$\hat{N}(\mathbf{r}_1,\mathbf{r}_2) = \tfrac{1}{2}a^\dagger(\mathbf{r}_1)a^\dagger(\mathbf{r}_2)a(\mathbf{r}_2)a(\mathbf{r}_1)1_N = \tfrac{1}{2}N(N-1)|\mathbf{r}_1\rangle|\mathbf{r}_2\rangle\langle\mathbf{r}_2|\langle\mathbf{r}_1|1_N, \tag{5.14}$$

with expectation

$$\langle \hat{N}(\mathbf{r}_1,\mathbf{r}_2)\rangle_N = \tfrac{1}{2}N(N-1)\operatorname{trace}(\rho_N|\mathbf{r}_1\rangle|\mathbf{r}_2\rangle\langle\mathbf{r}_2|\langle\mathbf{r}_1|) = f_N(2)(\mathbf{r}_1\mathbf{r}_2,\mathbf{r}_2\mathbf{r}_1), \tag{5.15}$$

the number of pairs of particles: one in state $|\mathbf{r}_1\rangle$, the other in state $|\mathbf{r}_2\rangle$.

6. EFFECTS OF DEGENERACY IN THE BASIS OF ONE-PARTICLE STATES

In the preceding sections we have required that the one-particle basis $|\mathbf{r}\rangle$ consist of simultaneous eigenvectors of a complete set of commuting observables $\hat{\mathbf{r}}$, so that the basis $|\mathbf{r}\rangle$ is nondegenerate. In this section we wish to consider the effects of degeneracy. Let the one-particle observable A_1 be an operator which commutes with the complete commuting set $\hat{\mathbf{r}}$. We write A_1 in diagonal form as

$$A_1 = \sum_\alpha |\alpha > a < \alpha|. \tag{6.1}$$

The set of eigenvectors $|\alpha\rangle$ is complete and orthonormal in the one-particle space,

$$\sum_\alpha |\alpha\rangle\langle\alpha| = 1_1, \quad \langle\alpha'|\alpha''\rangle = \delta(\alpha',\alpha''). \tag{6.2}$$

The basis vectors $|\mathbf{r}\rangle$ are also eigenvectors of A_1. They can be classified according to the value of α to which they belong. Let $|\mathbf{r}(\alpha)\rangle$ denote a vector $|\mathbf{r}\rangle$ belonging to the eigenvalue α. Then

$$|\alpha\rangle\langle\alpha| = \sum_{\mathbf{r}(\alpha)} |\mathbf{r}(\alpha)\rangle\langle\mathbf{r}(\alpha)| = 1_1(\alpha), \tag{6.3}$$

and the trace of $1_1(\alpha)$ gives

$$\operatorname{trace}[1_1(\alpha)] = \sum_{\mathbf{r}(\alpha)} 1 = g(\alpha), \tag{6.4}$$

the degeneracy of the eigenvalue α, the number of one-particle states $|\mathbf{r}(\alpha)\rangle$. The operator A_1, diagonal in the basis $|\mathbf{r}\rangle$, can also be written as in (4.2),

$$A_1 = \sum_{\mathbf{r}} |\mathbf{r}\rangle A_1(\mathbf{r},\mathbf{r})\langle\mathbf{r}| = \sum_\alpha \alpha 1_1(\alpha), \tag{6.5}$$

with

$$A_1(\mathbf{r},\mathbf{r}) = \langle\mathbf{r}|A_1|\mathbf{r}\rangle \quad \text{and} \quad A_1[\mathbf{r}(\alpha),\mathbf{r}(\alpha)] = \alpha. \tag{6.6}$$

It follows that

$$\operatorname{trace}[A_1 1_1(\alpha)] = \alpha \operatorname{trace}[1_1(\alpha)] = \alpha g(\alpha). \tag{6.7}$$

Unless A_1 itself represents a complete commuting set of observables (which must therefore be identical with the set $\hat{\mathbf{r}}$) the spectrum of A_1 is degenerate; $g(\alpha)$ is not equal to 1 for all α. However, $g(\alpha)$ cannot equal zero for any value of α since the $|\mathbf{r}\rangle$ basis is complete. Accordingly,

$$g(\alpha) \geq 1.$$

Now consider an additive N-particle property $A_N(1)$, from (4.8) for $n = 1$. We shall see how the degeneracy in the spectrum of A_1, the operator of (6.1), is reflected in that of $A_N(1)$. Define creation and destruction operators for the states $|\alpha\rangle$ in the petit space of N particles according to (2.9) and (2.10),

$$a^\dagger(\alpha) = N^{1/2}|\alpha\rangle, \quad a(\alpha) = N^{1/2}\langle\alpha|, \tag{6.8}$$

with the number operator

$$\hat{N}(\alpha) = a^\dagger(\alpha)a(\alpha) = N|\alpha\rangle\langle\alpha|. \tag{6.9}$$

Then $A_N(1)$ can be written in the diagonal forms, corresponding to (6.5),

$$A_N(1) = \sum_{\mathbf{r}} A_1(\mathbf{r},\mathbf{r})\hat{N}(\mathbf{r}) = \sum_\alpha \alpha\hat{N}(\alpha), \tag{6.10}$$

where $\hat{N}(\mathbf{r})$ was defined in (5.12), and from (6.3),

$$\hat{N}(\alpha) = \sum_{\mathbf{r}(\alpha)} \hat{N}[\mathbf{r}(\alpha)]. \tag{6.11}$$

When the one-particle operator A_1 is diagonal in the basis $|\mathbf{r}\rangle$, it follows from (6.10) that the additive operator $A_N(1)$ is diagonal in the occupation-number basis $|N\{N(\mathbf{r})\}\rangle$, since

$$\hat{N}(\mathbf{r}) = \sum_{\{N(\mathbf{r})\}(N)} |N\{N(\mathbf{r})\rangle N(\mathbf{r})\langle N\{N(\mathbf{r})\}|, \tag{6.12}$$

in the notation (3.12).

In the petit space of N particles the set of number operators $\hat{N}(\mathbf{r})$ for the one-particle states $|\mathbf{r}\rangle$ is a complete set of commuting observables, with simultaneous eigenvectors which are the non-degenerate basis vectors $|N\{N(\mathbf{r})\}\rangle$ in (3.10). The operators $\hat{N}(\alpha)$ and $A_N(1)$ commute with this set, according to (6.10), (6.11) and (6.12). We may write $A_N(1)$ in its own representation as

$$A_N(1) = \sum_{\mathfrak{A}} |\mathfrak{A}\rangle\mathfrak{A}\langle\mathfrak{A}|. \tag{6.13}$$

The set of eigenvectors $|\mathfrak{A}\rangle$ is complete and orthonormal in the petit space,

$$\sum_{\mathfrak{A}} |\mathfrak{A}\rangle\langle\mathfrak{A}| = 1_N, \quad \langle\mathfrak{A}'|\mathfrak{A}''\rangle = \delta(\mathfrak{A}',\mathfrak{A}''). \tag{6.14}$$

If the basis vectors $|N\{N(\mathbf{r})\}\rangle$ are classified according to the value of $\mathfrak{A}$ to which they belong, then

$$|\mathfrak{A}\rangle\langle\mathfrak{A}| = \sum_{\{N(\mathbf{r})\}(N,\mathfrak{A})} |N,\mathfrak{A},\{N(\mathbf{r})\}\rangle\langle N,\mathfrak{A},\{N(\mathbf{r})\}| = 1_N(\mathfrak{A}), \tag{6.15}$$

where the summation runs over those sets of occupation numbers satisfying both the petit-space condition (3.7),

$$N = \sum_{\mathbf{r}} N(\mathbf{r}) = \sum_\alpha \sum_{\mathbf{r}(\alpha)} N[\mathbf{r}(\alpha)] = \sum_\alpha N(\alpha), \tag{6.16}$$

and the "microensemble" condition,

$$\mathfrak{A} = \sum_{\mathbf{r}} A_1(\mathbf{r},\mathbf{r})N(\mathbf{r}) = \sum_\alpha \sum_{\mathbf{r}(\alpha)} A_1[\mathbf{r}(\alpha),\mathbf{r}(\alpha)]N[\mathbf{r}(\alpha)] = \sum_\alpha \alpha N(\alpha), \tag{6.17}$$

from (6.6). The degeneracy of the eigenvalue $\mathfrak{A}$ is given by

$$\text{trace } |\mathfrak{A}\rangle\langle\mathfrak{A}| = \sum_{\{N(\mathbf{r})\}\,(N,\mathfrak{A})} \text{trace}[|N,\mathfrak{A},\{N(\mathbf{r})\}\rangle\langle N,\mathfrak{A},\{N(\mathbf{r})\}|] = \sum_{\{N(\mathbf{r})\}\,(N,\mathfrak{A})} 1 = g(\mathfrak{A}). \tag{6.18}$$

The degeneracy $g(\mathfrak{A})$ is the number of basis vectors $|N\{N(\mathbf{r})\}\rangle$ of the petit space belonging to the set $|N,\mathfrak{A},\{N(\mathbf{r})\}\rangle$ consistent with a given eigenvalue $\mathfrak{A} = \sum_\alpha \alpha N(\alpha)$. $N(\alpha)$ is the occupation number for the degenerate eigenstate α, with $\sum_\alpha N(\alpha) = N$. Each different set of values $\{N(\alpha)\}$ will give some one of the eigenvalues $\mathfrak{A}$. Therefore, the degeneracy $g(\mathfrak{A})$ is equal to the number of distributions of N particles, with $N(\alpha)$ particle occupying the $g(\alpha)$ one-particle states belonging to eigenvalue α, consistent with the constraints (6.16) and (6.17). This is a well-known combinatorial problem.[11]

For bosons, the number of distributions with repetition of $N(\alpha)$ particles among $g(\alpha)$ states is

$$D_+\binom{N(\alpha)}{g(\alpha)} = \frac{[N(\alpha)+g(\alpha)-1]!}{N(\alpha)![g(\alpha)-1]!}. \tag{6.19}$$

For fermions, the number of distributions without repetition is

$$D_-\binom{N(\alpha)}{g(\alpha)} = \frac{g(\alpha)!}{N(\alpha)![g(\alpha)-N(\alpha)]!}. \tag{6.20}$$

$D_\pm$ are binomial coefficients given by the generating functions

$$(1+x)^g = \sum_{N=0}^{g} x^N D_-\binom{N}{g}, \qquad (1-x)^{-g} = \sum_{N=0} x^N D_+\binom{N}{g}. \tag{6.21}$$

The degeneracies $g(\mathfrak{A})$ in the spectrum of $A_N(1)$ are given accordingly as

$$g_\pm(\mathfrak{A}) = \sum_{\{N(\alpha)\}\,(N,\mathfrak{A})} \prod_\alpha D_\pm\binom{N(\alpha)}{g(\alpha)}, \tag{6.22}$$

In terms of the degeneracies $g(\alpha)$ in the spectrum of A_1. Applications are familiar to the equilibrium ensembles in the statistical mechanics of an ideal system whose Hamiltonian is an additive, $A_N(1)$ property.

---

*This work was supported by a grant from the Research Foundation of the State University of New York.

[1]V. Fock, Z. Phys. 75, 622 (1932).
[2]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, Evanston, Illinois, 1961), Chap. 6.
[3]A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), Chap. 1.
[4]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1947), 3rd ed., p. 57.
[5]N. N. Bogoliubov, *Lectures on Quantum Statistics, Volume 1: Quantum Statistics* (Gordon & Breach, New York, 1967).
[6]Cf. Ref. 2: p. 136, Eqs. (78c) for bosons; p. 139, Eq. (97) for fermions.
[7]Reference 1, Zweiter Teil, Eq. 7.
[8]Reference 2, p. 129, Eq. (38) and p. 138, Eq. (86).
[9]Reference 2, p. 131, Eqs. (47), (51); p. 139, Eqs. (89), (90).
[10]Reference 2, p. 132, Eqs. (63), (64); p. 139, Eqs. (95a), (95b).
[11]F. Perrin, *Mécanique statistique quantique, Fascicule V (Traité du calcul des probabilités et ses applications) par Émile Borel, Tome II* (Gauthier-Villars, Imprimeur-Editeur, Paris, 1939).