Functional calculus for the symmetric multigroup transport operator

William Greenberg

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
(Received 21 July 1975)

A rigorous treatment of the symmetric multigroup transport equation is given by developing the functional calculus for the transport operator. Von Neumann spectral theory is applied to nonorthogonal cyclic subspaces, and the isometries onto $C(N)$ are explicitly evaluated.

Hangelbroek and Larsen and Habetler have independently provided rigorous techniques for solving the time independent one speed linear transport equation.[1,2] While the Larsen—Habetler approach has the distinct advantage of demonstrating, as its central result, that the transport operator is spectral, when this result is already evident, as for example, with a self-adjoint kernel, the Hangelbroek approach appears to provide a most effective setting for understanding the underlying properties of the transport operator. In fact, Hangelbroek has succeeded in showing that the Wiener—Hopf factorization, which has been used as a basis for extending the solution of the full range problem to the half range, can be derived from a study of projections in the representation space of the full range theory.[3]

Recently, Zweifel has extended the Larsen—Habetler technique to the multigroup transport equation.[4,5] In this article we wish to show that for a symmetric kernel the functional calculus can be developed for the multigroup as by Hangelbroek for the one-speed equation.

Since it will be necessary to evaluate the isometries between subspaces of the solution space and the representation spaces explicitly, the von Neumann spectral theory will be applied to nonorthogonal cyclic subspaces. The subcritical case, $C < S$, is considered in detail first, with extensions in the last section to more general kernels.

1. THE ALGEBRA GENERATED BY $A^{-1}u$

We consider the Hilbert space $H = \oplus_{i=1}^{n} L^2(I)$, the direct sum of n copies of $L^2(I)$, where I is the real interval $[-1,1]$. For $1 \leq i \leq n$, let $e_i \in H$ be the zero function in each $L^2(I)$ except the ith copy, where it is the unit constant function. A vector $\psi \in H$ will be written $\psi = \{\psi_i\}_{i=1}^{n}$ with $(\psi,\phi)_H = \sum_{i=1}^{n} \int_{-1}^{1} du\, \psi_i(u)\, \bar{\phi}_i(u)$.

Let S be a positive, diagonal $n \times n$ matrix, C a real symmetric matrix, and assume for simplicity that $S - C > 0$. Throughout we will write σ_i for S_{ii}. Define the orthogonal projection $P: H \to H$ by

$$P\phi = \tfrac{1}{2} \sum_{i=1}^{n} (\phi, e_i)_H e_i,$$

and the bounded operators $A: H \to H$ and $M: H \to H$ by

$$A = S - CP,$$

$$(M\phi)(u) = u\phi(u), \quad u \in I.$$

Noting that P commutes with every constant matrix, the inverse of A is computed to be

$$A^{-1} = S^{-1} + S^{-1}C(S - C)^{-1}P.$$

Write K for the linear space H with inner product

$$\{\phi,\psi\} = (A\phi,\psi)_H. \tag{1}$$

On K, A is still positive, and additionally, $B = A^{-1}M$ is self-adjoint. Let $\mathcal{A}$ be the C^* algebra generated by B on K, $C(N)$ the C^* algebra of continuous, $\mathbb{C}$-valued functions on the spectrum $N = \sigma(B)$ of B with uniform norm, $P(H)$ [resp. $P(\mathcal{A})$, $P(C)$] the subspace of polynomials in each component in K [resp. $\mathcal{A}, C$], and E the unit constant function in $C(N)$. For each integer i, $1 \leq i \leq n$, define $M_i \subset K$ by $M_i = P(\mathcal{A})e_i$.

Lemma 1: If $T \in P(\mathcal{A})$, then $Te_i \in P(K)$ for each i, and the degree satisfies:

$$\deg T = \deg(Te_i, e_i)_H,$$

$$\deg T \geq \deg(Te_i, e_j)_H + 2, \quad j \neq i, \quad \deg T > 1.$$

Denoting by L.S. the linear manifold spanned, we have

$$M_i \cap \{\text{L.S.} \bigcup_{j \neq i} M_j\} = \phi,$$

$$\text{L.S.} \bigcup_{j=1}^{n} M_j = P(K).$$

Proof: If $T \in P(\mathcal{A})$ and $\deg T = n$, then

$$BTe_i = \sigma_i^{-1} uTe_i + S^{-1}C(S - C)^{-1}P(uTe_i),$$

and the first part follows by induction, with

$$Be_i = u\sigma_i^{-1}e_i, \quad B^2e_i = u^2\sigma_i^{-2}e_i + \tfrac{2}{3}S^{-1}C(S - C)^{-1}S^{-1}e_i.$$

For the last part it is sufficient that, for each i, all polynomials $\varphi \in K$ with $(\varphi, e_j) = 0$ unless $j = i$ can be obtained. Assume, by induction, that all such polynomials for all i and $\deg\varphi \leq N$ can be obtained, and assume $\psi = (a_{N+1}u^{N+1} + P_N(u))e_i$. If

$$a_{N+1}\sigma_i^{N+1}B^{N+1}e_i = a_{N+1}u^{N+1}e_i + Q_N(u)e_i + \sum_{j \neq i} R_{N-1,j}(u)e_j,$$

let

$$\sum_j S_{N,j}(B)e_j = Q_N(u)e_i,$$

$$\sum_j T_{N-1,j}(B)e_j = \sum_{j \neq i} R_{N-1,j}(u)e_j,$$

$$\sum_j U_{N,j}(B)e_j = P_N(u)e_i,$$

for polynomials Q_N, $R_{N-1,j}$, $S_{N,j}$, $T_{N-1,j}$, and $U_{N,j}$ of

indicated degree. Then

$$a_{N+1}\sigma_i^{N+1}B^{N+1}e_i+\sum_j[U_{N,j}(B)-S_{N,j}(B)-T_{N-1,j}(B)]e_j=\psi.$$

For each i, define $\pi_i:\mathcal{A}\to\bar{M}_i$ by π_i: $T\to Te_i$ and let κ: $\mathcal{A}\to C(N)$ be the Gelfand isomorphism. If $T\in\mathcal{A}$ and $\pi_i(T)=0$, then $T\mathcal{A}e_i=\mathcal{A}Te_i=0$ so $T=0$ on $\mathcal{A}e_i$. Since $T=0$ on $M_i^\perp$, this proves:

Lemma 2: π_i is a bounded linear isomorphism of $\mathcal{A}$ onto $\mathcal{A}e_i$.

Define F_i: $K\to C(N)$ by $F_i=\kappa\pi_i^{-1}$ on $\mathcal{A}e_i$ and $F=0$ on $M_i^\perp$. Let $C_i(N)$ be the linear space $C(N)$ with inner product

$$(\varphi,\psi)_i=\{F_i^{-1}\varphi,F_i^{-1}\psi\}.$$

Then F_i is an isometric linear isomorphism on $\mathcal{A}e_i$, and $F_i(e_i)=E$.

Lemma 3: For a unique positive Lebesgue–Stieltzes measure σ_i,

$$(\varphi,\psi)_i=\int_N\varphi(\nu)\bar{\psi}(\nu)\,d\sigma_i(\nu),$$

for $\varphi,\psi\in P(C)$.

Proof: Defining the linear functional l_i: $\varphi\to(\varphi,E)_i$ for $\varphi\in P(C)$, the estimate

$$\begin{aligned}|(\varphi,E)_i|&=|\{\kappa^{-1}\varphi e_i,e_i\}|\le\|\kappa^{-1}\varphi\|\{e_i,e_i\}\\&=\sup_{\nu\in N}|\varphi(\nu)|\{e_i,e_i\}\end{aligned}$$

proves $l_i\in P(C)^*$. Hence,

$$\begin{aligned}(\varphi,\psi)_i&=\{\varphi(B)e_i,\psi(B)e_i\}=\{\bar{\psi}(B)\varphi(B)e_i,e_i\}\\&=(\bar{\psi}\varphi,E)_i=\int_N\varphi(\nu)\bar{\psi}(\nu)\,d\sigma_i(\nu).\end{aligned}$$

If φ is positive, $\varphi(B)$ is a positive operator, by the spectral theorem for self-adjoint operators, so

$$\int_N\varphi(\nu)\,d\sigma_i(\nu)=(\varphi,E)_i=\{\varphi(B)e_i,e_i\}\ge 0.$$

By the Lemma, F_i extends to an isometric linear isomorphism of $\bar{M}_i$ onto $L^2(N,\sigma_i)$. Write $\hat{M}$ for the bounded operator on $L^2(N,\sigma_i)$,

$$(\hat{M}\varphi)(\nu)=\nu\varphi(\nu),\quad\varphi\in L^2(N,\sigma_i).$$

Corollary: For any $\varphi\in K$, $F_i(B\varphi)=\hat{M}F_i(\varphi)$.

For $F_i(B\varphi)=\nu\varphi(\nu)$ if $\varphi=\psi(B)e_i\in\bar{M}_i$, and $M_i^\perp$ is an invariant subspace of B.

2. SOLUTION OF THE EQUATION

Let $N_\pm$ represent the nonnegative/nonpositive subsets of N, and $P_\pm$ the orthogonal projections of $L^2(N,\sigma)$ onto $L^2(N_\pm,\sigma)$, viewed as subspaces of $L^2(N,\sigma)$, $P_\pm L^2(N_\mp,\sigma)=0$. The solution of the n-group isotropic nonhomogeneous linear transport equation is provided by the following theorem.

Theorem 1: Let $q:\mathbb{R}\to K$ be uniformly Hölder continuous, and $(q(x),q(x))_\mu$ uniformly bounded. Consider $A^{-1}q=\sum_{i=1}^n q_i$ with $q_i:\mathbb{R}\to\bar{M}_i$. Define

$$\varphi_{i\pm}(x)=\int_{-\infty}^{\infty}\exp[-(x-\xi)\nu^{-1}](P_\pm F_iq_i)(\xi)\,d\xi.\tag{2}$$

Then

$$\psi(x)=M^{-1}A\sum_{i=1}^n F_i^{-1}(\varphi_{i+}+\varphi_{i-})$$

is the unique solution of the transport equation

$$\frac{d}{dx}M\psi(x)=A\psi(x)+q(x),\tag{3}$$

satisfying $(M\psi(x),\psi(x))_\mu$ uniformly bounded.

The proof of the theorem is an immediate consequence of the following two lemmas.

Lemma 4: $\psi_i:\mathbb{R}\to K$ is a solution of

$$B\frac{d}{dx}\psi_i(x)=-\psi_i(x)+q_i(x),\tag{4}$$

satisfying $(M\psi_i(x),\psi_i(x))_\mu$ uniformly bounded if and only if $\varphi_i(x)=F_iA^{-1}M\psi_i$ is a solution of

$$\frac{d}{dx}\varphi_i(x)=-\hat{M}^{-1}\varphi_i(x)+F_iq_i(x),\tag{5}$$

satisfying $(\varphi_i(x),\varphi_i(x))_i$ uniform bounded.

Lemma 5[2]: Suppose $g:\mathbb{R}\to X$ is a uniformly Hölder continuous function from $\mathbb{R}$ to the Banach space X, and $\|g(x)\|$ is uniformly bounded. If $\beta<0$ and $-(T+\beta)$ is the generator of a bounded holomorphic semigroup,[6] then

$$\varphi(x)=\int_{-\infty}^{x}\left\{\frac{1}{2\pi i}\int\Gamma\exp[\lambda(x-\xi)]\frac{1}{-T+\lambda}\,d\lambda\right\}g(x)\,dx$$

for a contour Γ about $\sigma(T)$, is continuously differentiable in $\mathbb{R}$, and is the unique solution of

$$\frac{d}{dx}\varphi(x)=-T\varphi(x)+g(x),$$

satisfying $\|\varphi(x)\|$ uniformly bounded.

Lemma 4 is established immediately by the isomorphisms F_i, and Lemma 5, the generalization of a result from Hille-Yosida semigroup theory, is proved in Ref. 2.

If $\{\psi_i\}$ satisfies Eq. (4), then $\psi=\sum_{i=1}^n\psi_i$ is clearly a solution of the transport equation (3), so the problem reduces to a consideration of Eq. (5). Since $L^2(N_\pm,\sigma)$ are invariant subspaces of $\hat{M}$, it is sufficient to solve the equations on each of the subspaces. But on $L^2(N_\pm,\sigma)$, the restrictions of $\hat{M}^{-1}$ are semi-bounded self-adjoint operators, and Lemma 5 is applicable. Equation (2) results trivially from evaluating the contour integral in Lemma 5 for $T=1/\nu$.

3. EVALUATION OF THE ISOMORPHISMS F_i

It remains to compute the maps F_i and F_i^{-1} explicitly in order to apply Theorem 1, and it is desirable as well to derive the measure σ.

Lemma 6: Let ν_l denote the poles of $\Lambda(\lambda)^{-1}$ for $l=1,\ldots,m$ and $R(\nu_l)$ the residues. If $\varphi\in P(C)$, then

$$\begin{aligned}(F_i^{-1}\varphi)(\mu)=&\sum_{l=1}^m\varphi(\nu_l)(\nu_lI-\mu S^{-1})^{-1}R(\nu_l)e_i+(1/2\pi i)\\&\times P\int_{-1}^{1}d\nu\varphi(\nu)(\nu I-\mu S^{-1})^{-1}(\Lambda^{-1}(\nu)^--\Lambda^{-1}(\nu)^+)e_i\\&+\tfrac{1}{2}\tau_S(\varphi\Lambda^{-1+}+\varphi\Lambda^{-1-})(\mu)e_i,\end{aligned}$$

where $\Lambda^{-1\pm}$ are the boundary values of Λ^{-1} along the cut I, and for any function $W:\mathbb{R}\to L(K)$ from $\mathbb{R}$ to bounded

operators on K,

$(e_k, (\tau_S W)(x)\xi)_H = (e_k, W(x/\sigma_k)\xi)_H$,
for all $1 \leq k \leq n$ and $\xi \in K$.

Proof: Since $B - \lambda I = \Lambda(\lambda)(S^{-1}u - \lambda I)$, F_i^{-1} can be evaluated on analytic functions by a contour integral of the resolvent applied to the constant vectors e_i, i.e.,

$$(F_i^{-1}\varphi)(\mu) = (\varphi(B)e_i)(\mu)$$
$$= (1/2\pi i)\int_\Gamma d\lambda\, \varphi(\lambda)(\lambda I - S^{-1}\mu)^{-1}\Lambda^{-1}(\lambda)e_i. \qquad (7)$$

Computation of the contour integral is routine, contributions rising from the same terms as in the one group case.

For $\varphi, \psi \in K$, let

$$[\varphi, \psi] = \sum_{k=1}^{n} \varphi_k \psi_k.$$

Lemma 7: If $\varphi \in P(\mathcal{A})e_i$ and $\Omega: \mathbb{C} \to \mathcal{L}(K)$ satisfies

(a) $[\Lambda^{-1}(\lambda)e_i, (\tau_S^{-1}\Omega(\lambda))e_i] = 1$ on a neighborhood of N,

(b) $[e_k, \Omega(\lambda)e_i]$ analytic on the complement of I,

then

$$(F_i\varphi)(\nu) = (1/2\pi i)\int_\Gamma d\lambda[\varphi(\lambda), (\lambda I - \nu S)^{-1}\Omega(\lambda)e_i], \quad \nu \in N$$

for a closed contour Γ about N.

Proof: If Γ' is a contour about N inside Γ,

$$\hat{\varphi}(\nu) = (1/2\pi i)\int_{\Gamma'} d\lambda'[\varphi(\lambda'), (\lambda' I - \nu S)^{-1}\Omega(\lambda')e_i]$$
$$= (1/2\pi i)\int_\Gamma d\lambda (F\varphi)(\lambda)(1/2\pi i)\sum_{k=1}^{n}\int_{\Gamma'} d\lambda'(1/\sigma_k\lambda - \lambda')$$
$$\times [S\Lambda^{-1}(\lambda)e_i, e_k][(\lambda' I - \nu S)^{-1}\Omega(\lambda')e_i, e_k],$$

using Lemma 6. From property (b) and the known analytic behavior of $\Lambda(\lambda)$, the Γ' integral can be evaluated

$$\hat{\varphi}(\nu) = (1/2\pi i)\sum_{k=1}^{n}\int_\Gamma d\lambda (F\varphi)(\lambda)[S\Lambda^{-1}(\lambda)e_i, e_k]$$
$$\times [(S\lambda - \nu S)^{-1}(\tau_S^{-1}\Omega)(\lambda)e_i, e_k]$$
$$= (1/2\pi i)\int_\Gamma d\lambda (F\varphi)(\lambda)(1/\lambda - \nu)\sum_{k=1}^{n}[\Lambda^{-1}(\lambda)e_i, e_k]$$
$$\times [(\tau_S^{-1}\Omega)(\lambda)e_i, e_k) = (F\varphi)(\nu).$$

Corollary: $\Omega(\lambda)$ defined by

$$\Omega = (\tau_S\Lambda)^t,$$

satisfies Lemma 7.

Proof: For any $\lambda \in \mathbb{C}$,

$$[\tau_S^{-1}(\tau_S\Lambda)^t(\lambda)e_i, e_k] = [(\tau_S\Lambda)^t(\lambda\sigma_k)e_i, e_k]$$
$$= [\tau_S\Lambda(\lambda\sigma_k)e_k, e_i] = [\Lambda(\lambda)e_k, e_i].$$

The subspace spanned by $\{e_j\}_{j=1}^n$ is an invariant subspace of $\Lambda(\lambda)$, and the restriction $\Lambda_c(\lambda)$ of Λ is a constant matrix. Then for $\lambda \notin N$,

$$\sum_k [\Lambda(\lambda)^{-1}e_i, e_k][\Lambda(\lambda)e_k, e_i] = \sum_k \Lambda_c(\lambda)_{ik}\Lambda_c(\lambda)^{-1}_{ki} = I_{ii}.$$

Also,

$$\Lambda_c(\lambda)_{ik} = \left[\left(I + D\int_{-1}^{1} d\mu \frac{\sigma_k\mu}{\mu - \lambda\sigma_k}\right)e_k, e_i\right]$$
$$= \delta_{ik} + D_{ik}\int_{-1}^{1} d\mu \frac{\sigma_k\mu}{\mu - \lambda\sigma_k}$$

where $D = S^{-1}C(S - C)^{-1}$, so

$$(\Lambda_c\,(\lambda/\sigma_k)^t)_{ki} = \delta_{ik} + \sigma_k D_{ik}\int_{-1}^{1}(\mu/\mu - \lambda)\, d\mu,$$

which is analytic for $\lambda \in \mathbb{C}/I$.

Corollary: If $\varphi \in P(\mathcal{A})e_i$, then

$$(F_i\varphi)(\nu) = \tfrac{1}{2}[\tau_S^{-1}\varphi(\nu), \tau_S^{-1}(\Omega^+(\nu) + \Omega^-(\nu))e_i]\delta_{\nu\neq\nu_l}$$
$$+ (1/2\pi i)P\int_{-1}^{1} d\lambda[\varphi(\lambda), (\lambda I - \nu S)^{-1}(\Omega^-(\lambda) - \Omega^+(\lambda))e_i],$$

where

$$\delta_{\nu\neq\nu_l} = \begin{cases} 1, & -1 < \nu < 1 \\ 0, & \nu = \nu_l \end{cases}$$

and $(e_k, \tau_S^{-1}\varphi(\mu)e_i) = \varphi_k(\sigma_k\nu)$.

To compute σ conveniently, it is helpful to collect some properties of Λ.

Lemma 8: For $\lambda \in \mathbb{C}$,

$$\Lambda_c(\lambda)^- - \Lambda_c(\lambda)^+ = 2\pi i\lambda DSW_\lambda, \qquad (8)$$

where

$$(W_\lambda)_{kk} = \begin{cases} 1, & |\lambda| \leq 1/\sigma_k \\ 0, & |\lambda| > 1/\sigma_k \end{cases}$$

is diagonal. For $\lambda \notin N$,

$$\Lambda^{-1}(\lambda) = I - (I + DT_\lambda)^{-1}DPM(S^{-1}M - \lambda I)^{-1}, \qquad (9)$$

where

$$(T_\lambda)_{kk} = \tfrac{1}{2}\int_{-1}^{1}(\sigma_k\mu/\mu - \lambda\sigma_k)\, d\mu$$

is diagonal. Hence,

$$\Lambda^{-1}(0)^\pm = I - S^{-1}CP. \qquad (10)$$

Proof: Since $\Lambda(\lambda) = I + DP\mu(S^{-1}\mu - \lambda I)^{-1}$, $\Lambda_c(\lambda) = I + DT_\lambda$ and $I - DT_\lambda$ is invertible if $\lambda \notin N$. Multiplying Eq. (9) by Λ and suitably collecting terms proves the operator inversions:

$$\Lambda\Lambda^{-1} = I - \{(I + DT_\lambda)^{-1}D + D - DT_\lambda(I + DT_\lambda)^{-1}D\}P_\mu(S^{-1}\mu - \lambda I)^{-1}$$
$$= I - \{-(I + DT_\lambda)(I + DT_\lambda)^{-1} + I\}DP\mu(S^{-1}\mu - \lambda I)^{-1}$$
$$= I = \Lambda^{-1}\Lambda.$$

Since

$$\Lambda^{-1}(\lambda) = I - (I + DT_\lambda)^{-1}DSP - (I + DT_\lambda)^{-1}DS\lambda P(S^{-1}\mu - \lambda I)^{-1}$$

and $T_0 = S$,

$$\Lambda^{-1}(0) = I - (I + S^{-1}C(S - C)^{-1}S)^{-1}S^{-1}C(S - C)^{-1}SP$$
$$= I - (S^{-1}(S - C)C^{-1}S + I)^{-1}P.$$

Finally,

$$\Lambda(\lambda) = I + DPSS^{-1}\mu(S^{-1}\mu - \lambda I)^{-1} = I + DSP(S^{-1}\mu - \lambda I)^{-1},$$

and thus,

$$\Lambda_c(\lambda)^{\pm}_{jk} = (I + DS + \lambda DSP\int_{-1}^{1} d\mu\,(S^{-1}\mu - \lambda I)^{-1} \pm \lambda DS\pi i\delta(|\lambda| \le 1/\sigma_k))_{jk}$$

Theorem 9: For each $1 \le j \le n$, and $1 \le i \le n$,

$$\overline{(F_i e_j)}(\nu)\,d\sigma(\nu) = \begin{cases} \left(\frac{1}{\nu}(S-C)CS^{-1}D(\nu)e_i, e_j\right)_{\mathcal{H}}, & \nu = \nu_l, \\ ((S-C)C^{-1}S\Lambda^{-1}(\nu)^{-}DSW_\nu\Lambda^{-1}(\nu)^{+}e_i, e_j)_{\mathcal{H}}\,d\nu, & -1 < \nu < 1. \end{cases}$$

Proof: Since F_i^{-1} is an isometry, for all $\varphi \in P(\mathbb{C})$,

$$\begin{aligned}\int_N \varphi(\nu)F_i e_j(\nu)\,d\sigma(\nu) &= (\varphi, F_i e_j)_N = \{F_i^{-1}\varphi, F_i^{-1}F_i e_j\} \\ &= \{\varphi(B)e_i, e_j\} \\ &= (1/2\pi i)\int_\Gamma d\lambda\,\varphi(\lambda)\{(\lambda I - B)^{-1}e_i, e_j\}.\end{aligned}$$

Rewrite Λ as in the proof of Lemma 8 to obtain:

$$\begin{aligned}\{(\lambda I - B)^{-1}e_i, e_j\} &= \{(\lambda I - S^{-1}\mu)^{-1}\Lambda(\lambda)^{-1}e_i, e_j\} \\ &= ((\lambda I - S^{-1}\mu)^{-1}\Lambda(\lambda)^{-1}e_i, Ae_j)_{\mathcal{H}} \\ &= (1/\lambda)(P\lambda(\lambda I - S^{-1}\mu)^{-1}\Lambda(\lambda)^{-1}e_i, (S-C)e_j) \\ &= (1/\lambda)((C^{-1}S - I + P \\ &\quad - (C^{-1}S - I)\Lambda(\lambda))\Lambda(\lambda)^{-1}e_i, (S-C)e_j)_{\mathcal{H}} \\ &= (1/\lambda)((S-C)C^{-1}S\Lambda(\lambda)^{-1}e_i, e_j) \\ &\quad - (1/\lambda)((S-C)(C^{-1}S - I)e_i, e_j).\end{aligned}$$

Therefore, the integral can be expanded. The contour integration is completely analogous to that in the proof of Lemma 7, i.e.,

$$\begin{aligned}&\int_N \varphi(\nu)\overline{F_i e_j}(\nu)\,d\sigma(\nu) \\ &= \frac{1}{2\pi i}\int_\Gamma d\lambda\frac{\varphi(\lambda)}{\lambda}((S-C)C^{-1}S\,\Lambda^{-1}(\lambda)e_i, e_j) \\ &\quad - \frac{1}{2\pi i}\int_\Gamma d\lambda\frac{\varphi(\lambda)}{\lambda}((S-C)(C^{-1}S - I)e_i, e_j) \\ &= \sum_l \frac{\varphi(\nu_l)}{\nu_l}((S-C)C^{-1}SR(\nu_l))_{ji} + \frac{1}{2\pi i}P\int_{-1}^{1} d\nu\frac{\varphi(\nu)}{\nu} \\ &\quad \times((S-C)C^{-1}S(\Lambda^{-1}(\nu)^{-} - \Lambda^{-1}(\nu)^{+}))_{ji} \\ &\quad + \tfrac{1}{2}\varphi(0)((S-C)C^{-1}S(\Lambda^{-1}(0)^{+} \\ &\quad + \Lambda^{-1}(0)^{-})_{ji} - ((S-C)(C^{-1}S - I))_{ji}\varphi(0). \qquad (11)\end{aligned}$$

Equation (10) of Lemma 8 gives

$$\tfrac{1}{2}(S-C)C^{-1}S(\Lambda^{-1}(0)^{+} + \Lambda^{-1}(0)^{-}) = (S-C)(C^{-1}S - I)$$

so the last two terms of Eq. (11) cancel. Moreover,

$$\begin{aligned}\Lambda^{-1}(\nu)^{-} - \Lambda^{-1}(\nu)^{+} &= -\Lambda^{-1}(\nu)^{-}(\Lambda(\nu)^{-} - \Lambda(\nu)^{+})\Lambda^{-1}(\nu)^{+} \\ &= 2\pi i\nu\Lambda^{-1}(\nu)^{-}DSW_\nu\Lambda^{-1}(\nu)^{+}\end{aligned}$$

from Eq. (8). Since $P(\mathbb{C}) \subset L^2(N, \sigma)$ densely, the integrands are equal.

Note that Theorem 9 gives $\sigma(\nu)$, since $F_i e_i = e_i \equiv 1$.

4. EXTENSION TO SELF-ADJOINT KERNEL

The functional calculus approach which has been developed above can be extended in toto to the case of the general symmetric kernel: S diagonal, $C = C^*$, $\det(S - C) \neq 0$ (or C similar to a self-adjoint matrix, S invariant), by considering $\mathcal{H}$ as a Pontrjagin space with indefinite metric defined by Eq. (1). More precisely, for every pair of imaginary eigenvalues, a two-dimensional invariant subspace is split off from $\mathcal{H}$, and for each real eigenvalue such that $(A\varphi, \varphi)_{\mathcal{H}} < 0$, a one-dimensional eigenspace is removed. On the remainder, B is similar to a self-adjoint operator. This decomposition is due in general to Krein's Invariant Subspace theorem.

In the case of the 2-group, the assumption $C = C^*$ can also be dropped. For it is evident that C can always be symmetrized by a similarity transformation which leaves the diagonal matrix S invariant.

For details on these Pontrjagin techniques, see Refs. 3 and 7.

ACKNOWLEDGMENT

I am indebted to J. C. T. Pool, Associate Director of the Applied Mathematics Division, Argonne National Laboratory, for his kind hospitality during my visit, at which time part of this research was carried out.

[1] E. W. Larsen, and G. J. Habetler, Commun. Pure Appl. Math. **26**, 525 (1973).
[2] R. Hangelbroek, "A Functional Analytic Approach to the Linear Transport Equation," thesis, Groningen (1973).
[3] R. Hangelbroek, "Hilbert Space Approach to the Linear Transport Equation for Multiplying Media in Slab Geometry," in *The Fourth Conference in Transport Theory*, to appear.
[4] S. Sancaktar and P. Zweifel, "Multigroup Neutron Transport: Full-range," preprint.
[5] R. Bowden, S. Sancaktar, and P. Zweifel, "Multigroup Neutron Transport: Half-range," preprint.
[6] T. Kato, *Perturbation Theory For Linear Operators* (Springer, New York, 1966).
[7] W. Greenberg, "Pontrjagin Structure of the Transport Operator," in *The Fourth Conference in Transport Theory*, to appear.

The Case eigenfunction expansion for a conservative medium

William Greenberg

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060

Paul F. Zweifel*

Istituto di Matematica Applicata, Università degli Studi di Firenze, Firenze, Italy
(Received 25 August 1975)

By using the resolvent integration technique introduced by Larsen and Habetler, the one-speed, isotropic scattering, neutron transport equation is treated in the infinite and semi-infinite media. It is seen that the results previously obtained by Case's "singular eigenfunction" approach are in agreement with those obtained by resolvent integration.

I. INTRODUCTION

The linear transport equation with $c=1$ was treated by Shure and Natelson,[1] who used the Case singular eigenfunction approach.[2] Larsen and Habetler[3] later rederived Case's formulas using a contour integration technique which was not subject to some of the criticism which had been levelled at Case's approach through the years, mainly that the derivations were in fact only heuristic arguments. However, Larsen and Habetler were unable to treat the conservative case, $c=1$, but claimed (Ref. 3, p. 536) that the results for that specific case could be obtained by taking the limit $c\to 1$ in their derivation for $c\neq 1$. This contention has recently been attacked by Kaper,[4] and since Kaper's remarks seem to have merit, we present here the explicit analysis, along the lines developed in Ref. 3, for the case $c=1$. This case, incidentally, which corresponds to a critical half-space in neutron transport theory, has more physical significance in the context of radiative transport theory in stellar atmospheres, where it corresponds to a gray, conservative atmosphere in local thermodynamic equilbrium. (See Ref. 2, Sec. 10.5.)

An alternative to the Larsen–Habetler analysis was independently developed by Hangelbroek,[5] who proved that for $c<1$ the transport operator was similar to a self-adjoint operator, and so was able to apply von Neumann spectral theory. Lekkerkerker[6] has extended Hangelbroek's work to the case $c=1$ by defining a suitable subspace of the original Hilbert space, on which the transport operator is similar to a self-adjoint operator, obtaining a spectral theorem for the restriction of the transport operator to this subspace, and finally extending the results to the full space.

Our technique, following Larsen–Habetler, was inspired by Lekkerkerker. Specifically, the Larsen–Habetler technique fails for $c=1$ because the transport operator, K^{-1} in their notation, is not invertible. However, a suitable restriction of K^{-1} is invertible, and the entire Larsen–Habetler method of analysis can be carried out for this restriction. The extension of the results to the full space is then almost trivial. We feel that our analysis has some advantages over that of Ref. 6, in that it is considerably shorter and simpler, and in addition, is not restricted to a Hilbert space. Furthermore, the Larsen–Habetler technique appears to have some real advantages over both the Case and Hangelbroek methods in the analysis of the multigroup transport equation,[7,8] and it is planned to use techniques similar to those reported here to attempt to extend the multigroup results, which are so far restricted to the subcritical medium (but see Ref. 9).

II. THE RESOLVENT OPERATOR AND THE FULL RANGE EXPANSION

As in Ref. 6, we consider the one speed transport equation with isotropic scattering for a conservative medium, $c=1$, i.e.,

$$u\frac{\partial\psi}{\partial x}(x,u)+uK^{-1}\psi(x,u)=q(x,u),\quad u\neq 0 \tag{1a}$$

with

$$(K^{-1}f)(x,u)=(1/u)[f(x,u)-\tfrac{1}{2}\int_{-1}^{1}f(x,u')\,du']. \tag{1b}$$

A solution of Eq. (1) is understood to be a differentiable function $\psi:\mathbb{R}\to X_p$, $p>1$, where X_p is the Banach space of functions $f:[-1,1]\to\mathbb{C}$ satisfying

$$\|f\|_p=\left(\int_{-1}^{1}|uf(u)|^p\,du\right)^{1/p}<\infty,$$

and the vector $\psi(x)$ has been written $\psi(x,u)$. The nonhomogeneous term $q(x,u)$ is specified with $(1/u)q(x,u)\in X_p$.

Equation (1b) defines a densely defined, closed, unbounded, noninvertible operator $K^{-1}:X_p\to X_p$ with domain $D(K^{-1})=\{f\in X_p\,|\,f=ug,\ g\in X_p\}$. The choice of X_p-norm has the result that the operator $K^{-1}=u^{-1}A$ corresponds, for $p=2$, essentially to the product Au^{-1} of operators on L_2 used by Kaper[10] for a related problem in the kinetic theory of gases, rather than the product $u^{-1}A$ used by Lekkerkerker in Ref. 6. In fact, the unitary transformation $U:X_2\to L_2$ given by $(Uf)(u)=uf(u)$, transforms K^{-1} into $UK^{-1}U^{-1}=uu^{-1}Au^{-1}=Au^{-1}$. This avoids considerable difficulties encountered in Ref. 6; in particular, in our treatment, $D(u^{-1})=D(K^{-1})$.

In most of the remainder, explicit x-dependence will not appear, as the transport operator K^{-1} is studied in X_p. This notation agrees, except for minor variations, with that of Refs. 3 and 2, Sec. 6.9. Note that the extension of the analysis of Ref. 3 to X_p for $c\neq 1$ has been given in Ref. 11 for $p>1$. While it appears that the forthcoming analysis could be carried out in X_1, that

would require substantial alteration of the technique.[12]

The essence of the Larsen–Habetler technique is to invert K^{-1} to obtain K, calculate the resolvent $(zI-K)^{-1}$, and then integrate the resolvent along a contour surrounding the spectrum of K. Application of the Cauchy theorem yields the so-called Case completeness theorem. This technique fails in the present case because K^{-1} is not invertible on its range. In fact, $\lambda=0$ is an eigenvalue of K^{-1} with eigenvector e_0 defined by

$$e_0(u)=1, \quad -1<u<1. \tag{2}$$

Furthermore,

$$K^{-1}e_1=e_0, \tag{3a}$$

where

$$e_1(u)=u, \quad -1<u<1. \tag{3b}$$

We shall see that e_0 and e_1 span the $\lambda=0$ root linear manifold of K^{-1}.

As explained in the Introduction, we now define a subspace $Y_p\subset X_p$ such that $K^{-1}|\,Y_p$ is invertible. To this end, define

$$Y_p=\{f\in X_p \mid \int_{-1}^{1}u^i f(u)\,du=0,\ i=1,2\}$$

and

$$Y_{p0}=\mathrm{Sp}\{e_0,e_1\}$$

Theorem 1: The direct sum decomposition $X_p=Y_p+Y_{p0}$ reduces K^{-1}.

Proof: The linear functionals

$$\rho_0: f\to\tfrac{3}{2}\int_{-1}^{1}u^2 f(u)\,du, \tag{4a}$$

$$\rho_1: f\to\tfrac{3}{2}\int_{-1}^{1}u f(u)\,du, \tag{4b}$$

have the property

$$\rho_i(e_j)=\delta_{ij}, \quad i,j=0,1.$$

Hence,

$$P: f\to\rho_0(f)e_0+\rho_1(f)e_1$$

is a continuous projection onto Y_{p0}, and Y_p is its topological supplement. The computation $\rho_i(K^{-1}f)=0$ for $f\in Y_p$ follows immediately from Eq. (1b), and since $PD(K^{-1})=Y_{p0}\subset D(K^{-1})$, the subspaces are reducing.

Theorem 2: $K^{-1}|\,Y_p$ is invertible, and its bounded inverse K is given by

$$Kg=ug-\tfrac{3}{2}\int_{-1}^{1}s^3 g(s)\,ds.$$

Proof: Consider the equation

$$K^{-1}f=g \text{ with } g\in Y_p.$$

This may be written

$$f-\tfrac{1}{2}\int_{-1}^{1}f(s)\,ds=ug.$$

If the equation is multiplied by u^2 and integrated over u from -1 to 1, one obtains

$$\int_{-1}^{1}f(s)\,ds=-3\int_{-1}^{1}u^3 g(u)\,du,$$

and the result follows.

Theorem 3: For $z\in\mathbb{C}/[-1,1]$ and $g\in Y_p$,

$$(zI-K)^{-1}g=\frac{1}{z-u}\left\{g-[1/2\Lambda(z)]\int_{-1}^{1}[sg(s)/s-z]\,ds\right\},$$

where

$$\Lambda(z)=\left[1-\tfrac{1}{2}\int_{-1}^{1}(z/z-s)\,ds\right]$$

is the usual dispersion function[2] for $c=1$.

Proof: The analysis of Ref. 3 can be followed to arrive at the result

$$(zI-K)^{-1}g=(1/z-u)\left\{g+\tfrac{3}{2}\left[\int_{-1}^{1}\frac{s^3 g(s)}{s-z}\,ds\right]\times\left[1+\tfrac{3}{2}\int_{-1}^{1}\frac{t^3}{z-t}\,dt\right]^{-1}\right\}.$$

Then the identities

$$u^3/(z-u)=-u^2-uz+uz^2/(z-u)$$
$$=-u^2-uz-z^2+z^3/(z-u),$$

can be used to simplify the two integrals in the expression, yielding the stated result.

Note that this expression for the resolvent is identical to that obtained in Ref. 3, and so a great deal of the analysis given there can be taken as verbatim.

The spectrum of K can be obtained immediately from the expression for the resolvent in Theorem 3: $\sigma(K)=[-1,1]$. Although $\Lambda(z)\sim-1/3z^2$ for large z, $\int_{-1}^{1}(sg(s)/s-z)\,ds\sim 1/z^3$ in the same limit, so the resolvent $(zI-K)^{-1}$ converges to zero at infinity, reflecting of course the boundedness of K. Thus,

$$I=(1/2\pi i)\int_{\Gamma}(zI-K)^{-1}\,dz,$$

where Γ is any closed contour surrounding the cut $[-1,1]$.

Since the Hölder continuous functions are dense in Y_p by an easy application of the Weierstrass theorem, if

$$H_p=\{f\in Y_p \mid f \text{ is of class } H^*\},$$

then H_p+Y_{p0} is dense in X_p. It is also easy to see that $H_p\cap D(K^{-1})$ is dense in Y_p. Here by "of class H^*" is meant[2,13] that f is Holder continuous on the interior of $[-1,1]$, i.e.,

$$|f(u)-f(u')|\le\text{const}\times|u-u'|^{\alpha}, \quad \alpha>0,$$

and also that f near the endpoints $b=\pm1$ of the interval is a product of a function Hölder continuous on $[-1,1]$ and the function $(u-b)^{\beta}$, $\beta>-1$. The Larsen–Habetler analysis utilizes the pointwise evaluation of the boundary values of certain analytic functions of z in the domain of the resolvent $(zI-K)^{-1}$. For that reason it is necessary to stay on the manifold H_p, and extend the final results as in Ref. 11.

Alternatively, we may have chosen to "compute" on functions Hölder continuous on the entire interval $[-1,1]$, whence the Case transforms $A(v)$, as well as $\lambda(v)A(v)$, would have vanished at the endpoints b [by virtue of the fact that $\lambda(v)/N(v)\to0$ at the boundaries; see Eq. (6)]. However, this would lead to no simplification of the arguments.

In this manner, the analysis of Ref. 3 yields results analogous to the case of $c<1$; i.e., for each $f\in H_p$ there exists $A\in X_p$ of class H^* satisfying:

$$f(u)=\int_{-1}^{1}A(v)\phi_v(u)\,dv, \tag{6a}$$

$$A(v)=\frac{1}{N(v)}\int_{-1}^{1}uf(u)\phi_v(u)\,du, \tag{6b}$$

where ϕ_v is the usual Case "singular eigenfunction" corresponding to $c=1$; namely,

$$\phi_v(u)=(v/2)P(1/v-u)+\tfrac{1}{2}[\Lambda^+(v)+\Lambda^-(v)]\delta(v-u) \tag{7}$$

and

$$N(v)=v\Lambda^+(v)\Lambda^-(v)$$

converges to infinity at the endpoints ± 1. The notation is the same as that of Refs. 2,3 and 11. In the language of Ref. 2, we would say that every $f\in H_p$ can be expanded in terms of the Case continuum eigenfunctions alone.

To deal with $f\in Y_{p0}$, write

$$f=\tfrac{1}{2}a_0-\tfrac{1}{2}a_1 \tag{8}$$

where the factors $\pm\frac{1}{2}$ have been introduced to conform with standard notation. Multiplying Eq. (8) by u or u^2 and integrating, one finds

$$a_i=3\int_{-1}^{1}(-u)^{2-i}f(u)\,du. \tag{9}$$

Let $\lambda(v)$ denote

$$\lambda(v)=\tfrac{1}{2}[\Lambda^+(v)+\Lambda^-(v)]. \tag{10}$$

We wish to show that the linear transformation $F: f\to\lambda A$ defined by Eq. (6b) for f of class H^*,

$$(Ff)(v)=[\lambda(v)/N(v)]\int_{-1}^{1}uf(u)\phi_v(u)\,du,$$

extends to an isomorphism $F: Y_p\to X_p$. Define $F': \psi\to f$, the natural candidate for F^{-1}, by

$$(F'\psi)(u)=\int_{-1}^{1}[\psi(v)/\lambda(v)]\phi_v(u)\,dv$$

for any ψ of class H^*. Equations (6a) and (6b) establish the relationship $F'F=I$ on H_p, which is dense in Y_p. We must ascertain, however, that F' is a bounded transformation into Y_p, or else the extension of F to all of Y_p might no longer be invertible. Moreover, it is necessary to prove that the range of F is dense in X_p in order to insure that the solution of a transport problem solved in terms of the transformed function $A(v)$ will be the image under F of a vector in X_p.[14]

In Ref. 2 it is shown that if f is of class H^*, then A will be of class H^*, and hence so will $\lambda(v)A(v)$. Furthermore, any A of class H^* will yield a function f of class H^* via Eq. (6a), since

$$f(u)=\lambda(u)A(u)+\tfrac{1}{2}P\int_{-1}^{1}[vA(v)/v-u]\,dv, \tag{11}$$

and the boundary values of the Cauchy integral of a class H^* function are also of class H^*.

In Ref. 11, the inequality

$$\int_{-1}^{1}|v\lambda(v)A(v)|^p\,dv\le M_p\int_{-1}^{1}|uf(u)|^p\,du, \tag{12}$$

where M_p is a constant depending upon p, proves that $\lambda A\in X_p$ if $f\in H_p$, and that F is a bounded transformation. Let

$$H_p^1=\{A\in X_p\,|\,\lambda A\in X_p \text{ of class } H^*\}.$$

Then the same argument used to derive Eq. (12) also yields

$$\int_{-1}^{1}|uf(u)|^p\,du\le M_p'\int_{-1}^{1}|v\psi(v)|^p\,dv,$$

for $\psi\in H_p^1$, which implies that F is one-one on H_p. Combining these remarks, we obtain bounded transformations F and F' on Y_p and X_p, respectively, with $F'F=I$ on Y_p, and $FF'=I$ on $\mathrm{Ran}(F)$.

A direct computation shows $F'\psi\in Y_p$ for $\psi\in H_p^1$. For example,

$$\begin{aligned}\rho_0(F'\psi)&=\int_{-1}^{1}[v\psi(v)/\lambda(v)]P\int_{-1}^{1}(u^2/v-u)\,du\,dv\\&\quad+\int_{-1}^{1}[\psi(v)/\lambda(v)]v^2\lambda(v)\,dv\\&=0,\end{aligned}$$

since

$$P\int_{-1}^{1}(u^2/v-u)\,du=-2v\lambda(v).$$

Thus, to prove $\mathrm{Ran}(F)$ is dense in X_p, suppose $A\in H_p^1$ and

$$0=\int_{-1}^{1}A(v)\phi_v(u)\,dv. \tag{13}$$

Defining

$$n(z)=\int_{-1}^{1}A(v)v/(v-z)\,dv, \tag{14}$$

expanding Eq. (13) as in Eq. (11), and using the Plemelj formulas with Eq. (14), yields

$$(1/2\pi i)[n^+(u)-n^-(u)]\lambda(u)+\tfrac{1}{2}[n^+(u)+n^-(u)](u/2)=0.$$

With the substitution

$$u=(1/\pi i)[\Lambda^+(u)-\Lambda^-(u)],$$

and Eq. (10), this becomes

$$(1/2\pi i)(n^+\Lambda^+-n^-\Lambda^-)=0,\quad -1<u<1. \tag{15}$$

If $J(z)$ is defined by

$$J(z)=n(z)\Lambda(z),$$

then Eq. (15) proves that J is an entire function. But $\Lambda(\infty)=n(\infty)=0$, so by Liouville's Theorem, $J\equiv 0$, which proves $A(v)\equiv 0$. Hence $FF'=I$ and $F'=F^{-1}$. Using the density of H_p and H_p^1 in Y_p and X_p, the transformations in Eqs. (6a) and (6b) may be extended by continuity to all of X_p.

The above results can be summarized in Theorem 4.

Theorem 4: Let $f\in X_p$. Then f has an eigenfunction expansion of the form

$$f=\tfrac{1}{2}a_0-\tfrac{1}{2}a_1u+\int_{-1}^{1}A(v)\phi_v(u)\,dv, \tag{16}$$

where a_i are given by Eq. (9), $A(v)$ is given by Eq. (6b), and ϕ_v is the Case singular eigenfunction defined in Eq. (7). The linear transformation $F: f\to\lambda A$ is an isomorphism $F: Y_p\to X_p$.

III. HALF RANGE EXPANSION

Let X_p' be the Banach space of functions $f: [0,1]\to C$ with

$$\|f\|_p'=\left[\int_0^1|uf(u)|^p\,du\right]^{1/p}<\infty.$$

The object for the half range theory is to find an operator $E: X_p'\to X_p$ with certain analyticity properties given below. Then the full range expansion of Ef will correspond to the "half range expansion" of f (cf. Ref. 3, Sec. 4). It will in fact be necessary to restrict E to a subspace $Y_p'\subset X_p'$ such that $E|Y_p'$ will have its range in Y_p. Then the expansion of $(E|Y_p')f$ will give the half

range "continuum modes," while the discrete modes can be separately treated.

We require the operator E to have the properties:

(i) $(zI-K)^{-1}Ef$ analytic in z for all $\mathrm{Re}z<0$, $f\in Y_p'$.

(ii) $\rho_0(Ef)=0$ for all $f\in Y_p'$,

(iii) $\rho_1(Ef)=0$ for all $f\in X_p'$.

The first proper guarantees that the expansion of Ef contains only eigenfunctions ϕ_ν with $\nu>0$; the second and third guarantee that $Ef\in Y_p$; while the third also insures that the discrete coefficient a_1 of Ef vanishes.

Before the subspace Y_p' may be specified explicitly, let us recall some properties of the dispersion function Λ. The Wiener–Hopf factorization of $\Lambda(z)$ provides a function $X(z)$, analytic for $\mathrm{Re}z<0$, such that

$$X(z)X(-z)=3\Lambda(z). \tag{17}$$

Moreover,

$$X(z)=\int_0^1[\gamma(u)/u-z]\,du, \tag{18a}$$

where

$$\gamma(u)=\frac{u}{2}\frac{X^-(u)}{\Lambda^-(u)}. \tag{18b}$$

Now we may define $Y_p'\subset X_p'$ by

$$Y_p'=\{f\in X_p'\int_0^1 Y(\mu)f(\mu)\,d\mu=0\}.$$

By analogy with transport in absorbing media, we are led to study the transformation $E: X_p'\to X_p$, defined on $f\in X_p'$ of class H^* by

$$(Ef)(u)=\begin{cases}\dfrac{1}{X(u)}\dfrac{3}{2}\displaystyle\int_0^1\frac{sf(s)\,ds}{X(-s)(s-u)}, & u<0,\\ f(u), & u>0.\end{cases} \tag{19}$$

Since $X(u)$ is analytic and bounded away from zero for $u<0$, we see from the Hölder inequality that E extends to a bounded operator from X_p' to X_p.

Property (iii) is verified by Theorem 5.

Theorem 5: For all $f\in X_p'$, $\rho_1(Ef)=0$.

Proof: From Eq. (19),

$$\int_{-1}^1 u(Ef)(u)\,du=\int_0^1 uf(u)\,du+\frac{3}{2}\int_0^1 ds\,\frac{sf(s)}{X(-s)}\int_{-1}^0\frac{u\,du}{X(u)(s-u)}$$

for f of class H^*. Changing variable from u to $-u$ in the second term above and utilizing equations (18c) and (18a), the identity

$$\int_0^1\gamma(u)\,du=1, \tag{20}$$

and the continuity of E, the result follows.

Next we shall see that property (ii) is satisfied.

Theorem 6: For all $f\in X_p'$,

$$\rho_1(Ef)=\tfrac{2}{3}\int_0^1\gamma(u)f(u)\,du.$$

Hence, if $f\in Y_p'$, then $Ef\in Y_p$.

Proof: As in the proof of Theorem 5, we compute

$$\int_{-1}^1 u^2(Ef)(u)\,du=\int_0^1 u^2f(u)\,du+\frac{3}{2}\int_0^1\frac{sf(s)}{X(-s)}\,ds\int_{-1}^1\frac{u^2\,du}{X(u)(s-u)}.$$

The change of variable $u\to -u$ along with Eq. (18b) reduces this to

$$\int_{-1}^1 u^2(Ef)(u)\,du=\int_0^1 u^2f(u)\,du+\int_0^1\frac{sf(s)}{X(-s)}\,ds\int_0^1\frac{u\gamma(u)}{s+u}\,du.$$

Finally, writing $u/(s+u)=1-s/(s+u)$ and using Eq. (18a), we obtain the desired expression for $\rho_0(Ef)$.

This result, along with Theorem 5, proves that $Ef\in H_p$ if $f\in Y_p'$ and is of class H^*, since the Cauchy integral in Eq. (19) preserves Hölder continuity.

Let Y_{p0}' denote the subspace of X_p' spanned by $e_0'(u)\equiv 1$, $u\in[0,1]$. As a corollary of Theorem 6, we obtain

Corollary 1:

$X_p'=Y_p'+Y_{p0}'$. reduces E.

Proof: From equation (18) we obtain

$$X(u)=\tfrac{3}{2}\int_0^1 s\,ds/X(-s)(s-u), \tag{21}$$

and thus compute

$$(Ee_0')(u)=e_0(u).$$

Defining the bounded linear functional

$$\rho_0'\colon f\to\int_0^1\gamma(u)f(u)\,du,$$

and the projection

$$P'\colon f\to\rho_0'(f)e_0',$$

the identity equation (20) and Theorem 6 prove the reduction.

The remaining property of E to be confirmed is given by the following theorem.

Theorem 7: $(zI-K)^{-1}(Ef)(u)$ is analytic in z for $\mathrm{Re}z<0$, $f\in Y_p'$.

Proof: Analyticity is assured except for a possible branch cut $[-1,0]$. However, using Theorem 3 and Eq. (19), and applying Eqs. (17) and (18), yields for $u<0$, after some rearranging,

$$(zI-K)^{-1}(Ef)(u)=(1/z-u)\left\{\int_0^1 dt\,\gamma(t)\times f(t)\left[\frac{1}{X(u)(t-u)}-\frac{1}{X(z)(t-z)}\right]\right\}.$$

From this, the analyticity along $[-1,0]$ can be concluded.

The expansion of a function $f\in X_p'$ is accomplished by applying the full range expansion of Sec. II to Ef. In particular, let P'' represent the "projection" onto Y_p along Y_{p0}', $P''=(I-P)E$. Then

$$(Ef)(u)=\tfrac{1}{2}a_0+P''f(u),$$

since $a_1=0$ by Theorem 5. The expansion of $P''f$ is made as in Eq. (16), while a_0 can be calculated from Theorem 6. Thus,

$$P''f(u)=\int_0^1 A(v)\phi_v(u)\,dv, \tag{22a}$$

with

$$A(v)=[v/\gamma(v)N(v)]\int_0^1 f(u)\gamma(u)\phi_v(u)\,du \tag{22b}$$

and

$$a_0=2\int_0^1 \gamma(u)f(u)\,du/\int_0^1 \gamma(u)\,du. \tag{22c}$$

The solution of the half range neutron transport equation at $c=1$ may now be carried out as described in Refs. 3 and 11. The eigenfunction expansions developed here are used to choose a_0 and $A(v)$ to satisfy the boundary conditions at $x=0$ and $x\to\infty$, and the full solution is expressed in the form

$$\psi(u)=\tfrac{1}{2}a_0+\int_0^1 A(v)\phi_v(u)\exp(-x/v)\,dv.$$

For details, see the references cited.

IV. SOLUTION OF THE MILNE PROBLEM

We seek the solution $\psi_M(x,u)$, of the homogeneous transport equation in a half space subject to the conditions

$$\psi_M(0,u)=0,\quad u>0, \tag{23a}$$

and

$$\psi_M(x,u)\sim\tfrac{1}{2}x, \tag{23b}$$

as $x\to\infty$. The Milne problem is solved by

$$\psi_M(x,u)=\tfrac{1}{2}a_0+\tfrac{1}{2}(x-u)+\int_0^1 A(v)\phi_v(u)\exp(-x/v)\,dv, \tag{24}$$

where

$$z_o=-a_0=-\int_0^1 u\gamma(u)\,du/\int_0^1 \gamma(u)\,du \tag{25}$$

is the so-called "extrapolated endpoint," and

$$A(v)=[v/\gamma(v)N(v)]\tfrac{1}{2}\int_0^1 u\phi_v(u)\gamma(u)\,du. \tag{26}$$

It is trivial to verify that the first two terms of equation (24) do indeed satisfy Eq. (1), and since[11]

$$K^{-1}f=\int_0^1(1/v)A(v)\phi_v(u)\,dv,$$

for all $f\in Y_p'$, that the third term does also. The coefficient a_0 has been determined by setting $x=0$, multiplying both sides of equation (24) by $\gamma(u)$, integrating over u, and using the boundary condition (23a), as well as Theorem 6 to conclude that the integral does not contribute. Similarly, to solve for $A(v)$, imposing the boundary condition (23a) and using the fact that $a_0\in Y_{p0}'$, one obtains expression (26).

ACKNOWLEDGMENTS

The authors wish to express their deep appreciation to Dr. E. Larsen, who suggested the approach used here, and in fact spent considerable time in discussing the correct techniques. In addition, one of the authors (P. F. Z) expresses his appreciation to the University of Florence, and especially to Professor Aldo Belleni-Morante, for their hospitality.

*John Simon Guggenheim Memorial Foundation Fellow on leave from Virginia Polytechnic Institute & State University, Blacksburg, Va.

[1]F. Shure and M. Natelson, Ann. Phys. (N.Y.) 26, 274 (1964).

[2]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).

[3]E. W. Larsen and G. J. Habetler, Commun. Pure Appl. Math. 26, 525 (1973).

[4]H. G. Kaper, oral presentation, Fourth Conference on Transport Theory, Blacksburg, Va., March, 1975. Also, see H. G. Kaper and P. F. Zweifel, "A Report on the Fourth Conference on Transport Theory," Transp. Theory Stat. Phys. (to appear).

[5]R. J. Hangelbroek, "A Functional Analytic Approach to the Linear Transport Equation," thesis, Rijksuniversitat Groningen, Netherlands (1973).

[6]C. G. Lekkerkerker, "The Linear Transport Equation, The Degenerate Case $c=1$," submitted for publication.

[7]W. Greenberg, "Functional Calculus for the Symmetric Multigroup Transport Operator," J. Math. Phys. (to appear).

[8]R. L. Bowden, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. 17, xxx, (1976).

[9]W. Greenberg, "Two Group Equation and Its Pontrjagin Structure," in *Proceedings of the Fourth Conference on Transport Theory* (to appear).

[10]H. G. Kaper, "Full Range Boundary Value Problems in the Kinetic Theory of Gases," in *Proceedings of the Fourth Conference on Transport Theory* (to appear).

[11]E. W. Larsen, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. 16, 1117 (1975).

[12]E. W. Larsen, "Solution of Neutron Transport Problems in L_1," J. Math. Anal. Appl. (to appear).

[13]N. I. Muschelischwili, *Singulare Integralgleichungen* (Akademie-Verlag, Berlin, 1965), p. 25.

[14]For another approach see S. Sancaktar, thesis, Virginia Polytechnic Institute and State University (1975).

Inverse problem for complex "r^a-analytic" potentials of finite range*

Ilia Miodek

Département de Physique Mathématique,† Université des Sciences et Techniques du Languedoc, 34060-Montepellier Cedex, France
(Received 4 August 1975)

We present a simple solution to the inverse problem for the Schrödinger equation at fixed energy for complex "r^a-analytic" potentials of finite range. This is done via an interpolation formula for the regular radial solutions as functions of complex angular momentum. The interpolation formula is derived by Frobenius techniques and Cauchy's theorem. As an application we study an inverse problem for a spherically symmetric cold plasma perturbed by a small oscillatory electric potential of fixed (finite) frequency.

INTRODUCTION

In Sec. 1 we derive an interpolation formula for the regular solutions of the fixed energy radial Schrödinger equation as functions of complex angular momentum.[2] The result is valid for complex potentials which are analytic in r^a, $\mathrm{Re}a > 0$, and their generalizations of the form

$$u(r) = \sum_{\mathbf{n}} u_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}}, \quad \mathbf{n}\cdot\mathbf{a} = \sum_{j=1}^{J} n_j a_j,$$

$$n_j = 0, 1, 2, \ldots, \quad \mathrm{Re}a_j > 0.$$

This interpolation is then used to obtain an interpolation formula for the logarithmic derivatives at a fixed radius R. We show in Sec. 2 that this interpolation formula gives a simple solution to the fixed energy, finite range inverse problem of determining the potential from the logarithmic derivatives at finite radius. This much can be viewed as an extension of Loeffel's[3] results to a limited but large class of complex potentials. Our methods are, however, elementary. We use only Frobenius series and Cauchy's theorem. They are thus amenable to generalization. We give a method for constructing this interpolation for any finite number of logarithmic derivatives by using rational functions. We give a method for constructing these rational interpolations for the simplest case $J = 1$, but show that it is unstable and cannot be extended to an infinite data set. As far as we know this same instability problem (in constructing the interpolation) is also open in Loeffel's case.

In Sec. 3 we present an application of this technique to the inverse problem for the linearized plasma equation $\nabla\cdot[\omega_p^2(r) - \omega^2 - i\omega\nu]\nabla\Phi_\omega(x) = 0$, where $\omega_p^2(r) \propto$ ion density is to be constructed from the set of surface logarithmic derivatives $\{Rd_R \ln\Phi_l(R)\}$, where

$$\Phi_l(r) \propto \int d(\hat{x}) Y_{lm}^*(\hat{x})\Phi_\omega(x),$$

$$r = |x|, \quad x \in \mathbb{R}^D, \quad D = 2 \text{ or } 3, \quad \hat{x} = x/r.$$

Our solution[1] works with data at fixed finite frequency ω, as opposed to a previous solution by Degasperis[4] at $\omega \to \infty$. The basic idea used is that $[\omega_p^2(r) - \omega^2 - i\omega\nu]^{1/2} \times \Phi_\omega(x)$ satisfies an equation of the Schödinger type with a complex potential, and thus we can apply the above method. Our solution for this plasma problem is incomplete because we have not given sufficient conditions which will ensure that the ion density is positive.

1. THE INTERPOLATION FORMULA

In this section we derive the following interpolation formula, (1.1), for the regular solution of the fixed-energy radial Schrödinger equation with an "r^a-analytic" potential

$$u(r) \equiv (2m/\hbar^2)[r^2 v(r) - r^2 E] = \sum_{\mathbf{n}>0} u_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}}, \quad r < R,$$

where $u_{\mathbf{n}} \in \mathbb{C}$, $\mathbf{n}\cdot\mathbf{a} \equiv \sum_{j=1}^{J} n_j a_j$, $a_j \in \mathbb{C}$ with $\mathrm{Re}a_j > 0$, $n_j = 0, 1, 2, \cdots$; $\mathbf{n} > \mathbf{m}$ means $n_j - m_j \geq 0\ \forall\ j$ but $\mathbf{n} \neq \mathbf{m}$ and it partially orders $\{\mathbf{n}\}$. The range of the summation is controlled by $u_{\mathbf{n}} \equiv 0$, $\mathbf{n} \not> 0$, $0 \equiv (0, 0, \ldots, 0)$. These potentials are obviously a generalization of potentials analytic in r^k, where k is real. We will obtain the interpolation formula

$$h(\lambda; r) = 1 + \sum_{\mathbf{n}>0} \frac{c_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}} h(\mathbf{n}\cdot\mathbf{a}; r)}{\lambda + \mathbf{n}\cdot\mathbf{a}}, \quad c_{\mathbf{n}} \in \mathbb{C}, \quad r < R, \tag{1.1}$$

for solutions of

$$(rd_r + \lambda) r d_r h(\lambda; r) = u(r) h(\lambda; r), \tag{1.2}$$

$$\lim_{r\to 0} h(\lambda; r) = 1 \quad \text{(regular normalization condition for } \mathrm{Re}\,\lambda \geq 0) \tag{1.3}$$

by studying the Frobenius solutions in the λ plane. We will show that $\{c_{\mathbf{n}}\}$ completely characterizes the solution and the inverse problem will reduce to finding these constants[5] from appropriate data, i.e., the set of logarithmic derivatives $\{rd_r \ln h(\mathbf{n}\cdot\mathbf{a}; r)|_{r=R_1},\ R_1 < R, \mathbf{n} > 0\}$.

The usual form of the radial Schrödinger equation is obtained from (1.2) by substituting $h(\lambda_l; r) = r^{-l}\psi_l(r)$, $\lambda_l = 2l + D - 2$, $D = 2$ or 3 is the number of dimensions. (It is in fact valid for any integer $D \geq 2$.)

It is straightforward to verify that (1.1) satisfies the given radial equation and boundary condition providing that

$$u(r) = 2r d_r \sum_{\mathbf{n}>0} c_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}} h(\mathbf{n}\cdot\mathbf{a}; r), \quad r < R. \tag{1.4}$$

We will show that if

$$u(r) = \sum_{\mathbf{n}>0} u_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}}, \tag{1.5}$$

then the regular solution of (1.2), (1.3), which is well defined for $\mathrm{Re}\lambda \geq 0$, satisfies the interpolation formula (1.1). This will be done by using Frobenius techniques and Cauchy's theorem. It will then follow that $u(r)$ satisfies (1.4). Evaluating (1.1) at $\lambda = \mathbf{n}\cdot\mathbf{a}$, $\mathbf{n} > 0$, will give us a matrix equation in l^∞, from which we will show that $\{c_\mathbf{n}\}$ uniquely defines $\{h(\mathbf{n}\cdot\mathbf{a}; r)\}$ and thus determines $u(r)$ via (1.4). So the inverse problem will indeed be the determination of $\{c_\mathbf{n}\}$ from the data. [Logarithmic surface derivatives are appropriate data because they are independent of the normalization of the regular solutions and can be determined from surface measurements on any regular solution set. Clearly $u(r)$ is unchanged by changes in normalization of $h(\lambda_l; r)$.]

The derivation of (1.1) is exactly analogous to that of the simplest case $J = 1$. For that case we outline the standard procedure. Suppose $h(\lambda; r) = \sum_{n\geq 0} h_n(\lambda) r^{na}$, $h_0(\lambda) = 1$, and obtain the recursion relation

$$h_n(\lambda) = \frac{1}{na(na+\lambda)} \sum_{m<n} u_{n-m} h_m(\lambda), \quad n > 0.$$

Thus $h_n(\lambda)$ has poles at $\lambda = -ma$, $m = 1, 2, \ldots, n$. It follows that

$$|h_n(\lambda) r^{na}| \leq \frac{|u|(r)}{|na(na+\lambda)|} \sup_{m<n} |h_m(\lambda) r^{na}|,$$

where $|u|(r) \overset{\text{def}}{=} \sum_{n>0} |u_n r^{na}|$. From this bound deduce that, for any $\epsilon > 0$, $r \leq R_1 < R$,

(i) the Frobenius solution converges uniformly for $\lambda \in \mathbb{C}_\epsilon$, $\mathbb{C}_\epsilon \overset{\text{def}}{=} \{z \in \mathbb{C} \,|\, |z + ma| \geq \epsilon,\ m > 0\}$,

(ii) $h(\lambda; r) - 1 \to 0$ as $\lambda \to \infty$ in $\mathbb{C}_\epsilon$,

(iii) $\lim_{\epsilon \to 0} \epsilon h(\epsilon - na; r) = r^{na} b(na; r)$ and $b(na; 0)$ is finite.

All the results hold if we replace na with $\mathbf{n}\cdot\mathbf{a}$, ma with $\mathbf{m}\cdot\mathbf{a}$, and $m < n$ with $\mathbf{m} < \mathbf{n}$ since $u_{\mathbf{n}-\mathbf{m}} \equiv 0$ for $\mathbf{n} \not> \mathbf{m}$.

Since $h(\lambda; r)$ has simple poles at $\lambda = -\mathbf{n}\cdot\mathbf{a}$ for $\mathbf{n} > 0$ as its only possible singularities in the λ plane, and since $h(\lambda; r) - 1 \to 0$ as $\lambda \to \infty$ in $\mathbb{C}_\epsilon$ for any $\epsilon > 0$, Cauchy's theorem gives us

$$h(\lambda; r) - 1 = \sum_{\mathbf{n}>0} \frac{r^{\mathbf{n}\cdot\mathbf{a}} b(\mathbf{n}\cdot\mathbf{a}; r)}{\mathbf{n}\cdot\mathbf{a} + \lambda}, \quad r < R. \tag{1.6}$$

[For the general case the reader should convince himself that for ϵ small enough we can indeed find a sequence of contours in $\mathbb{C}_\epsilon$, at ever increasing distances such that Cauchy's theorem will give us (1.6).] If $\mathbf{n}\cdot\mathbf{a} = \mathbf{m}\cdot\mathbf{a}$, but $\mathbf{n} \neq \mathbf{m}$, we must only count one of them in (1.6), which is related to the following redundancy. $u_\mathbf{n}$ and $u_\mathbf{m}$ are not well defined if $\mathbf{n}\cdot\mathbf{a} = \mathbf{m}\cdot\mathbf{a}$ since they then multiply the same power of r. We eliminate this redundancy once and for all by ordering $\{\mathbf{n}\}$ and counting only the first such index that occurs. [e.g., extend the partial order to all $\mathbf{n}$ by defining $\mathbf{n} > \mathbf{m}$ whenever $\mathbf{n} \neq \mathbf{m}$ and either $\mathrm{Re}(\mathbf{n}-\mathbf{m})\cdot\mathbf{a} > 0$ or both $\mathrm{Re}(\mathbf{n}-\mathbf{m})\cdot\mathbf{a} = 0$ and $n_k > m_k$, $n_j = m_j \,\forall\, j > k$ hold.] We do this formally by putting $u_\mathbf{n} \equiv 0$ if $\mathbf{n} > \mathbf{m}$ and $\mathbf{n}\cdot\mathbf{a} = \mathbf{m}\cdot\mathbf{a}$, and omitting $\mathbf{n}$ from the summation $\sum_{\mathbf{n}>0}$.

It remains to show that $b(\mathbf{n}\cdot\mathbf{a}; r) = c_\mathbf{n} h(\mathbf{n}\cdot\mathbf{a}; r)$, $c_\mathbf{n} \in \mathbb{C}$. Clearly $r^{\mathbf{n}\cdot\mathbf{a}} b(\mathbf{n}\cdot\mathbf{a}; r) \equiv \lim_{\lambda \to -\mathbf{n}\cdot\mathbf{a}} (\lambda + \mathbf{n}\cdot\mathbf{a}) h(\lambda; r)$ satisfies (1.2) with $\lambda = -\mathbf{n}\cdot\mathbf{a}$. The operator identity $(rd_r - \mathbf{n}\cdot\mathbf{a}) \times rd_r r^{\mathbf{n}\cdot\mathbf{a}} \equiv r^{\mathbf{n}\cdot\mathbf{a}} (rd_r + \mathbf{n}\cdot\mathbf{a}) rd_r$ then shows that $b(\mathbf{n}\cdot\mathbf{a}; r)$ satisfies (1.2) with $\lambda = \mathbf{n}\cdot\mathbf{a}$. It is thus a linear combination of the regular and irregular solutions. But the irregular solution would change the normalization of $h(\lambda; r)$—cf. (iii) above (obtained by detailed examination of the residue) which shows directly that $b(\mathbf{n}\cdot\mathbf{a}; r)$ is regular. Thus $b(\mathbf{n}\cdot\mathbf{a}; r) = c_\mathbf{n} h(\mathbf{n}\cdot\mathbf{a}; r)$ (we extend our convention to $c_\mathbf{n} : c_\mathbf{m} \equiv 0$ if $\mathbf{m} > \mathbf{n}$, $\mathbf{m}\cdot\mathbf{a} = \mathbf{n}\cdot\mathbf{a}$), and we have verified (1.1), (1.2) in the following sense. The $h(\lambda; r)$ given by (1.1) is the unique solution of (1.2), (1.3) for $\mathrm{Re}\,\lambda \geq 0$ and it gives the analytic continuation to the whole λ plane. Equation (1.4) follows by applying (1.2) to (1.1).

It is perhaps worth noting that (1.1), (1.4), (1.5) must all have the same radius of convergence as the series[6] $\sum_{\mathbf{n}>0} c_\mathbf{n} r^{\mathbf{n}\cdot\mathbf{a}} \equiv q(r)$.

2. THE INVERSE PROBLEM

We show first that the $\{c_\mathbf{n}\}$ in (1.1) uniquely determines the vector $\mathbf{h}(r)$ with components $\mathbf{h}(r)_\mathbf{m} \equiv h(\mathbf{m}\cdot\mathbf{a}; r)$. Evaluating (1.1) at $\lambda = \mathbf{m}\cdot\mathbf{a}$, $\mathbf{m} > 0$, gives us the matrix equation

$$\mathbf{h}(r) = \mathbf{e} + MC(r)\mathbf{h}(r), \tag{2.1}$$

where $\mathbf{e}_\mathbf{n} = 1$, $\mathbf{M}_{\mathbf{n},\mathbf{m}} = 1/(\mathbf{n}+\mathbf{m})\cdot\mathbf{a}$, $C(r)_{\mathbf{n},\mathbf{m}} = \delta_{\mathbf{n},\mathbf{m}} c_\mathbf{n} r^{\mathbf{n}\cdot\mathbf{a}}$. As an operator in l^∞ the norm of $MC(r)$ is bounded by $\sum_{\mathbf{n}>0} |c_\mathbf{n} r^{\mathbf{n}\cdot\mathbf{a}}| / \mathrm{Re}(\mathbf{n}\cdot\mathbf{a} + a_1)$, $r < R$, where we have assumed that $0 < \mathrm{Re}a_1 \leq \mathrm{Re}a_2 \leq \cdots \leq \mathrm{Re}a_J$. This bound tends to zero like r^{a_1} as $r \downarrow 0$. Thus $MC(r)$ is contractive for r small enough, say $r < r_0$. Thus $\mathbf{h}(r)$ is uniquely determined by $C(r)$ for $r < r_0$, and so is $u(r)$ via (1.4). Analytic continuation then determines them uniquely for $r < R$. So we have shown that $\{c_\mathbf{n}\}$ determines $u(r)$ for $r < R$.[7]

We now want to construct $\{c_\mathbf{n}\}$ from $\{\beta(\mathbf{n}\cdot\mathbf{a})\}_{\mathbf{n}>0}$, where $\beta(\mathbf{n}\cdot\mathbf{a}) = rd_r \log h(\mathbf{n}\cdot\mathbf{a}; r)|_{r=R_1}$. Differentiate (1.1) and evaluate both (1.1) and its derivative at $r = R_1$.

$$h(\lambda; R_1) = 1 + \sum_{\mathbf{n}>0} b_\mathbf{n}/(\lambda + \mathbf{n}\cdot\mathbf{a}), \tag{2.2}$$

$$\beta(\lambda) h(\lambda; R_1) = \sum_{\mathbf{n}>0} [b_\mathbf{n}/(\lambda + \mathbf{n}\cdot\mathbf{a})][\beta(\mathbf{n}\cdot\mathbf{a}) + \mathbf{n}\cdot\mathbf{a}], \tag{2.3}$$

where $b_\mathbf{n} \equiv c_\mathbf{n} R_1^{\mathbf{n}\cdot\mathbf{a}} h(\mathbf{n}\cdot\mathbf{a}; R_1) \in \mathbb{C}$. If we can find $\{b_\mathbf{n}\}$ from $\{\beta(\mathbf{n}\cdot\mathbf{a})\}$, we will be done because (2.2) will then give us $h(\mathbf{n}\cdot\mathbf{a}; R_1)$ and $\{c_\mathbf{n}\}$ can then be obtained from the definition of $b_\mathbf{n}$.

Clearly (2.2) and (2.3) give us

$$\beta(\lambda) = \sum_\mathbf{n} \frac{b'_\mathbf{n}}{\lambda + \mathbf{n}\cdot\mathbf{a}} \Big/ \left(1 + \sum_\mathbf{m} \frac{b_\mathbf{m}}{\lambda + \mathbf{m}\cdot\mathbf{a}}\right) \tag{2.4}$$

where

$$b'_\mathbf{n}/b_\mathbf{n} = \beta(-\mathbf{n}\cdot\mathbf{a}) = \beta(\mathbf{n}\cdot\mathbf{a}) + \mathbf{n}\cdot\mathbf{a}. \tag{2.5}$$

It remains to construct this interpolation of $\{\beta(\mathbf{n}\cdot\mathbf{a})\}_{\mathbf{n}>0}$. Equations (2.4) and (2.5) are the necessary and sufficient conditions for the existence of a potential $u(r) = \sum_{\mathbf{u}>0} u_\mathbf{n} r^{\mathbf{n}\cdot\mathbf{a}}$, which will generate $\{\beta(\mathbf{n}\cdot\mathbf{a})\}_{\mathbf{n}>0}$ via the regular solutions of the radial equation.

Can we construct such an interpolation? The answer is yes for any finite data set, but no for an infinite data set.

Let us consider the finite interpolation

$$\beta^N(\lambda) = \sum_{0<\mathbf{n}<\mathbf{N}} \frac{b_{\mathbf{n}}'^N}{\lambda + \mathbf{n}\cdot\mathbf{a}} \Big/ \left(1 + \sum_{0<\mathbf{m}\leq\mathbf{N}} \frac{b_{\mathbf{n}}^N}{\lambda + \mathbf{m}\cdot\mathbf{a}}\right),$$

$$\mathbf{N} \equiv (N, N, \ldots, N). \tag{2.6}$$

This corresponds to the case $c_{\mathbf{n}} \equiv 0$ for $\mathbf{n} > \mathbf{N}$; a finite number of poles for $h(\lambda; r)$ at $\lambda = \mathbf{n}\cdot\mathbf{a}$, $0 < \mathbf{n} \leq \mathbf{N}$.

This $\beta^N(\lambda)$ is a rational function of λ. It is the ratio of two polynomials, $Q^{JN-1}/P^{NJ}(\lambda)$, of order $JN-1$ and NJ, respectively. These polynomials are in one-to-one correspondence with the $2NJ$ constants $\{b_{\mathbf{n}}'^N, b_{\mathbf{n}}^N\}$;

$$P^{NJ}(\lambda) \equiv p^{NJ}(\lambda)\left(1 + \sum_{\mathbf{n}} b_{\mathbf{n}}^N/(\lambda + \mathbf{n}\cdot\mathbf{a})\right),$$

$$Q^{NJ-1}(\lambda) \equiv p^{NJ}(\lambda) \sum_{\mathbf{n}} b_{\mathbf{n}}'^N/(\lambda + \mathbf{n}\cdot\mathbf{a}),$$

where $p^{NJ}(\lambda) \equiv \prod_{0<\mathbf{n}<\mathbf{N}}(\lambda + \mathbf{n}\cdot\mathbf{a})$. This rational function has $2NJ$ free parameters and can be constructed uniquely from its values at $\lambda = \pm\mathbf{n}\cdot\mathbf{a}$, $\mathbf{n} > 0$.

We need only interpolate $\{\beta^N(\mathbf{n}\cdot\mathbf{a}),\ \beta^N(-\mathbf{n}\cdot\mathbf{a}) = \beta^N(\mathbf{n}\cdot\mathbf{a}) + \mathbf{n}\cdot\mathbf{a}\}$ by $Q^{JN-1}(\lambda)/P^{NJ}(\lambda) = \beta^N(\lambda)$ to satisfy (2.4), (2.5). Then $b_{\mathbf{m}}^N = \lim_{\lambda\to-\mathbf{n}\cdot\mathbf{a}}(\lambda + \mathbf{n}\cdot\mathbf{a})P^{NJ}(\lambda)/p^{NJ}(\lambda)$ gives us a solution to the inverse problem for the finite data set $\{\beta^N(\mathbf{n}\cdot\mathbf{a})\}_{0<\mathbf{n}\leq\mathbf{N}}$. We cannot, however, guarantee the stability of these interpolations as $N\to\infty$, nor can we guarantee that the data generated by (2.6), $\{\beta^N(\mathbf{n}\cdot\mathbf{a})\}_{\mathbf{n}>\mathbf{N}}$, will be small. We do not expect it to be small in general.

In the simplest case $J=1$ we have studied this instability in more detail. Evaluating (2.6) at $\lambda = na$, $n = 1, 2, \ldots, N$, gives us the $N\times N$ matrix equation

$$\beta^N \mathbf{e}^N = [M^N - \beta^N H^N(\beta^N + \mu^N)^{-1}]\mathbf{k}^N, \tag{2.7}$$

where $\beta_{n,m}^N = \delta_{n,m}\beta^N(na)$, $H_{n,m}^N = 1/(n+m)$ in the $N\times N$ segment of the Hilbert matrix, $\mathbf{e}_n^N = 1$, and $\mathbf{k}_n^N = (1/a)\times[\beta(na) + na]b_n^N$, $\mu_{n,m}^N = \delta_{n,m}na$. The finite segments of the Hilbert matrix are invertable;

$$(H^N)^{-1} = A^N H^N A^N, \quad A_{n,m}^N = \delta_{n,m}\frac{(-)^{m-1}\Gamma(N+m+1)}{\Gamma(N-m+1)\Gamma(m)\Gamma(m+1)} \tag{2.8}$$

Thus

$$\mathbf{k}^N = [I^N - (H^N)^{-1}\beta^N H^N(\beta^N + \mu^N)^{-1}]^{-1}(H^N)^{-1}\beta^N \mathbf{e}^N, \tag{2.9}$$

where I^N is the $N\times N$ identity. The inverse of $[I^N - \ldots]$ can be constructed by finite matrix techniques for almost all data sets. Of course, $b_n^N = a[\beta(na) + na]^{-1}\mathbf{k}_n^N$ gives us b_n^N.

We cannot extend this technique to $N=\infty$ because the infinite Hilbert matrix does not have an inverse matrix. This is the cause of the instability.

Padé approximants may give a stable way to numerically construct these rational interpolations. The author is at present investigating this possibility.

It may be worth noting that the set $\{\beta(na)\}_{n=1}^{\infty}$ usually contains an infinite number of points which are not given by physical measurements. It may be possible to choose the values at the nonmeasurable points in such a way as to reproduce the measurable points at least to within experimental error.

We have based an extensive analysis of this stability problem on the "diagonalized" form of the infinite Hilbert matrix obtained by Rosenblum.[8] The difficulty with the infinite Hilbert matrix is that if $H^{(d)}(x)$ is the multiplication operator on $L^2(d\rho(x))$ such that the orthonormal basis $\{\varphi_n(x)\}$ turns $H^{(d)}(x)$ into the Hilbert matrix, $\int d\rho(x)\varphi_n^*(x)H^{(d)}(x)\varphi_m(x) = 1/(n+m)$, then $[H^{(d)}(x)]^{-1}$ has no matrix elements with respect to $\{\varphi_n(x)\}$. Rosenblum constructed $\{\varphi_n(x)\}$, $\rho(x)$, $H^{(d)}(x)$ explicitly. Using his results, we have translated the infinite version of (2.8) into a Fredholm equation of the third kind. Such equations can be shown to have singular (distributional) solutions in some cases. Extending work done by Bart and Warnock,[9] we have tried to solve this singular Fredholm equation. Our conclusion was that we would not be able to regenerate the matrix equations for these singular solutions because they are singular in just the same places as is $H^{(d)}(x)$ and thus do not have components relative to $\{\varphi_n(x)\}$.[10]

In short we have no positive results on a stable way to construct interpolations of the form (2.4) for infinite data sets.

Our results are: (i) The existence of our interpolation (2.4), (2.5) is a necessary and sufficient condition for the existence of $u(r) = \sum_{\mathbf{n}>0} u_{\mathbf{n}} r^{\mathbf{n}\cdot\mathbf{a}}$.

(ii) Interpolation of finite data sets by (2.6) is equivalent to interpolating $\{\beta(-\mathbf{n}\cdot\mathbf{a}), \beta(+\mathbf{n}\cdot\mathbf{a})\}_{0<\mathbf{n}\leq\mathbf{N}}$ by a rational function $\beta^N(\lambda)$ of appropriate order, given the condition $\beta(-\mathbf{n}\cdot\mathbf{a}) = \mathbf{n}\cdot\mathbf{a} + \beta(\mathbf{n}\cdot\mathbf{a})$.

(iii) Stability of such interpolations is an open question.

(iv) The methods used to obtain the interpolation formula are simple and amenable to further generalizations.

Brief comparison with related approaches

Our results should be compared with those of Loeffel[3] for a real potential of finite range satisfying much weaker assumptions than analyticity. As far as we know he too can give a necessary and sufficient interpolation formula without being able to construct it in a stable way.

The simplest version of our interpolation formula was already found by Sabatier[2] in 1967, but it was not fully exploited for the inverse problem. It was, however, by means of interpolations that he was able to develop an inverse method for spin–orbit potentials.[11]

The identity $1/(\lambda + na) = \int_0^1 (dt/t)t^{\lambda+na}$ for $\mathrm{Re}(\lambda + na) > 0$ connects our formulation to algebraic moment problems and also to generating formulas (with l-independent kernels)

$$h(\lambda_l; r) = 1 + \int_0^1 \frac{dt}{t} k(r, tr)t^{\lambda_l}, \quad k(r, y) = \sum_n c_n y^{na} h(na; r), \tag{2.11}$$

$$r^l h(\lambda_l; r) = r^l + \int_0^r \frac{dy}{y}\left[\left(\frac{y}{r}\right)^{(D-2)/2} k(r, \sqrt{ry})\right] y^l. \tag{2.12}$$

This is Loeffel's[3] transformation equation [his Eq. (27)]. Equations of the Gel'fand–Levitan, Regge–Newton

type[12] can be obtained from (2.12). It is interesting that this machinery is not necessary for this case.

We wish to note that for a very large class of real potentials Sabatier[13] gives a constructive solution to the inverse problem. To adapt his results, one would have to isolate the potentials of finite range in the large class of scattering-equivalent potentials produced by his method. Then one would have to find a new method to actually feed in the data for complex phase shifts. It is only in this last step that his formalism depends on the reality of the potential.

From the point of view of construction [Eq. (2.7)] our method resembles that of Newton[5] with the difference that his method replaces the powers of r with spherical Bessel functions. The matrix he obtains instead of our M has an inverse matrix, and thus he is able to construct solutions for infinite data sets. We do not completely understand why one case is stable but the other not.

3. INVERSE PLASMA PROBLEM

As an application of the results obtained in Secs. 1 and 2 we give here, in brief outline, the physical inverse problem which actually motivated the above study. Unfortunately, our method does not give a complete solution of the physical problem.

Consider a finite, cold, neutral, spherically symmetric plasma which is originally static and is then perturbed by a small electric potential applied at its surface. We want to solve the inverse problem of determining the distribution of the heavy ions inside the plasma from measurements of the angular dependence of the applied potential and the responding electric field at the plasma's surface. The magnetic field will be ignored. We will only be interested in one frequency component, $\exp(-i\omega t)\Phi_\omega(x)$, of the residual forced oscillations of the electric potential, $\Phi(x,t)$. The transient oscillations, which are damped out exponentially by collisions, will play no part in our approach to the inverse problem. It might reduce the class of equivalent potentials in a more general formulation if additional data could be fed in; but compatibility problems could also arise.

Following the analysis by Barston,[14] we obtain the following linearized plasma equations:

$$\nabla\cdot[\omega_p^2(r)+\partial_t^2+\nu\partial_t]\nabla\exp(-i\omega t)\,\Phi_\omega(x)$$
$$=\exp(-i\omega t)\nabla\cdot[\omega_p^2(r)-\omega^2-i\nu\omega]\nabla\Phi_\omega(x)=0. \tag{3.1}$$

Here $\omega_p^2(r)=\{[M_{\rm elec}/4\pi e^2]\,{\rm Ion}(r)\}$ is the square at the "plasma frequency" which, as is indicated, is proportional to the positive ion density ${\rm Ion}(r)$; $\Phi_\omega(x)=(1/2\pi)\times\int_{-\infty}^{\infty}dt\,\exp(i\omega t)\Phi(x,t)$ is the Fourier time transform of the residual electric potential $\Phi(x,t)$—these are forced oscillations coupled to an external driving field; finally ν is a positive constant called the collision frequency, which is introduced to represent the average damping effect of collisions. The ions are assumed to be too heavy to oscillate.

The inverse problem for this equation is to construct the ion density given the measurable surface data $\Phi_\omega(R\hat{x})$ and $\partial_R\Phi_\omega(R\hat{x})$, respectively the electric potential and the normal component of the electric field at the plasma's surface; $x\in\mathbf{R}^D$, $D=2$ or 3 for the cylindrical and spherical cases respectively; $|x|\equiv r$ is the radius variable and $\hat{x}=x/r$ is a unit vector representing the angle variable. Starting from these two functions of angle, we wish to construct $\omega_p^2(r)$ to within a normalization constant, say, $\omega_p(0)$ [or equivalently $\omega_p(R)$]. Clearly the potential $\Phi_\omega(x)$ cannot give us this information.

Spherical symmetry decouples the equations for the harmonic components of $\Phi_\omega(x)$, $\Phi_l(r)\propto\int d(\hat{x})Y_{lm}^*(\hat{x})\Phi_\omega(x)$, $l=0,1,2,\cdots$ [for $D=2$, $Y_{lm}(\hat{x})\to\exp(\pm il\theta)$], giving us the following equation for $f_l(r)\equiv r^{-l}\Phi_l(r)$:

$$[(rd_r+\lambda_l)\epsilon(r)rd_r+l\epsilon'(r)]f_l(r)=0,\quad f_l(0)=1, \tag{3.2}$$

where $\epsilon(r)\equiv\omega_p^2(r)-\omega^2-i\omega\nu$, $\epsilon'(r)\equiv rd_r\epsilon(r)$, $\lambda_l\equiv 2l+D-2$. Providing $\omega_p^2(r)-\omega_p^2(0)\sim r^a$, $a>0$ near $r=0+$, (3.2) defines $f_l(r)$ uniquely for ${\rm Re}\lambda_l\geq 0$. Notice that $f_0(r)\equiv 1$ is the only constant solution if $\epsilon(r)$ is not constant.

The implied equation for $|f_l(r)|^2$ is obtained by multiplying (3.2) by $f_l^*(r)$. The result when integrated gives us

$$\frac{1}{\omega}r^{\lambda_l}\epsilon(r)rd_r|f_l(r)|^2=\frac{1}{\omega}\int_0^r\frac{dy}{y}y^{\lambda_l}[\epsilon(y)|f_l'(y)|^2$$
$$-l\epsilon'(y)|f_l(y)|^2]. \tag{3.3}$$

Now $(1/\omega)\,{\rm Im}\epsilon(r)=\nu={\rm const}>0$. So the imaginary part of the right-hand side of (3.3) is strictly positive for $r>0$ and l real but nonzero. [If $\nu={\rm const}>0\to\nu(r)>0$, $\nu'(r)\leq 0$, this conclusion holds for $l>0$.] Noticing that $rd_r|f_l(r)|^2=2|f_l(r)|\,{\rm Re}[rd_r\ln f_l(r)]$, we deduce the following necessary consequence of $\omega_p^2(r)$ real and $\nu={\rm const}>0$:

$$|f_l(r)|>1,\ {\rm Re}rd_r\ln f_l(r)>0 \text{ for } r>0, \text{ and all real nonzero } l. \tag{3.4}$$

(This result has an analog for nonspherical plasmas.[15])

We would actually like to obtain necessary and sufficient conditions on $Rd_R\ln f_l(R)$ as a function of complex l which would ensure the reality and positivity of $\omega_p^2(r)$. We have not been able to do this yet, but it seems to us that more useful information should be obtainable from (3.3).

We now pose the inverse problem of constructing $\epsilon(r)$ to within a normalization constant given the set of logarithmic derivatives $\{Rd_R\ln f_l(R)\}_{l=1}^{\infty}$. (If the plasma is in a container with known dielectric properties, this data at the surface of the plasma can be generated from data at the surface of the container by solving a standard direct problem—Poisson's equation.)

Degasperis[4] solved this inverse problem at the asymptotic fixed frequency $\omega\to\infty$. There are unsatisfactory aspects about $\omega\to\infty$ from the points of view of measurements, validity of various physical approximations built into the equations and consistency problems which arise if interpolation—extrapolation proceedures are to be used to generate asymptotic data from data at finite values of ω. We discuss his approach and these problems more fully in Appendix A.

Our approach requires us to transform to Schrödinger form. The linearized plasma equation (3.1) has the form

$$\nabla\cdot\epsilon(r)\nabla\Phi_\omega(x)=0,\quad \epsilon(r)\equiv\omega_p^2(r)-\omega^2-i\omega\nu. \tag{3.5}$$

It is easy to verify that $\sqrt{\epsilon(r)}\Phi_\omega(x)\equiv\psi(x)$ satisfies

$$[\nabla^2-V(r)]\psi(x)=0,\quad V(r)\equiv[\epsilon(r)]^{-1/2}\{\nabla^2[\epsilon(r)]^{1/2}\}. \tag{3.6}$$

[Take $\omega_p^2(r)\in\mathbb{C}^2$ for example; we assume stronger conditions.] Now consider the inverse problem for the radial equation for $\psi(x)$. Define $h(\lambda_l;\,r)$, $\lambda_l=2l+D-2$, by

$$\begin{aligned}h(\lambda_l;\,r)&\equiv r^{-l}\psi_l(r)\equiv[\epsilon(r)/\epsilon(0)]^{1/2}r^{-l}\Phi_l(r)\\&\equiv[\epsilon(r)/\epsilon(0)]^{1/2}f_l(r).\end{aligned} \tag{3.7}$$

Its radial equation is

$$(rd_r+\lambda_l)rd_rh(\lambda_l;\,r)=u(r)h(\lambda_l;\,r),\quad h(\lambda_l;\,0)=1, \tag{3.8}$$

where

$$u(r)\equiv r^2V(r)=[\epsilon(r)]^{-1/2}(rd_r+\lambda_0)rd_r[\epsilon(r)]^{1/2} \tag{3.9}$$

This defines $h(\lambda_l;\,r)$ uniquely for $\mathrm{Re}\lambda_l\geq 0$. Notice that $h(\lambda_0;\,r)=[\epsilon(r)/\epsilon(0)]^{1/2}$. [From the point of view of the inverse problem we are constructing a set $\{h(\lambda_l;\,r)\}_{l=0}^{\infty}$ whose elements have the l-independent "property" $u(r)$ in common, and any one of the $h(\lambda_l;\,r)$ could be used to define it.] If $\omega_p^2(r)-\omega_p^2(0)=\sum_{n=1}^{\infty}\omega_n r^{na}$ (convergent for $r\leq R$), $a>0$, then $u(r)=\sum_{n=1}^{\infty}u_n r^{na}$ (also for $r\leq R$), $u_n\in\mathbb{C}$ if ω satisfies the sufficient condition

$$(\omega^2+\tfrac{1}{2}\nu^2)>\tfrac{1}{2}\nu^4+\sup_{r\leq R}[\omega_p^2(r)-\omega_p^2(0)]^2. \tag{3.10}$$

We can now apply the results of Secs. 1 and 2 (with $J=1$) providing the data set $\{\beta(na)\}_{n=1}^{\infty}$ can be deduced from the measured data $\{Rd_R\ln\Phi_l(R)\}_{l=1}^{\infty}$. From (3.7) we obtain

$$\begin{aligned}\beta(\lambda_l)&\equiv Rd_R\ln h(\lambda_l;\,R)\\&=Rd_R\ln\Phi_l(R)-l+\tfrac{1}{2}Rd_R\ln\epsilon(R)\quad\text{for } l=0,1,2,\ldots.\end{aligned} \tag{3.11}$$

From the interpolation formula or more general principles we know that $\beta(\lambda_l)\to 0$ as $l\to\infty$. Thus

$$\beta(\lambda_0)\equiv\tfrac{1}{2}Rd_R\ln\epsilon(R)=\lim_{l\to\infty}[l-Rd_R\ln\Phi_l(R)]. \tag{3.12}$$

It follows that the sets $\{\beta(\lambda_l)\}_{l=0}^{\infty}$ and $\{Rd_R\ln\Phi_l(R)\}_{l=1}^{\infty}$ can be uniquely determined from each other via Eqs. (3.11) and (3.12). [The fact that $\Phi_0(r)=\text{const}$ played an essential role.] So the inverse plasma problem has been reduced to interpolating the known values $\{\beta(\lambda_l),\beta(-\lambda_l)\}_{l=0}^{\infty}$, $\beta(-\lambda_l)=\beta(\lambda_l)+\lambda_l$ by equation (2.4) with $J=1$ provided that $\{\beta(\lambda_l)\}_{l=1}^{\infty}$ is included in $\{\beta(na)\}_{l=1}^{\infty}$. This is a restriction "a": both D/a and $2/a$ must be integers.

If we want to apply the constructive approximation procedure given in Sec. 2 following Eq. (2.6), we must introduce a further restriction on "a" for the case $D=2$. This is because then $\lambda_0=0$ but $\beta(0)\notin\{\beta(na),\beta(-na)\}_{n=1}^{N}$. We must now fit $\beta(0)$ by adjusting a nonphysical value in the set. It seems reasonable that this can be done but we have no proof. If, however, $\{\beta(na)\}_{n=1}^{N}$ is to contain nonphysical points for $D=2$ we must have $1/a$ integer.

The possibility of nonphysical points in these kinds of inverse problems was first noticed by Sabatier.[5] It may turn out that there is a way to adjust them so that the approximation procedure will become stable. The quastion is open as for as we know.

We have given necessary but not sufficient conditions for $\epsilon_p^2(r)$ to be real. The absence of sufficient conditions on $\beta(\lambda)$ which guarantee $\epsilon_p^2(r)$ real and positive leaves our solution incomplete. If $\nu(r)\overset{\text{def}}{=}\mathrm{Re}\{[\omega\nu(0)+i\omega_p^2(0)-i\omega^2]\times[h(\lambda_0;\,r)]^2\}\geq 0$, $\nu\equiv\nu(0)$, we could argue that the collision frequency depends on r. This is, of course, not really a satisfactory replacement for what is needed. [$\nu(0)$, $\omega_p(0)$ can be determined from $\nu(R)$, $\omega_p(R)$ but not from $\Phi_\omega(x)$.]

ACKNOWLEDGMENTS

We wish to thank Professor R.G. Newton and Professor P.C. Sabatier for their encouragement and extensive discussions. The hospitality and support of the mathematical physics departments of both Indiana University, Bloomington and U.S.T.L., Montpellier has been warmly appreciated. Our work at Indiana University was supported in part by the U.S. Army Research Office, Durham, North Carolina.

APPENDIX: RESUME OF DEGASPERIS' APPROACH AND DISCUSSION OF PHYSICAL APPROXIMATIONS

In 1970 Degasperis[4] solved this inverse problem for the two dimensional plasma at the "fixed" frequency $\omega\to\infty$. He expanded the radial logarithmic derivatives of the electric potential at fixed radius in an asymptotic Taylor series (essentially powers of ω^{-2}, about $\omega=\infty$). He was then able to show that the $(2l-1)$th algebraic moment of the ion density is determined by the lth harmonic component of the first order coefficient of that Taylor series. [In his paper the ion number density, $n(r)$, is often mistakenly called an electron density. The correct meaning is given in his appendix.] This algebraic moment problem is soluble and is in fact closely related to the inversion of the Hilbert matrix. Degasperis also showed that similar relations held for the higher order Taylor coefficients and thus that the inverse problem was overdetermined, given data at all frequencies.

The physical objections to working at the asymptotic limit $\omega\to\infty$ are that the neglect of the magnetic field and the use of the fluid approximation for the plasma are then invalid. We discuss this somewhat more fully in Sec. 2.0 of Ref. 1. It must be shown that there exists a range of frequencies ω where the asymptotic equations are valid simultaneously with the physical approximations embodied in the linearized plasma equation (3.1).

Another consideration of practical importance is that data must be measurable, at least in principle, with a good signal-to-noise ratio. The measured data must show enough sensitivity to the ion distribution to produce significant differences for different distributions. The very convergence of the asymptotic Taylor series itself shows that in this limit data tends to zero; in this limit the ion distribution hardly counts. In fact, the only way to get significant data for $\omega\to\infty$ would be to

interpolate and/or extrapolate data at finite frequencies in some self-consistent (and unknown) way.

For the plasma the most sensitive frequency range is controlled by the maximum and minimum (expected) values for the ion density. This is the range where resonant oscillations are likely to occur. However, if the interaction is too big, the linear approximations may be invalidated. The only way to check this seems to be self-consistency; i.e., check to see if the neglected terms are really small. Discussions of what is involved in the linear approximation are absent from the publications of Degasperis[4] and Barston.[14] Since this may be important in physical applications we refer the reader to Ref. 16 for more detailed discussion.

*This work has been done as a part of the "Recherche Coopérative sur Programme No. 264. Etude interdisciplinaire des problèmes inverses."

†Physique Mathèmatiques et Théorique, Equipe de recherche associée au CNRS No. 154.

[1]The results presented here are based on the author's Ph. D. dissertation: Inverse Methods from Scattering Theory to a Spherical Plasma, 1974, Indiana University, Bloomington, Indiana 47401.

[2]P. C. Sabatier, J. Math. Phys. 8, 1957 (1967). See, in particular, his Eqs. (1.24), (1,38), (1.41).

[3]J-J. Loeffel, Ann. Inst. Henri Poincaré 8, 339 (1968).

[4]A. Degasperis, J. Math. Phys. **11**, 3392 (1970).

[5](a) R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966); Sec. 20.3, Eq. (20.43), introduces a set of constants into a "symmetric" function via which one is able to feed in the data. This motivated our approach. (b) P. C. Sabatier, J. Math. Phys. 7, 1515, 2079 (1966); 8, 905 (1967), extended Newton's Method. R. G. Newton, J. Math. Phys. 8, 1566 (1966), contains related analysis. These authors are more concerned with properties as $r \to \infty$.

[6]This is the analog of the "symmetric" function introduced by Newton and is referred to in Ref. 5(a).

[7]This proof is the analog of that given in Ref. 5(a) to prove the existence and uniqueness of Newton's Eq. (20.46), p. 628. It amounts to proving existence and uniqueness for his equation (20.56) which is the analog of our interpolating equation with $\varphi_l^{(1)}(r)$ replacing r^l. Analyticity as a function of r plays a crucial role in both cases.

[8]M. Rosenblum, Proc. Am. Math. Soc. 9, 137, 581 (1958).

[9]G. R. Bart and R. L. Warnock, SIAM J. Math. Anal. 4, 609 (1973).

[10]More details are given in Appendices 4 and 5 of Ref. 1.

[11]P. C. Sabatier, J. Math. Phys. 8, 1241 (1968).

[12]We mean an equation of the type (20.46) in Ref. 5(a).

[13]P. C. Sabatier, J. Math. Phys. 13, 675 (1972).

[14]E. M. Barston, Ann. Phys. (N.Y.) 29, 282 (1964).

[15]Independent of spherical symmetry one can show that neither $\Phi_\omega(x)$ nor its normal derivative can vanish on any closed surface unless $\Phi_\omega(x) \equiv$ const inside that surface. This may be useful in other geometries.

[16]N. A. Krall and A. W. Trivelpiece, *Principles of Plasma Physics* (McGraw-Hill, New York, 1973), Chap. 8.

A new topology for curved space–time which incorporates the causal, differential, and conformal structures*

S. W. Hawking†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
and California Institute of Technology, Pasadena, California 91125

A. R. King‡ and P. J. McCarthy

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 16 May 1975)

A new topology is proposed for strongly causal space–times. Unlike the standard manifold topology (which merely characterizes continuity properties), the new topology determines the causal, differential, and conformal structures of space–time. The topology is more appealing, physical, and manageable than the topology previously proposed by Zeeman for Minkowski space. It thus seems that many calculations involving the above structures may be made purely topological.

1. INTRODUCTION

In 1964 Zeeman published a paper[1] showing that the causal structure of Minkowski space M, already implied its linear structure. Causality was defined by means of a partial ordering on M, and it was shown that the group of automorphisms of M preserving the ordering is generated by the inhomogeneous Lorentz group and dilatations. (This is the *homothecy* group H of M, comprising all affine automorphisms which preserve the Lorentz metric up to a *constant* factor.) He then proposed[2] that the conventional (positive definite) metric topology $\mathcal{M}$ of M should be replaced by a new topology $\mathcal{J}$ (the *fine* topology) which is related to the causal structure. $\mathcal{J}$ has the following properties[2]:

(1) $\mathcal{J}$ is defined to be the finest topology on M which induces the one-dimensional Euclidean topology on every straight timelike line, and the three-dimensional Euclidean topology on every spacelike hyperplane. Thus $\mathcal{J}$ is finer (and, in fact, strictly finer) than $\mathcal{M}$.

(2) $\mathcal{J}$ incorporates the (homothetic) Lorentz structure at the primitive level of topology (rather than, as is conventional, *after* imposing linear structure); the homeomorphism group of $\mathcal{J}$ is H.

(3) If the path of a particle is interpreted as a $\mathcal{J}$-continuous map γ of the unit interval I into M such γ preserves order, the image of γ is piecewise linear, consisting of a finite number of straight timelike line segments, like the path of a free particle undergoing a finite number of collisions.

(4) $\mathcal{J}$ is Hausdorff, connected and locally connected, but not normal, locally compact or first countable.

This new topology obviously has several advantages over the standard one, which merely characterizes the continuity of M. Its very definition (1) is more intuitively appealing than that of $\mathcal{M}$, since it requires a set to be open when (a) every inertial observer "times" it to be open,[7] and (b) every section of time simultaneity intersects it in an open set. The definition of $\mathcal{M}$ involves 4–spheres in space–time, which have no particular physical meaning. The idea of (2) incorporating causal, linear and even homothetic structure already in a topology $\mathcal{M}$ is certainly physically appealing, and the idea (3) that the requirement of *continuity* of a curve should already restrict it to be physically meaningful is fascinating. However, there are disadvantages:

(1*) A 3-dimensional section of simultaneity has no meaning in terms of physically possible experiments. Also, the use of *straight* timelike lines in defining $\mathcal{J}$ suggests that $\mathcal{J}$ from the beginning has been equipped with information involving *inertial* observers, so that the emergence of *linear* structure in (2) is less surprising. (Though in fact this is not the reason for its energence.)

(2*) While the *isometry* and *conformal* groups of M are certainly significant physical, it is not necessarily clear that this is so for the *homothecy* group of M.

(3*) The set of $\mathcal{J}$-continuous paths does not incorporate accelerating particles moving under forces in curved lines.[2]

(4*) $\mathcal{J}$ is technically complicated.[2] In particular, the fact that no point has a countable neighborhood basis makes $\mathcal{J}$ hard to calculate with.

Zeeman suggests[2] that criticism (3*) could be overcome by generalizing his theory to general relativity, where the image of γ should become piecewise geodesic (thus accounting for gravitational forces). This generalization has recently been carried out by Göbel[3] who, for strongly causal space–time manifolds, replaced (1) by replacing "time axis" and "spacelike hyperplane" by "timelike geodesic" and "spacelike hypersurface." He then proves that Zeeman's conjecture about the generalization of (3) is correct, and, with the help of a theorem of Hawking[4] relating causal to differential structure, that the homeomorphism group is again the homothety group.

However, even in general relativity, particles need not move along geodesics since, for example, they may be charged and an electromagnetic field may be present (and this applies in special relativity also).[8] Thus the generalization to general relativity of the topology $\mathcal{J}$

(we also call the generalizations $\mathcal{J}$) only partially answers (3*), and leaves the other criticisms (1*), (2*) and (4*) as before.

In this paper, we propose a new topology $\mathcal{P}$ for strongly causal space–times M which share the attractive features of the topologies $\mathcal{J}$, but which also answer some of the above criticisms and have additional attractive physical features. The topology $\mathcal{P}$, the *path* topology, has the following properties:

(1′) $\mathcal{P}$ is defined to be the finest topology on M which induces the Euclidean topology on *arbitrary* (not necessarily smooth) timelike curves.

(2′) The topology $\mathcal{P}$ incorporates the causal, differential, and smooth conformal structure; the homeomorphism group of $\mathcal{P}$ is the *conformal* group.

(3′) The set of $\mathcal{P}$-continuous paths incorporates *all* timelike paths. In fact, the set consists of all "Feynman paths" (described below).

(4′) $\mathcal{P}$ is still technically complicated, but less so than $\mathcal{J}$. Indeed, $\mathcal{P}$ is Hausdorff, connected and locally connected, but not normal or locally compact. However, every point has a countable neighborhood basis, and this makes $\mathcal{P}$ much easier to deal with than $\mathcal{J}$.

The definition (1′) has an immediate physical interpretation which is more appealing than (1): a set is open whenever every observer (accelerated or not) "times" it to be open. No nonphysical experiments are required. Also, (1′) does not require any smoothness properties to define $\mathcal{P}$, so that the emergence of smooth and conformal structure is more surprising. Again, (2′) answers criticism (2*), and (3′) shows that *all* possible observers, accelerated or not, are described by $\mathcal{P}$-continuous curves. The fact that $\mathcal{P}$ has a countable neighborhood basis at each point makes it much easier to deal with than $\mathcal{J}$. Thus the computation of the set of *all* $\mathcal{P}$-continuous paths becomes relatively easy (Zeeman only finds the order preserving paths for $\mathcal{J}$).[9] These paths are (possibly accelerated) "Feynman paths" which zigzag with respect to time orientation, like the Feynman track of an (accelerated) electron. This same "basis" property makes it relatively easy to find the general properties of $\mathcal{P}$. In fact, $\mathcal{P}$ is path connected and locally path connected, but not regular or paracompact, in addition to having the properties mentioned above. The relative ease of calculations with $\mathcal{P}$ suggest that it could be usefully applied to "practical" problems in general relativity. If one could show (using the basis property) that a space–time must admit a (local or global) $\mathcal{P}$-homeomorphism, this would mean that the space–time admits a (local or global) conformal diffeomorphism.

Thus we suggest that, while Zeeman's topology[2] and its general relativistic analogs[3] represent a radical and fascinating departure from the conventional schemes, our topology $\mathcal{P}$ has all of the required features, but is more intuitively appealing, manageable, and physical. Section 2 is devoted to standard results and definitions. In Sec. 3, timelike paths (not necessarily smooth) and Feynman paths are defined, and the topology is described. In Sec. 4 it is shown that $\mathcal{P}$ is not comparable to $\mathcal{J}$, and the important basis property is proved. This is used to show that the set of continuous paths is the set of Feynman paths, and to prove various general properties of $\mathcal{P}$. In Sec. 5 it is shown that $\mathcal{P}$ carries the chronological structure of M, and in Sec. 6, that $\mathcal{P}$ carries the causal, differential, and conformal structure of M. The final theorem is that the homeomorphism group of $\mathcal{P}$ is the group of conformal diffeomorphisms of M.

The burden of the argument is as follows. First we show that $\mathcal{P}$-homeomorphisms take timelike curves to timelike curves. Then we show that this implies that $\mathcal{P}$-homeomorphisms preserve causal relations. This is used to show that they are diffeomorphisms preserving the null cones, that is, conformal diffeomorphisms.

2. STANDARD DEFINITIONS AND RESULTS[5]

Space–time is taken throughout to be a connected, Hausdorff, paracompact, C^∞ real four-dimensional manifold M without boundary, with a C^∞ Lorentz metric (only a few orders of differentiability will actually be needed) and associated pseudo-Riemannian connection. M is taken to be time orientable throughout (that is, M admits a nonvanishing timelike vector field). For subsets A and B of M, the *chronological future* $I^+(A,B)$ *of* A *relative to* B is the set of all points in B which can be reached from A by a future directed smooth (i.e., C^∞) timelike curve in B of finite extent. The *causal future* $J^+(A,B)$ of A *relative to* B is the union of $A \cap B$ with the set of all points in B which can be reached from A by a future directed smooth causal curve (i.e., nonspacelike curve) in B. The *future horismos* $E^+(A,B)$ *of* A *relative to* B is defined as $J^+(A,B) - I^+(A,B)$. These definitions have duals, often regarded as self evident, in which "future" is replaced by "past" and "+" by "−." If A is the singleton set $\{p\}$ for a point $p \in M$, we write $I^+(p,B)$ rather than $I^+(\{p\},B)$, for example, and $I^+(p)$ for $I^+(p,M)$. The relations $p \in I^+(q)$, $p \in J^+(q)$, and $p \in E^+(q)$ will sometimes be written $q \ll p$, $q \leq p$, and $q \to p$, respectively. These last relations are respectively called, chronological, causal, and horismos relations. The definitions and results[5] below will be needed in this paper. They are only stated in the generality needed:

2.1 If $\mathcal{V}$ is an open set, $q \in I^+(p,\mathcal{V})$ implies $p \in I^-(q,\mathcal{V})$, and conversely. Similar results hold for J and E. $I^+(p,\mathcal{V})$ and $I^-(p,\mathcal{V})$ are open sets. In particular, these statements hold for $I^+(p)$ and $I^-(p)$.

2.2 If $\mathcal{V}$ is an open set, either

$$\left.\begin{aligned} &q \in J^+(p,\mathcal{V}), \quad r \in I^+(q,\mathcal{V}) \\ &q \in I^+(p,\mathcal{V}), \quad r \in J^+(q,\mathcal{V}) \end{aligned}\right\} \text{ imply } \quad r \in I^+(p,\mathcal{V}).$$

2.3 Let $T_p(M)$ denote the tangent space of $p \in M$, and $\exp: T_p(M) \to M$, the exponential mapping. Then there is an open neighborhood N of the origin of $T_p(M)$ such that $\mathcal{U} \equiv \exp(N)$ is an open convex normal neighborhood of $p \in M$. That is, every pair of points in $\mathcal{U}$ can be joined by a unique geodesic curve in $\mathcal{U}$, and geodesics in $\mathcal{U}$ through p are the images of straight lines through the origin in $N \subset T_p(M)$. Further, p possesses a neighborhood basis of open convex normal neighborhoods.

Furthermore, the normal neighborhoods may be taken to be normal neighborhoods of every point in them. Let $\epsilon > 0$ be sufficiently small so that the Euclidean open ball B of radius ϵ, centered at the origin, is contained in N. We define $B_u(p,\epsilon) = \exp(B)$, and whenever such a set is referred to, it is assumed that $\epsilon > 0$ is sufficiently small.[10]

2.4 Denote the normal coordinates of $q \in \mathcal{U}$ by $x^i(q)$, where $i = 0, 1, 2, 3$, and x^0 is the time coordinate, x^1, x^2, x^3 the space coordinates. Then $x^i(p) = 0$ and

$$J^+(p,\mathcal{U}) = \{q \in \mathcal{U} \mid (x^0(q))^2 - (x^1(q))^2 - (x^2(q))^2 - (x^3(q))^2 \geq 0,\ x^0(q) \geq 0\}.$$

$J^-(p,\mathcal{U})$ satisfies the same condition except that $x^0(q) \leq 0$, and $I^\pm(p,\mathcal{U})$ are defined analogously, except that all inequalities are strict. Define, for any open set $\mathcal{V}$, the cones (possibly) with and without origin by

$$C(p,\mathcal{V}) = I^+(p,\mathcal{V}) \cup I^-(p,\mathcal{V}),$$

$$K(p,\mathcal{V}) = C(p,\mathcal{V}) \cup \{p\},$$

and, for an open convex normal neighborhood $\mathcal{U}$ of p, define

$$L_u(p,\epsilon) = B_u(p,\epsilon) \cap K(p,\mathcal{U}).$$

Note that $L_u(p,\epsilon) - \{p\}$ is nonempty [the point with coordinates $(\frac{1}{2}\epsilon, 0, 0, 0)$ belongs to $B_u(p,\epsilon) \cap I^+(p,\mathcal{U})$] as is $B_u(p,\epsilon) - L_u(p,\epsilon)$ [take coordinates $(0,0,0,\frac{1}{2}\epsilon)$].

2.5 A space–time M is *chronological*, if there is no closed smooth future (or past) directed timelike curves in M. Equivalently, M is chronological, if and only if, $I^+(p) \cap I^-(p) = \phi$ for all $p \in M$. Analogously, M is *causal*, if there are no future (or past) directed causal curves.

2.6 Denote the manifold topology of M by $\mathcal{M}$. Consider the collection of all sets of the form $I^+(p) \cap I^-(q)$ for $p, q \in M$. These sets are open and (together with, possibly, the empty set ϕ) clearly form a basis for a topology on M. The topology is called the *Alexandroff topology* $\mathcal{A}$ of M, and in general, is coarser than $\mathcal{M}$.

2.7 A *causality neighborhood* D of a point $p \in M$ is an $\mathcal{M}$-neighborhood of p such that, whenever $\gamma: F \to M$ is a smooth causal path, $\gamma^{-1}(D)$ is connected. (Here F is a connected interval of the real line R.) M is *strongly causal at* p, if and only if, p has a neighborhood basis $\{D_\alpha(p) \mid \alpha \in A\}$ such that, for each $\alpha \in A$, $D_\alpha(p)$ is a causality $\mathcal{M}$-neighborhood. M is *strongly causal*, if it is strongly causal at each point. Another useful characterization of this property is given by the following consequence:

M is strongly causal $\Leftrightarrow M$ is $\mathcal{A}$-Hausdorff $\Leftrightarrow \mathcal{A} = \mathcal{M}$

2.8 If $\mathcal{U}$ is an open convex normal neighborhood of $p \in M$, $E^+(p,\mathcal{U})$ consists of future directed null geodesics in $\mathcal{U}$ from p, and $E^+(p,\mathcal{U}) \cup E^-(p,\mathcal{U})$ is the image of the null cone $N \cap N_p \subset T_p(M)$ in the neighborhood N of the tangent space $T_p(M)$ under exp.

2.9 The metric g at $p \in M$ is determined up to a constant by the tangent null cone $N_p \subset T_p(M)$.

2.10 The *isometry*, *homothecy* and *conformal* groups of M are those groups of C^∞ diffeomorphisms of M which preserve, respectively, the metric tensor, the metric tensor up to a *constant* factor, and the metric tensor up to a (possibly variable) factor.

3. THE TOPOLOGY $\mathcal{P}$

3.1 Let F be a connected interval of the real line R. (The singleton closed set $[x,x]$ is excluded. For our purposes we may take F to be bounded, for there is an order preserving C^∞ diffeomorphism $\tan^{-1}: R \to (-\pi/2, \pi/2)$. Thus F is a finite, closed, open, or half open half closed interval.) A map $\gamma: F \to M$ is called a *path*, and its image a *curve* (the same symbol γ often being used for either, it being clear from the context which is meant). A path γ is *continuous* if γ is continuous with respect to $\mathcal{M}$ and the topology on F induced from the standard one on R. A point $q \in M$ is said to be an *initial end point* of a continuous path $\gamma: F \to M$, if for every neighborhood N of q there is a $t_N \in F$, such that $t \in F$ and $t < t_N$ implies $\gamma(t) \in N$. If a continuous path γ has an initial end point $q \notin \gamma(F)$, one may find a new continuous path $\gamma': F \cup \{t_1\} \to M$ such that $\gamma'|_F = \gamma$ and $\gamma'(t_1) = q$ where t_1 is the greatest lower bound of F. We shall therefore assume without loss of generality that continuous paths contain both their initial and final endpoints if they have them.

3.2 A path $\gamma: F \to M$ is called[5] *future directed and timelike at* $t_0 \in F$, if and only if γ is continuous and there is a connected neighborhood U of t_0 in F, and an open convex normal neighborhood $\mathcal{U}$ of $p = \gamma(t_0)$ such that

$$t \in U \text{ and } t < t_0 \Rightarrow \gamma(t) \in I^-(p,\mathcal{U});$$

$$t \in U \text{ and } t > t_0 \Rightarrow \gamma(t) \in I^+(p,\mathcal{U}).$$

A path is called *future directed and timelike*, if it is future directed and timelike at each $t_0 \in F$. Similar dual definitions hold for "past directed." A path is *timelike at* t_0, if it is either future or past directed and timelike at t_0, and *timelike*, if it is either everywhere future directed, or everywhere past directed and timelike. A curve is timelike, if it is the image of a timelike path.

Proposition 3.3: Let $\gamma: F \to M$ be a continuous path which is timelike at each $t_0 \in \text{int}F$. Then γ is a timelike path.

Proof: Suppose γ is future directed at $t_0 \in \text{int}F$. Let U and $\mathcal{U}$ be as above. Suppose there were a $t_1 \in U$ with $t_1 > t_0$ such that γ was past directed at t_1. The coordinate x^0 of $\gamma(t)$ will be a continuous function of t. Thus there will be some t_2 such that $x^0(\gamma(t_2))$ is the maximum value of $x^0(\gamma([t_0, t_1]))$. Since γ is future directed at t_0, t_2 must be greater than t_0. Similarly, t_2 would have to be less than t_1. Consider the point $q = \gamma(t_2)$. On $I^+(q,\mathcal{U})$ the coordinate x^0 would be greater than its value at q. This would mean that $\gamma([t_0, t_1])$ could not be timelike at t_2. This shows that γ must be future directed for all $t \in U$ with $t \geq t_0$. A similar argument shows that γ must be future directed for all $t \in U$ with $t \leq t_0$. Thus, the set of points at which γ is future directed is open in int$F$. Similarly, the set on which γ is past directed is open. Because F is connected, γ must be either everywhere future directed or everywhere past directed in int$F$. Assume, without loss of generality, that γ is future directed. Suppose that $q = \gamma(t_1)$ is an initial endpoint.

Let U be a connected neighborhood of t_1 in F and let $\mathcal{U}$ be a convex normal neighborhood of q. Let $t_2 \in U$ with $t_2 > t_1$. Then $\gamma((t_1, t_2)) \subset I^-(r, \mathcal{U})$, where $r = \gamma(t_2)$. Therefore, by continuity $q \in \overline{I^-(r, \mathcal{U})} = J^-(r, \mathcal{U})$. Thus, $r \in J^+(q, \mathcal{U})$. However, by a similar argument one can find a $t_3 \in (t_1, t_2)$ such that $\gamma(t_3) \in J^+(q, \mathcal{U}) \cap I^-(r, \mathcal{U})$. Therefore, by 2.2, $r \in I^+(q, \mathcal{U})$. Hence γ is also future directed timelike at its endpoints. ∎ It follows from 2.2 and 3.2 that a timelike path is locally 1–1, that is, each $t_0 \in F$ has a neighborhood V such that $\gamma|_V$ is 1–1. Notice that timelike paths need not be smooth. The point of this definition is that $\mathcal{P}$ will be defined independently of smoothness properties, but nevertheless smooth structure will emerge from $\mathcal{P}$ (Theorem 5). Curiously enough, a path may be timelike and smooth without being timelike. Let $\gamma: R \to M$ be the path defined in Minkowski space (usual coordinates) by $\gamma(t) = (t, \sin t, 0, 0)$. Then γ is timelike and smooth, but not smooth timelike, since it is null at the points $t = n\pi$ for integral n. However, $I^+(A, B)$ and $I^-(A, B)$, as defined by smooth timelike curves, agree with $I^+(A, B)$ and $I^-(A, B)$ as defined by timelike curves.

3.4 We now define a class of paths which are similar to timelike paths, in that their curves are constrained to lie within local light cones, but which may zigzag with respect to time orientation. A path $\gamma: F \to M$ is a *Feynman path*, if γ is continuous and, for each $t_0 \in F$, there is an open connected neighborhood U of t_0, and an open convex normal neighborhood $\mathcal{U}$ of $p = \gamma(t_0)$ such that

$$\gamma(U) \subset K(p, \mathcal{U}).$$

A locally 1–1 Feynman path will be called a *Feynman track*. Suppose γ is a Feynman track, 1–1 in a neighborhood V of t_0. Let W be an open connected neighborhood of t_0 in $U \cap V$. Then, using the fact that W is connected, and that $\gamma|_W$ is 1–1 and continuous, it is easily shown that γ is either timelike at t_0 or $\gamma(W) \subset I^-(p, \mathcal{U}) \cup \{p\}$, or $\gamma(W) \subset I^+(p, \mathcal{U}) \cup \{p\}$. Obviously timelike paths are Feynman tracks, but there are many nontimelike Feynman tracks.

3.5 Suppose γ_1 is a timelike curve with future endpoint q, and γ_2 is a timelike curve with past endpoint q. Evidently the union $\gamma_1 \cup \gamma_2$ is also a timelike curve, which may be parametrized as a future or past directed timelike path. Any such path will be called a *product path* $\gamma_1\gamma_2$, and qualified with "future directed" or "past directed," according to the choice of the direction of the parameter. If γ_1 is as before, but with γ_2 now with future endpoint q, we may similarly define product paths, denoted $\widehat{\gamma_1\gamma_2} = \gamma$, which are timelike everywhere at $\gamma^{-1}(q)$. However, $\widehat{\gamma_1\gamma_2}$ is always a Feynman path.

3.6 Define a new topology (the *path* topology) $\mathcal{P}$ of M by specifying the collection $\mathcal{P}$ of open sets of the topology as follows: $\mathcal{P}$ is the finest topology satisfying the requirement that the induced topology on every timelike curve coincides with the topology induced from $\mathcal{M}$. Thus, if a set $E \subset M$ is $\mathcal{P}$-open, for every timelike curve γ there is an $O \in \mathcal{M}$ with

$$E \cap \gamma = O \cap \gamma.$$

Conversely, if E satisfies this condition, it is $\mathcal{P}$-open, and $\mathcal{P}$ is the largest collection of such sets. Obviously $O \in \mathcal{M}$ implies $O \in \mathcal{P}$, so $\mathcal{P}$ is finer than $\mathcal{M}$. We shall see below that it is strictly finer, and that $\mathcal{P}$ is not comparable to the general relatively analog of $\mathcal{J}$.

4. GENERAL PROPERTIES OF

Here we show that $\mathcal{P}$ is strictly finer than $\mathcal{M}$, but not comparable to $\mathcal{J}$, and find an explicit neighborhood basis for $\mathcal{P}$. Then we show that the $\mathcal{P}$-continuous paths are Feynman paths, and various general properties of $\mathcal{P}$ are proved.

Proposition 4.1: Let $\gamma: F \to M$ be a path. If γ is $\mathcal{P}$-continuous, γ is $\mathcal{M}$-continuous. If γ is timelike, γ is $\mathcal{P}$-continuous.

Proof: The first assertion follows since $\mathcal{P}$ is finer than $\mathcal{M}$. For the second, note first that $E \in \mathcal{P}$ implies $E \cap \gamma = O \cap \gamma$ for some $O \in \mathcal{M}$. Hence $\gamma^{-1}(E) = \gamma^{-1}(E \cap \gamma) = \gamma^{-1}(O \cap \gamma) = \gamma^{-1}(0)$. But γ is timelike and so, in particular, continuous. Therefore $\gamma^{-1}(E) = \gamma^{-1}(0)$ is open. ∎

Proposition 4.2: Sets of the form $K(p)$, $K(p, \mathcal{U})$, and $L_u(p, \epsilon)$ are $\mathcal{P}$-open.

Proof: Let γ be any timelike curve. Suppose first that $p \in \gamma$. Then by definition, $\gamma \subset I^+(p) \cup I^-(p) \cup \{p\} = K(p)$, so $\gamma \cup K(p) = \gamma \cup M$. Suppose next that $p \notin \gamma$. Then $\gamma \cap K(p) = \gamma \cap (I^+(p) \cup I^-(p))$. In either case, $\gamma \cap K(p) = \gamma \cap O$ for some $O \in \mathcal{M}$. The proof that $K(p, \mathcal{U})$ is $\mathcal{P}$-open is similar (replace M by $\mathcal{U}$), and the proof for $L_u(p, \epsilon)$ follows because $B_u(p, \epsilon)$, being $\mathcal{M}$-open, is *a fortiori* $\mathcal{P}$-open. ∎

This proposition shows that $\mathcal{P}$ is strictly finer than $\mathcal{M}$, since, for example, $p \in K(p, \mathcal{U})$ has no $\mathcal{M}$-neighborhood contained in $K(p, \mathcal{U})$. We will show further that $\mathcal{P}$ is not comparable to Zeeman-type topologies $\mathcal{J}$. Define S to be the set $[B_u(p, \epsilon) - [E^+(p, \mathcal{U}) \cup E^-(p, \mathcal{U})]]$, where $\epsilon > 0$ is smaller than $\frac{1}{2}$ [and also sufficiently small for $B_u(p, \epsilon)$ to make sense]. Let $\gamma \subset \mathcal{U}$ be the curve defined, in normal coordinates, by the timelike path $\gamma = [-\epsilon, \epsilon] \to \mathcal{U}$, with equation $\gamma(t) = (t, t^2, 0, 0)$. Consider the set $R = (S - \gamma) \cup \{p\}$. Then any timelike geodesic will have an open intersection with R, as will any spacelike hypersurface. Therefore R is $\mathcal{J}$-open. However, R is not $\mathcal{P}$-open since $R \cap \gamma = \{p\}$, which is closed. On the other hand, $K(p, \mathcal{U})$ is $\mathcal{P}$-open but not $\mathcal{J}$-open, since the intersection of any spacelike hypersurface containing p with $K(p, \mathcal{U})$ is $\{p\}$.

Theorem 1: Sets of the form $L_u(p, \epsilon)$ form a basis for the topology $\mathcal{P}$.

Proof: We must show that, for any $\mathcal{P}$-open set E and any $p \in E$, there is a $\mathcal{P}$-open neighborhood of p of the form $L_u(p, \epsilon)$ contained entirely in E. Suppose this to be false. Then there is an open convex normal neighborhood $\mathcal{U}$ of p such that, for every ball $B_u(p, \epsilon) \subset \mathcal{U}$ and corresponding $L_u(p, \epsilon)$, there is a $q \in L_u(p, \epsilon)$ with $q \notin E$. Fix such a set $L_u(p, \epsilon_1)$ and assume, without loss of generality, that it contains a q_1, not in E, with $q_1 \in I^+(p, \mathcal{U})$. [If there is no such q_1, all points of the required type lie in $I^-(p, \mathcal{U})$, and the proof is as before with I^+, I^- and "future," and "past" interchanged.] Since p belongs to the open set $I^-(q_1, \mathcal{U})$, we can find a $\delta > 0$ with $B_u(p, \delta) \subset I^-(p, \mathcal{U})$. Let ϵ_2 be any positive number satisfying $\epsilon_2 \leq \min(\epsilon, \frac{1}{2}\epsilon_1)$. There is a $p_2 \in L_u(p, \epsilon_2)$ with

$q_2 \notin E$. Assume without loss of generality that $q_2 \in I^+(p, \mathcal{U})$. [If all q_2 of the required type lie in $I^-(p, \mathcal{U})$, discard q_1 and start with q_2, interchanging I^+, I^- and "future," "past."] Construct $\epsilon_3 \leq \frac{1}{2}\epsilon_2$ and $q_3 \in L_{\mathcal{U}}(p, \epsilon_3)$ analogously, and inductively construct $\epsilon_{n+1} \leq \frac{1}{2}\epsilon_n \leq 2^{-n}\epsilon_1$ and $q_{n+1} \in L_{\mathcal{U}}(p, \epsilon_{n+1})$. This way, possibly after discarding a finite number of points, we obtain a sequence $S = \{q_n\}$ of distinct points $S \subset I^+(p, \mathcal{U})$ and $S \not\subset E$, with the property that consecutive points q_s and q_{s+1} can be joined by a (unique) timelike geodesic curve. Form a past directed timelike product path $\xi \equiv (q_1 q_2)(q_2 q_3) \cdots$ as in 3.5. Because $\epsilon_{n+1} \leq 2^{-n}\epsilon_1$, $\{q_n\}$ converges to p in the topology $\mathcal{M}$, and thus ξ has past endpoint p. From 3.1 there is, therefore, a unique timelike curve $\gamma = \xi \cup \{p\}$. Since E is $\mathcal{P}$-open, $E \cap \gamma = O \cap \gamma$ for some $O \in \mathcal{M}$. Since $S \not\subset E$, $S \not\subset E \cap \gamma = O \cap \gamma$, therefore $S \subset M - (O \cap \gamma) = (M - O) \cup (M - \gamma)$. But $S \subset \gamma$, so $S \not\subset M - \gamma$; therefore, $S = \{q_n\} \subset M - O$. But $M - O$ is $\mathcal{M}$-closed and $\{q_n\}$ is $\mathcal{M}$-convergent to p, therefore $p \in M - O$, so $p \notin O$. Hence, $p \notin O \cap \gamma = E \cap \gamma$. But p belongs to both E and γ, so we have a contradiction, and the theorem is proved. ∎

This basis property, which has no analogue in the fine topologies $\mathcal{J}$, makes the path topology $\mathcal{P}$-much more manageable. Below, it will be used to prove all the basic general properties of $\mathcal{P}$. But first, we shall find the set of all $\mathcal{P}$-continuous paths.

Theorem 2: A path $\gamma: F \to M$ is $\mathcal{P}$-continuous if and only if it is a Feynman path.

Proof: Suppose first that γ is a Feynman path. We must show that, for each $t_0 \in F$, the inverse image of each $\mathcal{P}$-neighborhood of $\gamma(t_0)$ is a neighborhood of t_0. Let $\mathcal{U}$ be the open convex normal neighborhood of $p = \gamma(t_0)$ used in the definition 3.2. Let $\{L_u(p, \epsilon)\}$ be a $\mathcal{P}$-neighborhood basis of p for all sufficiently small $\epsilon > 0$. Then $\gamma^{-1}(L_u(p, \epsilon)) = \gamma^{-1}(K(p, \mathcal{U}) \cap B_u(p, \epsilon)) = \gamma^{-1}(K(p, \mathcal{U})) \cap \gamma^{-1}(B_u(p, \epsilon))$. But $\gamma^{-1}(K(p, \mathcal{U}))$ contains an open connected neighborhood U of t_0, and $\gamma^{-1}(B_u(p, \epsilon))$ is an open neighborhood of t_0 (because γ is $\mathcal{M}$-continuous). Thus γ is $\mathcal{P}$-continuous.

Suppose next that γ is $\mathcal{P}$-continuous, and let $\mathcal{U}$ be an open convex normal neighborhood of $p = \gamma(t_0)$. Then $K(p, \mathcal{U})$ is a $\mathcal{P}$-neighborhood of p, so $\gamma^{-1}(K(p, \mathcal{U}))$ is a neighborhood W of t_0. Let U be an open connected neighborhood of t_0 in W. Then $\gamma(U) \subset K(p, \mathcal{U})$, and γ, being also $\mathcal{M}$-continuous, is a Feynman path.

Obviously $\mathcal{P}$ is first countable, since a countable neighborhood base at $p \in M$ is given by $\{L_u(p, 1/n); n = N, N+1 \cdots\}$ for some integer N. Because M admits a countable covering $\{\mathcal{U}_i\}$ by normal neighborhoods, $\mathcal{P}$ is also separable (take points with rational coordinates in $\mathcal{U}_i$). Note that sets of the form $K(p, \mathcal{U})$, $I^+(p, \mathcal{U})$, $I^-(p, \mathcal{U})$ and $L_u(p, \epsilon)$ are $\mathcal{P}$-path connected, since any pair of points in any of these sets can be joined by product curves $\gamma_1\gamma_2$ or $\widehat{\gamma_1\gamma_2}$ of the type discussed in 3.5. Thus $\mathcal{P}$ is locally path connected, because every $\mathcal{P}$-neighborhood E of $p \in M$ contains a neighborhood of the form $L_{\mathcal{U}}(p, \epsilon)$. Also it is not hard to show that $\mathcal{P}$ is path connected. Indeed, since M is $\mathcal{M}$-connected and M is a manifold, M is $\mathcal{M}$-path connected. If p, $q \in M$ are any pair of points, they may be joined by an $\mathcal{M}$-continuous path $\gamma: I \to M$ from the closed unit interval I into $\mathcal{M}$. It is then easy to approximate $\gamma(I)$ by a Feynman path, showing that M is $\mathcal{P}$-path connected. Before proving further general properties, we need:

Proposition 4.3: $\overline{L_u(p, \epsilon)}^{\mathcal{P}} = \overline{L_u(p, \epsilon)} - (\partial B_u(p, \epsilon) \cap \partial K(p, \mathcal{U}))$. (Here the lefthand side denotes the $\mathcal{P}$-closure, and all other topological symbols refer to $\mathcal{M}$.)

Proof: Certainly $\overline{L_u(p, \epsilon)}^{\mathcal{P}} \subset \overline{L_u(p, \epsilon)}$ since $\mathcal{P}$ is finer than $\mathcal{M}$. Now points $q \in \partial B_u(p, \epsilon) \cap \partial K(p, \mathcal{U})$, have $\mathcal{P}$-open neighborhoods of the form $L_u(q, \frac{1}{2}\epsilon)$ which do not meet $L_u(p, \epsilon)$, so $L_u(p, \epsilon)^{\mathcal{P}} \subset L_u(p, \epsilon) - (\partial B_u(p, \epsilon) \cap \partial K(p, \mathcal{U}))$. If $r \in M$ is in this latter set, every $\mathcal{P}$-neighborhood of r contains a $\mathcal{P}$-neighborhood of the form $L_u(r, \delta)$ which is easily seen to meet $L_u(p, \epsilon)$. ∎

Theorem 3: $\mathcal{P}$ is first countable and, separable. $\mathcal{P}$ is Hausdorff, path connected and locally path connected (and so *a fortiori* connected and locally connected). However, $\mathcal{P}$ is not regular, normal, locally compact or paracompact.

Proof: The first sentence has been dealt with above. $\mathcal{P}$ is Hausdorff because $\mathcal{P}$ is finer than $\mathcal{M}$. The connectivity properties have been dealt with above. To show that $\mathcal{P}$ is not regular, consider the $\mathcal{P}$-neighborhood $L_{\mathcal{U}}(p, \epsilon)$ and show that p has no $\mathcal{P}$-closed neighborhood S contained in $L_{\mathcal{U}}(p, \epsilon)$, using the basis property and 4.3. To show that $\mathcal{P}$ is not normal, consider the disjoint $\mathcal{P}$-closed subsets $\overline{L_{\mathcal{U}}(p, \epsilon)}^{\mathcal{P}}$ and $[\partial B_{\mathcal{U}}(p, \epsilon) \cap \partial K(p, \mathcal{U})]$, and show, using the basis property, that these sets cannot have disjoint $\mathcal{P}$-open neighborhoods. To show that $\mathcal{P}$ is not locally compact, use the fact that closed subspaces of compact sets are compact, the basis property and 4.3, and note that $\overline{L_u(p, \epsilon)}^{\mathcal{P}}$ is not $\mathcal{M}$-closed, so certainly not $\mathcal{M}$-compact, hence not $\mathcal{P}$-compact, since $\mathcal{P}$ is finer than $\mathcal{M}$. $\mathcal{P}$ cannot be paracompact because paracompact spaces are normal.[11] ∎

5. $\mathcal{P}$ AND CHRONOLOGICAL STRUCTURE

We wish to prove $\mathcal{P}$-homeomorphisms h, take timelike curves to timelike curves. Obviously h takes $\mathcal{P}$-continuous curves to $\mathcal{P}$-continuous curves, but, of course, $\mathcal{P}$-continuous curves (Feynman paths) need not be timelike. We single out a subclass of $\mathcal{P}$-continuous curves, by adding restrictions made only in terms of $\mathcal{P}$. This subclass will coincide with timelike curves. This will enable us to prove that for strongly causal spacetimes $\mathcal{P}$-homeomorphisms preserve or reverse causal relations.

Definition 5.1: A path $\gamma: F \to M$ is said to be *regular*, if and only if:

(A) γ is $\mathcal{P}$-continuous and locally 1–1.

(B) For every $t_0 \in F$, there is a connected neighborhood U of t_0 and a $\mathcal{P}$-neighborhood Π_0 of $p = \gamma(t_0)$, such that:

(1) $\gamma(U) \subset \Pi_0$

(2) Whenever $t_0 \in \text{int}F$ (the interior of F) and $a, b \in U$ satisfy $a < t_0 < b$, every $\mathcal{P}$-continuous curve in Π_0 joining $\gamma(a)$ to $\gamma(b)$ contains $p = \gamma(t_0)$.

Proposition 5.1: A $\mathcal{P}$-homeomorphism takes regular paths to regular paths.

Proof: The definition of regularity involves only set theoretical and $\mathcal{P}$-topological notions. Since h is a $\mathcal{P}$-homeomorphism, it preserves all required properties. ∎

Theorem 4: γ: $F \to M$ is timelike, if and only if, γ is regular.

Proof: Suppose first that γ is timelike and, for definiteness, future directed. By 4.1 γ is $\mathcal{P}$-continuous, so satisfies (A) of the regularity condition. For $t_0 \in \text{int}F$ let U and $\mathcal{U}$ be as in 3.2. Set $\Pi_0 = K(p, \mathcal{U})$ where $p = \gamma(t_0)$. From 4.2, Π_0 is a $\mathcal{P}$-neighborhood of $p = \gamma(t_0)$. Since γ is future directed and timelike, $\gamma(U) \subset \Pi_0$. Suppose next that $a, b \in U$ satisfy $a < t_0 < b$. Since γ is future directed $\gamma(a) \in I^-(p, \mathcal{U})$ and $\gamma(b) \in I^+(p, \mathcal{U})$. Let $\zeta = [r, s] \to \Pi_0$ be a $\mathcal{P}$-continuous path with $\gamma(a) = \zeta(r)$ and $\gamma(b) = \zeta(s)$. Since ζ is $\mathcal{P}$-continuous, ζ is, by 4.1, continuous, so $\zeta([r, s])$ is connected. If $p \notin \zeta([r, s])$, $\zeta([r, s]) \subset I^+(p, \mathcal{U}) \cap I^-(p, \mathcal{U})$, then $\zeta([r, s])$ is contained in a disjoint union of open sets, and meets both. This contradicts the connectivity of $\zeta([r, s])$, so in fact, $p \in \zeta([r, s])$ and γ is regular. An analogous proof holds if γ is past directed and timelike, so the first half of the proposition is proved.

Suppose next that γ is regular, and that Π_0 is a $\mathcal{P}$-neighborhood of $p = \gamma(t_0)$ satisfying the required conditions. Because γ is $\mathcal{P}$-continuous, γ is continuous. From Theorem 1, p has a $\mathcal{P}$-neighborhood of the form $L_{u'}(p, \epsilon)$ contained in Π_0. Choose an open convex normal neighborhood $\mathcal{U}$ of p with $\mathcal{U} \subset B_{u'}(p, \epsilon)$. Then the $\mathcal{P}$-neighborhood $K(p, \mathcal{U})$ of p is contained in $L_u(p, \epsilon)$. Since γ is $\mathcal{P}$-continuous, $\gamma^{-1}(K(p, \mathcal{U}))$ is a neighborhood of t_0. Let $U \subset \gamma^{-1}(K(p, \mathcal{U}))$ be a connected neighborhood of t_0 such that γ: $U \to M$ is 1—1. Then $\gamma(U) \subset K(p, \mathcal{U})$. Assume now that $t_0 \in \text{int}F$, and choose $a, b \in U$ with $a < t_0 < b$. Since γ is 1—1 on U, both $\gamma(a)$ and $\gamma(b)$ lie in $I^+(p, \mathcal{U}) \cup I^-(p, \mathcal{U})$. Assume for definiteness that $\gamma(a) \in I^-(p, \mathcal{U})$. Then $\gamma(b)$ cannot belong to $I^-(p, \mathcal{U})$ also, since otherwise, in view of the $\mathcal{P}$-path connectivity of $I^-(p, \mathcal{U})$, there would be a $\mathcal{P}$-continuous path in $I^-(p, \mathcal{U})$ which joins $\gamma(a)$ to $\gamma(b)$ but which does not contain p since $p \notin I^-(p, \mathcal{U})$. Therefore, $\gamma(a) \in I^-(p, \mathcal{U})$ and $\gamma(b) \in I^+(p, \mathcal{U})$. Let U_0^- and U_0^+ denote the disjoint connected intervals $\{t \in U | t < t_0\}$ and $\{t \in U | t > t_0\}$, respectively. Since γ is 1—1, $\gamma(U) \subset K(p, \mathcal{U})$, and $\gamma(U_0^-)$ and $\gamma(U_0^+)$ both belong to the disjoint union $I^-(p, \mathcal{U}) \cup I^+(p, \mathcal{U})$ of open sets. But γ is $\mathcal{P}$-continuous, hence continuous, therefore $\gamma(U_0^-)$ and $\gamma(U_0^+)$ are connected. However, we have just shown that $\gamma(U_0^-)$ meets $I^-(p, \mathcal{U})$ and $\gamma(U_0^+)$ meets $I^+(p, \mathcal{U})$, so in fact $\gamma(U_0^-) \subset I^-(p, \mathcal{U})$ and $\gamma(U_0^+) \subset I^+(p, \mathcal{U})$, therefore γ is future directed and timelike at $t_0 \in \text{int}F$ (or past directed and timelike at $t_0 \in \text{int}F$). Applying the same reasoning to each $t_0 \in \text{int}F$, we conclude that γ is 1—1 and timelike at each $t_0 \in \text{int}F$ and therefore, by proposition 3.3, γ is a timelike path ∎

Here and henceforth, M is always assumed chronological.

Proposition 5.3: A $\mathcal{P}$-homeomorphism h takes cones $C(p)$ bijectively onto cones; $h(C(p)) = C(h(p))$.

Proof: There is a timelike path γ joining q to p if and only if $q \in C(p)$. By proposition 5.1 and Theorem 4, the image path $h_0\gamma$ is timelike, and it joins $h(q)$ to $h(p)$. Therefore $h(q) \in C(h(p))$, and since h is bijective, it maps $C(p)$ bijectively onto $C(h(p))$.

Proposition 5.4: For a fixed $r \in M$, a $\mathcal{P}$-homeomorphism h maps $I^+(r)$ [respectively $I^-(r)$] bijectively onto either $I^+(h(r))$ [respectively $I^-(h(r))$] or $I^-(h(r))$ [respectively $I^+(h(r))$].

Proof: Suppose first that there is a $p \in I^+(r)$ with $h(p) \in I^+(h(r))$ and let $q \in I^-(r)$. Join p to r and r to q by past directed timelike paths η, respectively; and form a past directed product path $\gamma = \eta\zeta$. The image curves $h_0\eta$, $h_0\zeta$, and $h_0\gamma$ are timelike, and since $h(p) \in I^+(h(r))$, $h_0\eta$ is past directed. Therefore $h_0\gamma$ is past directed and timelike. Hence, $h(q) \in I^-(h(r))$ for every $q \in I^-(r)$. A similar construction starting with a given $q \in I^-(r)$ shows that $h(s) \in I^+(h(r))$ for all $s \in I^+(r)$. This completes the "either" part of the proof. If there is no $p \in I^+(r)$ with $h(p) \in I^+(h(r))$, a similar construction gives the remainder of the proof. ∎

Proposition 5.4: The previous proposition holds with "a fixed $r \in M$" replaced by "every $r \in M$."

Proof: Suppose that, for a given $r \in M$, h preserved time orientation, $h(I^+(r)) = I^+(h(r))$ and $h(I^-(r)) = I^-(h(r))$. Let $A \subset M$ be the set of points at which h preserves time orientation. Then certainly $p \in A$ whenever $C(p) \cap C(r) \neq \emptyset$. We assert that A is $\mathcal{M}$-open. Indeed, choose a point $s \in I^+(r)$. Then $r \in I^-(s)$ and since $I^-(s)$ is open, there is an open neighborhood U of r with $U \subset I^-(s)$. Then for any $q \in U$, $I^+(q) \cap I^+(r)$ contains s, so *a fortiori* $s \in C(q) \cap C(r)$. Therefore $q \in U \subset A$, and A is $\mathcal{M}$-open. The set $M - A$ of points at which h is time orientation reversing is also open. But M is connected and $r \in A$, so $M = A$. If there is no $r \in M$ satisfying the above condition, every point satisfies the opposite condition. ∎

Proposition 5.5: A $\mathcal{P}$-homeomorphism is an $\mathcal{A}$-homeomorphism.

Proof: This follows immediately from 5.4 and the definition of the Alexandroff topology.

6. $\mathcal{P}$ AND CAUSAL, DIFFERENTIAL AND CONFORMAL STRUCTURE

The fact that $\mathcal{P}$-homeomorphisms preserve or reverse chronological relations enable us to prove that, for strongly causal space—times, they locally preserve or reverse causal relations. In particular, they preserve null geodesics. It is then shown that $\mathcal{P}$-homeomorphisms are diffeomorphisms and, since they preserve null cones, conformal diffeomorphisms.

Proposition 6.1: Suppose now and henceforth M is strongly causal. Then a $\mathcal{P}$-homeomorphism h is $\mathcal{M}$-homeomorphism, and h maps null geodesic curves to null geodesic curves.

Proof: Since M is strongly causal, it is *a fortiori* chronological, hence h is a $\mathcal{A}$-homeomorphism. Strong causality also implies $\mathcal{A} = \mathcal{M}$, so h is an $\mathcal{M}$-homeomorphism. Let D be a causality neighborhood of $r \in M$ and $\mathcal{U} \subset D$ an open convex normal neighborhood of r. Then causality relations *restricted* to D, agree with causality relations *relative* to D. Now, in $\mathcal{U} \subset D$ the horismos relations can be expressed in terms of chronology

relations since, for $p, q \in \mathcal{U}$, $q \to p$ is and only if $[q \not\leq p$ and $z < q \Rightarrow z < p]$. Since h is a $\mathcal{H}$-homeomorphism, $h(\mathcal{U})$ and $h(D)$ are $\mathcal{H}$-open. Let $\mathcal{U}'$, D' be respectively, an open convex normal neighborhood and a chronology neighborhood of $h(r)$ with $\mathcal{U}' \subset D' \subset h(\mathcal{U})$. In $\mathcal{V} \equiv \mathcal{U} \cap h^{-1}(\mathcal{U}')$, h will preserve chronology (and hence horismos) relations or reverse them. Suppose γ is the unique null geodesic curve joining $s, t \in \mathcal{U} \cap h^{-1}(\mathcal{U}')$. Then $t \in \gamma$ if and only if $r \to t \to s$. Therefore $h(t)$ satisfies $h(r) \to h(t) \to h(s)$ or $h(s) \to h(t) \to h(r)$, that is $h(t)$ lies on a null geodesic joining $h(r)$ to $h(s)$. Hence, $h(\gamma)$ is a null geodesic curve in $h(\mathcal{U}) \cap \mathcal{U}'$. ∎

Theorem 5: A $\mathcal{H}$-homeomorphism $h: M \to M$, which takes null geodesic curves to null geodesic curves is a C^∞ diffeomorphism.

Proof: (This is a theorem of Hawking[4] which is given here in an improved from since it has never been published.) Let $\mathcal{U}$ be a convex normal neighborhood and $\gamma_i: F_i \to \mathcal{U} (i = 1\text{—}4)$ be four C^∞ null geodesic paths such that:

(1) For each $t_1 \in F_1$, there is a unique null geodesic curve λ in $\mathcal{U}$ joining the point $\gamma_1(t_1)$ to the null geodesic curve γ_2.

(2) For each $t_3 \in F_3$ there is a unique point $q \in \lambda$, such that q and $\gamma_3(t_3)$ lie on a null geodesic curve in $\mathcal{U}$.

(3) For each point $q \in \lambda$ there is a unique $t_4 \in F_4$, such that q and $\gamma_4(t_4)$ lie on a null geodesic curve in $\mathcal{U}$.

(4) The map $\psi: F_1 \times F_3 \to F_4$ defined by $\psi(t_1, t_3) = t_4$, where t_1, t_3 and t_4 are as in (1)—(3) above is C^∞, and is such that $\partial\psi/\partial t_1$ and $\partial\psi/\partial t_3$ are nonzero.

For a sufficiently small neighborhood $\mathcal{U}$, the metric differs by an arbitrarily small amount from that of Minkowski space. Comparison with Minkowski space shows that γ_i can be chosen to satisfy the above conditions. Condition (4) cannot be satisfied in less than three dimensions.

By the assumption of the theorem $h(\gamma_i)$ will be four null geodesic curves contained in $h(\mathcal{U})$. Thus one can find four C^∞ paths $\tilde{\gamma}_i: \tilde{F}_i \to h(\mathcal{U})$ which define the same null geodesic curves as $h(\gamma_i)$, but which may be parametrized differently. Let $\hat{h}_i: F_i \to \tilde{F}_i$ be defined by $\hat{h}_i = \tilde{\gamma}_i^{-1} h \gamma_i$. The maps $\hat{h}_i$ will be continuous and monotonic (because h preserves or reverses ordering). Therefore, by Lesbegue's theorem, they will be differentiable almost everywhere. Let $\tilde{\psi}: \tilde{F}_1 \times \tilde{F}_3 \to \tilde{F}_4$ be defined similarly to ψ. Then

$$\hat{h}_4(\psi(t_1, t_3)) = \tilde{\psi}(\hat{h}_1(t_1), \hat{h}_3(t_3)). \tag{I}$$

Differentiating (I) with respect to t_3 one has

$$h_4'(\psi(t_1, t_3)) \frac{\partial\psi}{\partial t_3} = \frac{\partial\tilde{\psi}}{\partial\tilde{t}_3} \hat{h}_3'(t_3). \tag{II}$$

Because $\hat{h}_3$ is differentiable almost everywhere, it follows from property (4) that $\hat{h}_4'$ exists and is continuous. Similarly, by choosing different combinations of null geodesic paths one can show that each $\hat{h}_i$ is C^1. Now, differentiating (II) with respect to t_1 gives

$$\hat{h}_4''(\psi) \frac{\partial\psi}{\partial t_1} \frac{\partial\psi}{\partial t_3} + \hat{h}_4'(\psi) \frac{\partial^2\psi}{\partial t_1 \partial t_3} = \frac{\partial^2\tilde{\psi}}{\partial\tilde{t}_1 \partial\tilde{t}_3} \hat{h}_1' \hat{h}_3'.$$

Therefore, $\hat{h}_4$ is C^2. By repeating the above process, it may be shown that each $\hat{h}_i$ is C^∞. In other words, h maps a C^∞ parameter on a null geodesic curve to a C^∞ parameter.

Let γ_i. $F_i \to \mathcal{U}$ be four C^∞ null geodesic curves, and $W \subset \mathcal{U}$ be a neighborhood such that the map Γ: $W \to$ {an open set of R^4} defined by $\Gamma(q) = \gamma_i^{-1}(I^+(q, \mathcal{U}))$ is a C^∞ diffeomorphism. (Comparison with Minkwoski's space shows that this is possible for W sufficiently small.) Pairs of the form (W, Γ) form a C^∞ atlas for M which is preserved by h. Thus h is a C^∞ diffeomorphism. ∎

Corollary: A $\mathcal{P}$-homeomorphism is a C^∞ diffeomorphism. ∎

Theorem 6: A $\mathcal{P}$-homeomorphism h is a smooth conformal diffeomorphism.

Proof: Since h is a diffeomorphism which, locally, preserves null cones, and the metric g at $p \in M$ is determined up to a constant by the tangent null cone, h preserves the metric up to a constant factor which must, since h is smooth, be smooth. ∎

Theorem 7: The group Homeo $(M, \mathcal{P})$ of $\mathcal{P}$-homeomorphisms of M coincides with the group G of conformal diffeomorphisms of M.

Proof: By Theorem 6, Homeo $(M, \mathcal{P}) \subset G$, and it remains to prove the opposite inclusion. Suppose $E \in \mathcal{P}$, so that whenever γ is timelike, $E \cap \gamma = O \cap \gamma$ for some $O \in \mathcal{H}$. Then if $g \in G$ $g(E) \cap g(\gamma) = gO \cap g\gamma$. But $g\gamma$ is timelike because h is conformal, and gO is $\mathcal{H}$-open. Therefore, $gE \in \mathcal{P}$, and g is $\mathcal{P}$-open. Similarly, g is $\mathcal{P}$-continuous so $g \in$ Homeo $(M, \mathcal{P})$. ∎

It is instructive to give an example of a manifold for which G is strictly larger than the homothecy group. This is not the case for Minkowski space because, though the infinitesimal conformal group is larger than the infinitesimal homothesy group, the infinitesimal conformal group cannot be exponentiated to give a global action on Minkowski space. However, consider the manifold N obtained by removing the following closed set S from Minkowski space M.

$$S = \{q \in M \mid [x^0(q)]^2 - [x^1(q)]^2 - [x^2(q)]^2 - [x^3(q)]^2 \geq 0\}.$$

The conformal group of this manifold is generated by the homogeneous Lorentz group (including space, time, and space—time reversal), dilatations, and the inversion I given, in coordinates, by

$$x^\mu(I(q)) = \frac{x^\mu(q)}{[x^0(q)]^2 - [x^1(q)]^2 - [x^2(q)]^2 - [x^3(q)]^2},$$

$(\mu = 0, 1, 2, 3)$.

In fact, infinitesimal conformal diffeomorphisms which are not infinitesimal isometries are rather rare. DeFrise-Carter[6] has shown that the infinitesimal conformal diffeomorphisms of Lorentz manifolds are with two exceptions, essentially isometries. The exceptions are Minkowski space and the "plane wave" space—times. In the former, there are five linearly independent infinitesimal conformal transformations which are not isometries (the dilatations and "accelerations"),

and in the latter, only one (the dilations). Only the homothecy group acts globally on Minkowski space, but N admits global conformal transformations which are neither isometries nor homothesies.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Rüdiger Göbel for an interesting seminar at Cambridge which stimulated this paper. One of us (P. J. M.) gratefully acknowledges the support of a Science Research Council fellowship. Another of us (A. R. K.) gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft during part of the time spent on this paper.

*Research supported in part by the National Science Foundation [MPS75-01398], by the United Kingdom Science Research Council, and by the Deutsche Forschungsgemeinschaft.
†Sherman Fairchild Distinguished Scholar at Caltech.
‡Present address: First Institute for Theoretical Physics, 2 Hamburg 36, West Germany.
[1]E. C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[2]E. C. Zeeman, Topology **6**, 161 (1967).
[3]R. Göbel, Physikal Teil (II) der Habilitationsschrift, Würzburg (1973), and "Zeeman Topologies on Space-Times of General Relativity Theory," to appear in Commun. Math. Phys.
[4]S. W. Hawking, "Singularities and the Geometry of Space–Time" (Adams Prize Essay, 1966, unpublished), p. 116.
[5]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U. P., London, 1973).
[6]L. DeFrise-Carter, Comm. Math. Phys. **40**, 273 (1975).
[7]*Added in proof*: We are grateful to Dr. M. Dobson for pointing out that the inverted commas on "times" are essential. The observor does not measure the length of a time interval—many experiments are required to determine whether a set is open.
[8]*Added in proof*: Rüdiger Göbel informed us that he has a modification of the general relativity analog of $\mathcal{F}$ which allows the effects of a *fixed* electromagnetic field to be incorporated. We feel it is preferable to use $\mathcal{P}$, thus allowing *all* timelike curves to be continuous (not just geodesics or particles with a fixed charge in a fixed field).
[9]*Added in proof*: Actually the Zeeman topology, and Göbel's generalization admit spacelike curves as continuous curves.
[10]*Added in proof*: We may also assume $\mathcal{U}$ to be an open convex normal neighborhood of each of its points.
[11]*Added in proof*: It may also be of interest to note that $\mathcal{P}$ is not metrizable, since it is separable but not regular, and neither can $\mathcal{P}$ arise from a uniformity, since it is not regular, therefore certainly not completely regular.

Kerr black holes in a magnetic universe

Frederick J. Ernst and Walter J. Wild

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 27 June 1975)

We present exact expressions for the electromagnetic fields associated with arbitrarily charged Kerr–Newman black holes in a magnetic universe. In the particular case of charge $Q = 2\ B_0\ J$, where B_0 is the magnetic field parameter and $J = ma$ is the angular momentum, exact expressions for the gravitational field are also presented, while for arbitrarily charged black holes metrical corrections of order B_0 are evaluated.

In this paper we shall employ a method devised[1] by Ernst in order to extend in two respects the study of Kerr black holes in uniform magnetic fields initiated[2] last year by Wald. First, we shall construct an *exact* solution of the Einstein–Maxwell field equations, which upon linearization in the magnetic field parameter B_0 reduces to Wald's approximate solution with charge $Q = 2B_0 J$, which he identified as describing the end product of charge accretion. In the second place, we shall find the exact electromagnetic fields in the case $Q \neq 2B_0 J$, but we shall work out the metrical corrections only to the first order in B_0. It should be mentioned that the exact solutions always approach asymptotically the Melvin magnetic universe,[3] for it is impossible to have a "uniform magnetic field" extending throughout all space while asymptotic flatness is preserved.

1. MAGNETIZING THE KERR-NEWMAN METRIC

The Kerr–Newman metric can be expressed in the form

$$ds^2 = \Sigma\left(\frac{dr^2}{\Delta} + d\theta^2 - \frac{\Delta}{A}\,dT^2\right) + \frac{A\sin^2\theta}{\Sigma}\left(d\phi - \frac{(2mr - e^2)\,a}{A}\,dT\right)^2, \tag{1.1}$$

where

$$\Delta = r^2 + a^2 - (2mr - e^2),\quad \Sigma = r^2 + a^2\cos^2\theta,$$
$$A = (r^2 + a^2)^2 - \Delta a^2\sin^2\theta. \tag{1.2}$$

By comparing our Eq. (1.1) with Eq. (2.1) of Ref. 1, we can identify the fields

$$d\zeta = 2^{-1/2}\left(\frac{dr}{\Delta^{1/2}} + i\,d\theta\right),\quad \rho = \Delta^{1/2}\sin\theta,$$
$$P = (A^{1/2}\sin\theta)^{-1},\quad f = -A\sin^2\theta/\Sigma,$$
$$\omega = (2mr - e^2)\,a/A. \tag{1.3}$$

The orthonormal components of the electromagnetic field for a "locally nonrotating observer" are derivable from a complex potential Φ such that

$$H_r + iE_r = P\frac{\partial\Phi}{\partial\theta},\quad H_\theta + iE_\theta = -P\Delta^{1/2}\frac{\partial\Phi}{\partial r}, \tag{1.4}$$

while the complex gravitational potential $\mathcal{E}$ is defined by

$$\mathcal{E} = f - |\Phi|^2 + i\varphi, \tag{1.5}$$

where φ is the twist potential. If one introduces the symbol

$$\nabla = \Delta^{1/2}\frac{\partial}{\partial r} + i\frac{\partial}{\partial\theta}, \tag{1.6}$$

the twist potential may be evaluated by using the equation

$$-\rho^{-1}f^2\nabla\omega = i\nabla\varphi + \Phi^*\nabla\Phi - \Phi\nabla\Phi^*. \tag{1.7}$$

For the Kerr–Newman metric the complex potentials Φ and $\mathcal{E}$, which were displayed in Ref. 1 also, may be written in the following form:

$$\Phi = -ie\cos\theta + \frac{ea}{r + ia\cos\theta}\sin^2\theta,$$
$$\mathcal{E} = -[(r^2 + a^2)\sin^2\theta + e^2\cos^2\theta] + 2mai\cos\theta(3 - \cos^2\theta) - 2a\frac{ma\sin^2\theta + ie^2\cos\theta}{r + ia\cos\theta}\sin^2\theta. \tag{1.8}$$

According to Ref. 1 magnetization of the Kerr–Newman metric will be achieved if we replace f and ω by f' and ω', where

$$f' = |\Lambda|^{-2}f, \tag{1.9}$$
$$\nabla\omega' = |\Lambda|^2\nabla\omega + \rho f^{-1}(\Lambda^*\nabla\Lambda - \Lambda\nabla\Lambda^*). \tag{1.10}$$

It is a simple matter to evaluate the complex field

$$\Lambda = 1 + B_0\Phi - \tfrac{1}{4}B_0^2\mathcal{E}, \tag{1.11}$$

but the determination of ω' entails considerable labor. Finally, to obtain the orthonormal components of the electromagnetic field, one must replace Φ in Eq. (1.4) by

$$\Phi' = \Lambda^{-1}(\Phi - \tfrac{1}{2}B_0\mathcal{E}). \tag{1.12}$$

2. EXACT SOLUTIONS ($Q = 2\,B_0 J$)

If one restricts attention to a vacuum metric such as the uncharged Kerr metric, Eq. (1.10) can be replaced by the simpler equation

$$d(\omega' - \omega) = \tfrac{1}{16}B_0^4[\varphi\,d\chi - (f^2 + \varphi)\,d\omega], \tag{2.1}$$

where χ is a new potential such that

$$-i\rho f^{-2}\nabla(f^2 + \varphi^2) = \nabla\chi, \tag{2.2}$$

and, of course,

$$-i\rho f^{-2}\nabla\varphi = \nabla\omega. \tag{2.3}$$

The existence of the potential χ is guaranteed by the vacuum field equations.[4]

In the case of the Kerr metric we have evaluated χ and have succeeded in integrating Eq. (2.1). The result is expressible in the form

$$\omega' = (\alpha - \beta\Delta)/(r^2 + a^2), \tag{2.4}$$

where

$$\begin{aligned}\alpha &= a(1 - B_0^4 m^2 a^2),\\ \beta &= \frac{a\Sigma}{A} + \frac{B_0^4}{16}\Big(-8mra\cos^2\theta(3-\cos^2\theta) - 6mra\sin^4\theta\\ &\quad + \frac{2ma^3\sin^6\theta}{A}[r(r^2+a^2)+2ma^2]\\ &\quad + \frac{4m^2a^3\cos^2\theta}{A}[(r^2+a^2)(3-\cos^2\theta)^2 - 4a^2\sin^2\theta]\Big).\end{aligned} \tag{2.5}$$

Since the field Λ is well-behaved as $\Delta \to 0$, the nonsingular nature of the event horizon can be established by finding a new coordinate system such that

$$\frac{dr^2}{\Delta} + d\theta^2 - \frac{\Delta}{A}\,dT^2 \quad \text{and} \quad d\phi - \omega' dT$$

are both replaced by expressions which are well-behaved as $\Delta \to 0$. Such a coordinate system is provided by (r, θ, ϕ', u), where

$$dT = du - \frac{r^2+a^2}{\Delta}\,dr \quad \text{and} \quad d\phi = d\phi' - \frac{\alpha}{\Delta}\,dr, \tag{2.6}$$

for

$$\frac{dr^2}{\Delta} + d\theta^2 - \frac{\Delta}{A}\,dT^2 = A^{-1}[-a^2\sin^2\theta\,dr^2 + A\,d\theta^2 + 2(r^2+a^2)\,du\,dr - \Delta\,du^2], \tag{2.7}$$

$$d\phi - \omega' dT = d\phi' - \omega' du - \beta dr. \tag{2.8}$$

Finally, the orthonormal components of the electromagnetic field for the locally nonrotating observer are given by

$$\begin{aligned}H_r + iE_r &= -\tfrac{1}{2}B_0\Lambda^{-2}P\frac{\partial\mathcal{E}}{\partial\theta}\\ &= B_0\Lambda^{-2}A^{-1/2}\left[(r^2+a^2)\cos\theta + mai\sin^2\theta \times\left(3 - \frac{4ai\cos\theta}{r+ia\cos\theta} + \frac{a^2\sin^2\theta}{(r+ia\cos\theta)^2}\right)\right],\\ H_\theta + iE_\theta &= \tfrac{1}{2}B_0\Lambda^{-2}P\Delta^{1/2}\frac{\partial\mathcal{E}}{\partial r}\\ &= -B_0\Lambda^{-2}\Delta^{1/2}A^{-1/2}\left[r - \frac{ma^2\sin^2\theta}{(r+ia\cos\theta)^2}\right]\sin\theta.\end{aligned} \tag{2.9}$$

These somewhat awkward expressions can be simplified by transforming to an alternative frame of reference moving along the ϕ direction at velocity v relative to the locally nonrotating observer's frame, where

$$v/c = \Delta^{1/2}\,a\sin\theta/(r^2+a^2). \tag{2.10}$$

In the new reference frame

$$\begin{aligned}H'_r + iE'_r &= A^{-1/2}[(r^2+a^2)(H_r + iE_r) - i\Delta^{1/2}\,a\sin\theta\,(H_\theta + iE_\theta)]\\ &= B_0\Lambda^{-2}\left(\cos\theta + \frac{ia\sin^2\theta}{r+ia\cos\theta} + \frac{ima\sin^2\theta}{(r+ia\cos\theta)^2}\right),\\ H'_\theta + iE'_\theta &= A^{-1/2}[(r^2+a^2)(H_\theta + iE_\theta) + i\Delta^{1/2}\,a\sin\theta\,(H_r + iE_r)]\\ &= -B_0\Lambda^{-2}\Delta^{1/2}\frac{\sin\theta}{r+ia\cos\theta}.\end{aligned} \tag{2.11}$$

3. COMPARISON WITH WALD'S SOLUTION

On the symmetry axis the electric field is given by

$$E'_r = 2B_0^3 ma/(1 + B_0^4 m^2 a^2)^2. \tag{3.1}$$

In the linearized theory the corresponding electric field would vanish. This suggests that our exact solution of the Einstein–Maxwell field equations should be identified not with Wald's $Q = 0$ solution, but rather with his $Q = 2B_0 J$ solution, the natural end product of the process of charge accretion.

Further evidence that this is the proper identification is provided by evaluating the integral[5]

$$\int *F = 4\pi Q \tag{3.2}$$

over a 2-sphere ($r = \text{const}$, $t = \text{const}$). To the first order in B_0 one gets precisely $Q = 2B_0 J$.

A detailed comparison of the electromagnetic fields shows that when all terms beyond the first order in B_0 are deleted our solution indeed reduces to Wald's approximate solution with $Q = 2B_0 J$. It should be noted, however, that even for small values of B_0 Wald's solution must break down as $r \to \infty$, since the linearized solution is asymptotically flat, while the exact solution resembles Melvin's magnetic universe asymptotically.

4. ELECTROMAGNETIC FIELDS ($Q \neq 2B_0 J$)

With a little more effort one can evaluate the *exact* electromagnetic fields for the solution which results from the application of the magnetizing technique to an arbitrarily charged Kerr–Newman black hole.

In the alternative frame of reference the electromagnetic fields are given by the expressions

$$\begin{aligned}H'_r + iE'_r &= \Lambda^{-2}\left[(1 + \tfrac{1}{4}B_0^2\mathcal{E})\frac{ie}{(r+ia\cos\theta)^2} + B_0(1 + \tfrac{1}{2}B_0\Phi)\left(\cos\theta + \frac{ia\sin^2\theta}{r+ia\cos\theta} + \frac{ima\sin^2\theta - e^2\cos\theta}{(r+ia\cos\theta)^2}\right)\right],\\ H'_\theta + iE'_\theta &= -B_0\Lambda^{-2}\Delta^{1/2}(1 + \tfrac{1}{2}B_0\Phi)\frac{\sin\theta}{r+ia\cos\theta},\end{aligned} \tag{4.1}$$

where the complex potentials are given in Eqs. (1.8) and Λ is defined in Eq. (1.11).

5. APPROXIMATE DETERMINATION OF ω' ($Q \neq 2B_0 J$)

Obtaining the metric in this case is extremely tedious, for while f' is easily evaluated using Eq. (1.9), the determination of ω' from Eq. (1.10) tries one's patience. Not having access to electronic symbol manipulation, we have not attempted to derive ω' exactly. Nevertheless, we found that it is not difficult to derive

the approximate expression for ω', from which it is possible to show that to the first order in B_0 the event horizon remains nonsingular.

In the linearized theory we may replace Λ by $1 + B_0\Phi$. Then Eq. (1.10) reduces to the simple equation

$$\nabla(\omega' - \omega) = 2B_0[(\mathrm{Re}\Phi)\nabla\omega + i\rho f^{-1}\nabla(\mathrm{Im}\Phi)]. \tag{5.1}$$

In the case of the Kerr–Newman metric this equation is easily integrated, and we find that ω' again has the form given in Eq. (2.4), where this time

$$\alpha = a - 2B_0 er, \quad \beta = (a/A)(\Sigma + 2B_0 era\sin^2\theta). \tag{5.2}$$

Because α is independent of θ, it is again possible to introduce new coordinates (r, θ, ϕ', u) in terms of which the nonsingular nature of the event horizon can be displayed.

While astrophysicists may be disappointed that our solutions are not asymptotically flat, general theorems[6] seem to indicate that it will be necessary to consider nonstationary fields if an asymptotically flat solution is to be constructed representing a black hole in an external magnetic field.

[1]F. J. Ernst, J. Math. Phys. **17**, 54 (1976).
[2]R. M. Wald, Phys. Rev. D **10**, 1680 (1974).
[3]W. B. Bonnor, Proc. Phys. Soc. Lond. A **67**, 225 (1953); M. A. Melvin, Phys. Rev. **139**, B225 (1965).
[4]W. Kinnersley called our attention to the possibility of formulating the vacuum field equations entirely in terms of divergence equations; F. J. Ernst subsequently showed that the electrovac field equations could also be expressed entirely in terms of divergence equations. It is, therefore, possible to generalize the potential χ for use in the electrovac case.
[5]Throughout this paper we refer to this solution as the $Q = 2B_0 J$ solution, although in fact there are contributions to Q of higher order in B_0.
[6]J. Ipser, Phys. Rev. Lett. **27**, 529 (1971); R. M. Wald, Ph. D. thesis, Princeton University, 1972 (unpublished); B. Carter, in *Black Holes*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1973).

The master analytic function and the Lorentz group. I. Reduction of the representations of O(3,1) in O(2,1) basis

Debabrata Basu and S. Datta Majumdar

Department of Physics, Indian Institute of Technology, Kharagpur, India
(Received 10 June 1975)

The reduction of the principal and supplementary series of representations of SL(2, C) in the SU(1,1) basis is carried out by using a basis function which formally resembles the coupled state of two angular momenta. The spectrum of the SU(1,1) representations contained in SL(2, C) and the transformation coefficients are obtained by expanding the SU(2) in terms of the SU(1,1) bases with the help of the Sommerfeld–Watson transformation. The orthogonality conditions for the principal and supplementary series are discussed. For the principal series this follows easily from the standard Sturm–Liouville theory of the second order differential equations. For the supplementary series the orthogonality condition is obtained from the fourth order differential equation satisfied by the Fourier transform of the basis function.

1. INTRODUCTION

The unitary infinite-dimensional representations of the Lorentz group [or SL(2, C)] were discovered in the forties [Gelfand and Naimark (1946, 1947), Harish-Chandra (1947)][1] and laid the foundation for the study of noncompact groups. These representations have been used extensively in the context of the local field theory for "infinite multiplets" [Feldman and Matthews (1966, 1967), Fronsdal (1967)],[2] in connection with the saturation of current algebras with infinite sets of particle states [Fubini (1967), de Alfaro (1967)],[3] and in the harmonic analysis of the forward elastic scattering amplitude [Hdjioannou (1966), Boyce (1967), Delbourgo *et al.* (1967), Toller, (1968)].[4]

Invariance under the Poincare group $P = \mathrm{SL}(2, C) \times T_4$, where T_4 is the group of space–time translations, is an essential prerequisite for all such investigations. For the analysis of scattering amplitude, the procedure which uses only the invariance and unitarity properties is equivalent to the usual partial wave expansion and is commonly called the "crossed partial wave analysis." The group which occurs most naturally in such analysis is the SU(1, 1) subgroup of SL(2, C). In recent years various kinds of partial wave expansions of the relativistic scattering amplitude have been proposed. The purpose of these expansions is to separate out the kinematical part of the amplitude in the form of basis functions of irreducible representations (IR) of P and then make various assumptions (like analyticity) which can be tested against experiment about the dynamical part. The functions occurring in the expansion are the bases of IR's of the little group of P, which is O(3) or O(2, 1) according as the eigenvalues of the Casimir operator p^2 of P are positive or negative. The O(2, 1) expansion is related to the analytic continuation of the O(3) expansion in the crossed channel at values of the parameters pertaining to the physical region of the direct channel. This is found to be identical with the Regge continuation of the amplitude by means of the Somerfeld–Watson (SW) transformation. The principal series of the O(2, 1) expansion is related to the background integral of the SW transform and the discrete series to the "nonsense channel" terms.

In view of possible applications to particle physics, such as those mentioned above, it is of interest to study the IR's of SL(2, C) reduced with respect to the SU(1, 1) subgroup. Mathematically, the problem is much harder than that treated in an earlier paper (I) [Basu and Majumdar (1973)][5] in which the SL(2, C) representations are reduced with respect to the compact SU(2) subgroup. The problem has been treated at length by Mukunda (M) [Mukunda (1968)][6] and by Sciarrino and Toller (ST) [Sciarrino and Toller (1967)][7] amongst others [Ström (1967)][8] and some important results have been established. These results are, to a large extent, complementary to one another. While M obtains the reduction of the principal series for integral j_0 [see Eq. (2. 5)] and also of the supplementary series, ST do this for the principal series only but for both integral and half-integral j_0. On the other hand, while ST determine the coefficients of transformation from the SU(2) to the SU(1, 1) basis in one particular case but do not derive the spectrum of the discrete j'-values in the decomposition of the principal series, M derives the spectrum by using the completeness relations for the basis functions but does not determine the transformation coefficients. In the case of the supplementary series, no basis functions are defined, and the nature of the spectrum is determined by expanding the kernel appearing in Naimark's definition [Naimark (1964)][9] of the scalar product in a function space with a nonlocal metric.

In the present paper we make a fresh attack on the problem along entirely different lines and not only derive all the results enumerated above in a simple and unitary way but also go, to some extent, beyond the work of our predecessors. For instance, while the principal series alone is considered by ST, we have succeeded in finding the transformation from the SU(2) to the SU(1, 1) basis for both the principal and the supplementary series. The success is attributable to the choice of the basis functions which are taken to be formally identical with the coupled states of two angular momenta j_1 and j_2. As shown by one of the authors [Majumdar (1958)][10] a long time ago, the coupled state ψ_{jm} can be written in a compact form in terms of the Gauss hypergeometric function (HGF) with j_1, j_2, j, m contained in it as parameters. This function or, more properly, its analytic continuation in $j_1, j_2, j,$

has been used already to study the IR's of SU(3) [Basu and Majumdar (1970, 1973)][11] and SL(2, C) reduced with respect to the SU(2) subgroup and has led to considerable simplification of the mathematical treatment. Encouraged by these successes we now use the same function for solving the problem of reduction of SL(2, C) in SU(1,1) basis and the Clebsch–Gordan problem for the group SU(1,1). Because of its usefulness in a variety of problems in the representation theory of the simplest Lie groups it can be legitimately called "The Master Analytic Function."

We conclude this section by outlining the procedure adopted for carrying out the reduction. It will be seen from Eqs. (2.10) and (2.11) that the basis states $\psi_{jm}^{SU(2)}$, $\psi_{jm}^{SU(1,1)}$ of the SU(2) and SU(1,1) subgroups have nearly the same form and are functions of the same variable $x = z\bar{z}$. For obtaining the reduction we can therefore expand $\psi_{jm}^{SU(2)}$ in a series of the functions $\psi_{jm}^{SU(1,1)}$. This is easily done by breaking up $\psi_{jm}^{SU(2)}$ into a power series and expanding each power of x in a series of HGF's of the appropriate type by means of the Burchnall–Chaundy formula. The series thus obtained yields the continuous and discrete spectra of the SU(1,1) representations after the SW transformation. Since, according to the present viewpoint, the two apparently unconnected problems mentioned in the last paragraph are seen to possess similar mathematical features, the same technique is applied to the CG problem discussed in Paper II. In order that the expansion coefficients found after the SW transformation may be the coefficients of a unitary transformation it is essential for the basis functions to fulfil the requirement of orthonormality. In the case of the principal series of representations of SL(2, C) it can be easily shown, with the usual definition of the scalar product, that the functions (2.11) do not form an orthogonal set. However, an orthogonal set is easily obtained by taking the linear combination (2.13) of ψ_{jm} and ψ_{-j-1m}. The proof of the orthogonality follows easily from the standard Strum–Liouville theory and is given in Sec. 5. In the case of the supplementary series the scalar product is defined in a function space with a nonlocal metric, and the proof of the orthogonality becomes much more difficult. The difficulty is resolved by adopting Naimark's alternative definition of the scalar product in the Fourier space. The differential equation satisfied by the Fourier transforms of the basis functions is easily set up and the orthogonality of the eigenfunctions of the resulting fourth order equation is established after elaborate calculations. From the form of the orthogonality condition the normalization factor of the basis function follows immediately. In the case of the CG coefficients of SU(1,1) the normalization factor is determined in an elementary way by comparing the coefficients of the direct and the inverse transformation.

2. CONSTRUCTION OF THE BASIS FUNCTIONS

The elements of the group SL(2, C) are complex 2×2 matrices of the form

$$a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{with } \det a = 1. \tag{2.1}$$

The generators $J_i = \frac{1}{2}\sigma_i$, $F_i = (i/2)\sigma_i$ $(i = 1, 2, 3)$ of this group satisfy the commutation relations

$$[J_i, J_j] = i\epsilon_{ijk}J_k, \quad [J_i, F_j] = i\epsilon_{ijk}F_k,$$
$$[F_i, F_j] = -i\epsilon_{ijk}J_k. \tag{2.2}$$

In the space of functions of four variables ξ_1, ξ_2, η_1, η_2 where (ξ_1, ξ_2), (η_1, η_2) are spinors transforming according to the fundamental representation and the complex conjugate representation, respectively, the generators can be written as first order differential operators of the form (2.3), (2.4) of I. Since all the irreducible representations of SL(2, C) including the infinite-dimensional ones can be generated by considering only homogeneous functions $f(\xi_1, \xi_2; \eta_1, \eta_2)$ of degree $2j_1$ in (ξ_1, ξ_2) and of degree $2j_2$ in (η_1, η_2) the differential operators can be written in terms of two variables $z, \bar{z}$ in the form

$$\begin{aligned}
F_3 &= iz\frac{\partial}{\partial z} + i\bar{z}\frac{\partial}{\partial \bar{z}} - i(\sigma - 1),\\
F_+ &= i\frac{\partial}{\partial z} - i\bar{z}^2\frac{\partial}{\partial \bar{z}} + i\bar{z}(\sigma - j_0 - 1),\\
F_- &= -iz^2\frac{\partial}{\partial z} + i\frac{\partial}{\partial \bar{z}} + iz(\sigma + j_0 - 1),\\
J_3 &= -z\frac{\partial}{\partial z} + \bar{z}\frac{\partial}{\partial \bar{z}} + j_0,\\
J_+ &= -\frac{\partial}{\partial z} - \bar{z}^2\frac{\partial}{\partial \bar{z}} + \bar{z}(\sigma - j_0 - 1),\\
J_- &= z^2\frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}} - z(\sigma + j_0 - 1)
\end{aligned} \tag{2.3}$$

where

$$J_\pm = J_1 \pm iJ_2, \quad F_\pm = F_1 \pm iF_2, \quad z = \xi_1/\xi_2, \quad \bar{z} = \eta_1/\eta_2.$$

Substitution of these into the eigenvalue equations of the Casimir operators

$$\begin{aligned}
&(-J_iJ_i + F_iF_i + \sigma^2 + j_0^2 - 1)f = 0,\\
&(F_iJ_i + 2ij_0\sigma)f = 0
\end{aligned} \tag{2.4}$$

gives

$$\sigma = j_1 + j_2 + 1, \quad j_1 - j_2 = j_0. \tag{2.5}$$

The values of σ and j_0 are restricted by the condition of unitarity of the representations. For the principal series of representations j_0 is an integer or a half-integer and σ is purely imaginary and for the supplementary series $j_0 = 0$ and σ is a real number in the interval $0 < \sigma < 1$.

The subgroup SU(2) of SL(2, C) is made up of all unimodular unitary matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \quad \text{with } |\alpha|^2 + |\beta|^2 = 1. \tag{2.6}$$

In the space of the monomials $\xi_1^a\xi_2^b$ (with $a + b = 2j$ = const) its generators $J_i = \frac{1}{2}\sigma_i$ can be represented as differential operators of the form

$$J_3 = \tfrac{1}{2}\left(\xi_1\frac{\partial}{\partial \xi_1} - \xi_2\frac{\partial}{\partial \xi_2}\right), \quad J_- = \xi_2\frac{\partial}{\partial \xi_1}, \quad J_+ = \xi_1\frac{\partial}{\partial \xi_2}\,. \tag{2.7}$$

Finally, the noncompact subgroup SU(1, 1) consists of all elements of SL(2, C) of the form

$$\begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} \quad \text{with } |\alpha|^2 - |\beta|^2 = 1. \tag{2.8}$$

Its generators F_1, F_2, J_3 can be represented as differential operators of the form (Barut and Fronsdal 1965)[12]

$$F_+ = F_1 + iF_2 = i\xi_1 \frac{\partial}{\partial \xi_2}, \quad F_- = F_1 - iF_2 = i\xi_2 \frac{\partial}{\partial \xi_1} \tag{2.9}$$

with J_3 given by (2.7).

Operating on $\xi_1^a \xi_2^b$ with these and using the hermiticity condition, one obtains the following four types of UIR's of SU(1, 1) [Bargmann (1947)][13]:

(a) The continuous nonexceptional (principal) series of representations

$$j = -\tfrac{1}{2} + i\lambda, \quad -\infty < \lambda < \infty, \quad m = 0, \ \pm\tfrac{1}{2}, \ \pm 1, \cdots.$$

(b) The exceptional (supplementary) series of representations

$$j = -\tfrac{1}{2} + s, \quad -\tfrac{1}{2} < s < \tfrac{1}{2},$$

$$m = 0, \ \pm 1, \ \pm 2, \cdots.$$

(c) The positive discrete class: j = a negative integer or half integer $m = -j, \ -j+1, \cdots$.

(d) The negative discrete class j = a negative integer or half integer $m = j, \ j-1, \cdots$.

A comparison of Eq. (2.3) of I with Eq. (2.7) shows that the generators of the SU(2) subgroup of SL(2, C) are formally identical to those for the coupling of a pair of angular momenta j_1, j_2. The coupled states[10]

$$g_m^{\,j} = z^{j_0 - m} \psi_m^{\,j} \tag{2.10}$$

where

$$\psi_m^{\,j} = (1 + z\bar{z})^{\sigma - j - 1} F(-j - m, \ j_0 - j; -2j; 1 + z\bar{z})$$

can, therefore, be taken to form the bases of UIR of SL(2, C) reduced explicitly with respect to the SU(2) subgroup. The operators F_1, F_2, J_3 of the subgroup SU(1, 1) are likewise formally identical to those for the coupling of two representations D^{j_1} and D^{j_2} of SU(1, 1). The basic states of UIR of SL(2, C) appropriate to the reduction of SU(1, 1) can therefore be written as

$$\varphi_m^{\,j'} = N_{j',m} z^{j_0 - m} (1 - z\bar{z})^{\sigma - j' - 1} F(-j' - m, j_0 - j'; -2j'; 1 - z\bar{z}) \tag{2.11}$$

where

$$N_{j',m} = \left(\frac{\Gamma(m - j')}{\Gamma(m + j' + 1)} \right)^{1/2}.$$

However, since the values of j_1, j_2 as given by Eq. (2.5) do not occur in the representations of SU(2) or SU(1, 1), the above construction of the basic states must be regarded as purely formal. Nevertheless they prove to be very helpful in studying the irreducible representations of SL(2, C).

The basis functions (2.11) are solutions of the equation

$$[F_1^2 + F_2^2 - J_3^2 + j'(j'+1)]\varphi_m j' = 0. \tag{2.12}$$

The functions (2.10) have been already used in I for finding the j-values of SU(2) representations contained in SL(2, C) and for determining the matrix elements of the generators and of finite transformations. Before using (2.11) for a similar purpose we must see that they fulfil the requirements of completeness and orthonormality. Although from the development of Sec. 3 it is evident that the functions form a complete set, a simple test shows that they are not orthogonal and hence are unsuitable for use as basis functions. However, as has been shown in Sec. 5, an orthonormal set can be easily constructed by taking instead of one solution, a linear combination of the first and second solutions of the hypergeometric equation to which Eq. (2.12) can be reduced. The appropriate linear combination turns out to be

$$\begin{aligned} u_m^{\,j'} &= \frac{\Gamma(m - j_0 + 1)\Gamma(2j' + 1)}{\Gamma(m + j' + 1)\Gamma(j' - j_0 + 1)} (1 - x)^{\sigma - j' - 1} \\ &\quad \times F(-j' - m, j_0 - j'; -2j'; 1 - x) \\ &\quad + \frac{\Gamma(m - j_0 + 1)\Gamma(-2j' - 1)}{\Gamma(m - j')\Gamma(-j' - j_0)} (1 - x)^{\sigma + j'} \\ &\quad \times F(j' + 1 - m, j_0 + j' + 1; 2j' + 2; 1 - x) \\ &= x^{m - j_0} (1 - x)^{\sigma - j' - 1} F(-j' + m, -j_0 - j'; m - j_0 + 1; x) \\ &\qquad \text{for } m - j_0 \geq 0 \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} v_m^{\,j'} &= \frac{\Gamma(j_0 - m + 1)\Gamma(2j' + 1)}{\Gamma(j' + 1 - m)\Gamma(j_0 + j' + 1)} (1 - x)^{\sigma - j' - 1} \\ &\quad \times F(-j' - m; j_0 - j'; -2j'; 1 - x) \\ &\quad + \frac{\Gamma(j_0 - m + 1)\Gamma(-2j' - 1)}{\Gamma(-j' - m)\Gamma(j_0 - j')} (1 - x)^{\sigma + j'} \\ &\quad \times F(j' + 1 - m, j_0 + j' + 1; 2j' + 2; 1 - x) \\ &= (1 - x)^{\sigma - j' - 1} F(-j' - m, j_0 - j'; 1 + j_0 - m; x) \\ &\qquad \text{for } m - j_0 < 0. \end{aligned} \tag{2.14}$$

The hypergeometric functions in Eqs. (2.13) and (2.14) converge within the unit circle $x = z\bar{z} = 1$. In the region outside the unit circle one must use a different pair of solutions of Eq. (2.12), namely

$$U_m^{j'} = x^{j' - j_0} (1 - x)^{\sigma - j' - 1} F(-j' + m, j_0 - j'; j_0 + m + 1; 1/x) \quad \text{for } j_0 + m \geq 0, \tag{2.15}$$

$$V_m^{j'} = x^{j' + m} (1 - x)^{\sigma - j' - 1} F(-j' - m, -j_0 - j'; 1 - j_0 - m; 1/x). \tag{2.16}$$

By the standard formulas for the analytic continuation of the hypergeometric function [Erdélyi (1953)][14] these can be similarly written as linear combinations of two hypergeometric functions of the form $F(a, b; c; 1 - 1/x)$.

3. REDUCTION OF THE PRINCIPAL SERIES OF REPRESENTATIONS OF 0(3,1)

For the determination of the complete spectrum of j-values appearing in the reduction by the technique of the following sections a knowledge of the discrete part

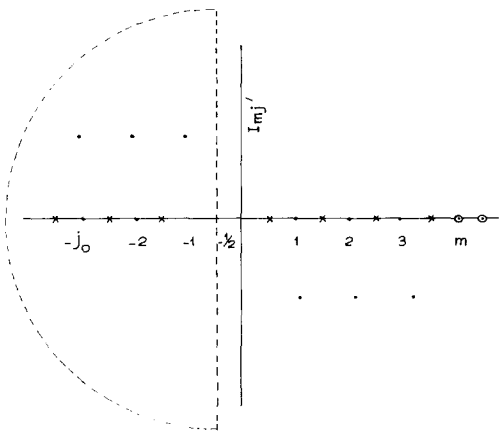

FIG. 1

seems to be necessary. This can be easily had by applying the operator F_3 to the function (2.11) and using the hermiticity condition:

$$F_3 f_m^{j'} = \frac{-i}{2}[(m-j')(m+j')]^{1/2}\alpha_{j'}(\sigma+j')\frac{a_{j'}}{a_{j'-1}}f_m^{j'-1} - \frac{im\sigma j_0}{j'(j'+1)}f_m^{j'} - 2i[(m-j'-1)(m+j'+1)]^{1/2} \times(\sigma-j'-1)\frac{a_{j'}}{a_{j'+1}}f_m^{j'+1} \tag{3.1}$$

where

$$f_m^{j'} = a_{j'}\varphi_m^{j'}, \quad \alpha_{j'} = \frac{j_0^2 - j'^2}{j'^2(4j'^2-1)}. \tag{3.2}$$

The condition

$$(F_3 f_m^{j'+1}, f_m^{j'}) = (f_m^{j'+1}, F_3 f_m^{j'}) \tag{3.3}$$

then gives

$$\left|\frac{a_{j'-1}}{a_{j'}}\right|^2 = \frac{|j'+\sigma|^2(j_0^2-j'^2)}{4j'^2(4j'^2-1)}. \tag{3.4}$$

For this to be positive j' must have the set of values $-j_0, -j_0+1, \cdots -\frac{1}{2}$ or 0.

Since j', $j'+1$, $j'-1$ cannot be simultaneously of the form $-\frac{1}{2}+is$ the above method obviously fails in the continuous case. To determine the continuous part of the spectrum of j' values, we expand the function ψ_m^j of the SU(2) basis in powers of $x = z\bar{z}$ and use the Burchnall–Chaundy formula [Erdélyi (1953)][15]

$$x^r = \sum_{n=0}^{\infty}(-)^n\frac{(a)_n(b)_n}{(c+n-1)_n n!}\,{}_3F_2\left[\begin{matrix}-r, & c+n-1, & -n\\ & a, b & \end{matrix}\right](1-x)^n \times F(a+n, b+n; c+2n; 1-x). \tag{3.5}$$

For $x<1$, $m \geq j_0$, $a = m+1-\sigma$, $b = 1-j_0-\sigma$, $c = 2(1-\sigma)$ the process yields

$$\psi_m^j = \frac{2^{\sigma-j-1}\exp[i\pi(m-j_0)]}{\Gamma(1+m-\sigma)\Gamma(1-j_0-\sigma)}\sum_{n=0}^{\infty}(-)^n \times\frac{\Gamma(m+1-\sigma+n)\Gamma(-j_0+1-\sigma+n)\Gamma(1-2\sigma+n)}{\Gamma(2(1-\sigma+n)-1)n!} \times S^n_{jmj_0}(1-x)^n F(1-\sigma+n-m, j_0+1-\sigma+n; 2(1-\sigma+n); 1-x) \tag{3.6}$$

where

$$S^n_{jmj_0} = \sum_{t=0}^{j-m}\frac{(\alpha)_t(\beta)_t}{(\gamma)_t t!}2^t\,{}_3F_2\left[\begin{matrix}-n, & 1-2\sigma+n, & -\sigma-t+j+1\\ & 1+m-\sigma, & 1-j_0-\sigma\end{matrix}; \tfrac{1}{2}\right], \tag{3.7}$$

$$\alpha = -j+m, \quad \beta = -j_0-j, \quad \gamma = -2j.$$

The expansion (3.6) of the SU(2) basis function has the desired analytic form but contains inadmissible values of j'. To circumvent this difficulty we express the sum (3.6) as a contour integral in the complex j' plane and apply the Sommerfeld–Watson transformation. The various terms in the sum are easily recognized as the residues at $j' = \sigma - n - 1$ of the analytic function

$$\chi(j') = \frac{\Gamma(m-j')\Gamma(-j_0-j')\Gamma(1+j'-\sigma)\Gamma(-\sigma-j')}{\Gamma(-2j'-1)}S^{j'}_{jmj_0}a_m^{j'} \tag{3.8}$$

where

$$S^{j'}_{jmj_0} = \sum_{t=0}^{j-m}\frac{(\alpha)_t(\beta)_t}{(\gamma)_t t!}2^t\,{}_3F_2\left[\begin{matrix}-j'-\sigma, & 1+j'-\sigma, & -\sigma-t+j+1\\ & 1+m-\sigma, & 1-j_0-\sigma\end{matrix}; \tfrac{1}{2}\right], \tag{3.9}$$

$$a_m^{j'} = (1-x)^{\sigma-j'-1}F(-j'-m, j_0-j'; -2j'; 1-x). \tag{3.10}$$

$S^{j'}_{jmj_0}$ and $a_m^{j'}$ are entire functions of j' and $\chi(j')$ is a meromorphic function going to zero rapidly as $|j'|$ tends to infinity in the region $\mathrm{Re}j' < 0$ (see Appendix). The singularities of $\chi(j')$ arise from the Γ-functions in the factor

$$\frac{\Gamma(m-j')\Gamma(-j_0-j')\Gamma(-\sigma-j')\Gamma(1+j'-\sigma)}{\Gamma(-2j'-1)}$$

and, as shown in Fig. 1, are located at the points

$$j' = \sigma-k-1, \ -\sigma+k, \ -j_0+k, \ m+k \quad (k=0,1,2,\cdots).$$

The singularities at $-\sigma+k$, $m+k$ for $k \geq 0$ and at $-j_0+k$ for $k > j_0$ lie in the region $\mathrm{Re}j' \geq 0$. The remaining singularities which are all simple poles lie in the region of interest $\mathrm{Re}j' < 0$. Because of the occurrence of $\Gamma(-2j'-1)$ in the denominator no singularities occur at the points $j' = 0, -\frac{1}{2}$. Let us now choose a contour C consisting of the infinite semicircle S on the left and the line $\mathrm{Re}j' = -\frac{1}{2}$. The singularities enclosed by the contour are the simple poles at $j' = \sigma-k-1$ $(k=0,1,2,\cdots)$ and at $j' = -j_0+l$ $(l=0,1,2,\cdots j_0-\frac{1}{2}$ or $j_0)$. Therefore, by Cauchy's theorem,

$$\frac{1}{2\pi i}\oint_C \chi(j')\,dj' = \sum_k \mathrm{Res}[\chi(j')]_{j'=\sigma-k-1} + \sum_l \mathrm{Res}[\chi(j')]_{j'=-j_0+l}. \tag{3.11}$$

Since the first term on the rhs equals ψ_m^j by our previous analysis and since the integral on the lhs vanishes on S according to the analysis given in Appendix, the equation can be written as

$$\psi_m^j = \frac{1}{2\pi i}\int_{-1/2-i\infty}^{-1/2+i\infty}\chi(j')\,dj' - \sum_l \mathrm{Res}[\chi(j')]_{j'=-j_0+l}. \tag{3.12}$$

Folding the integral about the real axis and evaluating

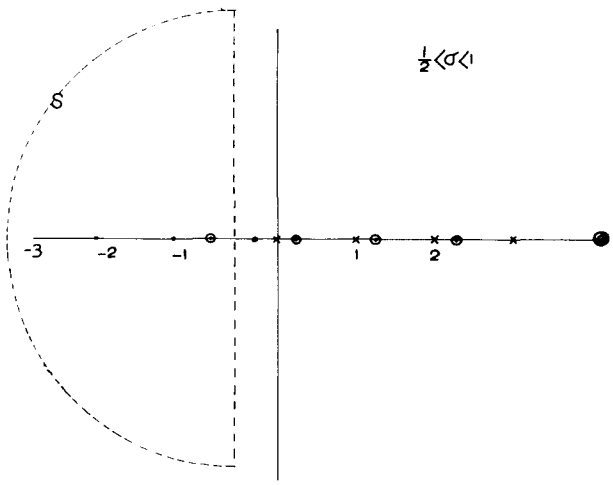

FIG. 2

the residues of the last term in Eq. (3.12), we have

$$\psi^j_m=\frac{2^{\sigma-j-1}\exp[i\pi(m-j_0)]}{\Gamma(m+1-\sigma)\Gamma(1-j_0-\sigma)}\frac{1}{\Gamma(m-j_0+1)}$$
$$\times\left(\frac{1}{2\pi i}\int_{-1/2}^{-1/2+i\infty}\frac{\Gamma(m-j')\Gamma(m+j'+1)\Gamma(-j_0-j')\Gamma(-j_0+j'+1)}{\Gamma(-2j'-1)}\right.$$
$$\left.\times\frac{\Gamma(-j'-\sigma)\Gamma(1+j'-\sigma)}{\Gamma(2j'+1)}S^{j'}_{jmj_0}u^{j'}_m\,dj'\right.$$
$$\left.+\sum_{j'=-1\text{ or }-3/2}^{-j_0}\frac{\Gamma(m-j')\Gamma(m-j_0+1)\Gamma(-\sigma-j')}{\Gamma(j'+j_0+1)}\right.$$
$$\left.\times\frac{\Gamma(1-\sigma+j')(-)^{j_0+j'}}{\Gamma(-2j'-1)}S^{j'}_{jmj_0}a^{j'}_m\right). \tag{3.13}$$

The absence of the identity representation in the decomposition is in agreement with the observation made by Fulling [Fulling (1974)][16] in a more general context.

In the second case $x=z\bar{z}<1$, $m<j_0$, a similar calculation shows that the discrete spectrum does not extend beyond $-m$ if $j_0>m>0$ and does not appear at all if $m<0$.

In deriving Eq. (3.13) we have used two different Kummer forms of the HGF occurring in ψ^j_m. This was necessary to reproduce the correct j'-values for the discrete part of the spectrum which is determined unambiguously by the unitarity condition (3.4). In fact, we need all the four Kummer forms of the HGF for obtaining the expansions and the spectrum of the j'-values in the four cases

(i) $x<1$, $m\geq j_0$ (ii) $x<1$, $m<j_0$

(iii) $x>1$, $j_0+m\geq 0$ (iv) $x>1$, $j_0+m<0$.

In case (iv) the expansion takes the form

$$\psi^j_m=\frac{2^{\sigma-j-1}\exp[i\pi(j-j_0)]}{\Gamma(-m+1-\sigma)\Gamma(1-j_0-\sigma)}\left(\frac{1}{2\pi i}\frac{1}{\Gamma(1-j_0-m)}\right.\times\int_{-1/2}^{-1/2+i\infty}dj'\frac{\Gamma(-m-j')\Gamma(-m+j'+1)\Gamma(-j_0+j'+1)\Gamma(-j_0-j')}{\Gamma(-2j'-1)}$$
$$\times\frac{\Gamma(-\sigma-j')\Gamma(-\sigma+j'+1)}{\Gamma(2j'+1)}(-)^{\sigma-j'-1}S^{j'}_{j-m\,j_0}V^{j'}_m+(-)^{j_0+\sigma-1}\sum_{j'=-1\text{ or }-3/2}^{-j_0}\frac{\Gamma(-m-j')\Gamma(j'+1-\sigma)\Gamma(-\sigma-j')}{\Gamma(-2j'-1)\Gamma(j'+j_0+1)}$$
$$\left.\times S^{j'}_{j-mj_0}a^{j'}_m\right) \tag{3.14}$$

and for the case (iii) the discrete spectrum, as before, terminates at $j'=-m$.

The above formulas show that in addition to the continuous nonexceptional series of SU(1, 1), in the first two cases (i) and (ii) the reduction yields only the positive discrete series while in the cases (iii) and (iv) we have only the negative discrete series.

4. REDUCTION OF THE SUPPLEMENTARY SERIES

For the supplementary series $j_0=0$ and σ is a real number lying in the interval $0<\sigma<1$. Since $j_0=0$ the spectrum of the SU(1, 1) representations do not contain any discrete part. Using the technique of the previous sections, we now have for $x<1$

$$\psi^j_m=\frac{2^{\sigma-j-1}(\mp)^m}{\Gamma(1+|m|-\sigma)\Gamma(1-\sigma)}\sum_{n=0}^{\infty}(-)^n$$
$$\times\frac{\Gamma(|m|+1-\sigma+n)\Gamma(1-\sigma+n)\Gamma(1-2\sigma+n)}{\Gamma(-1+2(1-\sigma+n))n!}S^n_{j|m|0}$$
$$\times(1-x)^nF(-m+1-\sigma+n,1-\sigma+n;2(1-\sigma+n);1-x) \tag{4.1}$$

where positive or negative sign is to be taken according as $m\gtrless 0$. The series on the rhs can be regarded as the sum of the residues at $j'=\sigma-n-1$ $(n=0,1,2,\cdots)$ of the analytic function $\chi(j')$ of Sec. 3 [Eq. (3.8)] with $j_0=0$. Besides these, the function, as shown in Fig. 2, possesses singularities at the points $j'=-\sigma+k$, $j'=m+k$, $j'=k$ $(k=0,1,2,\cdots)$. If $0<\sigma<\frac{1}{2}$ the only singularities that lie on the semiinfinite plane $\mathrm{Re}j'<-\frac{1}{2}$ and are enclosed by the contour C of Sec. 3 are the simple poles at $j'=\sigma-n-1$ $(n=0,1,2,\cdots)$. Since the integral on the semicircular part of C again vanishes, we have

$$\psi^j_m=\frac{2^{\sigma-j-1}(-)^m}{\Gamma(1-\sigma)\Gamma(m+1-\sigma)\Gamma(m+1)}\frac{1}{2\pi i}$$
$$\times\int_{-1/2}^{-1/2+i\infty}\frac{\Gamma(m-j')\Gamma(-j')\Gamma(m+j'+1)\Gamma(j'+1)}{\Gamma(2j'+1)}$$
$$\times\frac{\Gamma(j'+1-\sigma)\Gamma(-\sigma-j')}{\Gamma(-2j'-1)}u^{j'}_m S^{j'}_{jm0}\,dj'$$
$$\text{for } m\geq 0. \tag{4.2}$$

If, on the other hand, $\frac{1}{2}<\sigma<1$, then the pole at $j'=-1+\sigma$ nearest to the imaginary axis occurs between the points 0 and $-\frac{1}{2}$ and so lies outside C. For this range

of σ, however, one pole of $\Gamma(-\sigma-j')$, namely, the one on the extreme left at $j'=-\sigma$ lies inside C. The expansion of ψ_m^j, therefore, now contains two extra terms besides the integral on the line $\mathrm{Re}\,j'=-\frac{1}{2}$. Because of the relation (2.13) between the solutions of the hypergeometric equation the two terms combine to give $[u_m^{j'}]_{j'=-\sigma}$ and we have

$$\psi_m^j = 2^{\sigma-j-1}(-)^m\Bigg(\frac{\Gamma(m+\sigma)\Gamma(\sigma)}{\Gamma(m+1)\Gamma(2\sigma-1)}S_{jm}u_m^{-\sigma} + \frac{1}{\Gamma(m+1-\sigma)\Gamma(1-\sigma)\Gamma(m+1)}\frac{1}{2\pi i} \times\int_{-1/2}^{-1/2+i\infty}\frac{\Gamma(m-j')\Gamma(m+j'+1)\Gamma(-j')\Gamma(j'+1)}{\Gamma(2j'+1)} \times\frac{\Gamma(-\sigma-j')\Gamma(j'+1-\sigma)}{(-2j'-1)}u_m^{j'}S_{jm0}\,dj'\Bigg) \qquad (4.3)$$

for $m\geq 0$

where

$$S_{jm}=\sum_t\frac{(\alpha)_t(\beta)_t}{(\gamma)_t t!}2^t.$$

The extra term on the rhs corresponds to a representation $D_{-1+\sigma}$ of the exceptional series of SU(1, 1). Thus IR belonging to the supplementary series of O(3, 1) decomposes into IR's of the continuous nonexceptional series only of O(2, 1) if $0<\sigma<\frac{1}{2}$ and into IR's of the nonexceptional series and one member of the supplementary series of O(2, 1) if $\frac{1}{2}<\sigma<1$. The same conclusions are drawn by Mukunda [Mukunda (1968)] from entirely different considerations. The other cases, namely $x<1$, $m<0$; $x>1$, $m\gtrless 0$, can be investigated in an analogous manner and as they give similar results these need not be discussed separately.

5. ORTHOGONALITY OF THE BASIS FUNCTIONS AND THE TRANSFORMATION COEFFICIENTS

In the case of the principal series of the SL(2, C) representations the scalar product is defined in the traditional way and the orthogonality of the basis functions of the SU(1, 1) representations appearing in the reduction can be established easily from the standard Sturm–Liouville theory of second order differential equations. Since for the discrete class the HGF terminates and the orthogonality becomes trivial we shall consider here only the continuous nonexceptional class of SU(1, 1) representations corresponding to $j'=-\frac{1}{2}+is$. The differential equation satisfied by the real function,

$$\varphi_m^{j'}=(1-x)^{-j'-1}F(-j'-m, j_0-j'; j_0-m+1; x)$$

for $j_0-m\geq 0$, is

$$\frac{d}{dx}\left(x^{j_0-m+1}(1-x)^2\frac{d}{dx}\varphi_m^{j'}\right)-(j_0-m+1-mj_0)x^{j_0-m}\varphi_m^{j'} =-(\tfrac{1}{4}+s^2)x^{j_0-m}\varphi_m^{j'} \qquad (5.1)$$

If $\varphi_m^{l'}$, $l'=-\frac{1}{2}+it$, is another function satisfying the same equation, then we have

$$\int_0^1 x^{j_0-m}\varphi_m^{j'}\varphi_m^{l'}\,dx=\lim_{x\to 1}\frac{G_{j'l'}^m}{2}\Bigg(\frac{1}{i(s-t)}\left[(1-x)^{i(t-s)}-(1-x)^{-i(t-s)}\right]+\frac{1}{i(s+t)}\left[(1-x)^{i(t+s)}-(1-x)^{-i(t+s)}\right]\Bigg) \qquad (5.2)$$

where

$$G_{j'l'}^m=\frac{[\Gamma(j_0-m+1)]^2\Gamma(2j'+1)\Gamma(-2l'-1)}{\Gamma(j_0+j'+1)\Gamma(j'+1-m)\Gamma(j_0-l')\Gamma(-l'-m)}. \qquad (5.3)$$

We now set $(1-x)=\exp(-r)$ so that $\lim_{x\to 1}$ implies $\lim_{r\to\infty}$ and note that

$$\frac{1}{i(s-t)}[(1-x)^{i(t-s)}-(1-x)^{-i(t-s)}]=\lim_{r\to\infty}\int_{-r}^{r}\exp(iu(t-s))\,du =2\pi\delta(t-s).$$

Since σ is purely imaginary for the principal series of SL(2, C), this gives

$$\int_0^1 v_m^{j'}v_m^{l'*}x^{j_0-m+1}\,dx=\pi G_{j'l'}^m[\delta(t-s)-\delta(t+s)]. \qquad (5.4)$$

Similar orthogonality conditions hold for the functions $u_m^{j'}$, $U_m^{j'}$, $V_m^{j'}$ defined by the Eqs. (2.13), (2.15), (2.16) with the weight factors x^{m-j_0} and $x^{\pm(j_0+m)}$, respectively.

For the supplementary series of representations the Hilbert space is different and the above scalar product does not define a unitary representation of SL(2, C). Construction of the Hilbert space for this particular class of representations involves a nonlocal metric appearing in the definition (Naimark 1964)

$$(f_1,f_2)=\int\int f_1(z_1)f_2^*(z_2)K(z_1,z_2)\,dz_1\,dz_2 \qquad (5.5)$$

where

$$z_k=x_k+iy_k,\quad dz_k=dx_k\,dy_k\quad (k=1,2).$$

To avoid the complexity of nonlocality we formulate the orthogonality condition in terms of the Fourier transform of the basis function defined by

$$u_{j',m}(z,\bar z)=\frac{1}{2\pi}\iint\exp[i(ux+vy)]\varphi(u,v)\,du\,dv \qquad (5.6)$$

where

$$z=x+iy=r\exp(i\theta),\quad \bar z=r\exp(-i\theta),$$

$$u_{j',m}(z,\bar z)=z^{j_0-m}(1-z\bar z)^{\sigma-j'-1}F(-j'+m,-j';\,m+1;\,z\bar z). \qquad (5.7)$$

The generators $\tilde F_1, \tilde F_2, \tilde J_3$ of the SU(1, 1) subgroup in the Fourier space can be constructed in the usual way and these are given by

$$\tilde F_1=-\tfrac{1}{2}\left[u+\frac{\partial}{\partial u}\left(u\frac{\partial}{\partial u}+v\frac{\partial}{\partial v}+2\sigma\right)-i\frac{\partial}{\partial v}\tilde J_3\right],$$
$$\tilde F_2=-\tfrac{1}{2}\left[v+\frac{\partial}{\partial v}\left(u\frac{\partial}{\partial u}+v\frac{\partial}{\partial v}+2\sigma\right)+i\frac{\partial}{\partial u}\tilde J_3\right], \qquad (5.8)$$
$$\tilde J_3=i\left(v\frac{\partial}{\partial u}-u\frac{\partial}{\partial v}\right).$$

Since the algebraic structure of $\tilde J_3$ is the same as that of J_3, the Fourier transform $\varphi(u,v)$ must be of the form

$$\varphi(u,v)=\exp(-im\psi)g_{j',m}(\rho) \qquad (5.9)$$

where

$$u=\rho\cos\psi,\quad v=\rho\sin\psi,$$

$$g_{j',m}(\rho)=\int_0^1 J_m(\rho r)r^{m+1}(1-r^2)^{\sigma-j'-1}\times F(-j'+m,-j';\,m+1;\,r^2)\,dr. \tag{5.10}$$

For the sake of algebraic simplicity we shall treat here the special case $m=0$; the general case for arbitrary m can be treated essentially in the same way. The differential equation satisfied by $g_{j'}(\rho)\equiv g_{j',0}(\rho)$ can now be set up from the basic equation

$$[\tilde{F}_1^2+\tilde{F}_2^2-\tilde{J}_3^2+j'(j'+1)]g_{j'}(\rho)=0$$

and elaborate calculations using (5.8) lead to a differential equation of the fourth order. To make the equation self adjoint we multiply it by the weight factor $\rho^{1+2\sigma}$ and the resulting equation is given by

$$\frac{d^2}{d\rho^2}\rho^{3+2\sigma}\frac{d^2g_{j'}}{d\rho^2}+2\frac{d}{d\rho}\rho^{3+2\sigma}\frac{dg_{j'}}{d\rho}+[\rho^2+4(1+2\sigma)]\rho^{1+2\sigma}g_{j'}(\rho)=-4j'(j'+1)\rho^{1+2\sigma}g_{j'}(\rho). \tag{5.11}$$

If $g_{j'}$ and $g_{l'}$ be two solutions of the above equation belonging to the eigenvalues $j'=-\frac{1}{2}+is$ and $l'=-\frac{1}{2}+it$, respectively, the Eq. (5.11) then leads to

$$\begin{aligned}\int_0^\infty & g_{j'}(\rho)g_{l'}(\rho)\rho^{1+2\sigma}\,d\rho\\ &=\lim_{\rho\to\infty}\frac{(3+2\sigma)\rho^{2+2\sigma}}{4(s^2-t^2)}[g_{l'}(g_{j'}+g_{j'}'')-g_{j'}(g_{l'}+g_{l'}'')]\\ &\quad-\frac{\rho^{3+2\sigma}}{4(s^2-t^2)}\{[g_{l'}'(g_{j'}'+g_{j'}'')-g_{j'}'(g_{l'}'+g_{l'}'')]\\ &\quad-[g_{l'}(g_{j'}'+g_{j'}''')-g_{j'}(g_{l'}'+g_{l'}''')]\}.\end{aligned} \tag{5.12}$$

To evaluate the rhs of Eq. (5.12) a knowledge of the dominant term of the integral

$$g_{j'}(\rho)=\int_0^1 J_0(\rho r)r(1-r^2)^{\sigma-j'-1}F(-j',-j';\,1;\,r^2)\,dr \tag{5.13}$$

for arbitrarily large values of ρ is necessary. This can be obtained by noting Eq. (2.13), the standard formula [Ryzhik and Gradshtein (1951)],[17]

$$\int_0^1 r^{\nu+1}(1-r^2)^\mu J_\nu(\rho r)\,dr=2^\mu\Gamma(\mu+1)\rho^{-(\mu+1)}J_{\nu+\mu+1}(\rho) \tag{5.14}$$

and the asymptotic form of the Bessel functions [Erdélyi (1953)].[15]

To evaluate the derivatives appearing on the rhs of Eq. (5.12) we substitute, as before, Eq. (2.13) in (5.13) and assume the validity of the standard rules of differentiation under the integral sign. We write down here the final results which follow after differentiating the resulting equation and using the well-known recurrence relations

$$z^{-\nu}J_{\nu+1}(z)=-\frac{d}{dz}[z^{-\nu}J_\nu(z)],$$

$$zJ_{\nu-1}=\nu J_\nu(z)+zJ_\nu'(z)$$

in conjunction with the formula (5.14) and the asymptotic form of the Bessel functions

$$\begin{aligned}g_{j'}&\approx\left(\frac{2}{\pi\rho}\right)^{1/2}\left[A_{j'}\rho^{-\sigma+j'}\cos\left(\rho-\frac{\pi}{4}-\frac{\mu\pi}{2}\right)+A_{j'}^*\rho^{-\sigma-j'-1}\times\cos\left(\rho-\frac{\pi}{4}-\frac{\mu^*\pi}{2}\right)\right],\\ g_{j'}''&\approx-\left(\frac{2}{\pi\rho}\right)^{1/2}\left[A_{j'}\rho^{-\sigma+j'}\sin\left(\rho-\frac{\pi}{4}-\frac{\mu\pi}{2}\right)+A_{j'}^*\rho^{-\sigma-j'-1}\sin\left(\rho-\frac{\pi}{4}-\frac{\mu^*\pi}{2}\right)\right],\\ g_{j'}+g_{j'}''&\approx\left(\frac{2}{\pi\rho}\right)^{1/2}\left[(1+2\sigma-2j')A_{j'}\rho^{-\sigma+j'-1}\sin\left(\rho-\frac{\pi}{4}-\frac{\mu\pi}{2}\right)+(3+2j'+2\sigma)A_{j'}^*\rho^{-\sigma-j'-2}\sin\left(\rho-\frac{\pi}{4}-\frac{\mu^*\pi}{2}\right)\right],\\ g_{j'}'+g_{j'}'''&\approx\left(\frac{2}{\pi\rho}\right)^{1/2}\left[(1+2\sigma-2j')A_{j'}\rho^{-\sigma+j'-1}\cos\left(\rho-\frac{\pi}{4}-\frac{\mu\pi}{2}\right)+(3+2j'+2\sigma)A_{j'}^*\rho^{-\sigma-j'-2}\cos\left(\rho-\frac{\pi}{4}-\frac{\mu^*\pi}{2}\right)\right],\end{aligned} \tag{5.15}$$

where

$$A_{j'}=\frac{2^{\sigma-j'-1}\Gamma(2j'+1)\Gamma(\sigma-j')}{[\Gamma(j'+1)]^2},\qquad \mu=\sigma-j'. \tag{5.16}$$

It is now easy to verify that the first term on the rhs of Eq. (5.12) goes to zero as $1/\rho$ and on setting $\rho=e^\lambda$ the second term leads to

$$\frac{|G^\sigma_{j',l'}|}{\pi}\lim_{\lambda\to\infty}\left[\cosh(s+t)\frac{\pi}{2}\left(\frac{\exp[i\lambda(s-t)]-\exp[-i\lambda(s-t)]}{i(s-t)}\right)+\cosh(s-t)\frac{\pi}{2}\left(\frac{\exp[i\lambda(s+t)]-\exp[-i\lambda(s+t)]}{i(s+t)}\right)\right]$$

where

$$G^\sigma_{j',l'}=A_{j'}A_{l'}^*. \tag{5.17}$$

We therefore have the following orthogonality condition in the Fourier space:

$$\int_0^\infty g_{j'}(\rho)g_{l'}(\rho)\rho^{1+2\sigma}\,d\rho=2\,|G^\sigma_{j',l'}|\cosh\pi s[\delta(s-t)+\delta(s+t)]. \tag{5.18}$$

For arbitrary m the normalizor $G^{\sigma m}_{j'l'}$ can be obtained in a similar manner and its value is given by

$$G^{\sigma m}_{j'l'}=\frac{2^{2\sigma+l'-j'+1}[\Gamma(|m|+1)]^2\Gamma(2j'+1)\Gamma(\sigma-j')}{\Gamma(j'+1)\Gamma(j'+1+|m|)\Gamma(-l')\Gamma(-l'+|m|)}\times\Gamma(-2l'-1)\Gamma(\sigma+l'+1). \tag{5.19}$$

Using these orthogonality conditions for the principal and supplementary series it is now a straightforward matter to obtain the transformation coefficients from the SU(2) to SU(1, 1) basis. For example, for the principal series, we obtain from Eq. (3.13)

$$(\psi^{SU(2)}_{jm},\psi^{SU(1,1)}_{j'm})=(-)^{m-j_0}2^{\sigma-j-1}\Gamma(m-j_0+1)\times\frac{\Gamma(-\sigma-j')\Gamma(-\sigma+j'+1)\pi}{\Gamma(m+1-\sigma)\Gamma(-j_0+1-\sigma)}S^{j'}_{jmj_0}. \tag{5.20}$$

For the supplementary series we have similarly

$$(\psi^{SU(2)}_{jm}, \psi^{SU(1,1)}_{j'm'})$$
$$= 2^{3\sigma-j+1'-j'} \cos\left[(2j'+1)\frac{\pi}{2}\right]$$
$$\times \frac{\Gamma(m+1)\Gamma(\sigma-j')\Gamma(\sigma+j'+1)\Gamma(-\sigma-j')}{\Gamma(1-\sigma)}$$
$$\times \frac{\Gamma(-\sigma+j'+1)(-)^m}{\Gamma(m+1-\sigma)} S^{j'}_{jm0}. \tag{5.21}$$

APPENDIX

In what follows we shall investigate the asymptotic behaviour of the function

$$\chi(j') \equiv \chi^{j'}_{jm} = \frac{\Gamma(m-j')\Gamma(-j_0-j')\Gamma(-j'-\sigma)\Gamma(1+j'-\sigma)}{\Gamma(-2j'-1)}$$
$$\times (1-x)^{\sigma-j'-1} F(-j'-m, j_0-j'; -2j'; 1-x) S^{j'}_{jmj_0} \tag{A1}$$

on the infinite semicircle S (see Figs. 1 and 2) in the complex j'-plane. The behaviour of the HGF can be estimated by recalling Watson's formula [Erdélyi (1953)][15]

$$F(-j'-m, j_0-j'; -2j'; 1-x)$$
$$\approx \frac{2^{-2m+1}\Gamma(-2j')\Gamma(\frac{1}{2})(1+x)^{m+j'}}{\sqrt{-j'}\,\Gamma(j_0-j')\Gamma(-j_0-j')} \left(1+\frac{2\sqrt{x}}{1+x}\right)^{m+j'}. \tag{A2}$$

The function $S^{j'}_{jmj_0}$ which is a finite sum of ${}_3F_2\,(\frac{1}{2})$ functions,

$$S^{j'}_{jmj_0} = \sum_{t=0}^{j-m} \frac{(\alpha)_t(\beta)_t}{(\gamma)_t t!} 2^t\, {}_3F_2\left[\begin{matrix} j'+1-\sigma, -j'-\sigma, -\sigma-t+j+1 \\ m+1-\sigma, 1-\sigma-j_0 \end{matrix}; \tfrac{1}{2}\right], \tag{A3}$$

is an entire function of j' and its limiting value on S can be obtained by using the contour integral representation

$${}_3F_2\left[\begin{matrix} a, b, c \\ e, f \end{matrix}; z\right] = -\frac{\exp(-i\pi e)\Gamma(c)}{4\Gamma(e)\Gamma(e-c)\sin\pi c \sin\pi(e-c)}$$
$$\times \oint_p s^{c-1}(1-s)^{e-c-1} F(a, b; f; sz)\, ds \tag{A4}$$

where p stands for the Pochhammer contour. We now set

$$a = -\sigma - t + j + 1, \quad b = -j' - \sigma, \quad c = 1 + j' - \sigma, \quad e = m + 1 - \sigma,$$
$$f = -j_0 + 1 - \sigma, \quad f - a = t - j - j_0 = -r \tag{A5}$$

where r is a positive integer and note that by a Kummer transformation the HGF appearing on the rhs of Eq. (A4) can be transformed into $F(-r, -j_0+j'+1; f; sz)$ which for large $|j'|$ is given by

$$F(-r, -j_0+j'+1; f; sz)$$
$$\approx \exp(-i\pi r)\frac{\Gamma(f)}{\Gamma(a)}((1+j'-j_0)z)^r s^r. \tag{A6}$$

When this is substituted in Eq. (A4) the resulting integral then becomes identical with the Pochahammer contour integral representation of the HGF:

$$\oint_p s^{B-1}(1-s)^{C-B-1}(1-sz)^{-A}\, ds$$
$$= -\frac{4\Gamma(B)\Gamma(C-B)\sin\pi(C-B)\sin\pi B}{\Gamma(C)\exp(i\pi C)} F(A, B; C; z) \tag{A7}$$

with $A = b + r$, $B = c + r$, $C = e + r$. Using (A7) we obtain after some simplification

$${}_3F_2\left[\begin{matrix} a, b, c \\ e, f \end{matrix}; \tfrac{1}{2}\right] \approx \frac{P(j_0-j'-1)^r\Gamma(1+j'-\sigma+r)}{\Gamma(1+j'-\sigma)}$$
$$\times F(-\sigma+r-j', 1-\sigma+r+j'; e+r; \tfrac{1}{2}) \tag{A8}$$

where P is a numerical constant independent of j'. The asymptotic behaviour of the HGF appearing on the right can be estimated from the formula [Erdélyi (1953)][15]

$$F(\alpha-j', \beta+j'; d; \tfrac{1}{2}) \approx \frac{\Gamma(1-\beta-j')}{\sqrt{-j'}\,\Gamma(d-\beta-j')}$$
$$\times [P_1 \exp(-i(\pi/2)j') + P_2 \exp(i(\pi/2)j')] \tag{A9}$$

where P_1, P_2 are numerical constants. The formula (A9) shows that for $|j'| \to \infty$ the first or the second term will be dominant according as $Imj' \lesseqgtr 0$. Using (A9) we obtain after some calculation

$$S^{j'}_{jmj_0} \approx \frac{P_3(-j')^{j+j_0}\exp[\pm i(\pi/2)j']}{\sqrt{-j'}\,\Gamma(m-j')\Gamma(1+j'-\sigma)\sin\pi(\sigma-r-j')} \tag{A10}$$

where the positive or negative sign is to be taken according as $Imj' \lesseqgtr 0$. Using (A1), (A2), and (A10) we finally obtain the estimate

$$\chi(j') = \chi^{j'}_{jm} \approx P_4(-j')^{j-\sigma}\exp(\mp i\pi j'/2)(1-x)^{\sigma-j'-1}(1+x)^{m+j'}$$
$$\times \left(1+\frac{2\sqrt{x}}{1+x}\right)^{m+j}$$

where the upper or lower sign holds according as $Imj' \lesseqgtr 0$ and this goes to zero rapidly on S.

[1] I. M. Gelfand and M. A. Naimark, J. Phys. Moscow X, 93 (1946); Izv. Akad. Nauk. 11, 411 (1947); Harish-Chandra, Proc. Roy. Soc. 189 A, 372, (1947).
[2] G. Feldman and P. T. Mathews, Ann. Phys. 40, 19 (1966); Phys. Rev. 154, 1241 (1967); C. Fronsdal, Phys. Rev. 156, 1653 (1967).
[3] S. Fubini, Talk at Coral Gables, 1967; V. de Alfaro, Lectures at Schladming, Austria, 1967.
[4] T. Hdjioannou, Nuovo Cimento 44, 185 (1966); J. F. Boyce, J. Math. Phys. 8, 675 (1967); R. Delbourgo, A. Salam, and J. Strathdee, Phys. Lett. B 25, 230 (1967); M. Toller, Nuovo Cimento 54, 295 (1968).
[5] D. Basu and S. D. Majumdar, J. Phys. A 6, 1097 (1973).
[6] N. Mukunda, J. Math. Phys. 9, 50 (1968); 9, 417 (1968).
[7] A. Sciarrino and M. Toller, J. Math. Phys. 8, 1252 (1967).
[8] S. Ström, Ark. Fys. 29, 467 (1967).
[9] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, Oxford, 1964), pp. 170—9.
[10] S. D. Majumdar, Prog. Theor. Phys. (Japan) 20, 798 (1958).
[11] B. K. Basu and S. D. Majumdar, Ann. Phys. 56, 464 (1970); J. Math. Phys. 14, 1248 (1973).
[12] A. O. Barut and C. Fronsdal, Proc. Roy. Soc. A 287, 532 (1965).
[13] V. Bargmann, Ann. Math. 48, 568 (1947).
[14] A. Erdélyi, editor, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, pp. 107, 108.
[15] Ref. 14, pp. 187, 77; Vol. 2, p. 85.
[16] S. A. Fulling, J. Math. Phys. 15, 1567 (1974).
[17] I. M. Ryzhik and Gradshtein, *Tables of Integrals, Sums, Series, and Products* (Moscow, 1951).

The master analytic function and the Lorentz group. II. The Clebsch–Gordan problem for O(2,1)

S. Datta Majumdar

Department of Physics, Indian Institute of Technology, Kharagpur, India
(Received 11 June 1975)

The Clebsch–Gordan coefficients of the noncompact group O(2,1) representing Lorentz transformations in three-dimensional space-time are calculated in the compact O(2) basis. Considerable simplification is achieved by introducing a variable x and replacing all algebraic equations by differential equations. The coupled state appears in the theory as a solution of an ordinary differential equation reducible to the hypergeometric equation by a simple substitution. The coefficients in the Taylor–Laurent expansion of this solution in powers of x are shown to be identical with the Clebsch–Gordan coefficients. The inverse expansion, obtained by the use of certain identities for the hypergeometric function and the Sommerfeld–Watson transformation, yields the normalization factor and the values of j appearing in the reduction.

1. INTRODUCTION

The Clebsch—Gordan problem for the group O(2, 1) or its covering group SU(1, 1) was investigated by Holman and Biedenharn (1966, 1968),[1,2] Ferretti and Verde (1968),[3] and Kuo—Hsiang Wang (1970)[4] amongst others [Andrews and Gunson (1964),[5] Sannikov (1966, 1967),[6] Pukanszky (1961)[7]]. Holman and Biedenharn based their investigations on the fundamental recurrence relations satisfied by the Clebsch—Gordan coefficients (CGC). Their first paper was mainly concerned with the coupling of two representations of the discrete class, and other cases of coupling were considered in the second paper. Unnormalized CGC for the coupling of two representations of the continuous class were derived by Ferretti and Verde. Their investigations were based on the formula

$$d^{j_1}_{m_1' m_1}(z) d^{j_2}_{m_2' m_2}(z) = \sum_j C(j_1 j_2 j;\ m_1' m_2') C^*(j_1 j_2 j;\ m_1 m_2)\, d^j_{m_1'+m_2', m_1+m_2}(z)$$

where, the d's are representation functions, the C's are CGC, and $\sum$ denotes summation over discrete and integration over continuous j-values. The expansion was obtained by the use of the Burchnall—Chaundy (1944)[8] formula followed by the Sommerfeld—Watson transformation (SWT). The results of Ferretti and Verde were extended by Wang[4] to all cases of coupling except that of the supplementary series. All these authors used a compact basis for the irreducible representations (IR) of SU(1, 1). More recently, Mukunda and Radhakrishnan (1974)[9] have made a departure from the previous practice in evaluating the CG coefficients in a continuous noncompact basis. As a result of these investigations many important results have been established and the theory has reached a very satisfactory stage. The object of the present paper is, therefore, not so much to derive any new result as to approach the theory from a new angle and effect certain simplifications.

A major step towards the stated objective is taken by replacing the recurrence relations for the CGC by a pair of first order differential equations for the coupled states χ_{jm} and χ_{jm+1}. Elimination of χ_{jm+1} from the two equations leads to a second order equation for χ_{jm} which, by a simple substitution, can be reduced to the hypergeometric (HG) equation. These equations are formally the same as in the author's earlier paper [Majumdar (1958)][10] on angular momentum and only the values of j_1, j_2, j are different. This difference has the consequence of making the CG problem of SU(1, 1) much more complicated than the problem of coupling of angular momenta. As there are no simple CGC of SU(1, 1) for special values of the magnetic quantum numbers, there is no convenient starting point for the calculations, and it becomes difficult to see which particular solution of the HG equation would lead to the correct coupled state. The situation becomes clearer from the series solution $\sum a_{m_2} x^{m_2}$ of the equation for χ_{jm}. The coefficients a_{m_2}, as will be seen later, are identical with the unnormalized CGC. From the series solution the solution in terms of the hypergeometric function (HGF) is obtained by comparing coefficients.

To determine the normalization factor (NF) and the spectrum of j-values the simple power x^{m_2} representing a product state is expanded in a series of the χ_{jm}-functions. The expansion is brought into the right form and the j-values are determined by the SWT. If b_{m_2} are the expansion coefficients, then the NF is given by $(b^*_{m_2}/a_{m_2})^{1/2}$.

2. THE FUNDAMENTAL EQUATION AND ITS SERIES SOLUTION

As shown in Paper I, Sec. 2,[11] the generators $(f_1, f_2, j_3) = \frac{1}{2}(i\sigma_1, i\sigma_2, \sigma_3)$ of SU(1, 1) can be represented as differential operators of the form

$$-if_+ = -if_1 + f_2 = \xi_1 \frac{\partial}{\partial \xi_2}, \quad -if_- = -if_1 - f_2 = \xi_2 \frac{\partial}{\partial \xi_1},$$
$$j_3 = \tfrac{1}{2}\left(\xi_1 \frac{\partial}{\partial \xi_1} - \xi_2 \frac{\partial}{\partial \xi_2}\right). \tag{2.1}$$

For the product state

$$\xi_1^{j_1+m_1} \xi_2^{j_1-m_1} \eta_1^{j_2+m_2} \eta_2^{j_2-m_2} = (\xi_1 \xi_2)^{j_1} (\eta_1 \eta_2)^{j_2} \left(\frac{\xi_1}{\xi_2}\right)^m \left(\frac{\xi_2 \eta_1}{\xi_1 \eta_2}\right)^{m_2}$$
$$= u^{j_1} v^{j_2} \alpha^m x^{m_2}$$

the operators are

$$-iF_+ = -if_{1+} - if_{2+} = \alpha\left(x(1-x)\frac{\partial}{\partial x} + j_1 + j_2 x - \alpha \frac{\partial}{\partial \alpha}\right),$$

$$-iF_- = -if_{1-} - if_{2-} = \alpha^{-1}\left((1-x)\frac{\partial}{\partial x} + j_1 + \frac{j_2}{x} + \alpha\frac{\partial}{\partial\alpha}\right), \quad (2.2)$$

$$J_3 = j_{13} + j_{23} = \alpha\frac{\partial}{\partial\alpha}.$$

Operating on χ_{jm} and χ_{jm+1} these give

$$[x(1-x)D + j_1 + j_2 x - m]\chi_{jm} = (j-m)\chi_{jm+1}, \quad (2.3a)$$

$$\left[(1-x)D + j_1 + \frac{j_2}{x} + m + 1\right]\chi_{jm+1} = (j+m+1)\chi_{jm}, \quad (2.3b)$$

$$D \equiv \frac{\partial}{\partial x}.$$

Eliminating χ_{jm+1} from these equations, we have

$$\{x(1-x)^2D^2 + (1-x)[j_1 + j_2 - m + 1 + x(j_1 + j_2 + m - 1)]D + (j_1 - m)j_2/x + (j_1 + m)j_2 x + j_1(j_1+1) + j_2(j_2+1) - j(j+1)\}G = 0 \quad (2.4)$$

as the fundamental differential equation satisfied by the coupled state χ_{jm}. Although this is the same as Eq. (7) of Ref. 10, the fact that j_1, j_2, j can now take negative integral (half-integral) or complex values renders the task of obtaining a solution and interpreting it much more difficult. Writing $G = \sum a_{m_2} x^{m_2}$ we have the recurrence relations

$$(j_2 + m_2 + 1)(j_1 - m_1 + 1)a_{m_2+1} + (F_1 + F_2 - F + 2m_1 m_2)a_{m_2} + (j_2 - m_2 + 1)(j_1 + m_1 + 1)a_{m_2-1} = 0 \quad (2.5)$$

where, $F_1 = j_1(j_1+1)$, etc. A formal solution, for unrestricted j_1, j_2, j, is

$$a^{(1)}_{m_2} = \Gamma\begin{bmatrix} -j_2 + m_2, -j_1 - m_1, j_0 - m + 1 \\ j_2 + m_2 + 1, j_1 - m_1 + 1, -2j_2, -\sigma - m + 1 \end{bmatrix} {}_3F_2^{(1)}$$

with

$${}_3F_2^{(1)} = {}_3F_2\begin{bmatrix} -j_2 - m_2, -\sigma - j, -\sigma + j + 1; \\ -2j_2, -\sigma - m + 1; \end{bmatrix} 1, \quad (2.6)$$

$$j_0 = j_1 - j_2, \quad \sigma = j_1 + j_2 + 1.$$

That (2.6) satisfies (2.5) is easily verified by using the relation

$$-(d-a)(e-a)F(a-1) + [-bc - a(s-1) + (d-a)(e-a) + (a-b)(a-c)(1-z)]F(a) + [s - 1 + (-2a + b + c - 1)(1-z)]aF(a+1) + a(a+1)(1-z)F(a+2) = 0 \quad (2.7)$$

where $F(a) = {}_3F_2(a, b, c; d, e; z)$, $s = d + e - a - b - c$.

Evidently, it should be possible to obtain another formal solution for unrestricted j_1, j_2, j by the operation $\mathcal{J}$, that is, by the simultaneous interchanges $j_1 \leftrightarrow j_2$, $m_1 \leftrightarrow -m_2$. This second solution has the form

$$a^{(2)}_{m_2} = \Gamma\begin{bmatrix} -j_2 + m_2, -j_1 - m_1, -j_0 + m + 1 \\ j_2 + m_2 + 1, j_1 - m_1 + 1, -2j_1, -\sigma + m + 1 \end{bmatrix} {}_3F_2^{(2)}$$

where

$${}_3F_2^{(2)} = {}_3F_2\begin{bmatrix} -j_1 + m_1, -\sigma - j, -\sigma + j + 1; \\ -2j_1, -\sigma + m + 1; \end{bmatrix} 1. \quad (2.8)$$

Let us now assume that one of the representations coupled belongs to the discrete class. The assumption simplifies the mathematical problem a good deal and removes all uncertainties with regard to the convergence of the various series arising in the problem. If j_2 belongs to the positive discrete class D^+, then the coefficients a_{m_2} may all be taken to be zero for $m_2 < -j_2$ without violating (2.5). The series, in this case, begins with $m_2 = -j_2$ and the coefficient has the general form (2.6). If, on the other hand, j_1 belongs to the negative discrete class D^-, then the series begins with $m_2 = m - j_1$ and the general form of the coefficient is given by (2.8). Thus, the unnormalized CGC is obtained easily from the series solution of (2.4) when at least one of the representations coupled belongs to the discrete class. When j_2 belongs to D^+, j_1 belongs to D^- and m lies between $j_1 + j_2$ and $-j_1 - j_2$, we are at liberty to choose either form of the CGC. In this case it is easily shown [Slater (1966)][12] that

$${}_3F_2^{(1)} = \Gamma\begin{bmatrix} -2j_2, -\sigma - m + 1, -j_0 - j, -j + m \\ -2j_1, -\sigma + m + 1, j_0 - j, -j - m \end{bmatrix} {}_3F_2^{(2)}. \quad (2.9)$$

For the convergence of the various series arising in the subsequent sections it is necessary to assume that x is a real or complex variable lying in the region $0 < |x| < 1$. Nothing prevents us from making such an assumption for x may be regarded as a hypothetical variable introduced merely to reproduce the results of operation by the generators.

We conclude this section with a remark on the convergence of the expressions for the CGC for three continuous series. If j_1, j_2, j all belong to the continuous (nonexceptional) class D^c, then ${}_3F_2^{(1)}$ diverges for $m_1 > \frac{1}{2}$ and ${}_3F_2^{(2)}$ diverges for $m_2 < -\frac{1}{2}$. In this case one must use alternative expressions for the coefficients which remain convergent for all values of the magnetic quantum numbers. One such expression is

$$\Gamma\begin{bmatrix} -j_2 + m_2, -j_1 - m_1, \\ j_2 + m_2 + 1, -\sigma + j + 1, -j_2 - j - m_1 \end{bmatrix} \times {}_3F_2\begin{bmatrix} -j - m, j_0 - j, j_1 - m_1 + 1; \\ j_0 - m + 1, -j_2 - j - m_1; \end{bmatrix} 1. \quad (2.10)$$

That this is a solution of (2.5) is easily seen from the relation

$$(d-a)(e-1)F(a-1, e-1) + [a(e-a) + (e-1)(a-d) + (b-a)(c-a)z]F - a(e-a)(1-z)F(a+1) - \frac{a}{e}\cdot(e-b)(e-c)zF(a+1, e+1) = 0 \quad (2.11)$$

where

$$F = {}_3F_2(a, b, c; d, e; z), F(a-1, e-1) = {}_3F_2(a-1, b, c; d, e-1; z), \text{ etc.}$$

From (2.10) another expression for the coefficient is obtained by the operation $\mathcal{G}$. A third expression is

$$(-)^{j_2+m_2}\Gamma\begin{bmatrix} j_1+j_2-j+1 \\ j_2+m_2+1,\, j_1-m_2-j+1 \end{bmatrix} \times {}_3F_2\begin{bmatrix} -j_2-m_2,\, j_0-j,\, -j-m; \\ j_1-m_2-j+1,\, j_0-m+1; \end{bmatrix} 1. \tag{2.12}$$

3. THE DISCRETE PART OF THE SPECTRUM OF j-VALUES

For the determination of the spectrum of j-values appearing in the reduction it seems necessary to express the solution of (2.4) in terms of known functions of analysis. This is easily done by the substitution $G = x^{-j_2}(1-X)^{\sigma-j-1}F$ which reduces (2.4) to the HG equation

$$x(1-x)F'' + [2j + (1-x)(j_1-j_2-m-2j+1)]F' + (j+m)(j_0-j)F = 0. \tag{3.1}$$

Two solutions of interest to us are

$$G^{(1)} = x^{-j_2}E^{(1)} = x^{-j_2}(1-x)^{\sigma-j-1}F(-j-m, j_0-j, j_0-m+1, x) \tag{3.2}$$

and

$$G^{(2)} = x^{-j_1+m}E^{(2)} = x^{-j_1+m}(1-x)^{\sigma-j-1}F(-j+m, -j_0-j, -j_0+m+1, x). \tag{3.3}$$

These are identical with the series solutions of Sec. 2. Equation (3.2) corresponds to positive discrete j_2, and (3.3) to negative discrete j_1. If j_2, j_1 both belong to D^+ (D^-), then it follows from elementary considerations that the spectrum of j is purely discrete, of the positive (negative) class, and lies in the region $j \leq j_1 + j_2$. $j_0 - m + 1$, in this case, is negative but numerically greater than $j_2 + m_2$. If j_2 belongs to D^+ and j_1 to D^-, then one of the two cases must occur: Either (i) m_2 will attain the value $-j_2$ or (ii) m_1 will attain the value j_1. It is easily seen that $j_0 - m \geq 0$ in case (i) and ≤ 0 in case (ii). So, the denominator catastrophe does not happen when the HGF's in (3.2) and (3.3) are expanded.

The discrete part of the j-spectrum is determined easily by applying the operator $\bar{J}_3 = j_{1,3} - j_{2,3}$ to the coupled state and using the hermiticity condition. For positive discrete j the coupled state can be written in the alternative form

$$\Phi_{jm} = \alpha^m \chi_{jm} = \alpha^m (1-z)^{-j_2} z^{\sigma-j-1} F(-j-m, j_0-j, -2j, z)$$

with $z = 1-x$. (3.4)

Operating with

$$\bar{J}_3 = \alpha\frac{\partial}{\partial\alpha} + 2(1-z)\frac{\partial}{\partial z},$$

we have

$$\bar{J}_3\Phi_{jm} = (-j^2+m^2)\cdot\frac{(j_0^2-j^2)(\sigma+j)}{2j^2(4j^2-1)}\cdot\Phi_{j-1\ m} + \frac{mj_0\sigma}{j(j+1)}\cdot\Phi_{jm} + 2(\sigma-j-1)\cdot\Phi_{j+1\ m}. \tag{3.5}$$

This result is obtained by using the relations

$$(1-z)\frac{dF}{dz} = \frac{ab}{c}\cdot F - \frac{ab(c-b)(c-a)}{c^2(c+1)}\cdot zF(a+1, b+1, c+2) \tag{3.6}$$

and

$$(1-z)F = \frac{ab(c-a)(c-b)}{c^2(c^2-1)}\cdot z^2F(a+1, b+1, c+2) - \frac{c(c-a-b-1)+2ab}{c(c-2)}\cdot zF + F(a-1, b-1, c-2)$$

with $F = {}_2F_1(a, b, c, z)$, etc. (3.7)

The hermiticity of $\bar{J}_3$ now gives

$$\left|\frac{A_{jm}}{A_{j-1m}}\right|^2 = \frac{(\sigma^{*2}-j^2)4j^2(4j^2-1)}{(-j^2+m^2)\,|\sigma+j|^2(j_0^2-j^2)} \tag{3.8}$$

where A_{jm} is the normalization factor (NF) of Φ_{jm}. This equation determines the NF and the range of j-values, but with a degree of uncertainty. Since the remaining factor on the rhs of (3.8) is positive for $m^2 > j^2$, the ratio $|A_{jm}/A_{j-1\ m}|$ will be positive if $(\sigma^{*2}-j^2)/(j_0^2-j^2)$ is so. For discrete j_1, j_2 the latter quantity is positive in the region $j \leq j_1 + j_2$ and also in the region $j > -|j_0|$. The first region has an upper boundary at $j = j_1 + j_2$ which j cannot cross, and the second region has a lower boundary at $j = -|j_0|$. It is not possible to decide from (3.8) alone in which of the regions j will lie in any particular case of coupling of two discrete representations. The two regions are separated by no man's land stretching from $-|j_0|$ to $j_1 + j_2$. For the coupling of one discrete and one continuous representation the situation is more definite. Equation (3.8), in this case, permits all values of $j < -\frac{1}{2}$, -1 to appear in the reduction.

4. THE SOMMERFELD-WATSON TRANSFORMATION

For the determination of the NF and the complete spectrum of j-values it is necessary to expand $x^{j_2+m_2}$ (or $x^{j_1-m_1}$) in terms of the coupled states χ_{jm}. The expansion coefficients are complex conjugates of the CGC to be determined. We start from the identity [Erdélyi (1953)][13]

$$x^r = \sum_{n=0}^{\infty}(-)^n\frac{(a)_n(b)_n}{n!\,(c+n-1)_n}\cdot{}_3F_2\begin{bmatrix} -r,\, c+n-1,\, -n; \\ a,\, b; \end{bmatrix} 1 \times (1-x)^nF(a+n, b+n, c+2n, 1-x) \tag{4.1}$$

with

$$r = j_2+m_2, \quad n = \sigma-j-1, \quad a = -\sigma-m+1, \quad b = -2j_2, \quad c = -2\sigma+2. \tag{4.2}$$

In the case of coupling of $D^+_{j_2}$ and $D^+_{j_1}$, (4.1) and (4.2) give

$$x^{j_2+m_2} = \sum_{j=j_1+j_2}^{-\infty}\Gamma\begin{bmatrix} \sigma+m,\, j_0-j,\, -\sigma-j \\ j+m+1,\, -2j_2,\, -2j-1,\, \sigma-j \end{bmatrix} \times {}_3F_2^{(1)}(1-x)^{j_1+j_2-j}F(-j-m, j_0-j, -2j, 1-x) \tag{4.3}$$

$$= \sum_j\Gamma\begin{bmatrix} \sigma+m,\, j_0-j,\, -\sigma-j,\, -2j,\, -j_0+m \\ j+m+1,\, -2j_2,\, -2j-1,\, \sigma-j,\, -j+m,\, -j_0-j \end{bmatrix}$$

$$\times {}_3F_2^{(1)} E^{(1)} = \sum_j b_{m_2}^{(1)} E^{(1)}. \tag{4.4}$$

The values of j, therefore, decrease from $j_1 + j_2$ to $-\infty$ in steps of unity, the ratio of the NF's of $E^{(1)}$ and $x^{j_2+m_2}$ is

$$\left| \frac{A_j}{A_{m_2}} \right| = \left| \frac{b_{m_2}^{(1)*}}{a_{m_2}^{(1)}} \right|^{1/2} \quad \text{with}$$

$$a_{m_2}^{(1)} = \Gamma \begin{bmatrix} -j_2+m_2, -j_1+m_1, \sigma+m \\ j_2+m_2+1, j_1+m_1+1, -2j_2, -j_0+m \end{bmatrix} {}_3F_2^{(1)}$$

and the CGC (without the phase factor) is

$$(b_{m_2}^{(1)*} a_{m_2}^{(1)})^{1/2} = {}_3F_2^{(1)} (-2j-1)^{1/2} \cdot \frac{\Gamma(\sigma+m)}{\Gamma(-2j_2)} \times$$

$$\left(\Gamma \begin{bmatrix} j_0-j, -\sigma-j, -j_2+m_2, -j_1+m_1 \\ -j_0-j, \sigma-j, -j+m, j+m+1, j_2+m_2+1, j_1+m_1+1 \end{bmatrix} \right)^{1/2} \tag{4.5}$$

The CGC for $D_{j_2}^- \otimes D_{j_1}^-$ is obtained from this by the operation $\mathcal{G}$.

By using the expansion (4.1) we have obtained a complete solution of the CG problem for the coupling of two representations of the positive (or negative) discrete class. In other cases of coupling the rhs of (4.3) contains unwanted values of j, and, for obtaining the correct expansion it becomes necessary to apply the SWT. We write the rhs of (4.3) as a sum of residues at $j = \sigma - n - 1$ of the analytic function

$$2\pi i f(j) = \Gamma \begin{bmatrix} -j-m, j_0-j, -\sigma+j+1, -\sigma-j \\ -2j-1, -\sigma-m+1, -2j_2 \end{bmatrix}$$

$$\times {}_3F_2^{(1)} (1-x)^{\sigma-j-1} F(-j-m, j_0-j, -2j, 1-x). \tag{4.6}$$

This sum is equal to the value of $\int f(j)\, dj$ on a suitably defined contour in the j-plane. However, instead of this contour it is more convenient to choose a larger one consisting of the infinite semicircle on the left and the line $R1\ j = -\frac{1}{2}$. Besides the poles at $\sigma - n - 1$ the larger contour C may enclose other poles originating from the Γ-functions in the numerator of the integrand. If j_2 belongs to D^+ and j_1 to D^c, then the additional poles occur at $j = -m + k$ $(k = 0, 1, \cdots)$, and, hence, lie within C only if $m > k \geq 0$. The poles at $j_0 + k$, $-\sigma + k$, clearly, lie outside C. Hence,

$$\int f(j)\, dj = x^{j_2+m_2} + (\text{sum of the residues at } -m+k).$$

A study of the asymptotic behaviour of $f(j)$ for large $|j|$ (see Paper I, Appendix) shows that the integral vanishes for the semicircular part of C. Evaluating the residues at $-m+k$ and folding the integral about the real axis we, therefore, have

$$x^{j_2+m_2} = \left(\sum_{j=-1/2(-3/2)}^{-m} + i \int_{-1/2}^{-1/2+i\infty} \right) \eta(j) * {}_3F_2^{(1)} E^{(1)}$$

$$\times \Gamma \begin{bmatrix} j_0-j, -\sigma+j+1, -\sigma-j, j_0+j+1, -2j \\ -\sigma-m+1, -2j_2, j+m+1, -j+m, j_0-m+1, -2j-1 \end{bmatrix}$$

$$= \sum_j b_{m_2}^{(1)d} E^{(1)d} - i \int b_{m_2}^{(1)c} E^{(1)c}\, dj, \tag{4.7}$$

where

$$\eta(j) = 1 \quad \text{for discrete } j,$$

$$= -\cot \pi(j+m) \quad \text{for continous } j.$$

The factors involving the Γ – functions in $b_{m_2}^{(1)d}$ and $b_{m_2}^{(1)c}$ are positive, and

$$\begin{aligned} {}_3F_2^{(1)}(j_1^*, j) &= {}_3F_2^{(1)}(j_1^*, j^*) \\ &= {}_3F_2 \begin{bmatrix} -j_2-m_2, j_0-j, j_0+j+1; \\ j_0-m+1, -2j_2; \end{bmatrix} 1 \\ &= {}_3F_2^{(1)} \Gamma \begin{bmatrix} -j_1-m_1, j_0-m+1 \\ j_1-m_1+1, -\sigma-m+1 \end{bmatrix}, \end{aligned} \tag{4.8}$$

by Slater's identities. The quotient $b_{m_2}^{(1)*}/a_{m_2}^{(1)}$ is, therefore, positive for both discrete and continuous j, and the CGC, apart from a phase factor, has the value

$${}_3F_2^{(1)} [-(2j+1)\eta(j)]^{1/2} [\Gamma(-\sigma-m+1)\Gamma(-2j_2)]^{-1}$$

$$\times \left(\Gamma \begin{bmatrix} j_0-j, -\sigma+j+1, -\sigma-j, j_0+j+1, -j_2+m_2, \\ -j_1-m_1 \\ j_2+m_2+1, j_1-m_1+1, j+m+1, -j+m \end{bmatrix} \right)^{1/2} \tag{4.9}$$

In writing this a phase factor, equal to the square root of the coefficient of ${}_3F_2^{(1)}$ in (4.8), has been omitted to make the expression look like (4.5).

Equation (4.7) and the equation obtained from it by the operation $\mathcal{G}$ give the structure of the CG series for the products $D_{j_2}^+ \otimes D_{j_1}^c$:

$$D_{j_2}^+ \otimes D_{j_1}^c = \sum_{j=-1(-3/2)}^{-\infty} D_j^+ + \int_{-1/2}^{-1/2+i\infty} D_j^c\, dj. \tag{4.10}$$

Here, $\sum$ denotes a direct sum, $\int$ denotes a direct integral, and the values of j in the sum decrease from a maximum(-1 or $-\frac{3}{2}$) to $-\infty$ in steps of unity. Since the poles of $\Gamma(-j \pm m)$ and $\Gamma(-2j-1)$ cancel at the point $j = -\frac{1}{2}$, and since the point $j = 0$ lies outside C, the representations D_0 and $D_{-1/2}$ do not occur in the decomposition.

The last case of coupling to be considered is that of $D_{j_2}^+$ and $D_{j_1}^-$. To determine the range of j-values and the CGC in this case we have to use both the solutions of (2.4). We shall use Eqs. (2.6), (3.2), (4.6) if m_2 attains the value $-j_2$, and the corresponding equations obtained after the operation $\mathcal{G}$ if m_1 attains the value j_1. The function $f(j)$ and the function $\bar{f}(j)$ (obtained after the operation $\mathcal{G}$) have a set of simple and double poles lying along the real j axis. With the double poles and the poles in the right half of the j plane we are not concerned. The simple poles of $f(j)$ lie in the region

$$0 \text{ — } j_0 \quad \text{if } m \leq j_0 \leq 0$$

and

$$0 \text{ — } -m \quad \text{if } j_0 \geq m \geq 0.$$

The simple poles of $\bar{f}(j)$ lie in the region

$$0 \text{ — } -m \quad \text{if } j_0 \leq m \leq 0,$$

$$0 \text{ — } -j_0 \quad \text{if } m \geq j_0 \geq 0.$$

Therefore, the spectrum of j has a discrete part only if m and j_0 have the same sign, and the CG series has the form

$$D^+_{j_2} \otimes D^-_{j_1} = \sum_{j=-1(-3/2)}^{-(\text{smaller of } |j_0|, |m|)} D^+_j \text{ or } - + \int_{-1/2}^{-1/2+i\infty} D^c_j \, dj.$$

Expressions for the CGC are derived in the same manner as in the previous cases.

Unnormalized CGC for three continuous series have been already obtained in two different forms in Sec. 2 by solving the recurrence relations (2.5). The determination of the NF and the spectrum of j-values presents difficulties in this case because of the apparent absence of a compact analytic form of the coupled state. However, the difficulties can be circumvented and the above technique extended to this case also if the domain of variation of the complex variable x is suitably restricted. Calculations for this case are under way and the results will be communicated shortly.

[1]W. J. Holman and L. C. Biedenharn, Ann. Phys. 39, 1 (1966).
[2]W. J. Holman and L. C. Biedenharn, Ann. Phys. 47, 205 (1968).
[3]I. Ferretti and M. Verde, Nuovo Cimento 55, 110 (1968).
[4]K. H. Wang, J. Math. Phys. 11, 2077 (1970).
[5]M. Andrews and J. Gunson, J. Math. Phys. 5, 1391 (1964).
[6]S. S. Sannikov, Sov. Phys. Dokl. 11, 1045 (1967).
[7]L. Pukanszky, Trans. Am. Math. Soc. 100, 116 (1961).
[8]J. L. Burchnall and T. W. Chaundy, Proc. Lond. Math. Soc. 50, 56 (1944).
[9]N. Mukunda and B. Radhakrishnan, J. Math. Phys. 15, 1320 (1974); 15, 1332 (1974); 15, 1643 (1974); 15, 1656 (1974).
[10]S. D. Majumdar, Prog. Theor Phys. 20, 798 (1958).
[11]D. Basu and S. D. Majumdar, J. Math. Phys. 16, (1975), preceeding article.
[12]L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, 1966).
[13]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 187.

New identities on the Riemann tensor*

Kuo-shung Cheng

Department of Applied Mathematics, National Chiao-Tung University, Hsinchu, Taiwan 300, Republic of China
(Received 13 September 1974; revised manuscript received 12 May 1975)

A set of new identities which involve second covariant derivatives and quadratic forms of the Riemann tensor are proved. These new identities can be thought of as integrability conditions derived from the equations that define the Rieman tensor in terms of the affine connections.

1. INTRODUCTION

We all know the system of equations which define the Riemann tensor R^1_{ijk} in terms of the affine connections Γ^1_{ij}. These equations are

$$(\partial_j\Gamma^1_{ik} - \partial_k\Gamma^1_{ij}) + (\Gamma^a_{ik}\Gamma^1_{aj} - \Gamma^a_{ij}\Gamma^1_{ak}) = R^1_{ijk}. \tag{1}$$

It is obvious from Eqs. (1) that R^1_{ijk} is antisymmetric with respect to the indices j and k, that is

$$R^1_{ijk} = -R^1_{ikj}. \tag{2}$$

If we assume that the connections are symmetric, that is $\Gamma^1_{ij} = \Gamma^1_{ji}$ (from now on we all assume that the connections are symmetric), then we have

$$R^1_{ijk} + R^1_{jki} + R^1_{kij} = 0. \tag{3}$$

Equations (3) involve no derivative of the Riemann tensor. The famous Bianchi identities which involve first order covariant derivatives of the Riemann tensor are

$$R^1_{ijk;m} + R^1_{ikm;j} + R^1_{imj;k} = 0. \tag{4}$$

Equations (4) can be proved to be independent of Eqs. (2) and (3). That is, Eqs. (4) can not be obtained from the derivatives of Eqs. (2) and (3). We can give these identities some meanings if we regard R^1_{ijk} as given functions and Eqs. (1) as partial differential equations for Γ^1_{ij}.[1,2] From this point of view, Eqs. (2), (3), and (4) are integrability conditions for Eqs. (1). Now, besides Eqs. (2), (3) and (4), are there any other integrability conditions? The answer to this question is yes if the Riemannian space is of high dimensions. For example, in the two-dimensional space, Eqs. (3) and (4) are trivial. But in three-dimensional space, Eqs. (3) and (4) are not trivial at all. It is the purpose of this paper to derive new integrability conditions of Eqs. (1) for high-dimensional space and in turn they are the new identities on the Riemann tensor. We give the new identities in Sec. 2 and some discussions and comments in Sec. 3. We prove the identities in Appendix A.

2. THE NEW IDENTITIES

We can obtain the Bianchi identities, Eqs. (4), from the following procedure:

Step 1: Differentiate Eqs. (1) and get

$$\partial_m\partial_j\Gamma^1_{ik} - \partial_m\partial_k\Gamma^1_{ij} = \partial_m R^1_{ijk} + \partial_m(\Gamma^a_{ij}\Gamma^1_{ak} - \Gamma^a_{ik}\Gamma^1_{aj}). \tag{5}$$

Step 2: Change the order of indices j, k, m, from Eqs. (5) using the rule $j \to k \to m \to j$ and get

$$\partial_j\partial_k\Gamma^1_{im} - \partial_j\partial_m\Gamma^1_{ik} = \partial_j R^1_{ikm} + \partial_j(\Gamma^a_{ik}\Gamma^1_{am} - \Gamma^a_{im}\Gamma^1_{ak}). \tag{6}$$

Step 3: Repeat step 2 from Eqs. (6) and get

$$\partial_k\partial_m\Gamma^1_{ij} - \partial_k\partial_j\Gamma^1_{im} = \partial_k R^1_{imj} + \partial_k(\Gamma^a_{im}\Gamma^1_{aj} - \Gamma^a_{ij}\Gamma^1_{am}). \tag{7}$$

Step 4: Add Eqs. (5), (6), and (7) together and get

$$\text{lhs} = \text{rhs}.$$

The lhs is equal to zero since $\partial_m\partial_j\Gamma^1_{ik} = \partial_j\partial_m\Gamma^1_{ik}$, and the rhs contains first derivatives of the Riemann tensor. After some algebraic manipulations, we can convert the rhs into covariant form. The covariant form version of the rhs is the famous Bianchi identities.

Now we can use this procedure again.

Step 1: Differentiate Eqs. (5) and obtain

$$\partial_n\partial_m\partial_j\Gamma^1_{ik} - \partial_n\partial_m\partial_k\Gamma^1_{ij} = \partial_n\partial_m R^1_{ijk} + \partial_n\partial_m(\Gamma^a_{ij}\Gamma^1_{ak} - \Gamma^a_{ik}\Gamma^1_{aj}). \tag{8}$$

Step 2: Change the indices j,k,n, and m from Eqs. (8) using the rule $j \to k \to n \to m \to j$ and get

$$\partial_m\partial_j\partial_k\Gamma^1_{in} - \partial_m\partial_j\partial_n\Gamma^1_{ik} = \partial_m\partial_j R^1_{ikn} + \partial_m\partial_j(\Gamma^a_{ik}\Gamma^1_{an} - \Gamma^a_{in}\Gamma^1_{ak}) \tag{9}$$

Step 3: Repeat step 2 from Eqs. (9) and get

$$\partial_j\partial_k\partial_n\Gamma^1_{im} - \partial_j\partial_k\partial_m\Gamma^1_{in} = \partial_j\partial_k R^1_{inm} + \partial_j\partial_k(\Gamma^a_{in}\Gamma^1_{am} - \Gamma^a_{im}\Gamma^1_{an}). \tag{10}$$

Step 4: Repeat step 2 from Eqs. (10) and get

$$\partial_k\partial_n\partial_m\Gamma^1_{ij} - \partial_k\partial_n\partial_j\Gamma^1_{im} = \partial_k\partial_n R^1_{imj} + \partial_k\partial_n(\Gamma^a_{im}\Gamma^1_{aj} - \Gamma^a_{ij}\Gamma^1_{am}). \tag{11}$$

Step 5: Add Eqs. (8), (9), (10), and (11) all together and get

$$\text{lhs} = \text{rhs}.$$

The lhs is equal to zero since $\partial_n\partial_m\partial_j\Gamma^1_{ik} = \partial_j\partial_n\partial_m\Gamma^1_{ik}$. The right-hand side contains second order derivatives of the Riemann tensor. We can convert the rhs into covariant form. The covariant version of the rhs is

$$\begin{aligned} &R^1_{ijk;\,mn} + R^1_{ijk;\,nm} + R^a_{inj}R^1_{amk} + R^a_{jmk}R^1_{ian} \\ &\quad + R^1_{ikm;\,nj} + R^1_{ikm;\,jn} + R^a_{ijk}R^1_{anm} + R^a_{knm}R^1_{iaj} \\ &\quad + R^1_{imn;\,jk} + R^1_{imn;\,kj} + R^a_{ikm}R^1_{ajn} + R^a_{mjn}R^1_{iak} \\ &\quad + R^1_{inj;\,km} + R^1_{inj;\,mk} + R^a_{imn}R^1_{akj} + R^a_{nkj}R^1_{iam} = 0. \end{aligned} \tag{12}$$

Or, in short hand notation,

$$F(R^1_{ijk;\,mn} + R^1_{ijk;\,nm} + R^a_{inj}R^1_{amk} + R^a_{jmk}R^1_{ian}) = 0. \tag{12'}$$

Equations (12) are the new identities on the Riemann tensor.

3. DISCUSSIONS AND COMMENTS

(1) The identities, Eqs. (12), *may*[3] be derived from differentiation of the known identities, such as Eqs. (2), (3), and (4). These new identities can be thought of as integrability conditions derived from Eqs. (1).

(2) These identities become trivial when Eqs. (1) become linear, that is when the ($\Gamma\Gamma$ - $\Gamma\Gamma$) terms are absent. So these identities are a kind of measure of the nonlinearity of Eqs. (1).

(3) For three-dimensional space, Eqs. (12) can be derived from Eqs. (2), (3), and (4). (See Appendix B.)

(4) For five-dimensional space or higher, there are still new identities on the Riemann tensor.

(5) Perira[2] derived some identities on the Riemann tensor which from the point of view of the author are all dependent on Eqs. (2), (3), and (4).

APPENDIX A

We define the notation $F(\cdots)$ first. $F(\cdots)$ is an operator which acts on the indices j, k, m, and n with the following property

$$F(nmjk) = (nmjk) + (mjkn) + (jknm) + (knmj). \tag{13}$$

For example,

$$F(R^1_{ijk;\,mn}) = R^1_{ijkmn} + R^1_{ikn;\,jm} + R^1_{inm;\,kj} + R^1_{imj;\,nk}. \tag{14}$$

With this definition, we can easily derive two useful properties of the operator $F(\cdots)$:

$$\text{(i)}\quad F(A+B) = F(A) + F(B), \tag{15}$$

$$\text{(ii)}\quad F(nmjk) = F(mjkn) = F(jknm) = F(knmj). \tag{16}$$

Since

$$R^1_{ijk;\,mn} = \partial_n(R^1_{ijk;\,m}) + \Gamma^1_{an}R^a_{ijk;\,m} - \Gamma^a_{in}R^1_{ajk;\,m} - \Gamma^a_{jn}R^1_{iak;\,m} - \Gamma^a_{kn}R^1_{ija;\,m} - \Gamma^a_{mn}R^1_{ijk;\,a}, \tag{17}$$

we get

$$F(R^1_{ijk;\,mn} + R^1_{ijk;\,nm}) = F(\partial_n R^1_{ijk;\,m} + \Gamma^1_{an}R^a_{ijk;\,m} - \Gamma^a_{in}R^1_{ajkm} - \Gamma^a_{jn}R^1_{iak;\,m} - \Gamma^a_{kn}R^1_{ija;\,m} - \Gamma^a_{mn}R^1_{ijk;\,a} + \text{terms with } n \text{ and } m \text{ interchanged}). \tag{18}$$

But

$$F(\Gamma^a_{jn}R^1_{iak;\,m} + \Gamma^a_{kn}R^1_{ija;\,m} + \Gamma^a_{jm}R^1_{iak;\,n} + \Gamma^a_{km}R^1_{ija;\,n} + \Gamma^a_{mn}R^1_{ijk;\,a}) = F(\Gamma^a_{jn}R^1_{ika;\,m} + \Gamma^a_{nj}R^1_{iak;\,m}) + F(\Gamma^a_{mj}(R^1_{ina;\,k} + R^1_{iak;\,n} + R^1_{ikn;\,a})) = 0. \tag{19}$$

Thus Eqs. (18) become

$$F(R^1_{ijk;\,mn} + R^1_{ijk;\,nm}) = F(\partial_n R^1_{ijk;\,m} - \Gamma^1_{an}R^a_{ijk;\,m} - \Gamma^a_{in}R^1_{ajk;\,m} + \partial_m R^1_{ijk;\,n} + \Gamma^1_{am}R^a_{ijk;\,n} - \Gamma^a_{im}R^1_{ajk;\,n} - \Gamma^a_{nm}R^1_{ijk;\,a}). \tag{20}$$

Substituting the following equation

$$R^1_{ijk;\,m} = \partial_m R^1_{ijk} + \Gamma^1_{am}R^a_{ijk} - \Gamma^a_{im}R^1_{ajk} - \Gamma^a_{jm}R^1_{iak} - \Gamma^a_{km}R^1_{ija} \tag{21}$$

into Eqs. (20), we finally get[4]

$$F(R^1_{ijk;\,mn} + R^1_{ijk;\,nm} + R^a_{inj}R^1_{amk} + R^a_{mnj}R^1_{iak}) = 2F(\partial_n\partial_m R^1_{ijk} + \partial_n\partial_m(\Gamma^a_{ij}\Gamma^1_{ak} - \Gamma^a_{ik}\Gamma^1_{aj})) = 0. \tag{22}$$

APPENDIX B

For three-dimensional space, at least two of the indices n, m, j, k, must be equal. Let us say $n=m$, then the lhs of Eqs. (12) become

$$R^1_{ijk;\,mm} + R^1_{ijk;\,mm} + R^1_{ikm;\,jm} + R^1_{ikm;\,mj} + R^1_{imm;\,kj} + R^1_{imm;\,jk} + R^1_{imj;\,mk} + R^1_{imj;\,km} + R^a_{imj}R^1_{amk} + R^a_{jmk}R^1_{iam} + R^a_{imk}R^1_{ajm} + R^a_{kjm}R^1_{iam} + R^a_{ijm}R^1_{akm} + R^a_{mkm}R^1_{iaj} + R^a_{ikm}R^1_{amj} + R^a_{mmj}R^1_{iak} = (R^1_{ijk;\,m} + R^1_{ikm;\,j} + R^1_{imj;\,k})_{;\,m} \times 2 = 0.$$

Thus we prove the statement that Eqs. (12) are derivable from Eqs. (2), (3), and (4) in the three-dimensional case.

*Work partially supported by the National Science Foundation Council of Taiwan.

[1]C.M. Pereira, J. Math. Phys. 13, 1542 (1972).

[2]C.M. Pereira, J. Math. Phys. 15, 269 (1974).

[3]Thanks to Professor C.M. Pereira and the referee for pointing out this possibility.

[4]We can use normal coordinates to simplify the calculation.

The semi-Euclidean approach in statistical mechanics. I. Basic expansion steps and estimates*

Paul Federbush

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104
(Received 29 April 1975)

The semi-Euclidean formulation, developed in constructive quantum field theory to handle boson–fermion models, is adapted to the statistical mechanics setting.

INTRODUCTION

In Ref. 1 D. Brydges and I presented a formulation for treating boson—fermion models, taking powerful techniques for using Euclidean fields to study Boson models in constructive quantum field theory, and combining these with operator methods to handle the fermions. This, the semi-Euclidean approach, now seems like a useful framework to study such theories. In Ref. 2 Brydges studies the generalized Yukawa model using semi-Euclidean methods. In Ref. 3 techniques patterned after those in Refs. 1 and 2 were applied to prove the classic theorem of Dyson and Lenard[4] on the stability of matter. In the present series of papers I intend to introduce a full semi-Euclidean formalism into statistical mechanics, furthering the flow of ideas from constructive quantum field theory into more classical fields of physics.

Ginibre has made beautiful applications of functional integration techniques in statistical mechanics.[5] The formalism to be presented here will have many points in common with that of Ginibre. The interlacing we will see between the viewpoints of Refs. 1 and 5 seems very satisfying. The interactions in the stability of matter problem— $-1/r$ forces between positively and negatively charged particles—would not lead to stability in the absence of the kinetic energy generated by the exclusion principle for the negative charges.[6] This effect may be difficult to make explicit in the formalism of Ginibre, and the potentials considered in Ref. 5 exclude such forces. Our first long range goal will be to obtain the existence of the infinite volume correlation functions for the matter problem with the interaction $1/r$ modified to $\exp(-nr)/r$ (this interaction still is excluded in Ref. 5). The extension to the long range $1/r$ interaction is deferred.

Our avenue to the infinite volume correlation functions will hopefully be the adaptation of the cluster expansion of Glimm, Jaffe, and Spencer[7] to the present formalism. In the present paper the cluster expansion is not developed. However, we do present some of the expansion operations to be used—differentiation of the exponent and a pull-through formula. The pull-through formula is used to generate a Ginibre-like expansion for correlation functions. We also present operator estimates substituting for the "defermiation" estimates of Ref. 1. These estimates—to be used in the proof of convergence of the cluster expansion—convert operator expressions to a context with commutivity. As an example we estimate some terms in the expansion of a correlation function.

The field theory of boson models, the field theory of boson—fermion models, and the field theory of second quantized matter in statistical mechanics provide three types of field theories—with striking similarities and striking differences—that will develop with mutual enrichment.

1. NOTATION

We work with H of the form

$$H = H_0 + \tfrac{1}{2}\int dz_1 dz_2 : \bar{\psi}\psi(z_1)\, V(z_1, z_2)\, \bar{\psi}\psi(z_2): \qquad (1.1)$$

with V symmetric in z_1 and z_2 and H_0 the sum of a multiple of the number operator with the kinetic energy form for Dirichlet data on the boundary of a fixed volume V. ψ and $\bar{\psi}$ are annihilation and creation fields for a fermion or boson particle. The extension to more general potential interactions and more than one species of particle is straightforward (such as for protons and electrons moving in a fixed background charge). The objects from statistical mechanics we will study are of the form

$$\mathrm{Tr}[\exp(-\beta H)\, \bar{\psi}(x_1, t_1) \cdots \psi(x_n, t_n)], \qquad (1.2)$$

where the t_i correspond to imaginary times, the t_i decreasing $0 \leq t_i \leq \beta$, and the $\psi(x_i, t_i)$ are obtained from $\psi(x)$ by progagation under H for an imaginary time t_i. We now expunge this definition of the $\psi(x_i, t_i)$ from our memory and follow an alternate line of development.

We define $\psi(x, t) \equiv \psi(x)$, the t introduced only as a label to enable us to time order. $H(t)$ is H expressed in terms of the $\psi(t)$ and $\bar{\psi}(t)$:

$$H(t) = (-1/2M)\, \bar{\psi}(z, t)\, \nabla^2 \psi(z, t) + \mu\, \bar{\psi}(z, t)\, \psi(z, t) + \tfrac{1}{2} : \bar{\psi}\psi(z_1, t)\, V(z_1, z_2)\, \bar{\psi}\psi(z_2, t): \qquad (1.3)$$

(integrals are suppressed). (1.2) becomes

$$\mathrm{Tr}[T \exp[-\int_0^\beta H(t)\, dt]\, \bar{\psi}(x_1, t_1) \cdots \psi(x_n, t_n)] \qquad (1.4)$$

with T the time-ordering operation familiar to physicists. In fact (1.4) is essentially the interaction representation with the interaction taken to be the full Hamiltonian—so the interaction fields have no true time dependence, as our $\psi(x, t)$. Taking (1.4) as a serious expression to manipulate and perform estimates with is the heart of the semi-Euclidean approach.

We also want an expression for the path space measure generated by the one-particle free Hamiltonian

$$(\exp[-s(-1/2M\nabla^2 + \mu)]\psi)(y) = \int dx \int d\mu^s_{x,y}\, \psi(x), \qquad (1.5)$$

where $\int d\mu^s_{x,y}$ is a measure on paths $x_p(t)$, $0 \leq t \leq s$,

connecting x and y (and lying in V).

2. EXPANSION STEPS

In this paper we consider only two operations. The first of these, differentiation of the exponent, assumes $H(t)$ depends on a parameter λ, and so we write $H_\lambda(t)$. This happens by allowing $V(z_1, z_2)$ to depend on λ, and so, $V_\lambda(z_1, z_2)$. We write (1.4) as

$$\mathrm{Tr}[T \exp(-\int_0^\beta H_\lambda(t)\,dt)R]. \tag{2.1}$$

R a polynomial in fields.

Differentiation of the exponent

$$\frac{d}{d\lambda}\,\mathrm{Tr}[T \exp(-\int_0^\beta H_\lambda(t)\,dt)R]$$
$$= \int_0^\beta ds\ \mathrm{Tr}\left(T \exp(-\int_0^\beta H_\lambda(t)\,dt)\ \frac{dH_\lambda(s)}{d\lambda} R\right). \tag{2.2}$$

The other operation is a pull-through operation. (The pull-through operation used in Refs. 1 and 2, different from the one presented here, may be useful in some circumstances.)

Pull-through formula

$$T \exp(-\int_a^b H(t)\,dt)\,\bar\psi(x,a)$$
$$= T \int dy \int d\mu_{x,y}^{(b-a)}\ \bar\psi(y,b)$$
$$\times \exp\left(-\int_a^b [H(t) + V(x_p(t), z)\,\bar\psi\psi(z)]\,dt\right). \tag{2.3}$$

We will not write the similar expression for ψ and more general potentials.

This pull-through formula provides the connection between the semi-Euclidean formulation and the work of Ginibre, as we will see in the next section.

To prove (2.3), we consider the equality

$$T \int dy \int d\mu_{x,y}^{(b-a)}\ \bar\psi(y,b) \exp\left(-\int_a^b [H(t) + V(x_p, z)\bar\psi\psi(z)]\,dt\right)$$
$$- T \exp(-\int_a^b H(t)\,dt)\,\bar\psi(x,a)$$
$$= \int_a^b ds\ \frac{d}{ds}\ \left[T \exp(-\int_s^b H(t)\,dt) \int dy \int d\mu_{x,y}^{(s-a)}\,\bar\psi(y,s)\right.$$
$$\left.\times \exp\left(-\int_a^s (H(t) + V(x_p, z)\bar\psi\psi)\,dt\right)\right] \tag{2.4}$$

and verify that the differentiation with respect to s in the brackets gives zero.

3. GINIBRE-TYPE EXPANSION

For clarity we confine our attention to the following correlation function $F(x_2, t_2, x_1, t_1)$:

$$F = \mathrm{Tr}(T \exp[-\int_0^\beta H(t)\,dt]\,\psi(x_2, t_2)\,\bar\psi(x_1, t_1)) \tag{3.1}$$

with $\beta > t_2 > t_1 > 0$. We now use the pull-through formula (2.3) to move $\bar\psi(x,t)$ to the left. We may also use the relation

$$\bar\psi(x,0) = \bar\psi(x,\beta), \tag{3.2}$$

following from commutativity of the trace—see (4.1) in the next section—giving rise to well-known periodicity of our correlation functions. As $\bar\psi$ is continuously moved to left, and re-entered at right by (3.2), a series of terms is generated by the possibility of contraction with ψ after any number of sweeps.

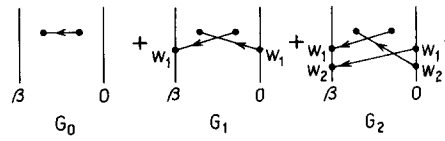

FIG. 1. Graphical representation of G_i.

$$\psi(x)\,\bar\psi(y) - \epsilon\bar\psi(y)\,\psi(x) = \delta(x-y) \tag{3.3}$$

with $\epsilon = 1$ for bosons and $\epsilon = -1$ for fermions. (3.1) then becomes

$$F = G_0 + G_1 + G_2 + \cdots + G_N + R_{N+1}. \tag{3.4}$$

The G_i are represented graphically in Fig. 1. Expressions for G_0, G_1, and G_N follow:

$$G_0 = \mathrm{Tr}\left[T \int d\mu_{x_1,x_2}^{(t_2-t_1)} \exp\left(-\int_0^\beta H(t)\,dt\right)\right.$$
$$\left.\times \exp\left(-\int_{t_1}^{t_2} [V(x_p(t), z)\bar\psi\psi(z)]\,dt\right)\right], \tag{3.5}$$

$$G_1 = (\epsilon)\,\mathrm{Tr}\left[T \int dw_1 \int d\mu_{x_1,w_1}^{(\beta-t_1)} \int d\mu_{w_1,x_2}^{t_2} \exp(-\int_0^\beta H(t)\,dt)\right.$$
$$\times \exp\left(-\int_0^{t_1} V(x_{p_1}, z)\bar\psi\ \psi(z)\,dt\right) \exp\left(-\int_{t_2}^\beta V(x_{p_0}, z)\bar\psi\psi(z)\,dt\right)$$
$$\left.\times \exp\left(-\int_{t_1}^{t_2} [V(x_{p_0}, z)\bar\psi\psi(z) + V(x_{p_1}, z)\bar\psi\psi(z) + V(x_{p_0}, x_{p_1})]\,dt\right)\right], \tag{3.6}$$

$$G_N = (\epsilon)^N\,\mathrm{Tr}\left\{T \int dw_1 \cdots dw_N \int d\mu_{x_1,w_1}^{(\beta-t_1)} \cdots d\mu_{w_i,w_{i+1}}^{\beta}\right.$$
$$\cdots d\mu_{w_N,x_2}^{t_2} \exp(-\int_0^\beta H(t)\,dt)$$
$$\times \exp\left[-\int_0^{t_1} \left(\sum_{j=1}^N V(x_{p_j}, z)\,\bar\psi\psi(z) + \sum_{1\le i<j}^N V(x_{p_i}, x_{p_j})\right)\,dt\right]$$
$$\times \exp\left[-\int_{t_2}^\beta \left(\sum_{j=0}^{N-1} V(x_{p_j}, z)\,\bar\psi\psi(z)\right.\right.$$
$$\left.\left. + \sum_{0\le i<j}^{N-1} V(x_{p_i}, x_{p_j})\right)dt\ \right]$$
$$\left.\times \exp\left[-\int_{t_1}^{t_2} \left(\sum_{j=0}^N V(x_{p_j}, z)\,\bar\psi\psi(z) + \sum_{0\le i<j}^N V(x_{p_i}, x_{p_j})\right)\,dt\right]\right\}. \tag{3.7}$$

This expansion may be compared to the expansions of Ginibre in Ref. 5. We will not here make explicit an expression for R_N. However, in Appendix C an explicit expression for R_1 is given, and estimated as an example of the operator estimates given in the next section. In Appendix B G_0 is similarly estimated. Future applications of the semi-Euclidean formalism depend on our ability to control estimates--Appendix B and Appendix C are simpler than the estimates needed in the cluster expansion, but use the same basic techniques.

4. OPERATOR ESTIMATES

We collect here, first, well-known estimates we will use concerning traces; recall an operator estimate from Ref. 1; and finally present a cute new estimate for traces in a theorem.

We begin with the well-known facts:

$$\mathrm{Tr}(AB)=\mathrm{Tr}(BA), \tag{4.1}$$

$$\begin{cases} A\geq B \Rightarrow \mathrm{Tr}(e^{-A})\leq \mathrm{Tr}(e^{-B}), & (4.2)\\ A\geq 0 \Rightarrow \mathrm{Tr}(A)\geq 0, & (4.3)\\ A\geq B \Rightarrow \mathrm{Tr}(CAC^*)\geq \mathrm{Tr}(CBC^*), & (4.4)\end{cases}$$

$$|\mathrm{Tr}(AC)|\leq [\mathrm{Tr}(AA^*)]^{1/2}\cdot[\mathrm{Tr}(CC^*)]^{1/2}, \tag{4.5}$$

$$|\mathrm{Tr}(ABC)|\leq [\mathrm{Tr}(AA^*)]^{1/2}\cdot\|B\|\cdot[\mathrm{Tr}(CC^*)]^{1/2}. \tag{4.6}$$

We recall from Ref. 1:

Estimate: Let $H(t)\geq C(t)$ with $C(t)$ a numerical function. Then

$$\| T\exp(-\int_0^\beta H(t)\,dt)\|\leq \exp(-\int_0^\beta C(t)\,dt). \tag{4.7}$$

The following theorem is an estimate similar to this last estimate, but for traces.

Theorem: Let $H(t)\geq \bar{H}+C(t)$ with $C(t)$ a numerical function and $\bar{H}$ a t-independent operator. Then

$$|\mathrm{Tr}\,(T\exp[-\int_0^\beta H(t)\,dt])|\leq \mathrm{Tr}(\exp(-\beta\bar{H})\cdot\exp(-\int_0^\beta C(t)\,dt). \tag{4.8}$$

A proof of this theorem is presented in Appendix A.

APPENDIX A: PROOF OF TIME–ORDERED TRACE INEQUALITY

The inequality follows by taking limits in a discrete form of the result, stated in the following lemma, and using (4.2). The lemma is a special case of Corollary 3.2 in Ref. 8.

Lemma:

$$|\mathrm{Tr}(A_1\cdots A_{2^N})|\leq \prod_i \{\mathrm{Tr}[(A_iA_i^*)^{2^{N-1}}]\}^{1/2^N}. \tag{A1}$$

(The restriction of the product of A_i's to contain a power of two elements is a function of our method of proof.)

Proof: We prove the result by induction on N. First for $N=1$:

$$|\mathrm{Tr}(A_1A_2)|\leq [\mathrm{Tr}(A_1A_1^*)]^{1/2}\cdot[\mathrm{Tr}(A_2A_2^*)]^{1/2}. \tag{A2}$$

This is just (4.5). Assume (A1) holds for N and proceed to the $N+1$ case:

$$|\mathrm{Tr}(A_1\cdots A_{2^{N+1}})|=|\mathrm{Tr}(B_1\cdots B_{2^N})| \tag{A3}$$

with

$$B_i=A_{2i-1}A_{2i}$$

$$\leq \prod_i \{\mathrm{Tr}[(B_iB_i^*)^{2^{N-1}}]\}^{1/2^N} \tag{A4}$$

by the induction hypothesis. We consider one of these terms, $i=1$, for notational simplicity:

$$\mathrm{Tr}[(B_1B_1^*)^{2^{N-1}}]=\mathrm{Tr}[(A_1^*A_1A_2A_2^*)^{2^{N-1}}]. \tag{A5}$$

(4.1) was here used. By the induction hypothesis,

$$\leq \{\mathrm{Tr}[(A_2A_2^*)^{2^N}]\}^{1/2}\{\mathrm{Tr}[(A_1^*A_1)^{2^N}]\}^{1/2}\ . \tag{A6}$$

Substituting (A6) back in (A4) we obtain the $(N+1)$th relation, completing the induction.

APPENDIX B: ESTIMATE FOR G_0

G_0 is given in (3.5). To obtain an estimate for G_0, we require two estimates; we write as follows:

$$H_0/2+\tfrac{1}{2}\int dz_1\,dz_2:\bar{\psi}\psi(z_1)\,V(z_1,z_2)\,\bar{\psi}\psi(z_2):\geq K_0, \tag{B1}$$

$$H_0/2+\tfrac{1}{2}\int dz_1\,dz_2:\bar{\psi}\psi(z_1)\,V(z_1,z_2)\,\bar{\psi}\psi(z_2):$$
$$+\int dz\,V(x,z)\,\bar{\psi}\psi(z)\geq K_1. \tag{B2}$$

We do not discuss evaluation of K_0 and K_1 here, as depending on V, the volume, and the interaction. However, it is important to note that in these evaluations the kinetic energy arising in the Fermion case may be exploited (see Ref. 3). We now use (4.8) to obtain

$$|G_0|\leq \mathrm{Tr}[\exp(-\beta H_0/2)]\cdot\exp[-K_1(t_2-t_1)]$$
$$\times\exp[-K_0(\beta-(t_2-t_1))]$$
$$\cdot\frac{(M/2\pi)^{3/2}}{|t_2-t_1|^{3/2}}\exp\left(-\frac{M}{2}\frac{|x_2-x_1|^2}{(t_2-t_1)}\right)\ . \tag{B3}$$

APPENDIX C: Estimate for R_1

We graphically represent R_1 in Fig. 2. For definiteness we require $2/3\beta>t_2>t_1>\beta/3$, this is not essential. Having pulled-through $\bar{\psi}$ past t_2 without contraction with ψ, we stop at $2/3\beta$ and then pull-through ψ to $1/3\beta$. The resulting expression for R_1 is

$$R_1=\epsilon\int dw_1\,dw_2\int d\mu_{x_1,w_1}^{(2/3\beta-t_1)}\int d\mu_{w_2,x_2}^{(t_2-1/3\beta)}$$
$$\times\mathrm{Tr}\Big[T\exp(-\int_{2/3\beta}^{\beta}H(t)\,dt)\,\bar{\psi}(w_1,2/3\beta)$$
$$\cdot\exp\Big(-\int_{t_2}^{2/3\beta}[H(t)+V(x_{p_1},z)\bar{\psi}\psi(z)]\,dt\Big)$$
$$\cdot\exp\Big(-\int_{t_1}^{t_2}[H(t)+V(x_{p_1},z)\bar{\psi}\psi(z)$$
$$+V(x_{p_2},z)\bar{\psi}\psi(z)+V(x_{p_1},x_{p_2})]\,dt\Big)$$
$$\cdot\exp\Big(-\int_{\beta/3}^{t_1}[H(t)+V(x_{p_2},z)\,\bar{\psi}\psi(z)]\,dt\Big)$$
$$\times\psi(w_2,1/3\beta)\exp(-\int_0^{\beta/3}H(t)\,dt)\Big]. \tag{C1}$$

We abbreviate this as

$$R_1=\epsilon\int dw_1\,dw_2\int d\mu_1\,d\mu_2\exp\Big(-\int_{t_1}^{t_2}V(x_{p_1},x_{p_2})\,dt\Big)$$
$$\times\mathrm{Tr}[TE_1\,\bar{\psi}(w_1)E_2E_3E_4\psi(w_2)E_5], \tag{C2}$$

where the E_i are the obvious exponentials, except that, as indicated, a portion of E_3 has been separated out explicitly. We use (4.6):

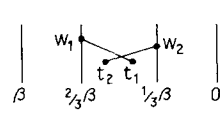

FIG. 2. Graphical representation of R_1.

$$|R_1| \leq \int dw_1 dw_2 \int d\mu_1 d\mu_2 \exp\left(-\int_{t_1}^{t_2} V(1,2)\,dt\right)$$
$$\times \{\mathrm{Tr}[(TE_1)\,\bar{\psi}(w_1)\,\psi(w_1)(T^*E_1^*)]\}^{1/2}\ \|E_2 E_3 E_4\|$$
$$\times \{\mathrm{Tr}[(T^*E_5^*)\,\bar{\psi}(w_2)\,\psi(w_2)(TE_5)]\}^{1/2} \tag{C3}$$

using

$$2xy \leq x^2 + y^2, \tag{C4}$$

we get

$$2\,|R_1| \leq A + B$$

with

$$A = \int dw_1 \int dw_2 \int d\mu_1 \int d\mu_2 \exp\left(-\int_{t_1}^{t_2} V(1,2)\,dt\right)$$
$$\times \mathrm{Tr}[(TE_1)\,\bar{\psi}(w_1)\,\psi(w_1)(T^*E_1^*)]\ \|E_2 E_3 E_4\|$$

and a similar expression for B.

Let

$$\left|\int d\mu_1 \int d\mu_2 \exp\left(-\int_{t_1}^{t_2} V(1,2)\,dt\right)\right| \leq C \tag{C5}$$

for all w_1 and w_2. This inequality, a statement concerning two particles interacting by a mutual potential, can be estimated by standard methods. We get

$$A \leq C \int \exp\{-[(2/3\beta - t_2)K_1 + (t_2 - t_1)K_2 + (t_1 - \beta/3)K_1]\}$$
$$\times \mathrm{Tr}[\exp(-2/3\beta H_0)N]\ . \tag{C6}$$

K_2 is defined similarly to K_0 and K_1. The methods of Appendices B and C can be applied in much more complicated situations, and can, very essentially, include localization.

*This work was supported in part by NSF Grant GP-17523.

[1]D. Brydges and P. Federbush, J. Math. Phys. 15, 730 (1974).
[2]D. Brydges, J. Math. Phys. 16, 1649 (1975).
[3]P. Federbush, A new approach to the stability of matter Problem. I and II, J. Math. Phys. 16, 347, 706 (1975).
[4]F. J. Dyson and A. Lenard, J. Math. Phys. 8, 423 (1967); 9, 698 (1968).
[5]J. Ginibre, "Some Applications of Functional Integration in Statistical Mechanics," in *Statistical Mechanics and Quantum Field Theory, Les Houches* 1970, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971).
[6]F. J. Dyson, J. Math. Phys. 8, 1538 (1967).
[7]J. Glimm, A. Jaffe, and T. Spencer, "The Cluster Expansion, in: Constructive Quantum Field Theory," in *The* 1973 *"Ettore Majorana" International School of Mathematical Physics*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973).
[8]R. A. Kunze, Trans. Amer. Math. Soc. 89, 519 (1958).

The semi-Euclidean approach in statistical mechanics. II. The cluster expansion, a special example*

Paul Federbush

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104
(Received 29 April 1975)

A form of the Glimm–Jaffe–Spencer cluster expansion, adapted to the statistical mechanics setting, is shown to converge for certain two-body potential interactions. The theory treated corresponds to negatively charged fermions and positively charged bosons interacting by a modified Coulomb interaction—the $1/r$ potential, cutoff at high and low momenta, becoming $(1/r)(e^{-\alpha r}-e^{-\gamma r})$.

INTRODUCTION

In Ref. 1 a formalism was presented to adapt the semi-Euclidean approach from constructive quantum field theory to statistical mechanics. In the present paper the Glimm—Jaffe—Spencer cluster expansion[2] is shown to have a close analog in this setting. The theory treated corresponds to negatively charged fermions and positively charged bosons interacting by a modified Coulomb interaction—the $1/r$ potential, cutoff at high and low momenta, becoming $(1/r)(e^{-\alpha r}-e^{-\gamma r})$. For suitable values of the parameters the cluster expansion patterned after Ref. 2 will be shown to converge. We hope to come back in a later paper to the $\gamma=\infty$ case, whose treatment should require only technical improvements over the present procedure. The methods of Ref. 3 apply to the interaction treated in the present paper, but not to the $\gamma=\infty$ situation. We have no definite ideas on how to treat the physically interesting case with $\alpha=0$, $\gamma=\infty$.

A knowledge of Ref. 2 is required to read the present paper. Space is cut into unit cubes—the greater flexibility of allowing other size cubes is sacrificed to agree most completely with Ref. 2. Before presenting further details we point out the following similarities and differences to orient the reader.

(a) The potential $(1/r)(e^{-\alpha r}-e^{-\gamma r})$ is interpolated by exactly the same procedure as in Ref. 2.

(b) Unlike Ref. 2 we here interpolate Hamiltonians rather than covariances.

(c) The interaction portion of the Hamiltonian is interpolated as indicated in (a). The kinetic energy portion is interpolated by erecting parameter dependent barriers—an infinite barrier giving Dirichlet data.

(d) Estimates of integrals of functions on Euclidean space are replaced by operator estimates from Ref. 1, characteristic of the semi-Euclidean approach.

We anticipate extension of the present work to more general interactions, as well as the development of a cluster expansion for boson–fermion field theory models within the semi-Euclidean formalism.

1. THE CLUSTER EXPANSION

Space is filled with unit cubes $\{\Delta_i\}$. Each Δ_i is of the form $a_i \leq x_i \leq a_i+1$, $i=1,2,3$ for some set of integers (a_1, a_2, a_3). A union of such cubes is said to be intimately connected if it remains connected after removal of all edges and vertices (i.e., contacts must take place across faces). The set of faces is denoted by $\{F_i\}$, the indexing of Δ_i and F_i being unrelated. If W is a union of cubes, we define

$$\hat{W}=\{\mathbf{x}\in W \mid d(\mathbf{x}, \partial W)\geq \delta\} \tag{1.1}$$

for a small parameter δ.

With a given W as above, we associate the Foch Hilbert space, H_W, constructed of functions vanishing outside $\hat{W}$. Tr_W is the trace over this Hilbert space. Our interaction will be made up of Yukawa interations e^{-nr}/r, in the infinite volume. For a union of cubes W we define $y_W^n(x,y)$, x and y in W, by

$$(-\Delta+n^2)_x y_W^n(x,y)=(-\Delta+n^2)_x \exp(-n|x-y|)/|x-y| \tag{1.2}$$

and

$$y_W^n(x,y)=0 \quad \text{if } y\in \partial W. \tag{1.3}$$

Note that

$$y_{R3}^n(x,y)=\exp(-n|x-y|)/|x-y|. \tag{1.4}$$

With a union of cubes W, we associate Δ_W, the Laplacian on functions defined in $\hat{W}$ satisfying Dirichlet data on $\partial\hat{W}$. Λ is a fixed finite intimately connected union of cubes. $\bar{\psi}$, ψ and $\bar{\phi}$, ϕ are fields describing fermions and bosons, respectively. $H=H^\Lambda$ is the Hamiltonian defined on H_Λ:

$$H^\Lambda=H_0^\Lambda+V^\Lambda, \tag{1.5}$$

$$H_0^\Lambda=-\frac{1}{2M}\bar{\psi}\Delta_\Lambda\psi-\frac{1}{2m}\bar{\phi}\Delta_\Lambda\phi+\frac{\mu}{\beta}\bar{\psi}\psi+\frac{\mu}{\beta}\bar{\phi}\phi, \tag{1.6}$$

$$V^\Lambda=\tfrac{1}{2}:(\bar{\psi}\psi-\bar{\phi}\phi)v^\Lambda(\bar{\psi}\psi-\bar{\phi}\phi):, \tag{1.7}$$

$$v^\Lambda=q^2(y_\Lambda^\alpha-y_\Lambda^\gamma). \tag{1.8}$$

Integrals have been suppressed. If it is clear, the Λ indicators may be omitted. For any union of cubes, W say, a similar expression H^W may be constructed. Using the notation of Ref. 1, we want to consider an expression of the form

$$\mathrm{Tr}_\Lambda[T\exp(-\textstyle\int_0^\beta H^\Lambda(t)\,dt)A] \tag{1.9}$$

with A a polynomial in the fields. For simplicity we will also assume all the fields in A localized in a fixed cube Δ_0.

We now proceed to consider the interpolated Hamiltonians. A chanracteristic function for a neighborhood of the face F_i is defined as follows:

$$\chi_i = \{\mathbf{x} \mid d(\mathbf{x}, F_i) < \delta\} \tag{1.10}$$

With each F_i is associated a parameter s_i, $0 \leq s_i \leq 1$. We define

$$h(s) = (1/s) - 1 \tag{1.11}$$

so that $h(0) = \infty$ and $h(1) = 0$. The barrier potentials are given by

$$B_s = \sum B_{is}, \tag{1.12}$$

$$B_{is} = (\bar{\psi}\psi + \bar{\phi}\phi)\chi_i h(s_i). \tag{1.13}$$

v_s is v interpolated by the same definition as in Ref. 2. This makes sense since $\exp(-nr)/r$ has a Fourier transform with the same form as a covariance. Thus we have

$$V_s^\Lambda(t) = \tfrac{1}{2} : (\bar{\psi}\psi - \bar{\phi}\phi) v_s^\Lambda (\bar{\psi}\psi - \bar{\phi}\phi) : . \tag{1.14}$$

Time dependences are introduced as in Ref. 1. Finally we have the interpolating Hamiltonians:

$$H_s(t) = H_0(t) + V_s(t) + B_s(t). \tag{1.15}$$

Note $H_s(t) = H(t)$ if, for all i, $s_i = 1$.

We are now set to define the cluster expansion. For any union of cubes X define

$$X^c = \Lambda - X \tag{1.16}$$

and Z_W by

$$Z_W = \mathrm{Tr}_W[\exp(-\beta H^W)]. \tag{1.17}$$

Our substitute for (3.15) of Ref. 2 is the following:

$$\mathrm{Tr}_\Lambda[T \exp(-\textstyle\int_0^\beta H^\Lambda(t)\,dt)A]/Z_\Lambda$$

$$= \sum_{\substack{X \text{ i. c.} \\ \Delta \subset X \\ \Gamma \subset X}} \int \partial^\Gamma \mathrm{Tr}_X[T \exp(-\textstyle\int_0^\beta H_s^X(t)\,dt)A]\,ds(\Gamma) \cdot Z_X c / Z. \tag{1.18}$$

The notation i. c. means that X must be intimately connected. Γ, as in Ref. 2, is a subset of the faces in the interior of X, such that the faces not in Γ but in the interior of X do not separate the interior of X. The result we claim is that for fixed β, m, M, and $\gamma - \alpha$, if μ and α are large enough and g^2 and δ are small enough then (1.18) converges uniformly in the volume Λ. From this result follows analogs of all the results in Ref. 2.

2. UNIFORM STABILITY OF THE POTENTIALS

The two-body potential $v(x, y)$ is said to be stable if for any set of N points $\{x_i\}$

$$\sum_{i<j}^{N} v(x_i, x_j) \geq -CN \tag{2.1}$$

for some constant C independent of N and the $\{x_i\}$. Applying this definition to our problem, we claim there is a constant L such that

$$\tfrac{1}{2} \int \int_W dx\,dy : (\bar{\psi}\psi - \bar{\phi}\phi)(x) v_s^W(x, y) (\bar{\psi}\psi - \bar{\phi}\phi)(y) :$$

$$\geq -Lq^2 \int_W dx (\bar{\psi}\psi + \bar{\phi}\phi)(x). \tag{2.2}$$

We say the v_s^W are uniformly stable. L is independent of W, q, and $\{s_i\}$. In fact, L can be picked equal to $\frac{1}{2}(\gamma - \alpha)$. Without loss of generality we set $q = 1$ in the remainder of this section.

For orientation we consider the free case, $W = R^3$ and $s_i = 1$, all i. Then $v = \exp(-\alpha|x-y|)/|x-y| - \exp(-\gamma|x-y|)/|x-y|$. There follows

$$\tfrac{1}{2} \iint dx\,dy : (\bar{\psi}\psi - \bar{\phi}\phi) v (\bar{\psi}\psi - \bar{\phi}\phi) :$$

$$= \tfrac{1}{2} \iint dx\,dy (\bar{\psi}\psi - \bar{\phi}\phi) \int \frac{dk}{2\pi^2} \exp(ik \cdot x) \exp(-ik \cdot y)$$

$$\times \left(\frac{1}{k^2 + \alpha^2} - \frac{1}{k^2 + \gamma^2} \right) (\bar{\psi}\psi - \bar{\phi}\phi)$$

$$- \tfrac{1}{2} \int dx (\bar{\psi}\psi + \bar{\phi}\phi)(x)(\gamma - \alpha)$$

$$\geq -\tfrac{1}{2}(\gamma - \alpha) \int dx (\bar{\psi}\psi + \bar{\phi}\phi)(x). \tag{2.3}$$

We have used the fact that a positively weighted integral of operators times their conjugates is positive.

In the general case v_s^W is constructed as a positively weighted sum, with total weight one, of $\{v_{\tilde{s}}^W\}$, in which the values of the s_i are restricted to one and zero. It is sufficient to prove the result for a single $v_{\tilde{s}}^W$. Say the s_i in $\tilde{s}$ equal to zero are those for which $i \in I$. Let $\{\phi_j\}$ be the set of eigenvectors of Δ in W satisfying Dirichlet data on ∂W and F_i for $i \in I$:

$$\Delta\phi_j = -\lambda_j \phi_j,$$

$$\phi_j = 0 \quad \text{on} \quad \partial W \cup \{F_i \mid i \in I\}. \tag{2.4}$$

Assume the ϕ_j are normalized:

$$\int_W \phi_j^2 = 1. \tag{2.5}$$

Then

$$v_{\tilde{s}}^W = 4\pi \sum_j \phi_j \phi_j \left(\frac{1}{\lambda_j + \alpha^2} - \frac{1}{\lambda_j + \gamma^2} \right) \tag{2.6}$$

and

$$\tfrac{1}{2} \int \int_W dx\,dy : (\bar{\psi}\psi - \bar{\phi}\phi) v_{\tilde{s}}^W (\bar{\psi}\psi - \bar{\phi}\phi) :$$

$$\geq -\tfrac{1}{2} \int_W dx (\bar{\psi}\psi + \bar{\phi}\phi) v_{\tilde{s}}^W(x, x) \tag{2.7}$$

as in (2.3). There follows

$$\geq -\tfrac{1}{2} \int dx (\bar{\psi}\psi + \bar{\phi}\phi) \cdot \sup_x v_{\tilde{s}}^W(x, x). \tag{2.8}$$

We are reduced to estimating $v_{\tilde{s}}^W(x, x)$:

$$v_{\tilde{s}}^W = \frac{4\pi}{-\Delta + \alpha^2} - \frac{4\pi}{-\Delta + \gamma^2} \tag{2.9}$$

for the Laplacian satisfying the data in (2.4):

$$= 4\pi \int_0^\infty [\exp(-\alpha^2 t) - \exp(-\beta^2 t)] \exp(\Delta t)\,dt \tag{2.10}$$

So

$$v_{\tilde{s}}^W(x, x) = 4\pi \int_0^\infty [\exp(-\alpha^2 t) - \exp(-\beta^2 t)] \int d\mu_{x,x}^t\,dt,$$

$$\text{paths } (W, s), \tag{2.11}$$

where $d\mu^t_{x,x}$ is the path space measure constructed for mass 1/2 and the paths summed over lie in W and avoid ∂W and $\cup_{i\in I} F_i$. Clearly this is $\leq$ the value of the sum over all paths. So

$$v^W_\beta(x, x) \leq \gamma - \alpha. \tag{2.12}$$

3. A GRAND CANONICAL ESTIMATE

We define local number operators N_i:

$$N_i = \int_{\Delta_i} (\bar\psi\psi + \bar\phi\phi)\, dx. \tag{3.1}$$

For a set of integers $\{\alpha_i\}$ we want to estimate

$$G(\alpha) = \mathrm{Tr}_W[\exp(-\beta H_0^W/2)\prod_i N_i^{\alpha_i} e^N]. \tag{3.2}$$

We claim if $|W|$ is the volume of W, and if μ is large enough, there is a c_1 such that

$$G(\alpha) \leq \left(\prod_i \alpha_i^{\alpha_i}\right) \cdot \exp(c_1 |W|). \tag{3.3}$$

Letting $D(z, V, T)$ be the grand canonical ensemble partition function, we see

$$\begin{aligned} D &= \sum_0^\infty z^N Q_N(V, T) \\ &= (ze^2)^N Q_N(V, T)\exp(-N)\exp(-N) \\ &= (ze^2)^N Q_N(V, T)\left(\prod_i \exp(-N_i)\right)\exp(-N). \end{aligned} \tag{3.4}$$

(3.3) follows from

$$N_i^{\alpha_i}\exp(-N_i) \leq \alpha_i^{\alpha_i} \tag{3.5}$$

and $D(ze^2, V, T) \leq \exp(c_1 V)$ for z small enough.

4. CONVERGENCE

Convergence is achieved basically the same way as in Sec. 10 of Ref. 2. In (10.1) of Ref. 2 one must estimate

$$\int \partial^\Gamma \Pi_i^\eta \phi(x_i)\exp[-\lambda V(\Lambda)]\, d\phi_{s(\Gamma)}\, ds(\Gamma); \tag{4.1}$$

we have been led to consider

$$\int \partial^\Gamma \mathrm{Tr}_X[TA\exp(-\textstyle\int_0^\beta H_s^X(t)\,dt)]\, ds(\Gamma). \tag{4.2}$$

To compare the two expressions note:

(a) The Λ in (4.1) is an arbitrary volume, and the use of X instead of Λ would have been clearer.

(b) Our A abbreviates $\Pi_i^\eta\phi(x_i)$ (rather an essentially equivalent expression); in fact, for the sake of convergence we can pick $A = 1$ as this factor enters unessentially.

(c) The trace substitutes for the Euclidean space integration $\int d\phi_{s(\Gamma)}$.

The differentiations ∂^Γ are performed in (4.2) using (2.2) of Ref. 1. Thereafter, a polynomial in the fields is downstairs in the trace. We now use (2.3) of [1] to move the fields according to the following steps.

Step 1: One at a time move each annihilation field ϕ, ψ to the right [using (3.2) of Ref. 1 to re-enter at left] until either it contracts with a creation field, or moves without contracting from $t = \beta$ to $t = \beta/2$, in which case stop the field at $t = \beta/2$.

Step 2: One at a time move each creation field $\bar\phi$, $\bar\psi$ to the left until either it contracts (with one of the annihilation fields at $t = \beta/2$), or else having passed $t = 0$ just once, stop the field at $t = \beta/2$, to the left of all annihilation fields at $t = \beta/2$.

At the end of this finite, noninductive, process all the fields downstairs in the trace are at $t = \beta/2$, with creation fields all to the left of annihilation fields. With a field that has moved from (x_1, t_1) to (x_2, t_2) and then contracted (for simplicity without passing $t = 0$) there is associated

$$\int d\mu^{(t_2 - t_1)}_{x_1, x_2}. \tag{4.3}$$

With a field $\bar\phi$ that has moved from x_1, t_1 left to reentry at $t = 0$, and then brought to rest at $t = \beta/2$ at $(x_2, \beta/2)$ there is associated

$$\int dw \int d\mu^{(\beta - t_1)}_{x_1, w} \int d\mu^{\beta/2}_{w, x_2}\, \bar\phi(x_2, \beta/2). \tag{4.4}$$

These are representative of all possibilities. As in Ref. 2, we localize—but only the space variables. Thus for (4.3) there are two localization indices (j_1, j_2). That is $x_1 \in \Delta_{j_1}$, $x_2 \in \Delta_{j_2}$. For (4.4) there are three localization indices (j_1, j_2, j_3), with $x_1 \in \Delta_{j_1}$, $w \in \Delta_{j_2}$, $x_2 \in \Delta_{j_3}$.

There are now sums (over j's, partitions of Γ, and contractions) and integrals (over s's, x's, t's, and path space), the portion of the integrand we now consider is of the form

$$\mathrm{Tr}_W\left(E_1 \overset{s}{\Pi}\bar\psi(x_i)\overset{s}{\Pi}\psi(y_i)E_2\right). \tag{4.5}$$

Here E_1 and E_2 are time-ordered exponentials from $\beta/2$ to β and from 0 to $\beta/2$ respectively. The $\bar\psi$'s (ψ's) stand for $\bar\psi$'s and $\bar\phi$'s (ψ's and ϕ's). In doing estimates we will take absolute values of the integrands, and use

$$\begin{aligned} &\left|\mathrm{Tr}_W\left(E_1 \overset{s}{\Pi}\bar\psi(x_i)\overset{s}{\Pi}\psi(y_i)E_2\right)\right| \\ &\leq \tfrac{1}{2}\mathrm{Tr}_W\left(E_1\overset{s}{\Pi}\bar\psi\psi(x_i)E_1^*\right) + \tfrac{1}{2}\mathrm{Tr}_W\left(E_2^*\overset{s}{\Pi}\bar\psi\psi(y_i)E_2\right). \end{aligned} \tag{4.6}$$

Of these terms, that can be estimated alike, we consider the first term. All the $\bar\psi(x_i)$ in this term have associated to them, by (4.4), a path space integral

$$\int d\mu^{\beta/2}_{w_i, x_i}. \tag{4.7}$$

From the integrals to be performed we isolate the following portion of present interest:

$$\mathrm{Tr}_W\left[E_1 \overset{s}{\underset{i=1}{\Pi}}\left(\int_{\Delta_{j_{3_i}}} dx_i \int d\mu^{\beta/2}_{w_i, x_i}\bar\psi\psi(x_i)\right)E_1^*\right]. \tag{4.8}$$

We have used the key fact, largely motivating our development to now, that E_1 does not depend on the integrals in (4.7). There are c_2 and c_3 such that

$$\sup_{\substack{w_i\in\Delta_{j_{2_i}}\\ x_i\in\Delta_{j_{3_i}}}} \int d\mu^{\beta/2}_{w_i, x_i} \leq c_2\exp[-c_3 d(\Delta_{j_{2_i}}, \Delta_{j_{3_i}})]. \tag{4.9}$$

(Exponential falloffs are good enough for us; we sacrifice the actual Gaussian falloff.) Using (4.7) from Ref. 1 and (3.3), we find that (4.8) is less than

$$c_2^s \exp\left(-c_3 \sum_i d(\Delta_{j_{2_i}}, \Delta_{j_{3_i}})\right) \exp(c_1 |W|) \prod_i \alpha_i^{\alpha_i} \exp(c_5 |\Gamma|), \tag{4.10}$$

where $\Pi_i \alpha_i^{\alpha_i}$ is the product over squares Δ_i, and α_i is the number of j_3's in Δ_i. By changing the free parameters as stated at the end of Sec. 1, c_2 and c_5 can be made arbitrarily small and c_3 arbitrarily large. c_1 is not chosen to vary. The factor e^N in (3.2) and $\exp(c_5|\Gamma|)$ arise as estimates of the exponent

$$H(t) \geq \tfrac{1}{2}H_0 - (N + 8|\Gamma|)\tfrac{1}{2}(\gamma - \alpha)g^2 \tag{4.11}$$

using the uniform stability of the interaction. $8|\Gamma|$ is an estimate of the maximum number of "particles" added to the exponent by (2.3) of Ref. 1. We have kept $\frac{1}{2}H_0$ in (4.11) rather than H_0 to anticipate a development for more general interactions.

Prop. 5.1, Prop. 5.2, and Prop. 8.1 (for v_s^W) are the same here as in Ref. 2. The completion of estimates for convergence are parallel to those in Ref. 2, Sec. 10. There is a mild novelty in the treatment of the barrier potentials.

ACKNOWLEDGMENT

I would like to thank Professor J. Rauch for the idea of interpolating Dirichlet data by barrier potentials and Professor B. Simon for forcing me to read about the cluster expansion.

*This work was supported in part by NSF Grant GP-17523.

[1] P. Federbush, J. Math. Phys. **16**, xxxx (1975).

[2] J. Glimm, A. Jaffe, and T. Spencer, "The Cluster Expansion," in *Constructive Quantum Field Theory, The* 1973 *"Ettore Majorana" International School of Mathematical Physics*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973).

[3] J. Ginibre, "Some Applications of Functional Integration in Statistical Mechanics," in *Statistical Mechanics and Field Theory, Les Houches* 1970, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971).

Stability of nonlinear parametric-decay interactions in finite homogeneous plasma

V. Fuchs and G. Beaudry

Direction Sciences de Base, Institut de Recherche de l'Hydro-Québec, Varennes, Québec, Canada
(Received 3 July 1975)

We prove the existence and uniqueness of a stable steady state of stimulated backscattering from a bounded homogeneous lossless interaction region, with boundary conditions corresponding to a steady incident pump wave-mode Poynting flux and zero-flux input to the decay wave modes. In steady state, once the excitation of the interaction region exceeds a certain critical value, the boundary value problem is characterized by a finite number of eigenvalues, and associated nontrivial eigenfunctions equilibria of the system, corresponding to mutually distinct states of anomalous reflection of the pump wave. A stability analysis of these equilibria with respect to small phase and amplitude perturbations reveals that (i) in the vicinity of the nonfundamental equilibria the phase perturbations exhibit singularities, preventing phase locking from occurring, and (ii) in the vicinity of the fundamental equilibrium both the phase and amplitude perturbations asymptotically vanish. A WKBJ phase-integral stability condition is derived to show that growing normal modes of the amplitude-perturbation boundary-value problem cannot propagate in the potential formed by the field of the depleted, spatially inhomogeneous, pump of the fundamental equilibrium.

I. INTRODUCTION

We will study the process of stimulated backscattering of radiation from a bounded homogeneous nondissipative plasma, resonant for the parametric decay of an incident electromagnetic wave, $\mathcal{E}_1$, into a backward electromagnetic wave, $\mathcal{E}_2$, and a forward electrostatic plasma wave, $\mathcal{E}_3$. For coherent plane waves,

$$\mathcal{E}_\alpha = \mathbf{E}_\alpha(x,t)\exp(\omega_\alpha t - \mathbf{k}_\alpha \cdot \mathbf{n}_x x); \quad \alpha = 1,\ 2,\ 3, \tag{1}$$

interacting at resonance,

$$\omega_1 = \omega_2 + \omega_3, \quad \mathbf{k}_1 = \mathbf{k}_2 + \mathbf{k}_3, \tag{2}$$

the slow space-time modulations $E_\alpha(x,t)$ due to coupling of the electric field amplitudes and phases are described, to lowest nonlinear order, by the system of coupled equations,[1–4]

$$\frac{\partial E_1}{\partial t} + v_1 \frac{\partial E_1}{\partial x} = -M_1 E_2 E_3, \tag{3a}$$

$$\frac{\partial E_2^*}{\partial t} + v_2 \frac{\partial E_2^*}{\partial x} = M_2 E_1^* E_3, \tag{3b}$$

$$\frac{\partial E_3^*}{\partial t} + v_3 \frac{\partial E_3^*}{\partial x} = M_3 E_1^* E_2. \tag{3c}$$

Here, the positive x direction is oriented along $\mathbf{k}_1$, which is parallel to $\mathbf{k}_3$ but antiparallel to $\mathbf{k}_2$. Thus, for the wave group velocities, v_α, determined by the respective mode dispersion relations, we have $v_1 > 0$, $v_2 < 0$, $v_3 > 0$. The M_α are positive coupling coefficients. Their explicit form for stimulated Raman and Brillouin backscattering are presented, for example, in Refs. 5–6. The plasma is bounded between $x = 0$ and $x = L$, with free space elsewhere. At the boundaries we require E_α real,

$$E_1(0,t) = E_{10} > 0, \quad E_3(0,t) = 0, \tag{4}$$

and we specify

$$E_2(L,t) = 0, \tag{5}$$

in order to eliminate convective backward amplification of nonzero background thermal noise.[7]

The given two-point boundary value problem has been studied analytically as well as numerically by several authors, using various approximations. Linear theory is due to Bobroff and Haus[8] who, assuming a constant pump wave-amplitude, $E_1 = E_{10}$, predict absolute instability of the decay waves $E_{2,3}$ when the gain of the interaction region exceeds a certain threshold. Namely, when

$$E_{10} L (M_2 M_3 / |v_2 v_3|)^{1/2} \equiv \Gamma > \pi/2. \tag{6}$$

Manheimer[7] shows that this threshold is identical with that for steady reflection, as described by nonlinear theory.

The nonlinear case has been studied analytically in steady state by Andersson and Wilhelmsson,[5] and Andersson,[9] who shows that the boundary value problem leads to a countable spectrum of eigenvalues and associated steady eigenfunctions. The analysis is then continued to obtain a reflection coefficient, under the assumption that there is no way to single out one of the eigenfunctions as the unique steady-state solution. As a consequence, the reflection coefficient is then defined as a mean over a certain distribution of eigenvalues. The transient numerical analysis of Harvey and Schmidt[10] presents a different picture of the interaction. First, the solution evolves in time towards an aperiodic monotonic structure in space, independent of the initial conditions. Second, surprisingly enough, even when the amplitudes were placed initially into one of the available periodic steady eigenfunctions, they again evolved towards the aperiodic steady solution. Since amongst the steady eigenfunctions there is just one which is aperiodic, the authors concluded that it is just that which represents uniquely the steady state. Finally, to further support this conclusion, the equations were linearized around the eigenfunctions and noise was added. In the aperiodic case the noise subsided, whereas in the periodic one the perturbations grew.

The results of Harvey and Schmidt indicate that the actual, physically observable, state of anomalous reflection is represented by the fundamental, aperiodic, equilibrium. In the present study we therefore, propose to support this assertion analytically, by analyzing the stability of the spatial equilibria of system (3) with respect to phase and amplitude perturbations. The result is that the fundamental equilibrium is asymptotically stable, whereas the nonfundamental equilibria are unstable. In Sec. II, we describe the equilibria. In Sec. III, we clarify the concept of stability to be applied, and we show that phase perturbations are not coupled to the amplitude perturbations. The analysis of phase-perturbations, in Sec. IV, shows that in the vicinity of the nonfundamental equilibria the wave phases cannot lock. In Sec. V, we derive a WKBJ phase-integral stability condition to show that the amplitude-perturbations subside in the vicinity of the fundamental equilibrium.

II. STEADY STATE SOLUTIONS

In this section we present, for further reference, the steady state solutions of Eqs. (3). Writing $E_\alpha = A_\alpha \exp\Phi_\alpha$, with E_α and Φ_α real, Eqs. (3) become

$$\dot{A}_\alpha + v_\alpha A'_\alpha = \sigma_\alpha M_\alpha A_\beta A_\gamma \cos\Phi, \tag{7}$$

$$\dot{\Phi}_\alpha + v_\alpha \Phi'_\alpha = M_\alpha (A_\beta A_\gamma / A_\alpha) \sin\Phi, \tag{8}$$

$$(\alpha,\ \beta,\ \gamma) = (1,\ 2,\ 3)\ \text{cycl.},$$

where $(\ )^{\cdot} \equiv \partial/\partial t$, $(\ )' \equiv \partial/\partial x$, $\sigma_1 \equiv -1$, $\sigma_2 = \sigma_3 = 1$, and $\Phi = \Phi_2 + \Phi_3 - \Phi_1$. We now assume $(\ )^{\cdot} = 0$, rename A_α by S_α to distinguish between the transient and steady state amplitudes, and introduce the normalizations

$$u_\alpha = S_\alpha (M_\beta M_\gamma / |v_\beta v_\gamma|)^{1/2}. \tag{9}$$

We obtain, with $s_1 = s_2 = -1$ and $s_3 = 1$,

$$u'_\alpha = s_\alpha u_\beta u_\gamma \cos\Phi, \quad \Phi'_\alpha = -(u_\beta u_\gamma / u_\alpha) \sin\Phi. \tag{10}$$

The solution of Eqs. (10), subject to conditions (4) and (5), is

$$\Phi_\alpha = 0, \tag{11}$$

$$u_1 = u_{10}\,\mathrm{dn}(\xi, k), \quad u_2 = u_{20}\,\mathrm{cn}(\xi, k), \quad u_3 = u_{20}\,\mathrm{sn}(\xi, k), \tag{12}$$

with

$$\xi = u_{10} x, \quad k = u_{20}/u_{10}.$$

The modulus k of the Jacobian elliptic solutions is seen to be equal to the (transverse action) reflectivity of the slab, and is given by boundary condition (5), that is,

$$\mathrm{cn}(\Gamma, k) = 0, \quad \Gamma = u_{10} L. \tag{13}$$

Equation (13) is only satisfied when

$$\Gamma = (2n+1) K(k_n), \tag{14}$$

where $K(k_n)$, the quarter period of the Jacobian elliptic functions, is equal to the complete elliptic integral of the first kind. Since $K(k)$ increases monotonically from $\pi/2$ to ∞, as k runs from 0 to 1, we see that if $\Gamma < \pi/2$, then $k = 0$. If, on the other hand, $\Gamma > \pi/2$, there exist unique solutions k_n of Eq. (14) such that

$$1 > k_0 > k_1 > \cdots > k_{N-1} > 0, \tag{15}$$

where $2N - 1$ = highest odd integer part of $(2\Gamma/\pi)$.

The fundamental solution $u_\alpha^{(0)}$, associated with k_0, consists of one quarter period of the Jacobian functions.

Remark: The fundamental mode is the only eigenfunction which is defined and is continuous for all $\Gamma \in (0, \infty)$. We recall that for $\Gamma \in (0, \pi/2)$ the solution is trivial. For $\Gamma > \pi/2$, as $\Gamma \to \infty$ and $k \to 1$, the fundamental mode evolves continuously into

$$u_{1,2} \to u_{10}\,\mathrm{sech} u_{10} x, \quad u_3 \to u_{10} \tanh u_{10} x, \tag{16}$$

which is the unique solution of the boundary value problem for a semi-infinite plasma, $L \to \infty$.

III. STABILITY–BASIC CONSIDERATIONS

A number of stability concepts developed for an autonomous system of ordinary differential equations,

$$\dot{Y}_m = F_m(Y_1,\ Y_2, \ldots, Y_n), \quad m = 1, \ldots, n, \tag{17}$$

can be transferred directly to the (hyperbolic) system of first-order partial differential equations (7) and (8). Let us therefore recall[11,12] that the stability of system (17) is related to the stability of its singular points $\dot{Y}_m = 0$, given by $F_m = 0$.

In the following, we will analyze the stability of systems (7) and (8) in the spirit of Lyapunov's first method, that is, we study the linearized equations in the vicinity of the equilibria $\dot{A}_\alpha = \dot{\Phi}_\alpha = 0$. Let us first denote the equilibrium solutions of (7) and (8) by $Y_m^{(n)}$, $m = 1, \ldots, 6$, $n = 0, \ldots, N-1$, with N given by Eq. (15). The vector Y_m is defined as

$$Y_m = (A_\alpha, \Phi_\alpha), \quad m = 1, \ldots, 6, \quad \alpha = 1, 2, 3. \tag{18}$$

Correspondingly,

$$Y_m^{(n)} = (S_\alpha^{(n)}, 0), \tag{19}$$

the solutions being given by (11), and (12) complemented by (9). Then, in terms of y_α, a variable measured from equilibrium,

$$y_m = Y_m - Y_m^{(n)}, \quad m = 1, \ldots, 6, \tag{20}$$

the variational equations corresponding to (7) and (8), that is, to

$$\dot{Y}_m + v_m Y'_m = f_m(Y_1, \ldots, Y_6), \tag{21}$$

are (in matrix form),

$$\dot{\mathbf{y}} + \mathbf{B}\mathbf{y}' = \mathbf{J}^{(n)}\mathbf{y}. \tag{22}$$

Here, $\mathbf{J}^{(n)}$ is the Jacobian matrix,

$$J_{ij}^{(n)} = \left.\frac{\partial f_i}{\partial Y_j}\right|_{Y_j = Y_j^{(n)}}, \tag{23}$$

and

$$\mathbf{B} = \mathrm{diag}(\mathbf{V}, \mathbf{V}), \quad \mathbf{V} = \mathrm{diag}(v_1, v_2, v_3). \tag{24}$$

The matrix $\mathbf{J}$ is quasidiagonal,

$$\mathbf{J} = \mathrm{diag}(\mathbf{J}_a, \mathbf{J}_\phi), \tag{25}$$

where

$$\mathbf{J}_a = \begin{bmatrix} 0 & -M_1 S_3 & -M_1 S_2 \\ M_2 S_3 & 0 & M_3 S_1 \\ M_3 S_2 & M_3 S_1 & 0 \end{bmatrix} \tag{26}$$

and

$$\mathbf{J}_\phi = \begin{bmatrix} C_1 & -C_1 & -C_1 \\ C_2 & -C_2 & -C_2 \\ C_3 & -C_3 & -C_3 \end{bmatrix}, \tag{27}$$

with

$$C_\alpha = M_\alpha(S_\beta S_\gamma / S_\alpha), \quad \alpha,\ \beta,\ \gamma = (1,\ 2,\ 3)\ \text{cycl.} \tag{28}$$

We thus obtain the important result that in the vicinity of the equilibria the amplitude variations are not coupled to the phase variations. Namely, with

$$y_m = (a_\alpha, \phi_\alpha), \quad m = 1, \ldots, 6, \quad \alpha = 1,\ 2,\ 3, \tag{29}$$

Equation (22) becomes

$$\dot{\mathbf{a}} + \mathbf{V}\mathbf{a}' = \mathbf{J}_a \mathbf{a}, \tag{30a}$$

$$\dot{\phi} + \mathbf{V}\phi' = \mathbf{J}_\phi \phi, \tag{30b}$$

subject to the boundary conditions,

$$a_{1,3}(0,t) = a_2(L,t) = 0, \tag{31a}$$

$$\phi_{1,3}(0,t) = \phi_2(L,t) = 0, \tag{31b}$$

and to arbitrary, but not all identically vanishing, initial conditions consistent with (31), that is,

$$\lim_{x \to 0_+} s_{1,3}(x,0) = \lim_{x \to L_-} a_2(x,0) = 0, \tag{32a}$$

$$\lim_{x \to 0_+} \phi_{1,3}(x,0) = \lim_{x \to L_-} \phi_2(x,0) = 0. \tag{32b}$$

The variational Eqs. (30) represent a significant simplification with respect to the nonlinear system of Eqs. (7) and (8), solely because of the decoupling of the a_α from the ϕ_α. Otherwise, since $\mathbf{J}$ is a function of x, exact integration is still not possible. Furthermore, general stability criteria for hyperbolic systems, to the best of our knowledge, have not as yet been developed, except for the case of identical characteristics ($v_1 = v_2 = v_3$) treated by Zubov.[12] Consequently, we have to resort to approximations. In this respect, it is useful to realize that since the interaction region is finite, with no excitation at the boundaries and initial conditions identically zero outside $\langle 0, L\rangle$, asymptotic stability is guaranteed whenever the perturbations remain bounded at all times, that is when the systems (30) are not absolutely unstable, and do not possess singular solutions. Then, namely, since the systems (30) are linear hyperbolic, the response to an initial perturbation eventually propagates out of the interaction region.

IV. PHASE-LOCKING

In the vicinity of nonfundamental modes, $S_\alpha^{(n)}$, $n > 0$, the hyperbolic system (30b),

$$\dot{\phi}_\alpha + v_\alpha \phi_\alpha' = M_\alpha(S_\beta S_\gamma / S_\alpha)\phi \equiv -C_\alpha \phi, \tag{33}$$

$$\alpha,\ \beta,\ \gamma = (1,\ 2,\ 3)\ \text{cycl}, \quad \phi = \phi_2 + \phi_3 - \phi_1,$$

cannot have continuous bounded solutions[13] inside $x \in \langle 0, L\rangle$, since the coefficients C_α are discontinuous and singular inside the interaction region. The easiest way how to demonstrate the singularity of the solutions is to solve the Cauchy problem[13] in the absence of nondiagonal terms in $\mathbf{J}_\phi$. We obtain

$$\phi_\alpha(x,t) = g_\alpha(x - v_\alpha t)\, S_\alpha(x)/S_\alpha(x - v_\alpha t), \tag{34}$$

where

$$g_\alpha(x) = \begin{cases} \phi_\alpha(x,0) \not\equiv 0, & x \in \langle 0, L\rangle, \\ 0, & \text{elsewhere.} \end{cases} \tag{35}$$

We now consider the fundamental mode. We Laplace-transform (33), integrate with respect to x, and apply the boundary conditions (31b). We obtain

$$\bar{\phi}_{1,3}(x,p) = \frac{1}{v_{1,3}} \int_0^x \exp\left(-\frac{p}{v_{1,3}}(x - \xi)\right) \times [p g_{1,3}(\xi) - \bar{\phi}(\xi,p) C_{1,3}(\xi)]\, d\xi, \tag{36a}$$

$$\bar{\phi}_2(x,p) = \frac{1}{v_2} \int_x^L \exp\left(-\frac{p}{v_2}(x - \xi)\right) [p g_2(\xi) - \bar{\phi}(\xi,p) C_2(\xi)]\, d\xi. \tag{36b}$$

The functions $C_\alpha(x)$ are now definite in sign and bounded inside $(0, L)$, so that the integrals (36a) and (36b) are guaranteed to exist. The inverse Laplace-transforms of $\bar{\phi}_\alpha$ therefore also exist and we may interchange the order of integration. It follows that:

$$\phi_{1,3}(x,t) = g_{1,3}(x - v_{1,3} t) - \frac{1}{v_{1,3}} \times \int_0^x \phi\left(t - \frac{x - \xi}{v_{1,3}}, \xi\right) C_{1,3}(\xi)\, d\xi, \tag{37a}$$

$$\phi_2(x,t) = g_2(x - v_2 t) + \frac{1}{v_2} \times \int_x^L \phi\left(t - \frac{x - \xi}{v_2}, \xi\right) C_2(\xi)\, d\xi. \tag{37b}$$

On the other hand, we have for $\phi = \phi_2 + \phi_3 - \phi_1$, as follows from (33):

$$\dot{\phi} + Q(x)\phi = -v_2 \phi_2' - v_3 \phi_3' + v_1 \phi_1', \tag{38}$$

$$Q(x) = C_2 + C_3 - C_1 = \frac{u_1 u_2}{u_3} v_3 + \frac{u_3}{u_1 u_2} v_1 (u_1^2 - u_2^2), \tag{39}$$

with the u_α given by (12). For the fundamental mode, $Q(x)$ is positive definite on $\langle 0, L\rangle$. It follows that ϕ is not a growing mode, and neither are the ϕ_α as given by (37). Therefore, there exists a time asymptotic solution, and it is easily obtained as follows. Let $t \gg L/|v_\alpha|$. Then $g_\alpha(x - v_\alpha t) = 0$, since the argument falls into a region where originally $g_\alpha(x)$ vanishes identically, and the temporal argument of ϕ under the integral sign in (37) becomes just t. Thus,

$$\phi_\alpha^{(\text{as})\prime} = -\frac{1}{v_\alpha} \phi^{(\text{as})} C_\alpha(x). \tag{40}$$

The only solution of (40), satisfying the boundary conditions (31b) is,

$$\phi^{(\text{as})} S_1 S_2 S_3 = 0, \tag{41}$$

so that finally $\phi^{(\text{as})} = 0$. We conclude that in the vicinity of the fundamental mode the phases relock.

V. AMPLITUDE PERTURBATIONS

Since the instability of the nonfundamental equilibria already follows from the behavior of phase perturbations, it remains to be shown that the amplitude pertur-

bations are stable in the vicinity of the fundamental mode. In what follows, we establish a sufficient condition for the stability of system (30a), in terms of a WKBJ phase-integral stability criterion for the decay wave perturbations a_2 and a_3. This procedure enables us to ascertain that the nonlinearly depleted pump wave S_1, appearing in Eqs. (30a) as the background driving field of the decay wave-perturbations, cannot sustain growing modes.

Let us first note that a rigorous stability analysis of the system (30a),

$$\dot{a}_1 + v_1 a_1' = -M_1 S_3 a_2 - M_1 S_2 a_3, \tag{42a}$$

$$\dot{a}_2 + v_2 a_2' = M_2 S_3 a_1 + M_2 S_1 a_3, \tag{42b}$$

$$\dot{a}_3 + v_3 a_3' = M_2 S_2 a_1 + M_3 S_1 a_2, \tag{42c}$$

cannot be carried out in closed analytical form. A sufficient stability condition can, however, be obtained if we notice that, due to the absence of diagonal terms in J_a, stability is guaranteed if pairwise coupling of the equations is stable. Let us consider, for example, the coupling of the decay wave-amplitude perturbations a_2 and a_3. From the point of view of the respective Eqs. (42b) and (42c), a_1 appears as a source term, which, as is clear from (42a), is not a growing mode if a_2 and a_3 are not growing modes. Thus, in the first iteration, if we establish stability of the system (42b) and (42c) in the absence of the source, subsequent iterations will not lead to growing modes. Consequently, a sufficient and necessary condition for the stability of the system,

$$\dot{y}_2 + v_2 y_2' = M_2 A_{10} f(x) y_3, \tag{43a}$$

$$\dot{y}_3 + v_3 y_3' = M_3 A_{10} f(x) y_2, \tag{43b}$$

$$y_2(L,t) = y_3(0,t) = 0, \tag{44a}$$

$$f(x) = \mathrm{dn}(u_{10}x, k), \quad K(k) = u_{10}L, \tag{44b}$$

is a sufficient condition for the stability of the system (42).

To prove that the system (43) admits as solutions only temporally decaying modes, we first have to find the dispersion relation for the normal modes of the system. Let us therefore Laplace-transform (43), neglecting initial conditions, eliminate $\bar{y}_2$, and substitute

$$\bar{y}_3(x,p) = Z\sqrt{f}\exp\left[-\frac{p}{2}\left(\frac{1}{v_2}+\frac{1}{v_3}\right)x\right]. \tag{45}$$

We obtain

$$Z'' + \kappa_3 Z = 0, \tag{46}$$

$$\kappa_3(x,p) = u_{10}^2 f^2 + \tfrac{1}{2}(\ln f)'' - \tfrac{1}{4}(f'/f + \alpha p)^2,$$

with

$$u_{10}^2 = -A_{10}^2 M_2 M_3 / v_2 v_3, \quad \alpha = 1/v_3 - 1/v_2.$$

We now subject the WKBJ solutions of (46) to the boundary conditions (44). The requirement $\mathrm{Re}\, p < 0$, then entails a stability threshold condition. To simplify the analysis, we will look for solutions with p real, in the vicinity of $p = 0$. The construction of the WKBJ solutions of (46) in the vicinity of $p = 0$ depends on the degree, k, of pump depletion, as given by (44b). This is seen as follows: The function $\kappa = \kappa_3(x, 0)$ is

$$\kappa = \tfrac{1}{4}u_{10}^2(3\,\mathrm{dn}^2 - 1 - k'^2 + 3k'/\mathrm{dn}^2) \equiv \tfrac{1}{4}u_{10}^2\psi, \tag{47}$$

with $k' = (1-k^2)^{1/2}$ the complementary modulus. The function ψ, defined on $\xi \in \langle 0, K\rangle$ and $k' \in (0,1\rangle$, has the following properties:

(1) ψ is even with respect to $\xi = K/2$. For $k' = 1$, $\psi(\xi, k') = 4$. When $k' < 1$, ψ has a minimum at $\xi = K/2$, equal to $\psi_m = 6k' - 1 - k'^2$. Thus:

(2) When $k' > k_c' = 3 - \sqrt{8}$, then $\psi > 0$. When $k' < k_c'$, $\psi(\xi)$ has two zeros, which we denote by $\psi(a) = \psi(b) = 0$, $a < b$. Clearly, $a + b = K$. The function $\psi(\xi)$ is shown in Fig. 1.

Therefore,

(i) when $k' > k_c'$, the boundary conditions can be directly applied to the WKBJ solution[14]

$$Z = \kappa_3^{1/4}[A\exp(i\Omega_p) + B\exp(-i\Omega_p)], \tag{48}$$

where

$$\Omega_p(0,x) = \int_0^x \kappa_3^{1/2}(x_1, p)\,dx_1. \tag{49}$$

(ii) when pump depletion is strong, $k' < k_c'$, Eq. (46) has two turning points and we must apply connection techniques.

Let us first consider case (i). Applying the boundary conditions (44a) to (45), with Z given by (48), and making use of the fact that $f'(L) = 0$, we obtain the dispersion relation

$$\tfrac{1}{2}[p\alpha - \tfrac{1}{2}(\ln\kappa_3)_L']\sin\Omega_p(0,L) + (\kappa_3^{1/2})_L\cos\Omega_p(0,L) = 0, \tag{50}$$

with $\Omega_p(0,L)$ given by (49). More explicitly,

$$(q/2)R(q)\sin\Omega_q(0,L) + S(q)\cos\Omega_q(0,L) = 0, \tag{51}$$

with $q = \alpha p$ and

$$R(q) = u_{10}^2 - u_{10}^2 k^2/4 - q^2/4, \tag{52a}$$

$$S(q) = (u_{10}^2 - u_{10}^2 k^2/2 - q^2/4)^{3/2}. \tag{52b}$$

It follows from (51) that when $q = 0$, then

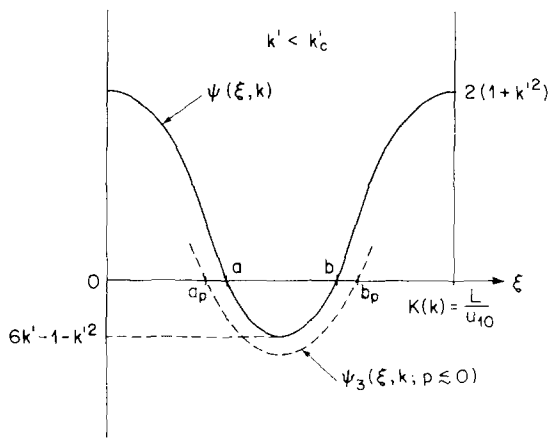

FIG. 1. The null-growth ($p = 0$) potential $\psi(\xi, k)$, as given by Eq. (47), for the propagation of the decay wave-amplitude perturbations in the field of the spatially inhomogeneous pump-wave Eq. (44b). For weaker reflection (or pump depletion), specifically when $k' > k_c' = 3 - \sqrt{8}$, ψ is positive-definite. A small negative growth rate shifts the potential downwards.

$$\Omega(0,L)\equiv\Omega_{q=0}(0,L)=(2n-1)\pi/2,\quad n=1,\ 2,\ \ldots\ .$$

Let us first establish bounds for $\Omega(0,L)$, to show that the threshold excitation is just $\pi/2$. The function ψ, defined by (47), satisfies the inequality

$$\psi^2/4\leq\psi\leq 2\,\mathrm{dn}^2+2k'^2/dn^2,$$

so that further,

$$\psi/2\leq\psi^{1/2}\leq\sqrt{2}(\mathrm{dn}+k'/\mathrm{dn}).$$

Upon integration over $(0,L)$, we obtain

$$[6E(k)-(1+k'^2)K(k)]/4\leq\Omega(0,L)<\pi/\sqrt{2},$$

where $E(k)$ is the complete elliptic integral of the second kind. Thus,

$$0.74<\Omega(0,L)<\pi/\sqrt{2} \tag{53}$$

for all $k'>k'_c$. We note that the bounds (53) for $\Omega(0,L)$ are established very conservatively, since they are derived from inequalities valid locally. The phase-integral $\Omega(0,L)$ thus lies in the neighborhood of $\pi/2$ and, due to the continuity of the functions $S(q)$, $R(q)$ and Ω_q, the only acceptable solutions, q_r, of Eq. (51), are those which lie in the vicinity of $q=0$, corresponding to values of $\Omega(0,L)$ in the vicinity of $\pi/2$. Now, we obtain the stability threshold condition as follows. Writing

$$\Omega_q=\Omega+(d\Omega_q/dq)q\equiv\Omega+Dq,$$

we define an implicit function of q and Ω, given by (51). Thus,

$$\left.\frac{dq}{d\Omega}\right|_{q=0,\ \Omega=\pi/2}=\frac{S(0)}{\frac{1}{2}R(0)-DS(0)}\,.$$

This expression can be shown to be always positive (Appendix A), so that finally $q<0$ if and only if

$$\Omega(0,L)<\pi/2. \tag{54}$$

We now turn to the case (ii) $k'<k'_c$. Equation (46) now has two turning points, which we denote a_q and b_q, corresponding, respectively, to the zeros $\psi(a)$ and $\psi(b)$, as shown on Fig. 1. The regions $\mathrm{I}\equiv\langle 0,a_b)$ and $\mathrm{III}\equiv(b_q,L\rangle$ are "wells" and $\mathrm{II}\equiv(a_q,b_q)$ is a barrier. Proceeding by standard techniques,[14] we join together the WKBJ approximations in regions I and III through a_q and b_q, and apply the boundary conditions (44a). We obtain, in analogy with Eq. (51),

$$\tfrac{1}{2}qR(q)(D_1\sin\gamma-D_2)+D_1S(q)\cos\gamma=0, \tag{55}$$

where

$$\gamma/2=\Omega_q(0,a_q)=\Omega_q(b_q,L), \tag{56}$$

$$D_{1,2}=\exp(-s)/4\pm\exp s, \tag{57}$$

$$s=\int_{a_q}^{b_q}(-\kappa_3)^{1/2}dx. \tag{58}$$

Since $D_1>0$ and $D_2<0$, it is easy to show, similarly as in case (i), that $q<0$ if and only if

$$2\Omega(0,a)<\pi/2. \tag{59}$$

This concludes the first part of the stability analysis.

It remains to be shown that conditions (54) and (59) are indeed satisfied. Since the condition (59) is more easily satisfied than is (54), it suffices to prove the latter, that is,

$$\Omega(k)=\tfrac{1}{2}\int_0^{K(k)}\psi^{1/2}(\xi,k)\,d\xi\leq\pi/2, \tag{60}$$

with ψ given by (47). To start with, when $k=0$, then $\psi=4$, and $\Omega(k)=K(0)=\pi/2$. This is seen to be the stability threshold condition (6) for the constant pump approximation. Now let $k>0$. Making use of Schwarz's inequality,

$$[\int_a^b fg\,dx]^2\leq[\int_a^b f^2dx]\,[\int_a^b g^2dx],$$

we obtain, with $f=\psi^{1/2}$ and $g=1$,

$$\Omega(k)<\tfrac{1}{2}[K(k)\int_0^{K(k)}k\,dx]^{1/2}\equiv I(k). \tag{61}$$

The integral $I(k)$, evaluated as

$$I(k)=\tfrac{1}{2}K^{1/2}[6E-(1+k'^2)K]^{1/2}, \tag{62}$$

decreases on the interval $k\in\langle 0,k_c)$, and $I(0)=\pi/2$. This concludes the analysis.

VI. CONCLUDING REMARKS

From a linear point of view, the asymptotic stability of the fundamental mode guarantees that the transient solution of the system will approach this equilibrium from any given initial state. This, unfortunately, is not necessarily true for a nonlinear system. However, inasmuch as we can rely on physical intuition, the initial conditions for the decay interaction should lie in the region of attraction of the fundamental mode. Namely, let $\Gamma>\pi/2$ with an initially uniform pump and zero decay wave-amplitudes. Since at these conditions the decay-waves are driven absolutely unstable, they will grow within the entire extent of the interaction region but not uniformly, since the boundary conditions define the electrostatic wave at the front edge of the slab, and the backscattered electromagnetic wave at the back edge of the slab. Thus, the decay waves will build up in the directions, respectively, of their group velocities. In turn, transverse action conservation requires the pump-wave to be depleted in the direction of its propagation. The transient response thus evolves towards the fundamental mode.

There is no doubt that the steady state is accessible in the case of weak excitation, when the fundamental mode lies in the vicinity of the initial conditions. However, the perturbation method applied in this study to analyze the stability in the small of the equilibria is inadequate to provide any information pertaining to the behavior of the system initially far from equilibrium. A more profound understanding of the interaction, based on Lyapunov's direct method, is therefore required.

ACKNOWLEDGMENTS

We wish to thank Dr. R.C. Neufeld, Dr. T.W. Johnston and Dr. J.P. Novak for valuable discussions and comments, and Dr. C. Richard for his continuing encouragement in the course of this work.

APPENDIX

We will prove that if $k'>k_c=3-\sqrt{8}$, then

$$\tfrac{1}{2}R(0)-DS(0)>0, \tag{A1}$$

where $R(0)$ and $S(0)$ are given by (52a) and (52b) respectively, and

$$D = \frac{d\Omega_q(0,L)}{dq}\bigg|_{q=0}. \tag{A1}$$

Let us first calculate D. We have, with Ω_q given by (49),

$$D = -\tfrac{1}{4}\int_0^L \kappa^{-1/2}(\ln f)'\,dx, \tag{A3}$$

and with κ given by (47),

$$D = \frac{1}{2u_{10}\sqrt{3}}\int_{k'}^{1}\left(f^4 - \frac{1+k'^2}{3}f^2 + k'^2\right)^{-1/2} df. \tag{A4}$$

When $(1+k'^2)/3 < 2k'$, which is just equivalent to the assumption $k' > k'_c$, the elliptic integral (A4) is equal to[15]

$$D = [(3k')^{-1/2}/2u_{10}][K(\bar{k}) - F(\alpha_1, \bar{k})], \tag{A5}$$

where F is the incomplete elliptic integral of the first kind, with

$$\alpha_1 = \arccos[(1-k')/(1+k')], \quad \bar{k}^2 = (6k' - 1 - k'^2)/12. \tag{A6}$$

We can now prove (A1). We first estimate[15] that when $k' > k'_c$, then $K(\bar{k}) - F(\alpha_1, \bar{k}) < 1$. Therefore,

$$D < (3k')^{-1/2}/2u_{10} \equiv H/u_{10}, \tag{A7}$$

and consequently

$$\tfrac{1}{2}R(0) - DS(0) > \tfrac{1}{2}R(0) - HS(0)/u_{10}. \tag{A8}$$

But

$$R(0) - 2HS(0)/u_{10} > u_{10}^2[1 - 2H(1 - k^2/2)^{1/2}] > 0, \tag{A9}$$

since the inequality $1 - 4H^2(1 - k^2/2) > 0$, is equivalent to the assumption $k' > k'_c$. This completes the proof.

[1] P. A. Sturrock, Ann. Phys. (N.Y.) **9**, 422 (1960).
[2] V. Tsytovich, *Non-linear Effects in Plasmas* (Plenum, New York, 1970), Chap. III.
[3] J. A. Armstrong, Sudhanshu S. Jha, and N. S. Shiren, IEEE J. Quantum Electron. **QE-6**, 123 (1970).
[4] T. J. M. Boyd and J. G. Turner, J. Phys. A **5**, 881 (1972).
[5] D. Andersson and H. Wilhelmsson, Nucl. Fusion **14**, 537 (1974).
[6] C. N. Lashmore-Davies, Plasma Phys. **17**, 281 (1975).
[7] W. N. Manheimer, Phys. Fluids **17**, 1634 (1974).
[8] D. L. Bobroff and H. A. Haus, J. Appl. Phys. **33**, 390 (1967).
[9] D. Andersson, J. Appl. Phys. **45**, 5482 (1974).
[10] R. Harvey and G. Schmidt, Phys. Fluids **18**, 1395 (1975), and Bull. Am. Phys. Soc. **19**, 882 (1974).
[11] K. Ogata, *State Space Analysis of Control Systems* (Prentice-Hall, Englewood Cliffs, N.J., 1967), Chap. 8.
[12] V. I. Zubov, *Methods of A.M. Lyapunov and Their Application* (Hafner, New York, 1964), Chap. V.
[13] P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964), Chaps. 1, 4.
[14] J. Heading, *Phase Integral Methods* (Wiley, New York, 1962), Chaps. IV, V.
[15] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, New York, 1971).

Gurtin-type properties associated with wave propagation in a superfluid

J. C. Murray

Mathematics Department, University of Southampton, Southampton, England SO9 5NH
(Received 2 July 1975)

Variational and reciprocity principles of the Gurtin type are established for the initial-boundary value problem associated with wave propagation in a superfluid.

1. INTRODUCTION

Variational and reciprocity principles of the Gurtin type have been obtained by several authors[1–4] for the linear initial-boundary value problems associated with the classical wave equation. These properties have also been derived for the initial-boundary value problems which arise in other areas of mathematical physics such as those in which the governing equation is the Schrödinger[5] or Klein–Gordon[3] equation. The results obtained for the classical wave equation in Refs. 1–4 are applicable to the acoustic problem for an inviscid gas. However, at low temperatures quantum effects become important and the gas exhibits superfluidity properties. It then becomes necessary to employ two equations to characterize the wave motion.

It is the purpose of this paper to establish variational and reciprocity principles for the linear initial-boundary value problem associated with superfluid acoustics.

2. FORMULATION OF INITIAL-BOUNDARY VALUE PROBLEM

The propagation of sound in a superfluid is governed by the equations[6]

$$\alpha\frac{\partial^2 p}{\partial t^2}-\Delta p-\beta\frac{\partial^2 T}{\partial t^2}=f(x,t),\quad (x,t)\in D\times(0,\infty) \tag{2.1}$$

and

$$\gamma\frac{\partial^2 T}{\partial t^2}-\Delta T-\mu\frac{\partial^2 p}{\partial t^2}=g(x,t),\quad (x,t)\in D\times(0,\infty), \tag{2.2}$$

where p and T represent the small changes in pressure and temperature from their constant equilibrium values. The coefficients α,β,γ,μ are the constant equilibrium quantities $(\partial\rho_0/\partial p_0)_{T0}$, $[(\rho_n)_0/(\rho_s)S_0^2](\partial S_0/\partial T_0)_{p_0}$, $-(\partial\rho_0/\partial T_0)_{p_0}$, $-[(\rho_n)_0/(\rho_s)_0S_0^2](\partial S_0/\partial p_0)_{T_0}$, where $\rho_0=\rho_n+\rho_s$ and $(\rho_n)_0,(\rho_s)_0$, S_0, T_0, p_0 are the equilibrium values of normal density, superfluid density, entropy, temperature, and pressure. The possible effects of sources are accounted for by the prescribed functions $f(x,t)$ and $g(x,t)$.

The quantities p and T are functions of position $x\equiv(x_1,x_2,\ldots,x_n)$ and time t defined on $\bar{D}\times[0,\infty)$, where D is a bounded n-dimensional domain in the Euclidean space R^n with closure $\bar{D}$ and a boundary ∂D which consists of the disjoint union of two $(n-1)$-dimensional domains ∂D_1 and ∂D_2. The outward normal to ∂D is denoted by ν and the n-dimensional gradient and Laplacian operators are denoted by ∇ and Δ, respectively. In what follows we will employ well-behaved real-valued functions defined on $\bar{D}\times[0,\infty)$, and we will assume that ∂D is sufficiently smooth to justify all mathematical operations throughout the analysis.

To complete the formulation of the initial-boundary value problem, we impose the following initial and boundary conditions on p and T:

$$p(x,0)=f_1(x),\quad T(x,0)=f_2(x),\quad x\in\bar{D}, \tag{2.3}$$

$$\frac{\partial p(x,0)}{\partial t}=g_1(x),\quad \frac{\partial T(x,0)}{\partial t}=g_2(x),\quad x\in\bar{D}, \tag{2.4}$$

$$p(x,t)=h_1(x,t),\quad (x,t)\in\partial D\times(0,\infty), \tag{2.5}$$

$$T(x,t)=h_2(x,t),\quad (x,t)\in\partial D_1\times(0,\infty), \tag{2.6}$$

$$\frac{\partial T}{\partial\nu}+kT=h_3(x,t),\quad (x,t)\in\partial D_2\times(0,\infty), \tag{2.7}$$

where f_i,g_i, $i=1,2$, h_j, $j=1,2,3$, are prescribed and k (≥ 0) is a prescribed function of x.

3. VARIATIONAL PRINCIPLES

It is not difficult to show that Eqs. (2.1)–(2.4) are equivalent to the pair of integro–differential equations

$$\alpha p-t^*\Delta p-\beta T=F,\quad (x,t)\in D\times(0,\infty), \tag{3.1}$$

$$\gamma T-t^*\Delta T-\mu p=G,\quad (x,t)\in D\times(0,\infty), \tag{3.2}$$

where the convolution of any two functions $v(x,t)$, $w(x,t)$ is defined in the usual way by

$$v^*w(x,t)=\int_0^t v(x,t_1)w(x,t-t_1)\,dt_1, \tag{3.3}$$

and

$$F(x,t)=t^*f(x,t)+\alpha[tg_1(x)+f_1(x)]-\beta[tg_2(x)+f_2(x)], \tag{3.4}$$

$$G(x,t)=t^*g(x,t)-\mu[tg_1(x)+f_1(x)]+\gamma[tg_2(x)+f_2(x)]. \tag{3.5}$$

We introduce the functional $\Phi_t(p,T)$ defined on some function space L by

$$\begin{aligned}\Phi_t(p,T)=&\int_D\mu[\tfrac{1}{2}(\alpha p^*p+t^*\nabla p^*\cdot\nabla p)-\beta p^*T-F^*p]\,d\tau\\&+\int_D\beta[\tfrac{1}{2}(\gamma T^*T+t^*\nabla T^*\cdot\nabla T)-G^*T]\,d\tau\\&-\int_{\partial D}\mu t^*(p-h_1)^*\frac{\partial p}{\partial\nu}\,d\sigma-\int_{\partial D_1}\beta t^*(T-h_2)^*\frac{\partial T}{\partial\nu}\,d\sigma\\&+\int_{\partial D_2}\tfrac{1}{2}\beta kt^*(T-2h_3)^*T\,d\sigma,\quad t\in[0,\infty),\end{aligned} \tag{3.6}$$

where, for any two functions $v(x,t)$ and $w(x,t)$,

$$\nabla v^* \cdot \nabla w \equiv \sum_{i=1}^{n} \frac{\partial v^*}{\partial x_i} \frac{\partial w}{\partial x_i}.$$

A variational principle can now be stated in the form

$$\delta\Phi_t(p,T)=0 \quad \text{on } L, \quad t\in[0,\infty), \tag{3.7}$$

for a particular pair of functions (p,T) if and only if (p,T) is a solution of the initial-boundary value problem (2.1)–(2.7).

Proof: By proceeding in the usual manner of the variational calculus and using the divergence theorem we find

$$\begin{aligned}\delta\Phi_t(p,T) &= \mu\int_D (\alpha p - t^*\Delta p - \beta T - F)^*\,\delta p\,d\tau \\ &+ \beta\int_D (\gamma T - t^*\Delta T - \mu p - G)^*\,\delta T\,d\tau \\ &- \mu\int_{\partial D} t^*(p-h_1)^*\frac{\partial}{\partial\nu}(\delta p)\,d\sigma \\ &- \beta\int_{\partial D_1} t^*(T-h_2)^*\frac{\partial}{\partial\nu}(\delta T)\,d\sigma \\ &+ \beta\int_{\partial D_2} t^*\left(\frac{\partial T}{\partial\nu}+kT-h_3\right)^*\delta T\,d\sigma, \quad t\in[0,\infty).\end{aligned} \tag{3.8}$$

If (p,T) satisfies the initial-boundary value problem, then (3.7) holds. Conversely, if (3.7) holds, then, setting $\delta p\not\equiv 0$ on $D\times[0,\infty)$, $\delta T\not\equiv 0$ on $D\times[0,\infty)$, $(\partial/\partial\nu)(\delta p)=0$ on $\partial D\times[0,\infty)$, $(\partial/\partial\nu)(\delta T)=0$ on $\partial D_1\times[0,\infty)$, and $\delta T=0$ on $\partial D_2\times[0,\infty)$, we obtain

$$\int_D (\alpha p - t^*\Delta p - \beta T - F)^*\,\delta p\,d\tau = 0, \quad t\in[0,\infty), \tag{3.9}$$

so that by the fundamental theorem of the variational calculus[1] Eq. (3.1) follows. Similarly Eq. (3.2) can be obtained by choosing $\delta T\not\equiv 0$ on $D\times[0,\infty)$, $(\partial/\partial\nu)(\delta p)=0$ on $\partial D\times[0,\infty)$, $(\partial/\partial\nu)(\delta T)=0$ on $\partial D_1\times[0,\infty)$, and $\delta T=0$ on $\partial D_2\times[0,\infty)$, and again appealing to the fundamental theorem of the variational calculus. Also by the appropriate choice of $(\partial/\partial\nu)(\delta p)$, $(\partial/\partial\nu)(\delta T)$ and δT on $\partial D\times[0,\infty)$ we can obtain the boundary conditions (2.5)–(2.7) and the proof is complete. We remark that all the boundary and initial conditions are natural for this variational principle.

It has been shown recently[4] that variational principles of the Gurtin type can be simplified by reducing the number of convolutions used in the construction of the appropriate functional. For the problem under investigation in this paper we can accomplish this by introducing the functional $\chi_t(p,T)$, where

$$t^*\chi_t \equiv \Phi_t. \tag{3.10}$$

Since $\delta\Phi_t=0$ if and only if $\delta\chi_t=0$, the variational principle (3.7) can be stated in the same form as in (3.7) with Φ_t replaced by χ_t.

To obtain χ_t, we differentiate (3.6) twice with respect to t and find

$$\begin{aligned}\chi_t(p,T) &= \int_D \mu\{\tfrac{1}{2}[\alpha\mathring{p}^*\mathring{p}+2\alpha p(x,0)\dot{p}(x,t)+\nabla p^*\circ\nabla p] \\ &-\beta[p(x,0)\mathring{T}(x,t)+T(x,0)\mathring{p}(x,t)-\dot{p}^*\mathring{T}] \\ &-F^*\dot{p}-p(x,0)\dot{F}(x,t)-F(x,0)\dot{p}(x,t)\}\,d\tau \\ &+\int_D \beta\{\tfrac{1}{2}[\gamma\mathring{T}^*\mathring{T}^*+2\gamma T(x,0)\dot{T}(x,t)+\nabla T^*\circ\nabla T] \\ &-\dot{G}^*\dot{T}-G(x,0)\dot{T}(x,t)-T(x,0)\dot{G}(x,t)\}\,d\tau \\ &-\int_{\partial D}\mu(p-h_1)^*\frac{\partial p}{\partial\nu}\,d\sigma-\int_{\partial D_1}\beta(T-h_2)^*\frac{\partial T}{\partial\nu}\,d\sigma \\ &+\int_{\partial D_2}\tfrac{1}{2}\beta T^*(T-2h_3)\,d\sigma, \quad t\in[0,\infty),\end{aligned} \tag{3.11}$$

where the dot denotes differentiation with respect to time.

4. RECIPROCITY PRINCIPLES

We can set $p=\Pi+q$ and $T=\theta+u$, where the functions Π and θ satisfy the equations

$$\alpha\frac{\partial^2\Pi}{\partial t^2}-\Delta\Pi-\beta\frac{\partial^2\theta}{\partial t^2}=0, \quad (x,t)\in D\times(0,\infty), \tag{4.1}$$

$$\gamma\frac{\partial^2\theta}{\partial t^2}-\Delta\theta-\mu\frac{\partial^2\Pi}{\partial t^2}=0, \quad (x,t)\in D\times(0,\infty), \tag{4.2}$$

together with the conditions (2.3)–(2.7) and the functions q and u satisfy the initial-boundary value problem

$$\alpha\frac{\partial^2 q}{\partial t^2}-\Delta q-\beta\frac{\partial^2 u}{\partial t^2}=f(x,t), \quad (x,t)\in D\times(0,\infty), \tag{4.3}$$

$$\gamma\frac{\partial^2 u}{\partial t^2}-\Delta u-\mu\frac{\partial^2 q}{\partial t^2}=g(x,t), \quad (x,t)\in D\times(0,\infty), \tag{4.4}$$

$$q(x,0)=u(x,0)=0, \quad x\in\bar{D}, \tag{4.5}$$

$$\frac{\partial q(x,0)}{\partial t}=\frac{\partial u(x,0)}{\partial t}=0, \quad x\in\bar{D}, \tag{4.6}$$

$$q(x,t)=0, \quad (x,t)\in\partial D\times(0,\infty), \tag{4.7}$$

$$u(x,t)=0, \quad (x,t)\in\partial D_1\times(0,\infty), \tag{4.8}$$

$$\frac{\partial u}{\partial\nu}+ku=0, \quad (x,t)\in\partial D_2\times(0,\infty). \tag{4.9}$$

The solution of the initial-boundary value problem for q and u is unique[7] so that if (q_i,u_i) is the solution of equations (4.3)–(4.9) associated with the source pair (f_i,g_i), $i=1,2$, then a reciprocity principle can be stated in the form

$$\int_D(\mu f_1^* q_2+\beta g_1^* u_2)\,d\tau=\int_D(\mu f_2^* q_1+\beta g_2^* u_1)\,d\tau. \tag{4.10}$$

Moreover, in the special cases $f_i\equiv 0$, $g_i\not\equiv 0$, $i=1,2$ or $f_i\not\equiv 0$, $g_i\equiv 0$, $i=1,2$, the reciprocity principle reduces to the form

$$\int_D f_1^* q_2\,d\tau=\int_D f_2^* q_1\,d\tau \tag{4.11}$$

or

$$\int_D g_1^* u_2\,d\tau=\int_D g_2^* u_1\,d\tau. \tag{4.12}$$

These latter statements of the reciprocity principle are similar to those associated with the classical wave equation.[2,3]

To prove (4.10), we use the pair of integro–differential equations equivalent to (4.3)–(4.6) which have the form

$$\alpha q - t^* \Delta q - \beta u = t^* f, \quad (x,t) \in (0,\infty), \tag{4.13}$$

and

$$\gamma u - t^* \Delta u - \mu q = t^* g, \quad (x,t) \in (0,\infty). \tag{4.14}$$

We have

$$\int_D (\mu \alpha q_1^* q_2 - \mu t^* \Delta q_1^* q_2 - \mu \beta u_1^* q_2)\, d\tau = \int_D \mu t^* f_1^* q_2\, d\tau, \tag{4.15}$$

$$\int_D (\mu \alpha q_2^* q_1 - \mu t^* \Delta q_2^* q_1 - \mu \beta u_2^* q_1)\, d\tau = \int_D \mu t^* f_2^* q_1\, d\tau. \tag{4.16}$$

Subtracting (4.15) from (4.16) and using the divergence theorem together with the boundary condition (4.7), we find

$$\int_D \mu\beta(u_1^* q_2 - u_2^* q_1)\, d\tau = \int_D \mu t^* (f_2^* q_1 - f_1^* q_2)\, d\tau. \tag{4.17}$$

Similarly, by using (4.14), we obtain

$$\int_D \mu\beta(u_2^* q_1 - u_1^* q_2)\, d\tau = \int_D \beta t^* (g_2^* u_1 - g_1^* u_2)\, d\tau. \tag{4.18}$$

By differentiating Eqs. (4.17) and (4.18) twice with respect to t and adding, we obtain (4.10), and the reciprocity principle is established.

[1] M.E. Gurtin, Quart. Appl. Math. **22**, 252 (1964).
[2] Ll. G. Chambers, Quart. Appl. Math. **28**, 391 (1970).
[3] J.C. Murray, J. Inst. Math. Appl. **10**, 305 (1972).
[4] I. Herrera and J. Bielak, Arch. Ratl. Mech. Anal. **53**, 131 (1974).
[5] M.E. Gurtin, J. Math. Phys. **6**, 1506 (1965).
[6] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, London, 1959)
[7] J.C. Murray, Phys. Lett. A (1975) (to appear).

A geometric formulation of the Taylor theorem for curves in affine manifolds*

Stanisław L. Bażański

Institute of Theoretical Physics, Warsaw University, Warsaw, Poland
(Received 14 July 1975)

The paper contains a theorem from differential geometry which exhibits the geometrical content of the classical Taylor expansion theorem applied to curves in a differentiable manifold. It also presents a review of a slightly modified version of the calculus of affine extensions, necessary for the proof of the theorem.

1. FORMULATION OF THE THEOREM

In certain applications of differential geometry to general relativity one meets Taylor expansions of the form

$$x^\alpha(\tau)=x^\alpha(\tau_0)+\left(\frac{dx^\alpha}{d\tau}\right)_0(\tau-\tau_0)+\frac{1}{2!}\left(\frac{d^2x^\alpha}{d\tau^2}\right)_0(\tau-\tau_0)^2+\cdots, \tag{1.1}$$

where $x^\alpha(\tau)$ $(\alpha=1,\ldots,n)$ are coordinates of a point $p=\gamma(\tau)$ on an analytic curve γ, parametrized with τ belonging to a certain interval $I\subset\mathbb{R}$, in an n-dimensional differentiable manifold $\mathcal{A}_n$. The subscript 0 in (1.1) denotes that the coefficients are evaluated at the point $q=\gamma(\tau_0)$, i.e., for $\tau=\tau_0$. A particular example of such a situation will be given in a subsequent paper on the concept of generalized geodesic deviations and their dynamical meaning in the theory of general relativity.

In general the coefficients of $(\tau-\tau_0)^k$ in (1.1), $(d^kx^\alpha/d\tau^k)_0$ for $\alpha=1,\ldots,n$, are not components of any geometric object at the point q. This is inconvenient for some applications of the expansion (1.1) despite the fact that all coefficients up to the kth taken together are components of an inhomogeneous geometric object—the k-jet $j^k_q\gamma$. The purpose of this paper is to reformulate the Taylor theorem in a form which exhibits its geometric content and removes the above mentioned inconvenience. Its essential result is expressed in the following theorem:

Theorem: Let $\gamma\colon\mathbb{R}\to\Omega\subset\mathcal{A}_n$ be an analytic curve in an open region Ω of an n-dimensional differentiable manifold $\mathcal{A}_n$ endowed with a symmetric affine connection, analytic in Ω. Then in the tangent space T_q at a point $q=\gamma(\tau_0)\in\Omega$ there exists a vector curve $l\colon\mathbb{R}\to T_q$ defined by

$$\tau\mapsto l(\tau)=l(q)+\frac{1}{2!}\left(\frac{\mathcal{J}l}{d\tau}\right)_q(\tau-\tau_0)+\cdots+\frac{1}{(k+1)!}\left(\frac{\mathcal{J}^kl}{d\tau^k}\right)_q(\tau-\tau_0)^k+\cdots \tag{1.2}$$

— where $(\mathcal{J}^kl/d\tau^k)_q\in T_q$ $(k=1,2,\cdots)$ is the absolute kth affine extension along γ at the point q of the vector l tangent to γ and the series (1.2) is convergent for $\tau,\tau_0\in I\subset\mathbb{R}$—such that the exponential map

$$\exp_q(\tau-\tau_0)l(\tau)=\gamma(\tau) \tag{1.3}$$

maps the vectors $(\tau-\tau_0)l(\tau)\in T_q$ into points $\gamma(\tau)$ of the curve γ for τ,τ_0 belonging to a certain interval $I'\subset I$.

Remark 1: The concept of the absolute affine extension of a geometric object was introduced to differential geometry by O. Veblen and T. Y. Thomas.[1,2] A short review of definitions and properties concerning this concept will be given in Sec. 2 of this article. The coefficients of the series (1.2) may also be defined by induction as vectors with the following components in the natural frame corresponding to coordinates x^α $(\alpha=1,\ldots,n)$:

For $k=1$: $\dfrac{\mathcal{J}t^\alpha}{d\tau}=\dfrac{Dt^\alpha}{d\tau}$

$(D/d\tau$—absolute derivative along $\gamma)$; for $k=m$:

$$\frac{\mathcal{J}^mt^\alpha}{d\tau^m}=\frac{Dt^\alpha}{d\tau^m}-\sum_{j=0}^{m-2}\sum_{i=0}^{m-j-i-2}\binom{m-j-1}{i}\times\frac{\mathcal{J}^{m-j-i-1}}{d\tau^{m-j-i-1}}(\Gamma^\alpha_{\rho\mu})\frac{\mathcal{J}^i}{d\tau^i}\left(\frac{D^jt^\rho}{d\tau^j}t^\mu\right), \tag{1.4}$$

where

$$\frac{\mathcal{J}^r\Gamma^\alpha_{\rho\mu}}{d\tau^r}=\sum_{\{i,j,h,\cdots,k\}}C^r_{ijh\cdots k}N^\alpha_{\rho\mu\alpha_1\cdots\alpha_i\beta_1\cdots\beta_j\cdots\kappa_1\cdots\kappa_k}t^{\alpha_1}\cdots t^{\alpha_i}\times\frac{\mathcal{J}t^{\beta_1}}{d\tau}\cdots\frac{\mathcal{J}t^{\beta_j}}{d\tau}\cdots\frac{\mathcal{J}^lt^{\kappa_1}}{d\tau^l}\cdots\frac{\mathcal{J}^lt^{\kappa_k}}{d\tau^l} \tag{1.5}$$

with

$$C^r_{ijh\cdots k}:=\left(\frac{1}{2!}\right)^j\left(\frac{1}{3!}\right)^h\cdots\left(\frac{1}{l!}\right)^k\frac{r!}{i!j!h!\cdots k!}$$

and $N^\alpha_{\rho\mu\alpha_1\cdots k_\kappa}$ being the components of normal tensors (cf. Ref. 3 and Sec. 2) which are known functions of the curvature tensor and its covariant derivatives. The summation above runs over all the sets $\{i,j,h,\ldots,k\}$ of nonnegative integer solutions of the equation

$$i+2j+3h+\cdots+lk=r.$$

From (1.4) and (1.5) a few first coefficients in (1.2) may be found as

$$\frac{\mathcal{J}t^\alpha}{d\tau}=\frac{Dt^\alpha}{d\tau}$$

$$\frac{\mathcal{J}^2t^\alpha}{d\tau^2}=\frac{D^2t^\alpha}{d\tau^2},$$

$$\frac{\mathcal{J}^3t^\alpha}{d\tau^3}=\frac{D^3t^\alpha}{d\tau^3}+R^\alpha{}_{\rho\mu\nu}t^\rho\frac{Dt^\mu}{d\tau}t^\nu,$$

$$\frac{\mathcal{J}^4t^\alpha}{d\tau^4}=\frac{D^4t^\alpha}{d\tau^4}+\left(\frac{7}{3}\frac{D^2t^\mu}{d\tau^2}t^\rho t^\nu+3\frac{Dt^\rho}{d\tau}\frac{Dt^\mu}{d\tau}t^\nu\right)R^\alpha{}_{\rho\mu\nu}+\tfrac{2}{3}$$

$$\times R^{\alpha}{}_{\rho\mu\nu;\sigma} t^{\rho} t^{\mu} \frac{Dt^{\nu}}{d\tau} t^{\sigma}. \tag{1.6}$$

Remark 2: Readers not familiar with the concept of the exponential map may treat (1.3) as a shorthand way of saying that the geodesic Γ_l which for $\tau = \tau_0$ passes through the point q and for which its tangent vector at q (i.e, for $\tau = \tau_0$) is equal to $l(\tau)$ will intersect the curve γ at the point p with coordinates $x^{\alpha}(\tau)$ for the value τ of the affine parameter along Γ_l. If $\mathcal{A}_n$ is additionally endowed with a Riemannian (or pseudo-Riemannian) structure, it means the geodesic distance between q and p along Γ_l is equal to $(g_{\alpha\beta} l^{\alpha} l^{\beta})^{1/2} (\tau - \tau_0)$.

The proof of the theorem will follow in Sec. 3.

2. THE AFFINE EXTENSION OF A GEOMETRIC OBJECT

In the proof of the theorem formulated above, certain properties of the notion of the affine extension, introduced to differential geometry some fifty years ago,[1,2] will occur. Nowadays, however, they are not widely known and will, for reference purposes, be here shortly reviewed, leaving aside the rather simple proofs which may be found in the literature.[3,4] The notion of the affine extension was originally introduced by means of a construction connected with a special coordinate system and it will be here referred to in the same way, but the whole construction and its results are coordinate independent. One may also execute the construction in a wholy coordinate independent manner, as will be pointed out in the concluding remarks of this section.

Let $\{x^{\alpha}\}$ be a certain coordinate map valid in a neighborhood $\Omega \subset \mathcal{A}_n$ of a point $q \in \mathcal{A}_n$ and $\Gamma^{\alpha}_{\beta\gamma}(x^{\mu})$ a field of components of the symmetric affine connection in the natural frame corresponding to this map. This field is assumed to be analytic in Ω.

As is known, an affine geodesic Γ in $\mathcal{A}_n$ is a curve described in the map $\{x^{\alpha}\}$, after an appropriate choice of its parameter τ, by the equations

$$\frac{d^2 \xi^{\alpha}}{d\tau^2} + \Gamma^{\alpha}_{\beta\gamma}(\xi^{\mu}) \frac{d\xi^{\beta}}{d\tau} \frac{d\xi^{\gamma}}{d\tau} = 0, \tag{2.1}$$

where

$$\xi^{\alpha}(\tau) := x^{\alpha} \circ \Gamma(\tau), \quad \Gamma^{\alpha}_{\beta\gamma}(\xi^{\mu}) := \Gamma^{\alpha}_{\beta\gamma}(\xi^1, \xi^2, \ldots, \xi^n);$$

all Greek indices range from 1 to n and summation is indicated by repeated indices.

Differentiating (2.1) $r-2$ times ($r = 2, 3, \cdots$) with respect to τ, we get

$$\frac{d^r \xi^{\alpha}}{d\tau^r} + \Gamma^{\alpha}_{\beta_1\beta_2\cdots\beta_r} \frac{d\xi^{\beta_1}}{d\tau} \frac{d\xi^{\beta_2}}{d\tau} \cdots \frac{d\xi^{\beta_r}}{d\tau} = 0, \tag{2.2}$$

where

$$\Gamma^{\alpha}_{\beta_1\beta_2\cdots\beta_r} := \Gamma^{\alpha}_{(\beta_1\beta_2\cdots\beta_{r-1},\beta_r)} - (r-1)\Gamma^{\sigma}_{(\beta_1\beta_2} \Gamma^{\alpha}_{\beta_3\cdots\beta_r)\sigma} \tag{2.3}$$

and thus for any integer $r > 2$ the field $\Gamma^{\alpha}_{\beta_1\beta_2\cdots\beta_r}$ is defined in Ω by induction. The comma represents partial differentiation with respect to x^{α} and the round brackets denote the complete symmetrization over all indices inside them.

Let us take a geodesic Γ passing through a point q with coordinates $\xi^{\alpha}(\tau_0) =: \xi^{\alpha}_0$ and let us consider another point p on Γ characterized by the value τ of the parameter. For $q, p \in \Omega$ the following Taylor expansion holds:

$$\xi^{\alpha}(\tau) = \xi^{\alpha}_0 + \left(\frac{d\xi^{\alpha}}{d\tau}\right)_0 (\tau - \tau_0) + \frac{1}{2!}\left(\frac{d^2\xi^{\alpha}}{d\tau^2}\right)_0 (\tau - \tau_0)^2 + \frac{1}{3!}\left(\frac{d^3\xi^{\alpha}}{d\tau^3}\right)_0 (\tau - \tau_0)^3 + \cdots. \tag{2.4}$$

The index 0 denotes that all derivatives of ξ^{α} are evaluated for $\tau = \tau_0$, i.e., at the point q. From (2.2) and (2.3) they can be expressed by means of the values of the fields $\Gamma^{\alpha}_{\beta_1\cdots\beta_r}$ at the point q and by means of the components $l^{\alpha} := (d\xi^{\alpha}/d\tau)_0$ of the vector l tangent to Γ at the point q. Therefore,

$$\xi^{\alpha}(\tau) = \xi^{\alpha}_0 + l^{\alpha}(\tau - \tau_0) - (1/2!)\Gamma^{\alpha}_{\beta_1\beta_2}(q) \times l^{\beta_1} l^{\beta_2} (\tau - \tau_0)^2 - (1/3!)\Gamma^{\alpha}{}_{\beta_1\beta_2\beta_3}(q) \times l^{\beta_1} l^{\beta_2} l^{\beta_3} (\tau - \tau_0)^3 - \cdots. \tag{2.5}$$

For given $\Gamma^{\alpha}_{\beta\gamma}$ in Ω all the coefficients in the series above are given. The series (2.5) determines then a solution of (2.1) as a function of τ in terms of the initial conditions ξ^{α}_0 and l^{α} and is convergent in a certain neighborhood U of q, where $U \subset \Omega$. It is a result of the analyticity of $\Gamma^{\alpha}_{\beta\gamma}$ in the region Ω and of the Cauchy theorem on analyticity of solutions of a set of ordinary differential equations.

If one introduces in U a new coordinate map $\{y^{\alpha}\}$ such that

$$x^{\alpha} - x^{\alpha}_0 = y^{\alpha} - (1/2!)\Gamma^{\alpha}_{\beta_1\beta_2}(q) y^{\beta_1} y^{\beta_2} - (1/3!)\Gamma^{\alpha}_{\beta_1\beta_2\beta_3}(q) y^{\beta_1} y^{\beta_2} y^{\beta_3} - \cdots, \tag{2.6}$$

Eq. (2.5) will, in this new map, take the form

$$\eta^{\alpha}(\tau) := y^{\alpha} \circ \Gamma(\tau) = (\tau - \tau_0) l^{\alpha}. \tag{2.7}$$

The coordinates y^{α} are called *normal coordinates* at the point q (or with the point q as their origin). They are valid in a neighborhood U of q. Generally, a normal coordinate system at q may be defined as one in which all geodesics passing through the point q are described (possibly after a suitable change of their parameter) by a linear equation (2.7). The formula (2.7) also implies that any point $p \in U$ may be connected with q by only one, geodesics, i.e., U is a star-shaped neighborhood of q.

Here is a list of basic properties of normal coordinates.

(i) The pair (U, y^{α}) belongs, under the assumptions of this section, to the analytic subatlas on $\mathcal{A}_n$.

(ii) A coordinate map $\{y^{\alpha'}\}$ will be normal at q if and only if it can be obtained from a normal coordinate map $\{y^{\alpha}\}$ at q by means of a nondegenerate linear transformation with coefficients being constant in U.

A normal coordinate map $\{y^{\alpha}\}$ is called *associated* with a map $\{x^{\alpha}\}$ at the point q if at q

$$y^{\alpha} = 0 \quad \text{and} \quad \left(\frac{\partial y^{\alpha}}{\partial x^{\beta}}\right)_q = \delta^{\alpha}_{\beta}. \tag{2.8}$$

Geometrically the coordinate lines of the normal coordinate system associated with a system $\{x^{\alpha}\}$ are

geodesics tangent at q to the coordinate lines of the system $\{x^\alpha\}$.

(iii) Let $\{y^\alpha\}$ and $\{y^{\alpha'}\}$ be the normal coordinate systems associated, respectively, with systems $\{x^\alpha\}$ and $\{x^{\alpha'}\}$ at the point q; then

$$y^{\alpha'} = \left(\frac{\partial x^{\alpha'}}{\partial x^\beta}\right)_q y^\beta \tag{2.9}$$

holds for the normal coordinates of any point $p \in U$—the star-shaped neighborhood of q.

(iv) In the normal coordinate system $\{y^\alpha\}$ at q at any point $p \in U$

$$\begin{array}{l} \Gamma^{*\alpha}_{\beta\gamma} y^\beta y^\gamma = 0 \\ \text{----------} \\ \Gamma^{*\alpha}_{\beta_1\beta_2\cdots\beta_r} y^{\beta_1} y^{\beta_2} \cdots y^{\beta_r} = 0 \end{array} \tag{2.10}$$

for any integer $r > 2$. The star is to indicate that the components of considered fields are taken in a normal coordinate system. (The first of these equations is also sufficient to characterize the normal coordinates.)

Expanding (2.10) in Taylor series around q, we immediately get that

(v) At the point q, in the normal coordinates at q,

$$\begin{array}{l} \Gamma^{*\alpha}_{\beta\gamma}(q) = 0 \\ \text{----------} \\ \Gamma^{*\alpha}_{(\beta_1\cdots\beta_r,\gamma_1\cdots\gamma_s)}(q) = 0 \end{array} \tag{2.11}$$

for any integers $r \geq 2$ and $s \geq 0$.

Both the map $\{x^\alpha\}$ and the normal map $\{y^\alpha\}$ associated with it at q define in the tangent space T_q at the point q the same natural frame

$$\left(\frac{\partial}{\partial x^\alpha}\right)_q = \left(\frac{\partial}{\partial y^\alpha}\right)_q, \quad \alpha = 1, \ldots, n.$$

From this observation and from (2.9) the next property follows.

(vi) If at the point q, in the normal coordinate system $\{y^\alpha\}$ associated with a system $\{x^\alpha\}$, a set of components Λ^*_L (L is a collective index) is given which under a transformation to another normal coordinate system $\{y^{\alpha'}\}$, transform like components of a relative tensor, then Λ^*_L are numerically equal to components Λ_L in the map $\{x^\alpha\}$ of a geometric object Λ defined at the point q of $\mathcal{A}_n$. The object Λ transforms under a change of coordinates $\{x^\alpha\} \to \{x^{\alpha'}\}$ (where $\{y^{\alpha'}\}$ is associated with $\{x^{\alpha'}\}$ at q) like a relative tensor of the same weight and kind as Λ^* under $\{y^\alpha\} \to \{y^{\alpha'}\}$.

This last property provides the foundation for the following constructions.

Let T be a relative tensor field defined in a neighborhood of a point $q \in \mathcal{A}_n$ and $T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}$ be its components in the natural frame corresponding to coordinates $\{x^\alpha\}$. The rth *affine extension* $\mathcal{I}^r T(q)$ of T is defined at the point q as a geometric object whose components $\mathcal{I}_{\gamma_r\cdots\gamma_1} T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}(q)$ in the coordinate map $\{x^\alpha\}$ are defined as

$$\mathcal{I}_{\gamma_r\cdots\gamma_1} T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l} := \left(\frac{\partial^r T^{*\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}}{\partial y^{\gamma_r}\cdots\partial y^{\gamma_1}}\right)_q, \tag{2.12}$$

where $T^{*\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}$ are components of the field T in the normal coordinate map $\{y^\alpha\}$ associated with $\{x^\alpha\}$ at q.

Thus, the rth affine extension $\mathcal{I}^r T$ of a relative tensor field T of the type (k,l) is a relative tensor of the same weight as T and of the type $(k, l+r)$; further

$$\mathcal{I}_{\gamma_r\cdots\gamma_1} T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l} = \mathcal{I}_{(\gamma_r\cdots\gamma_1)} T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}. \tag{2.13}$$

The procedure of constructing the affine extension may be applied not only to relative tensor fields but also to fields of some other geometric objects. One of the more important examples is provided by the affine connection field of the manifold. Since under a change of normal coordinates $\{y^\alpha\} \to \{y^{\alpha'}\}$ the $\Gamma^\alpha_{\beta\gamma}$ are transforming like components of a tensor field, the quantity $\mathcal{I}^{r-2}\Gamma(q)$ ($r = 3, 4, \cdots$) defined at the point q in a coordinate system $\{x^\alpha\}$ by the components

$$N^\alpha_{\beta_1\cdots\beta_r}(q) := \left(\frac{\partial^{r-2}\Gamma^{*\alpha}_{\beta_1\beta_2}}{\partial y^{\beta_r}\circ\cdots\circ\partial y^{\beta_3}}\right)_q, \quad r = 3, 4, \cdots, \tag{2.14}$$

where $\Gamma^{*\alpha}_{\beta\gamma}$ are the components of the affine connection field in the normal coordinate system $\{y^\alpha\}$ associated to $\{x^\alpha\}$ at q, is a tensor called the $(r-2)$*th normal tensor*.

It fulfills the following symmetry conditions

$$N^\alpha_{\beta_1\beta_2\beta_3\cdots\beta_r} = N^\alpha_{(\beta_1\beta_2)\beta_3\cdots\beta_r} = N^\alpha_{\beta_1\beta_2(\beta_3\cdots\beta_r)} = N^\alpha_{(\beta_1\beta_2\beta_3\cdots\beta_r)}. \tag{2.15}$$

The last of them follows from (2.11).

Another important object of this kind is provided by the concept of the *absolute extension* of a relative tensor field in the direction of a given curve. Let $\gamma: I \to \mathcal{A}_n$ ($I \subset \mathbb{R}$) be a curve passing through a point $q \in \mathcal{A}_n$ and let $x^\alpha(\sigma) := x^\alpha \circ \gamma(\sigma)$ be coordinates in a map $\{x^\alpha\}$ of points on γ which belong to a star-shaped neighborhood U of q. In the normal coordinates $\{y^\alpha\}$ associated with $\{x^\alpha\}$ at q the coordinates of the same points on γ will be denoted by $y^\alpha(\sigma)$. Let T be a relative tensor field defined in a neighborhood of q. The rth *absolute extension* of T in the direction of γ at the point $q \in \gamma$ is a relative tensor $(\mathcal{I}^r T/d\sigma^r)(q)$ at q, of the same weight and type as T, whose components in the coordinate map $\{x^\alpha\}$ are defined as

$$\frac{\mathcal{I}^r T^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l}}{d\sigma^r}(q) := \left(\frac{d^r}{d\sigma^r} T^{*\alpha_1\cdots\alpha_k}_{\beta_1\circ\cdots\circ\beta_l}(y^1(\sigma), y^2(\sigma), \ldots, y^r(\sigma))\right) \tag{2.16}$$

The formulas (2.12) and (2.14) have defined the objects $\mathcal{I}^r T(q)$ and $\mathcal{I}^{r-2}\Gamma(q)$ at a point q. Application of these definitions at every point of a convex region of $\mathcal{A}_n$ leads to corresponding fields of affine extensions $\mathcal{I}^r T$ and $\mathcal{I}^{r-2}\Gamma$ in that region. These fields may be expressed in terms of the field T and its covariant derivatives as well as of the curvature tensor of the manifold and its covariant derivatives. The corresponding formulas are given in the literature.[3,4] Here only some more important examples will be quoted.

The first normal tensor: The following relation between the curvature tensor and the first normal tensor

$$R^\alpha{}_{\beta\gamma\delta} = N^\alpha_{\beta\delta\gamma} - N^\alpha_{\beta\gamma\delta} \tag{2.17}$$

follows from the definition of $R^{\alpha}{}_{\beta\gamma\delta}$ in the normal coordinates and from the tensor character of all the quantities involved. Equations (2.17) and (2.15) for $N^{\alpha}{}_{\beta\gamma\delta}$ may be solved algebraically to yield

$$N^{\alpha}{}_{\beta\gamma\delta} = -\tfrac{1}{3}(R^{\alpha}{}_{\beta\gamma\delta} + R^{\alpha}{}_{\gamma\beta\delta}) = -\tfrac{1}{3}(2R^{\alpha}{}_{\beta\gamma\delta} + R^{\alpha}{}_{\delta\beta\gamma}). \tag{2.18}$$

The second normal tensor: By differentiation of the Riemann tensor and by Eq. (2.14),

$$R^{\alpha}{}_{\beta\gamma\delta;\epsilon} = N^{\alpha}{}_{\beta\delta\gamma\epsilon} - N^{\alpha}{}_{\beta\gamma\delta\epsilon}, \tag{2.19}$$

which together with (2.15) determines

$$6N^{\alpha}{}_{\beta\gamma\delta\epsilon} = -5R^{\alpha}{}_{\beta\gamma\delta;\epsilon} - 4R^{\alpha}{}_{\beta\delta\epsilon;\gamma} - 3R^{\alpha}{}_{\epsilon\beta\delta;\gamma} - 2R^{\alpha}{}_{\delta\epsilon\gamma;\beta} - R^{\alpha}{}_{\gamma\epsilon\delta;\beta}. \tag{2.20}$$

It may be shown by induction that the nth normal tensor is determined by the curvature tensor and by its covariant derivatives up to the order $n-1$.

The extension of a vector: By definition the components of the first affine extension of a vector T are

$$\mathcal{I}_{\mu} T^{\alpha} = \left(\frac{\partial T^{*\alpha}}{\partial y^{\mu}}\right)_q . \tag{2.21}$$

However,

$$\frac{\partial T^{*\alpha}}{\partial y^{\mu}} = (\nabla_{\mu} T^{\alpha})^* - \Gamma^{*\alpha}_{\mu\sigma} T^{*\sigma}. \tag{2.22}$$

Since the second term here vanishes at the origin q of normal coordinates, we see that the first extension of a vector (as well as of any other tensor) equals its covariant derivative

$$\mathcal{I} T = \nabla T. \tag{2.23}$$

For the second extension, inserting the right-hand side of (2.22) into the definition, we obtain

$$\mathcal{I}_{\nu\mu} T^{\alpha} = \left(\frac{\partial^2 T^{*\alpha}}{\partial y^{\nu} \partial y^{\mu}}\right)_q = \left(\frac{\partial}{\partial y^{\nu}} (\nabla_{\mu} T^{\alpha})^*\right)_q - \left(\frac{\partial}{\partial y^{\nu}} \Gamma^{*\alpha}_{\mu\sigma} T^{*\sigma}\right)_q = \nabla_{\nu\mu} T^{\alpha} - N^{\alpha}{}_{\sigma\mu\nu} T^{\sigma}. \tag{2.24}$$

This formula shows (since $\nabla_{\nu\mu} T^{\alpha} = \mathcal{I}_{\nu} \mathcal{I}_{\mu} T^{\alpha}$) that $\mathcal{I}_{\mu\nu} \neq \mathcal{I}_{\mu} \mathcal{I}_{\nu}$, which was also evident from (2.18) and (2.20). Similarly

$$\mathcal{I}_{\rho\nu\mu} T^{\alpha} = \nabla_{\rho\nu\mu} T^{\alpha} - N^{\alpha}{}_{\sigma\mu\nu} \nabla_{\rho} T^{\sigma} - N^{\alpha}{}_{\sigma\rho\nu} \nabla_{\mu} T^{\sigma} - N^{\alpha}{}_{\sigma\rho\mu} \nabla_{\nu} T^{\sigma} - N^{\sigma}{}_{\rho\nu\mu} \nabla_{\sigma} T^{\alpha} - N^{\alpha}{}_{\sigma\rho\nu\mu} T^{\sigma}. \tag{2.25}$$

The absolute extension of a vector: The same procedure applied to the absolute extension of a vector along a curve γ (parametrized by σ) leads to (t being the tangent vector to γ)

$$\frac{\mathcal{I} T^{\alpha}}{d\sigma} = \frac{D T^{\alpha}}{d\sigma} = t^{\mu} \nabla_{\mu} T^{\alpha}, \tag{2.26}$$

$$\frac{\mathcal{I}^2 T^{\alpha}}{d\sigma^2} = \frac{D^2 T^{\alpha}}{d\sigma^2} - N^{\alpha}{}_{\rho\mu\nu} T^{\rho} t^{\mu} t^{\nu}, \tag{2.27}$$

$$\frac{\mathcal{I}^3 T^{\alpha}}{d\sigma^3} = \frac{D^3 T^{\alpha}}{d\sigma^3} - N^{\alpha}{}_{\rho\mu\nu\sigma} T^{\rho} t^{\mu} t^{\nu} t^{\sigma} - N^{\alpha}{}_{\rho\mu\nu} T^{\rho} t^{\mu} \frac{D t^{\nu}}{d\sigma} - 2N^{\alpha}{}_{\rho\mu\nu} \frac{D}{d\sigma} (T^{\rho} t^{\mu}) t^{\nu} - N^{\alpha}{}_{\rho\mu\nu} \frac{D T^{\rho}}{d\sigma} t^{\mu} t^{\nu}. \tag{2.28}$$

Generally, for the nth absolute extension, using the Leibniz rule for the nth derivative of a product, we get

$$\frac{\mathcal{I}^n T^{\alpha}}{d\sigma^n} = \frac{\mathcal{I}^{n-1}}{d\sigma^{n-1}} \frac{D T^{\alpha}}{d\sigma} - \frac{\mathcal{I}^{n-1}}{d\sigma^{n-1}} (\Gamma^{\alpha}_{\rho\mu} T^{\rho} t^{\mu}) = \frac{\mathcal{I}^{n-1}}{d\sigma^{n-1}} \frac{D T^{\alpha}}{d\sigma} - \sum_{i=0}^{n-2} \binom{n-1}{i} \frac{\mathcal{I}^{n-i-1}}{d\sigma^{n-i-1}} (\Gamma^{\alpha}_{\rho\mu}) \frac{\mathcal{I}^{i}}{d\sigma^{i}} (T^{\rho} t^{\mu}).$$

Now the formula above can be applied to its first term. Repeating this procedure $n-2$ times (in the last step $\mathcal{I}/d\sigma = D/d\sigma$), we come to

$$\frac{\mathcal{I}^n T^{\alpha}}{d\sigma^n} = \frac{D^n T^{\alpha}}{d\sigma^n} - \sum_{j=0}^{n-2} \sum_{i=0}^{n-j-2} \binom{n-j-1}{i} \frac{\mathcal{I}^{n-j-i-1}}{d\sigma^{n-j-i-1}} (\Gamma^{\alpha}_{\rho\mu}) \times \frac{\mathcal{I}^{i}}{d\sigma^{i}} \left(\frac{D^j T^{\rho}}{d\sigma^j} t^{\mu}\right). \tag{2.29}$$

Here, for any integer r, $1 \leq r \leq n-1$, we should insert

$$\frac{\mathcal{I}^r}{d\sigma^r} (\Gamma^{\alpha}_{\rho\mu}) = \sum_{\{i,j,h,\ldots,k\}} \left(\frac{1}{2!}\right)^j \left(\frac{1}{3!}\right)^h \cdots \left(\frac{1}{l!}\right)^k \frac{r!}{i!\,j!\,h! \cdots k!} \times N^{\alpha}{}_{\rho\mu\alpha_1 \cdots \alpha_i \beta_1 \cdots \beta_j \cdots \kappa_1 \cdots \kappa_k} t^{\alpha_1} \cdots t^{\alpha_i} \times \frac{\mathcal{I} t^{\beta_1}}{d\sigma} \cdots \frac{\mathcal{I} t^{\beta_j}}{d\sigma} \cdots \frac{\mathcal{I}^l t^{\kappa_1}}{d\sigma^l} \cdots \frac{\mathcal{I}^l t^{\kappa_k}}{d\sigma^l}, \tag{2.30}$$

where the summation runs over all the sets $\{i,j,h,\ldots,k\}$ of nonnegative integer solutions of the equation

$$i + 2j + 3h + \cdots + lk = r.$$

In the considerations above use has been made of the fact that the Leibniz rule and the formula for the nth derivative of a composed function,[5] which led to (2.30), apply without any change to the nth absolute extension.

Formula (2.29) is then a recurrence relation for the nth absolute extension in terms of extensions of lower order applied to the vector T, its first $n-2$ absolute derivatives and the tangent vector t. For a particular case of $T = t$ it reduces to (1.4) and (1.6). Neither (2.29) nor (1.4) has been published before.

In the review presented in this section special coordinates have been extensively used. Such an approach has made it possible to perform the desired calculations. For the very definition of the objects introduced here, however, it is not necessary to call for special coordinates. These objects may be defined purely geometrically, for instance, in the following way.

Consider the exponential map at the point q,

$$\exp_q : O(T_q) \to \mathcal{A}_n,$$

that is defined as a mapping of an open neighborhood $O(T_q)$ of the zero element of the tangent space T_q into $\mathcal{A}_n$ under which for any $t \in O(T_q)$

$$t \mapsto \exp_q(t) = \Gamma_t(1),$$

where Γ_t is a geodesic such that $\Gamma_t(0) = q$ with t being its tangent vector at q. As it is known,[6] $\exp_q$ is a diffeomorphic mapping of $O(T_q)$ onto an open neighborhood $U \subset \mathcal{A}_n$ of the point q. Under its inverse mapping, $\ln_q$, tensor fields on U are dragged along onto tensor fields on $O(T_q)$ (since $d \ln_q$ is one to one for the domain considered here). Thus to any tensor field T on U there is

assigned a tensor field T^* on $O(T_q)$. Since the partial differentiation of tensor fields on a vector space leads to tensor fields, the value of the nth partial derivative of the field T^* evaluated at the origin of T_q gives the nth affine extension of T at q. Similarly, the nth derivative of the field T^* taken in the direction of a curve $\ln_q\gamma$ (which is passing through the origin of T_q) and evaluated there leads to the nth absolute extension of T in the direction of γ at q.

3. THE PROOF OF THE THEOREM

Since (1.1) holds in any coordinates from an analytic subatlas, it may be written in the normal coordinates $\{y^\alpha\}$ associated at the point q with the coordinates $\{x^\alpha\}$ originally used in (1.1). Thus, denoting by $y^\alpha(\tau)$: $=y^\alpha\circ\gamma(\tau)$, for a point $p=\gamma(\tau)$ such that $p\in U\subset\Omega$, where U is a neighborhood of $q=\gamma(\tau_0)$ in which (2.5) is convergent, we have

$$y^\alpha(\tau)=t_0^{*\alpha}(\tau-\tau_0)+\frac{1}{2!}\left(\frac{dt^{*\alpha}}{d\tau}\right)_0(\tau-\tau_0)^2+\cdots+\frac{1}{(n+1)!}\left(\frac{d^nt^{*\alpha}}{d\tau^n}\right)_0(\tau-\tau_0)^{n+1}+\circ\circ\circ; \quad (3.1)$$

here $t_0^{*\alpha}:=(dy^\alpha(\tau)/d\tau)_0$ are components in the normal map $\{y^\alpha\}$ of the vector t tangent to γ at the point q and besides that according to (2.8) the coordinates of the point q are equal $y^\alpha(\tau_0)=0$.

A geodesic Γ_l passing for the value σ_0 of its affine parameter σ through the point q with l as its tangent vector at q is, according to (2.7), in the map $\{y^\alpha\}$ described by the equations

$$\eta^\alpha(\sigma)=(\sigma-\sigma_0)l^{*\alpha}, \quad (3.2)$$

where $\eta^\alpha(\sigma):=y^\alpha\circ\Gamma_l(\sigma)$.

The question for which l the point $p=\gamma(\tau)$ will be lying on Γ_l leads to the equation

$$\eta^\alpha(\sigma)=y^\alpha(\tau), \quad (3.3)$$

which may have solutions for a set of values of the parameters σ and τ. We shall limit the number of solutions by demanding that $\sigma-\sigma_0=\tau-\tau_0$. Then (3.3) together with (3.1) and (3.2) determines

$$l^{*\alpha}=t_0^{*\alpha}+\frac{1}{2!}\left(\frac{dt^{*\alpha}}{d\tau}\right)_0(\tau-\tau_0)+\cdots+\frac{1}{(n+1)!}\left(\frac{d^nt^{*\alpha}}{d\tau^n}\right)_0(\tau-\tau_0)^n+\cdots. \quad (3.4)$$

The series (3.4) is convergent if and only if (3.1) is convergent. Formula (3.4) gives a relation between components in the map $\{y^\alpha\}$ of vectors defined at the point q. Due to the definition (2.16) it is equivalent to the following vector relation at q

$$l=t(q)+\frac{1}{2!}\frac{\nabla t}{d\tau}(q)(\tau-\tau_0)+\cdots+\frac{1}{(n+1)!}\frac{\nabla^n t}{d\tau^n}(q)(\tau-\tau_0)^n+\circ\circ\circ, \quad (3.5)$$

and that completes the proof.

ACKNOWLEDGMENTS

The author is very much indebted to Professor Joshua N. Goldberg and Professor Andrzej Trautman for critical reading of the manuscript and for their comments.

*Research supported in part by the National Science Foundation, Contract GF-36217.

[1]O. Veblen and T. Y. Thomas, Trans. Am. Math. Soc. **25**, 551 (1923).
[2]O. Veblen and T. Y. Thomas, Trans. Am. Math. Soc. **26**, 373 (1924).
[3]O. Veblen *Inwariants of Quadratic Differential Forms*, Cambridge Tracts No. 24 (Cambridge U. P., Cambridge, 1927).
[4]J. A. Schouten *Ricci Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.
[5]E. Goursat, *Cours d' analyse mathématique, I* (Gauthier-Villars, Paris, 1943), 5th ed., p. 78.
[6]S. Helgason *Differential Geometry and Symmetric Spaces* (Academic, New York and London, 1962).

Multiple time scale analysis of an anharmonic crystal

I. C. Khoo

Department of Physics, The University of Rochester, Rochester, New York 14627

Y. K. Wang

Department of Physics, Northeastern University, Boston, Massachusetts 02115
(Received 2 January 1975; revised manuscript received 24 March 1975)

We use the multiple time scale perturbation method to study the lattice dynamics of an anharmonic crystal. The Heisenberg equations of motion for the creation and annihilation operators are solved, and the frequency shift and the decay constants are found to the second order. We also discuss briefly how to apply our solution to calculate the correlation function.

I. INTRODUCTION

We consider here the problem of a three-dimensional anharmonic crystal. In the second quantized form, the phonon creation and annihilation operators obey a system of coupled nonlinear equations. Techniques of studying these coupled phonon problems are plentiful. For example, the phonon frequency shifts and lifetimes have been calculated by Kokkedee,[1] using a diagrammatic technique, by Maradudin and Fein,[2] using phonon propagators, and by Cowley,[3] using the thermodynamic Green's function. Later, Wallace[4] developed a method using undetermined coefficients to renormalize the phonon creation operators to first order, and it was found that the method worked well and was simple compared to some of the complicated diagrams summations.[1-3]

In the quest for simple yet powerful techniques to solve this many body problem, it is surprising how rarely one resorts to methods that abound in nonlinear mechanics. The authors[5] have recently shown how an almost trivial zeroth order solution of the equations of motion lead to the Wigner—Weisskopff approximation frequently used in quantum optics. Our primary purpose here is to apply a method from nonlinear mechanics, that is, the so-called multiple time scale analysis,[6,8] to the anharmonic crystal problem and show how the method easily gives the phonon frequency shifts and lifetimes to second order, in comparison with the complexity of the methods employed in previous treatment.[1-4] Working directly with the operator equations of motion, we show how the successive higher order solution may be achieved without encountering the usual secular terms in Hamiltonian perturbation theory. Incidentally, the subject of secular terms and its removal by a new perturbation theory has recently been studied by Helleman and Montroll.[7] The multiple time scale analysis has been applied to many problems. Lee, Lee, and Chang[8] applied it to the spontaneous radiation process. Varga and Aks[9] applied it to the ϕ^4 model of quantum field theory and obtained a first order renormalized Hamiltonian, while Frieman[10] applied it to irreversible approach to equilibrium in gases. We attempt to fill in the gap here and apply the method to solids and specifically to coupled phonons in an anharmonic crystal.

In Sec. II, we shall illustrate the principle of multiple time scale analysis by applying it to a problem of a nonlinear oscillator so that our paper is more or less self-contained. The equations of motion for the creation and annihilation operators of the crystal are solved in Sec. III. As an application of our general solution, we shall apply it to calculate the correlation function in Sec. IV.

II. THE MULTIPLE TIME SCALE PERTURBATION EXPANSION (MTSPE)

Before we go to the more complex analysis of the anharmonic lattice problem, let us give a simple example to illustrate the principles of MTSPE applied to nonlinear oscillations. The example is well known, but we present it here to make the paper more self-contained. Consider a nonlinear equation

$$\frac{d^2f(t)}{dt^2}+\omega^2 f(t)=\epsilon f(t)^2, \tag{2.1}$$

with ω being the frequency (real) and ϵ a perturbative parameter.

It is well known that the ordinary perturbation theory cannot be applied to (2.1) due to the so-called secular terms of the form $t\exp(i\omega t)$, $t^2\exp(i\omega t)$, $\cdots$, etc., in the higher order expansions. This divergent difficulty can be avoided by introducing the MTSPE as we shall illustrate. Let us replace the original single time variable t by a collection of variables $\tau=(\tau_0,\tau_1,\tau_2,\cdots)$ defined by

$$\tau_0=t,$$
$$\tau_n=\epsilon^n t+x_n,\quad n\geq 1. \tag{2.2}$$

The $\{x_n\}$ in (2.2) are considered independent of each other, and hence the new variables $\{\tau_n\}$ are also independent of each other. We shall now generalize $f(t)$ in (2.1) into a function of many variables $f(\tau)\equiv f(\tau_0,\tau_1,\cdots)$, and require it to obey the following equation:

$$\frac{\partial^2}{\partial t^2}f(\tau_0,\tau_1,\cdots)+\omega^2 f(\tau_0,\tau_1,\cdots)=\epsilon f(\tau_0,\tau_1,\cdots)^2. \tag{2.3}$$

If such a function $f(\tau)$ is found, then, by simply setting all $x_n=0$, or equivalently all $\tau_n=\epsilon^n t$, we can recover the original $f(t)$. This is the basic principle of MTSPE. The generalization to $f(\tau)$ gives us a partial differential equation (2.3), which itself cannot determine $f(\tau)$ uniquely. In fact, in the search for such an $f(\tau)$, we have the freedom to impose certain extra conditions upon f. As we shall see later, if we choose these conditions properly, the secular terms can be systematically removed and give rise to a renormalized frequency.

Let us expand $f(\tau)$ into a power series of ϵ,

$$f(\tau)=f_0(\tau)+\epsilon f_1(\tau)+\cdots, \tag{2.4}$$

then,

$$\frac{\partial f}{\partial t}=\frac{\partial}{\partial \tau_0}f+\epsilon\frac{\partial}{\partial \tau_1}f+\cdots$$

and

$$\begin{aligned}\frac{\partial^2 f}{\partial t^2}=&\frac{\partial^2}{\partial \tau_0^2}f+\epsilon\frac{\partial^2}{\partial \tau_0\partial \tau_1}f+\epsilon^2\frac{\partial^2}{\partial \tau_0\partial \tau_2}f+\cdots\\&+\epsilon\frac{\partial^2}{\partial \tau_1\partial \tau_0}f+\epsilon^2\frac{\partial^2}{\partial \tau_1^2}f+\cdots\\&+\cdots.\end{aligned} \tag{2.5}$$

Substituting (2.5) into (2.3) and comparing the coefficients to each order of ϵ, we get

$$O(\epsilon^0):\ \frac{\partial^2}{\partial \tau_0^2}f_0+\omega^2 f_0=0, \tag{2.6}$$

$$O(\epsilon^1):\ \frac{\partial^2}{\partial \tau_0^2}f_1+2\frac{\partial^2}{\partial \tau_0\partial \tau_1}f_0+\omega^2 f_1=f_0^2, \tag{2.7}$$

$$\begin{aligned}O(\epsilon^2):\ &\frac{\partial^2}{\partial \tau_0^2}f_2+2\frac{\partial^2}{\partial \tau_0\partial \tau_1}f_1+2\frac{\partial^2}{\partial \tau_0\partial \tau_2}f_0\\&+\frac{\partial^2}{\partial \tau_1^2}f_0+\omega^2 f_2=2f_0f_1,\end{aligned} \tag{2.8}$$

$\cdots$

The solution for (2.6) is simple, namely,

$$f_0(\tau)=A(\tau_1,\tau_2,\cdots)\exp(i\omega\tau_0)+\text{c.c.}, \tag{2.9}$$

where $A(\tau_1,\tau_2,\cdots)$ is an arbitrary complex function of $\tau_1,\tau_2,\cdots$, etc., and c.c. means its complex conjugate. To the next order $O(\epsilon)$, we shall substitute (2.9) into (2.7), and find

$$\begin{aligned}\frac{\partial^2}{\partial \tau_0^2}f_1+\omega^2 f_1=&[A(\tau_1,\tau_2,\cdots)\exp(i\omega\tau_0)+\text{c.c.}]^2\\&-2\left(\frac{\partial A}{\partial \tau_1}i\omega\exp(i\omega\tau_0)+\text{c.c.}\right).\end{aligned} \tag{2.10}$$

By integrating (2.10), we can see that the term $\exp(i\omega\tau_0)$ will contribute a secular term of the form $\tau_0\exp(i\omega\tau_0)$ in the f_1. To remove this term, we are forced to impose an additional condition on A,

$$\frac{\partial A}{\partial \tau_1}=0, \tag{2.11}$$

which implies $A=A(\tau_2,\tau_3,\cdots)$. With (2.11), the f_1 can be integrated to yield a well-behaved function,

$$f_1=-\frac{A^2}{3\omega}\exp(i2\omega\tau_0)+\frac{|A|^2}{\omega^2}+\text{c.c.} \tag{2.12}$$

Proceeding to the next order $O(\epsilon^2)$, we shall substitute (2.9) and (2.12) into (2.8),

$$\begin{aligned}\frac{\partial^2 f_2}{\partial \tau_0^2}+\omega^2 f_2=&\left(\frac{10}{3}\frac{|A|^2}{\omega^2}A\exp(i\omega\tau_0)-\frac{2A^3}{3\omega^2}\exp(i3\omega_0\tau)+\text{c.c.}\right)\\&+\left[-2i\omega\left(\frac{\partial A}{\partial \tau_2}\right)\exp(i\omega\tau_0)+\text{c.c.}\right].\end{aligned} \tag{2.13}$$

Again, the terms proportional to $\exp(i\omega\tau_0)$ will give rise to secular terms. Hence we shall require the coefficient to $\exp(i\omega\tau_0)$ equal to zero,

$$\frac{5}{3}\frac{|A|^2}{\omega^2}A-i\omega\frac{\partial A}{\partial \tau_2}=0. \tag{2.14}$$

The solution of (2.14) is in the form,

$$A(\tau_2,\tau_3,\cdots)=B(\tau_3,\tau_4,\cdots)\exp(i\alpha\tau_2),$$

where

$$\tfrac{5}{3}|B|^2/\omega^2=-\omega\alpha$$

or

$$A(\tau_2,\tau_3,\cdots)=B(\tau_3,\tau_4,\cdots)\exp[-i\tfrac{5}{3}(|B|^2/\omega^3)\tau_2] \tag{2.15}$$

From (2.9), (2.12), and (2.14), the function $f(\tau)$ becomes

$$\begin{aligned}f(\tau)=&f(\tau_0,\tau_2,\tau_3,\cdots)\\=&(B(\tau_3\cdots)\exp\{i\omega[\tau_0-\tfrac{5}{3}(|B|^2/\omega^4)\tau_2]\}+\text{c.c.})\\&+\epsilon(|B|^2/\omega-(B^2/3\omega^2)\exp\{i2\omega[\tau_0-\tfrac{5}{3}(|B|^2/\omega^4)\tau_2]\}\\&+\text{c.c.})+O(\epsilon^2).\end{aligned} \tag{2.16}$$

The perturbation scheme can be carried out in a similar fashion to any order, although the complexity is formidable. To the second order, we can set all higher order time scale variables $\tau_3,\tau_4,\cdots$, etc., equal to zero, and $B(\tau_3,\tau_4,\cdots)$ becomes a constant (independent of $\tau_3,\tau_4,\cdots$, etc.). The perturbative equations (2.6), (2.7), $\cdots$, etc., are correct to $O(\epsilon^2)$ only.

The lowest order term f_0 becomes

$$f_0(\tau)=B\exp\{i\omega[\tau_0-\tfrac{5}{3}(|B|^2/\omega^4)\tau_2]\}+\text{c.c.} \tag{2.17}$$

Setting $x_2=0$, we have

$$f_0(t)=B\exp[i(\omega+\Delta\omega)t]+\text{c.c.} \tag{2.18}$$

with

$$\Delta\omega=-\epsilon^2\tfrac{5}{3}|B|^2/\omega^3.$$

The constant B can be obtained by initial condition. Let us suppose

$$f(0)=1,\quad \dot f(0)=1;$$

then

$$B=\tfrac{1}{2}+O(\epsilon)$$

and, therefore,

$$\Delta\omega=-\epsilon^2\frac{5}{12}\frac{1}{\omega^3}+O(\epsilon^3). \tag{2.19}$$

We thus see that, by properly choosing these additional conditions in MTSPE, we can systematically remove the secular terms. Also, from the second order solution (2.16), we see that there are two time scales involved in the function $f(t)$, namely the inverse of frequency ω^{-1} and shift $\Delta\omega^{-1}$. The removal of the secular terms defines the behavior of $f(t)$ in these time scales. Higher order τ_k's, $k\geq 3$, corresponding to larger time scale can be set equal to zero as far as the calculation is carried out up to second order only. In the next section, we shall use this MTSPE method to the more complex problem of anharmonic lattice dynamics.

III. THE MTSPE ANALYSIS OF THE ANHARMONIC LATTICE

The diagonalization of the Hamiltonian of an anharmonic crystal can be seen in the standard texts of Born and Huang[11] or Peierls.[12] We follow here the notation of the paper by Wallace.[4] The total Hamiltonian H, neglecting the quartic interaction in Wallace's treatment, is given by

$$H = H_0 + H_1, \tag{3.1}$$

where, in second quantized form,

$$H_0 = \sum_{\kappa} \hbar\omega_\kappa a_\kappa^\dagger a_\kappa \tag{3.2}$$

and

$$H_1 = \lambda \sum_{\kappa\kappa'\kappa''} B_{\kappa\kappa'\kappa''}(a_\kappa + a_{-\kappa}^\dagger)(a_{\kappa'} + a_{-\kappa'}^\dagger)(a_{\kappa''} + a_{-\kappa''}^\dagger). \tag{3.3}$$

a_κ and $a_\kappa^\dagger$ are the usual phonon annihilation and creation operators; ω_κ's are the normal modes frequencies of the crystal; $\kappa = (\mathbf{K}, s)$, where $\mathbf{K}$ is the phonon wave vector and s is the phonon polarization; $B_{\kappa\kappa'\kappa''}$ are the coupling coefficients and depend on the type of crystal structure under study. It is completely symmetric in its indices, and contains a factor $\delta(\mathbf{K}+\mathbf{K}'+\mathbf{K}'')$. Furthermore, $B_{\kappa\kappa'\kappa''} = B^*_{-\kappa-\kappa'-\kappa''}$; λ is a perturbation parameter to be set equal to unity eventually.

In the absence of the cubic interaction H_1, the normal modes are not coupled and the phonons propagate in the crystal independent of one another. The introduction of nonlinearity leads to scattering between the phonons and each phonon is associated with a life time Γ_κ^{-1}. In this problem there are thus two distinct time scales: One is associated with the period of oscillation ω_κ^{-1}, and the other is the phonon's lifetime, and we denote them as τ_0 and τ_2 respectively.

In the Heisenberg picture, the equations of motion for a_κ and $a_\kappa^\dagger$ are

$$\dot{a}_\kappa^\dagger = i\omega_\kappa a_\kappa^\dagger + \frac{3i}{\hbar}\lambda \sum_{\kappa'\kappa''} B_{\kappa\kappa'\kappa''}(a_{\kappa'} + a_{-\kappa'}^\dagger)(a_{\kappa''} + a_{-\kappa''}^\dagger), \tag{3.4}$$

$$\dot{a}_\kappa = -i\omega_\kappa a_\kappa - \frac{3i}{\hbar}\lambda \sum_{\kappa'\kappa''} B^*_{\kappa\kappa'\kappa''}(a_{\kappa''}^\dagger + a_{-\kappa''})(a_{\kappa'}^\dagger + a_{-\kappa'}). \tag{3.5}$$

Following the MTSPE, we introduce the variables $\tau = (\tau_0, \tau_1, \cdots)$ as before,

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \tau_0} + \lambda\frac{\partial}{\partial \tau_1} + \lambda^2\frac{\partial}{\partial \tau_2} + \cdots, \tag{3.6}$$

and expand the operator $a_\kappa(\tau)$ in powers of λ:

$$a_\kappa(\tau) = \sum_{n=0}^{\infty} \lambda^n a_\kappa^{(n)}(\tau). \tag{3.7}$$

Substituting Eqs. (3.6) and (3.7) into (3.4) and (3.5), and equating like powers of λ, we get

$$O(\lambda^0):\ \frac{\partial}{\partial \tau_0} a_\kappa^{(0)}(\tau) = -i\omega_\kappa a_\kappa^{(0)}(\tau), \tag{3.8}$$

$$O(\lambda):\ \frac{\partial}{\partial \tau_1} a_\kappa^{(0)}(\tau) + \frac{\partial}{\partial \tau_0} a_\kappa^{(1)}(\tau) = -i\omega_\kappa a_\kappa^{(1)}(\tau) - \frac{3i}{\hbar}\sum_{\kappa'\kappa''} B^*_{\kappa\kappa'\kappa''}[a_{\kappa''}^{(0)\dagger}(\tau) + a_{-\kappa''}^{(0)}(\tau)][a_{\kappa'}^{(0)\dagger}(\tau) + a_{-\kappa'}^{(0)}(\tau)], \tag{3.9}$$

$$O(\lambda^2):\ \frac{\partial}{\partial \tau_2} a_\kappa^{(0)}(\tau) + \frac{\partial}{\partial \tau_1} a_\kappa^{(1)}(\tau) + \frac{\partial}{\partial \tau_0} a_\kappa^{(2)}(\tau) = -i\omega_\kappa a_\kappa^{(2)}(\tau) - \frac{3i}{\hbar}\sum_{\kappa'\kappa''} B^*_{\kappa\kappa'\kappa''} \times\{[a_{\kappa''}^{(0)\dagger}(\tau) + a_{-\kappa''}^{(0)}(\tau)][a_{\kappa'}^{(1)\dagger}(\tau) + a_{-\kappa'}^{(1)}(\tau)] + [a_{\kappa''}^{(1)\dagger}(\tau) + a_{-\kappa''}^{(1)}(\tau)][a_{\kappa'}^{(0)\dagger}(\tau) + a_{-\kappa'}^{(0)}(\tau)]\}. \tag{3.10}$$

We can immediately integrate Eq. (3.8) to give the zeroth order solution:

$$a_\kappa^{(0)}(\tau) = a_\kappa^{(0)}(0, \tau_1, \tau_2, \cdots)\exp(-i\omega_\kappa\tau_0) \equiv a_\kappa^{(0)}(0)\exp(-i\omega_\kappa\tau_0), \tag{3.11}$$

where we have abbreviated $a_\kappa^{(0)}(0, \tau_1, \tau_2, \cdots)$ by $a_\kappa^{(0)}(0)$ as long as no confusion arises. If we substitute (3.11) into the first order equation (3.9), we can integrate it to get $a_\kappa^{(1)}(\tau)$. To avoid the complication that some of the denominator in the expression $a_\kappa^{(1)}(\tau)$ may go to zero, let us introduce a small imaginary part ϵ to the frequencies ω_κ, i.e.,

$$\omega_\kappa \to \omega_\kappa + i\epsilon, \quad \text{for all } \kappa,$$

and let $\epsilon \to 0$ in the final result of our calculation. Substitute (3.11) into (3.9) with ω_κ replaced by $\omega_\kappa + i\epsilon$, we find

$$a_\kappa^{(1)}(\tau) = a_\kappa^{(1)}(0, 0, \tau_2, \cdots)\exp(-i\omega_\kappa\tau_0) - \frac{3i}{\hbar}\sum_{\kappa'\kappa''} B^*_{\kappa\kappa'\kappa''} \times\left\{\frac{a_{\kappa''}^{(0)\dagger}(0)a_{\kappa'}^{(0)\dagger}(0)[\exp[i(\omega_{\kappa'}+\omega_{\kappa''})\tau_0] - \exp(-i\omega_\kappa\tau_0)]}{i(\omega_\kappa + \omega_{\kappa''} + \omega_{\kappa'} + 3i\epsilon)}\right. + a_{\kappa''}^{(0)\dagger}(0)a_{-\kappa'}^{(0)}(0)\frac{\exp[i(\omega_{\kappa''} - \omega_{-\kappa'})\tau_0] - \exp(-i\omega_\kappa\tau_0)}{i(\omega_\kappa + \omega_{\kappa''} - \omega_{-\kappa'} + i\epsilon)} + a_{-\kappa''}^{(0)}(0)a_{\kappa'}^{(0)\dagger}(0)\frac{\exp[i(-\omega_{-\kappa''} + \omega_{\kappa'})\tau_0] - \exp(-i\omega_\kappa\tau_0)}{i(\omega_\kappa - \omega_{-\kappa''} + \omega_{\kappa'} + i\epsilon)} + \left. a_{-\kappa''}^{(0)}(0)a_{-\kappa'}^{(0)}(0)\frac{\exp[i(-\omega_{-\kappa''} - \omega_{-\kappa'})\tau_0] - \exp(-i\omega_\kappa\tau_0)}{i(\omega_\kappa - \omega_{-\kappa''} - \omega_{-\kappa'} - i\epsilon)}\right\}. \tag{3.12}$$

Note that, with the introduction of ϵ in the frequencies ω_κ, the rhs of (3.9) does not have secular terms, so that, in deriving (3.12), we must set $(\partial/\partial\tau_1)a_\kappa^{(0)}(0, \tau_1, \cdots)$ equal to zero. In other words, $a_\kappa^{(0)}(\tau)$ does not depend on τ_1:

$$a_\kappa^{(0)}(\tau) = a_\kappa^{(0)}(0, 0, \tau_2, \tau_3, \cdots)\exp(-i\omega_\kappa\tau_0), \tag{3.11'}$$

which implies $(\partial/\partial\tau_1)a_\kappa^{(1)}(\tau) = 0$. The operator $a_\kappa^{(1)\dagger}(\tau)$ can be obtained in a similar fashion by integrating the corresponding complex conjugate of Eq. (3.9). The solution for $a_\kappa^{(1)\dagger}(\tau)$ is, of course, just the complex conjugate of Eq. (3.12).

To proceed to the second order $O(\lambda^2)$, let us consider Eq. (3.10). The second term on the rhs of Eq. (3.10) contains terms of the form $a_{\kappa''}^{(0)\dagger}(\tau)a_{\kappa'}^{(1)\dagger}(\tau)$, $a_{\kappa''}^{(0)\dagger}(\tau)a_{-\kappa'}^{(1)}(\tau)$, $\cdots$, etc. In expanding them by (3.12), there appear terms behaving like $\exp(-i\omega_\kappa\tau)$. Hence, after integrating (3.10), these terms will give rise to terms proportional to $\tau_0\exp(-i\omega_\kappa\tau_0)$ (i.e., the secular terms). The method of MTSPE enables us to choose $(\partial/\partial\tau_2)a_\kappa^{(0)}(\tau)$ so as to exactly cancel these terms. In particular, consider the term $a_{\kappa''}^{(0)\dagger}(\tau)a_{\kappa'}^{(1)\dagger}(\tau)$ on the rhs of (3.10). Using the Hermitian conjugate of Eq. (3.12) for $a_\kappa^{(1)\dagger}(\tau)$, we have

$$B^{*}_{k\kappa',\kappa''}a^{(0)\dagger}_{\kappa''}(\tau)a^{(1)\dagger}_{\kappa'}(\tau)$$

$$=\frac{3i}{\hbar}B^{*}_{k\kappa',\kappa''}a^{(0)\dagger}_{\kappa''}(0)\sum_{\kappa_1\kappa_2}B_{\kappa'\kappa_1\kappa_2}a^{(0)}_{\kappa_2}(0)a^{(0)}_{\kappa_1}(0)$$

$$\times\frac{\exp[i(-\omega_{\kappa_1}-\omega_{\kappa_2}+\omega_{\kappa''})]}{-i(\omega_{\kappa'}+\omega_{\kappa_1}+\omega_{\kappa_2}+3i\epsilon)}$$

+ other terms which will not give $\exp(-i\omega\tau_0)$ dependence. (3.13)

In the above expression, the $\exp(-i\omega\tau_0)$ terms are obtained if ($\kappa_1=\kappa''$ and $\kappa_2=\kappa$) or ($\kappa_2=\kappa''$ and $\kappa_1=\kappa$), and are equal to

$$\frac{6i}{\hbar}B^{*}_{k\kappa',\kappa''}B_{\kappa''\kappa',\kappa}a^{(0)\dagger}_{\kappa''}(0)a^{(0)}_{\kappa''}(0)a^{(0)}_{\kappa}(0)$$

$$\times\exp(-i\omega_\kappa\tau_0)/[-i(\omega_\kappa+\omega_{\kappa'}+\omega_{\kappa''}+3i\epsilon)], \quad (3.14)$$

which is one of the undesirable secular terms that we intend to remove. Similarly, the contribution to the secular terms from $a^{(0)\dagger}_{\kappa''}(\tau)a^{(1)}_{-\kappa}(\tau)$ is

$$-\frac{6i}{\hbar}B^{*}_{k\kappa',\kappa''}B^{*}_{-\kappa,-\kappa',-\kappa''}a^{(0)\dagger}_{\kappa''}(0)a^{(0)}_{\kappa''}(0)a^{(0)}_{\kappa}(0)$$

$$\times\exp(-i\omega_\kappa\tau_0)/i(\omega_{\kappa'}-\omega_{\kappa''}-\omega_\kappa-i\epsilon) \quad (3.15)$$

and so on. Proceeding through similar arguments, we can collect those secular terms from the rhs of (3.10) which we shall denote by $\hat{S}_2$.

$$\hat{S}_2=-\left(\frac{3i}{\hbar}\right)\left(\frac{6i}{\hbar}\right)\sum_{\kappa'\kappa''}B^{*}_{k\kappa',\kappa''}a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)$$

$$\times\left(B_{\kappa''\kappa'\kappa}\frac{\hat{n}_{\kappa''}+\hat{n}_{-\kappa'}+1}{i(-\omega_{\kappa'}-\omega_\kappa-\omega_{\kappa''}-3i\epsilon)}\right.$$

$$-B^{*}_{-\kappa,-\kappa',-\kappa''}\frac{\hat{n}_{\kappa''}-\hat{n}_{\kappa'}}{i(\omega_{\kappa'}-\omega_{\kappa''}-\omega_\kappa-i\epsilon)}$$

$$+B_{\kappa''\kappa'\kappa}\frac{\hat{n}_{-\kappa''}-\hat{n}_{-\kappa'}}{i(-\omega_{\kappa'}-\omega_\kappa+\omega_{-\kappa''}-i\epsilon)}$$

$$\left.-B^{*}_{-\kappa',-\kappa'',-\kappa}\frac{\hat{n}_{-\kappa''}+\hat{n}_{\kappa'}+1}{i(\omega_{-\kappa'}+\omega_{-\kappa''}-\omega_\kappa+i\epsilon)}\right), \quad (3.16)$$

where $\hat{n}_\kappa=a^{(0)\dagger}_{\kappa}(0)a^{(0)}_{\kappa}(0)$. Now, there are still other secular terms on the rhs of Eq. (3.10), namely when

$$-\kappa''=\kappa \quad \text{or} \quad -\kappa'=\kappa$$

in the expression

$$-\frac{3i}{\hbar}\sum_{\kappa'\kappa''}B^{*}_{k\kappa',\kappa''}\{a^{(0)}_{-\kappa''}(\tau)[a^{(1)\dagger}_{\kappa'}(\tau)+a^{(1)}_{-\kappa'}(\tau)]$$

$$+[a^{(1)\dagger}_{\kappa''}(\tau)+a^{(1)}_{-\kappa''}(\tau)]a^{(0)}_{-\kappa'}(\tau_0)\}.$$

Using $a^{(1)\dagger}_{\kappa'}$ and $a^{(1)}_{-\kappa'}$ from Eq. (3.12), we get the secular term in the case of $-\kappa''=\kappa$, to be

$$-\frac{9}{\hbar^2}\sum_{\kappa'\kappa''}B^{*}_{k\kappa',-\kappa}a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)$$

$$\times\left[\left(B_{\kappa',-\kappa''\kappa''}+B^{*}_{-\kappa',-\kappa''\kappa''}\right)\right.$$

$$\left.\times\frac{\hat{n}_{\kappa''}}{i(\omega_{\kappa'}+i\epsilon}+\frac{\hat{n}_{-\kappa''}+1}{i(\omega_{\kappa'}+i\epsilon)}\right)\Bigg]. \quad (3.17a)$$

On the other hand, when $-\kappa'=\kappa$, the secular term can be similarly obtained as

$$-\frac{9}{\hbar^2}\sum_{\kappa'\kappa''}B^{*}_{-\kappa\kappa''}a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)$$

$$\times\left[(B_{\kappa''-\kappa'\kappa'}+B^{*}_{-\kappa''-\kappa'\kappa'})\left(\frac{\hat{n}_{\kappa'}}{i(\omega_{\kappa''}+i\epsilon)}+\frac{\hat{n}_{-\kappa'}+1}{i(\omega_{\kappa''}+i\epsilon)}\right)\right]. \quad (3.17b)$$

Collecting (3.17a) and (3.17b), we get the contribution to secular terms after some rearranging as $\hat{S}'_2$:

$$\hat{S}'_2=-\frac{9}{\hbar^2}a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)\sum_{\kappa'\kappa''}\left(B^{*}_{k\kappa',-\kappa}\right.$$

$$\times(B_{\kappa',-\kappa''\kappa''}+B^{*}_{-\kappa',-\kappa''\kappa''})\frac{\hat{n}_{\kappa''}+\hat{n}_{-\kappa''}+1}{i(\omega_{\kappa'}+i\epsilon)}$$

$$\left.+B^{*}_{\kappa-\kappa\kappa''}(B_{\kappa''-\kappa'\kappa'}+B^{*}_{-\kappa''-\kappa'\kappa'})\frac{\hat{n}_{\kappa'}+\hat{n}_{-\kappa'}+1}{i(\omega_{\kappa''}+i\epsilon)}\right). \quad (3.18)$$

Using Eqs. (3.16) and (3.18) for $\hat{S}_2$ and $\hat{S}'_2$ respectively, we can write the equation of motion (3.10) as follows:

$$\frac{\partial}{\partial\tau_0}a^{(2)}_{\kappa}(\tau)+\frac{\partial}{\partial\tau_2}a^{(0)}_{\kappa}(\tau)=-i\omega_\kappa a^{(2)}_{\kappa}(\tau)+\hat{S}_2+\hat{S}'_2$$

+ terms which do not give rise to secular terms. (3.19)

To remove the secular behavior in $a^{(2)}_{\kappa}(\tau)$, the MTSPE consists in choosing the term $(\partial/\partial\tau_2)a^{(0)}_{\kappa}(\tau)$ so as to cancel $\hat{S}_2+\hat{S}'_2$ in Eq. (3.19),

$$\frac{\partial}{\partial\tau_2}a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)=-i\hat{\xi}_\kappa a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0), \quad (3.20)$$

where we have written

$$\hat{S}_2+\hat{S}'_2=-i\hat{\xi}_\kappa a^{(0)}_{\kappa}(0)\exp(-i\omega_\kappa\tau_0)$$

and the operator $\hat{\xi}_\kappa$ is given by

$$\hat{\xi}_\kappa=-\frac{18}{\hbar^2}\sum_{\kappa'\kappa''}B^{*}_{k\kappa',\kappa''}\left(B_{\kappa''\kappa'\kappa}\frac{\hat{n}_{\kappa''}+\hat{n}_{-\kappa'}+1}{(\omega_\kappa+\omega_{\kappa'}+\omega_{\kappa''}+3i\epsilon)}\right.$$

$$+B^{*}_{-\kappa,-\kappa',-\kappa''}\frac{\hat{n}_{\kappa''}-\hat{n}_{\kappa'}}{(\omega_{\kappa'}-\omega_{\kappa''}-\omega_\kappa-i\epsilon)}$$

$$+B_{\kappa''\kappa'\kappa}\frac{\hat{n}_{-\kappa''}-\hat{n}_{-\kappa'}}{(\omega_{\kappa'}+\omega_\kappa-\omega_{-\kappa''}+i\epsilon)}$$

$$\left.+B_{-\kappa',-\kappa''-\kappa}\frac{\hat{n}_{-\kappa''}+\hat{n}_{\kappa'}+1}{(\omega_{-\kappa'}+\omega_{-\kappa''}-\omega_\kappa+i\epsilon)}\right)$$

$$-\frac{9}{\hbar^2}\sum_{\kappa'\kappa''}\left(B^{*}_{k\kappa',-\kappa}(B_{\kappa',-\kappa''\kappa''}+B^{*}_{-\kappa',-\kappa''\kappa''})\frac{\hat{n}_{\kappa''}+\hat{n}_{-\kappa''}+1}{i(\omega_{\kappa'}+i\epsilon)}\right.$$

$$\left.+B^{*}_{-\kappa\kappa''}(B_{\kappa''-\kappa'\kappa'}+B^{*}_{-\kappa''-\kappa'\kappa'})\frac{\hat{n}_{\kappa'}+\hat{n}_{-\kappa'}+1}{i(\omega_{\kappa''}+i\epsilon)}\right). \quad (3.21)$$

In the thermodynamic limit of $N\to\infty$, the operator $\hat{\xi}_\kappa$ commutes with $a^{(0)}_{\kappa}(0)$. Hence Eq. (3.20) can be integrated to give

$$a^{(0)}_{\kappa}(0,0,\tau_2,\tau_3,\cdots)=a^{(0)}_{\kappa}(0,0,0,\tau_3,\cdots)\exp(-i\hat{\xi}_\kappa\tau_2). \quad (3.22)$$

As before, to the second order $O(\lambda^2)$, we can set $\tau_k=0$, for $k\geq 3$, and

$$a^{(0)}_{\kappa}(\tau)=a^{(0)}_{\kappa}(\tau=0)\exp(-i\hat{\xi}_\kappa\tau_2-i\omega_\kappa\tau_0). \quad (3.23)$$

The expectation value of $\hat{\xi}_\kappa$ is, of course, dependent on the state of the system.

For a general state $|\{n_\kappa\}\rangle$, we have

$$\langle\hat{\xi}_\kappa\rangle = -\frac{18}{\hbar^2}\sum_{\kappa'\kappa''} B_{\kappa\kappa'\kappa''}B_{-\kappa-\kappa'-\kappa''}$$
$$\times\left(\frac{n_{\kappa''}-n_{-\kappa'}+1}{\omega_\kappa+\omega_{\kappa'}+\omega_{\kappa''}+3i\epsilon}+\frac{n_{-\kappa''}-n_{-\kappa'}}{\omega_\kappa+\omega_{\kappa'}-\omega_{\kappa''}+i\epsilon}\right.$$
$$\left.+\frac{n_{\kappa'}-n_{\kappa''}}{\omega_\kappa+\omega_{\kappa''}-\omega_{\kappa'}+i\epsilon}-\frac{n_{-\kappa''}+n_{\kappa'}+1}{\omega_\kappa-\omega_{\kappa'}-\omega_{\kappa''}-i\epsilon}\right)$$
$$-\frac{36}{\hbar^2}\sum_{\kappa'\kappa''} B_{\kappa-\kappa\kappa'}B_{\kappa''-\kappa''-\kappa'}\frac{n_{\kappa''}+n_{-\kappa''}+1}{(\omega_{\kappa'}+i\epsilon)}, \quad (3.24)$$

where we have used the facts that $B^*_{\kappa\kappa'\kappa''}=B_{-\kappa-\kappa'-\kappa''}$ and that $B_{\kappa\kappa'\kappa''}$ are completely symmetric in their indices.

We shall now let $\epsilon\to 0$; to take care of this singular behavior, we use the identity

$$\lim_{\epsilon\to 0}[1/(x\pm i\epsilon)]=(1/x)_p\mp i\pi\delta(x)$$

with p denoting the principal part. The operator $\hat{\xi}_\kappa$ can therefore be written as

$$\hat{\xi}_\kappa=\hat{\Delta}_\kappa - i\hat{\Gamma}_\kappa \quad (3.25)$$

and its average as

$$\langle\hat{\xi}_\kappa\rangle=\langle\hat{\Delta}_\kappa\rangle - i\langle\hat{\Gamma}_\kappa\rangle.$$

From (3.24) we get

$$\langle\hat{\Delta}_\kappa\rangle = -\frac{18}{\hbar^2}\sum_{\kappa'\kappa''} B_{\kappa\kappa'\kappa''}B_{-\kappa-\kappa'-\kappa''}$$
$$\times\left(\frac{n_{\kappa''}+n_{-\kappa'}+1}{(\omega_\kappa+\omega_{\kappa'}+\omega_{\kappa''})_p}+\frac{n_{-\kappa''}-n_{-\kappa'}}{(\omega_\kappa+\omega_{\kappa'}-\omega_{\kappa''})_p}\right.$$
$$\left.+\frac{n_{\kappa'}-n_{\kappa''}}{(\omega_\kappa+\omega_{\kappa''}-\omega_{\kappa'})_p}-\frac{n_{-\kappa''}+n_{\kappa'}+1}{(\omega_\kappa-\omega_{\kappa'}-\omega_{\kappa''})_p}\right)$$
$$-\frac{36}{\hbar^2}\sum_{\kappa'\kappa''} B_{\kappa-\kappa\kappa'}B_{\kappa''-\kappa''\kappa'}\frac{n_{\kappa''}+n_{-\kappa''}+1}{(\omega_{\kappa'})_p} \quad (3.26)$$

and

$$\langle\hat{\Gamma}_\kappa\rangle=\frac{18}{\hbar^2}\pi\sum_{\kappa'\kappa''} B_{\kappa\kappa'\kappa''}B_{-\kappa-\kappa'-\kappa''}$$
$$\times[(n_{\kappa''}+n_{-\kappa'}+1)\delta(\omega_\kappa+\omega_{\kappa'}+\omega_{\kappa''})$$
$$+(n_{-\kappa''}-n_{-\kappa'})\delta(\omega_\kappa+\omega_{\kappa'}-\omega_{\kappa''})$$
$$+(n_{\kappa'}-n_{\kappa''})\delta(\omega_\kappa+\omega_{\kappa''}-\omega_{\kappa'})$$
$$+(n_{-\kappa''}+n_{\kappa'}+1)\delta(-\omega_\kappa-\omega_{\kappa'}-\omega_{\kappa''})]. \quad (3.27)$$

The above results for the frequency shift ($\langle\hat{\Delta}_\kappa\rangle$) and the phonon lifetime ($\langle\hat{\Gamma}_\kappa\rangle$) agrees with the results obtained by Wallace[4] and others.[13] It is obvious that the present formalism can be employed to solve the same problem with cubic and quartic interaction. Equation (3.22) now becomes

$$a_\kappa^{(0)}(t)=a_\kappa^{(0)}(0)\exp[-i(\omega_\kappa+\hat{\Delta}_\kappa)t]\exp(-\hat{\Gamma}_\kappa t), \quad (3.28)$$

which shows that, in the long time limit, the phonon operators decay exponentially with decay constant $\langle\hat{\Gamma}_\kappa\rangle$, while, in the time scale ($\tau_0\sim\omega_\kappa^{-1}$), they are oscillatory.

IV. CORRELATION FUNCTION

As a simple application of our general formulation, let us calculate the correlation function of the phonon creation and annihilation operators in the von Hove's "$\lambda^2 t$" limit.[14] By definition, we want to calculate the ensemble average at temperature T of the quantity $\langle a_\kappa^\dagger(t)a_\kappa(0)\rangle$ in the limit that the coupling constant λ is very small ($\lambda\to 0$), but time t is very large ($t\to\infty$), such that $\lambda^2 t$ is finite. Other correlation functions such as momentum–momentum correlation function can be expressed in terms of them through the normal coordinate transformation.

The simplification in the $\lambda^2 t$ limit is enormous, because, in that limit, $a_\kappa^\dagger(t)$ becomes

$$a_\kappa^\dagger(t)\to a_\kappa^{(0)\dagger}(t)=a_\kappa^{(0)\dagger}(0)\exp[i(\omega_\kappa+\hat{\Delta}_\kappa)t]\exp(-\hat{\Gamma}_\kappa t) \quad (4.1)$$

and all higher order terms go to zero. Furthermore, the ensemble average $\langle a_\kappa^\dagger(t)a_\kappa(0)\rangle$ can be evaluated using the unperturbed Hamiltonian H_0; i.e.,

$$\lim_{\lambda^2 t}\langle a_\kappa^\dagger(t)a_\kappa(0)\rangle$$
$$=\mathrm{Tr}\{a_\kappa^\dagger(0)a_\kappa(0)\exp[-i(\omega_\kappa+\hat{\Delta}_\kappa)t-\hat{\Gamma}_\kappa t]\exp(-\beta H_0)\}. \quad (4.2)$$

Note that we can change the order of $a_\kappa(0)$ in the above formula in the limit $N\to\infty$.

Let $|\{n_\kappa\}\rangle$ be the eigenstate of H_0, so that

$$a_\kappa^\dagger(0)a_\kappa(0)\,|\{n_\kappa\}\rangle=n_\kappa\,|\{n_\kappa\}\rangle,$$
$$\hat{\Delta}_\kappa\,|\{n_\kappa\}\rangle=\langle\hat{\Delta}_\kappa\rangle\,|\{n_\kappa\}\rangle,$$
$$\hat{\Gamma}_\kappa\,|\{n_\kappa\}\rangle=\langle\hat{\Gamma}_\kappa\rangle\,|\{n_\kappa\}\rangle, \quad (4.3)$$

where $\langle\hat{\Delta}_\kappa\rangle$ and $\langle\hat{\Gamma}_\kappa\rangle$ are given by (3.26) and (3.27). The trace (4.2) can be written in the form

$$\langle a_\kappa^\dagger(t)a_\kappa(0)\rangle$$
$$=\sum_{\{n_\kappa\}} n_\kappa\exp[-i(\omega_\kappa+\langle\hat{\Delta}_\kappa\rangle)t-\langle\hat{\Gamma}_\kappa\rangle t]\exp[-\beta\textstyle\sum\hbar\omega_\kappa n_\kappa]. \quad (4.4)$$

To $O(\lambda^2)$, both $\langle\hat{\Delta}_\kappa\rangle$ and $\langle\hat{\Gamma}_\kappa\rangle$ are linear functions of the occupation number $\{n_\kappa\}$. The summation in (4.4) can therefore be carried out easily. Let us write

$$\langle\hat{\Delta}_\kappa\rangle=\sum_{\kappa'}\Delta_1(\kappa,\kappa')n_{\kappa'}+\Delta_2(\kappa),$$
$$\langle\hat{\Gamma}_\kappa\rangle=\sum_{\kappa'}\Gamma_1(\kappa,\kappa')n_{\kappa'}+\Gamma_2(\kappa) \quad (4.5)$$

and substitute (4.5) into (4.4); we get

$$\langle a_\kappa^\dagger(t)a_\kappa(0)\rangle=\bar{n}_\kappa\exp[i[\omega_\kappa+\Delta_2(\kappa)]t-\Gamma_2(\kappa)t$$
$$\times\prod_{\kappa'} f(\kappa,\kappa' t) \quad (4.6)$$

with

$$f(\kappa\kappa' t)$$
$$=\frac{1-\exp(-\beta\hbar\omega_{\kappa'})}{1-\exp[i\Delta_1(\kappa\kappa')t-\Gamma_1(\kappa\kappa')t-\beta\hbar\omega_{\kappa'}]} \quad (4.7)$$

and $\bar{n}_\kappa=[\exp(\beta\hbar\omega_\kappa)-1]^{-1}$ is the average number of phonon at temperature T. The expression (4.7) is the general formula for the correlation function, and is very complicated. We shall try to analyze it for some very simple one-dimension model later.

In conclusion, we have presented here a very simple

nondiagrammatic method to solve the equations of motion for an anharmonic lattice. In our opinion, the method is not only straightforward but very useful.

ACKNOWLEDGMENT

We are most grateful to Professor E. W. Montroll for introducing us to the theory of nonlinear oscillations, and giving us many inspiring discussions.

[1]J. J. Kokkedee, Physica **268**, 374 (1962).
[2]A. A. Maradudin and A. E. Fein, Phys. Rev. **128**, 2589 (1962).
[3]R. A. Cowley, Advan. Phys. **12**, 421 (1963).
[4]D. C. Wallace, Phys. Rev. **152**, 247 (1966).
[5]Y. K. Wang and I. C. Khoo, Optics Commun. **11**, 323 (1974).
[6]N. M. Krylov and N. N. Bogoliubov, *Introduction to Nonlinear Mechanics*, translated by S. Lefshetz (Princeton U. P., Princeton, N. J., (1947).
[7]R. H. G. Helleman and E. W. Montroll, Physics **74**, 22 (1974).
[8]P. S. Lee, Y. C. Lee and C. T. Chang, Phys. Rev. A **8**, 1722 (1973).
[9]B. B. Varga and S. O. Aks, J. Math. Phys. **15**, 149 (1974).
[10]E. Frieman, J. Math. Phys. **4**, 410 (1963).
[11]M. Born and K. Huang, *Dynamical Theory of Crystal Lattice* (Clarendon, Oxford, 1954).
[12]R. E. Peierls, *Quantum Theory of Solids* (Clarendon, Oxford, 1955).
[13]R. A. Cowley, Advan. Phys. **12**, 421 (1963); W. Ludwig, J. Phys. Chem. Solids **4**, 283 (1958); D. C. Wallace, Phys. Rev. **133**, A153 (1964).
[14]L. van Hove, Physics **21**, 517 (1955); **23**, 441 (1957); also see L. van Hove, in *La théorie des gazes neutre et ionises*, edited by C. Dewitt and J. G. Deteuf (Wiley, New York, 1959).

Simple supersymmetries*

Peter G. O. Freund and Irving Kaplansky

University of Chicago, Chicago, Illinois 60637
(Received 6 October 1975)

Two infinite families of simple graded Lie algebras (GLA's) over the complex numbers are described: the special linear algebras SL($m|n$) [whose Bose sector is the direct sum of a one-dimensional algebra with the ordinary Lie algebra SL(m)×SL(n)], and the orthosymplectic algebras OSp($2r|s$) [with Bose sector Sp($2r$)×O(s)]. The GLA's of physics fit into these two families either directly or via Inönü–Wigner contraction. These algebras along with further exceptional GLA's constitute all the classical GLA's, i.e., all GLA's with a nondegenerate metric (not necessarilly Killing) form. The existence of infinite families of hyperexceptional GLA's, i.e., of GLA's that are simple but not classical is pointed out.

1. INTRODUCTION

The supersymmetries discovered in dual models and in quantum field theories are expressed in terms of graded Lie algebras (GLA's) of conserved charges.[1-4] The graded Lie bracket of two charges is not always a commutator as for ordinary Lie algebras; for two fermionic charges it is an anticommutator. Supersymmetries mix bosons with fermions. Progress in the field of supersymmetries has been somewhat hampered by the lack of information on the structure of (at least) the finite-dimensional GLA's. Since all work on supersymmetries has thus far been quite divorced from experiment, a classification of GLA's may spur the choice of a more realistic candidate for the purposes of particle physics. Motivated by this and by the purely mathematical interest of the problem we describe two infinite families of simple GLA's (i.e., GLA's that have no invariant subalgebras) over the complex numbers. The extent to which these account for all simple GLA's is discussed in Sec. 4.

2. SPECIAL LINEAR GLA's

An algebra of this family is determined by two positive integers m and n and is written SL($m|n$). We write the members of SL($m|n$) as $(m+n)\times(m+n)$ matrices, divided into blocks as indicated:

$$\left(\begin{array}{c|c} m\times m & m\times n \\ \hline n\times m & n\times n \end{array}\right).$$

A boson (even element) B has the form

$$B=\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix},$$

$$\mathrm{tr}B\equiv \mathrm{tr}a-\mathrm{tr}d=0.$$

A fermion (odd element) F has the form

$$F=\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}.$$

The bracket of two B's or a B and a F is a commutator,

$$[B_1B_2]=\begin{pmatrix} a_1a_2-a_2a_1 & 0 \\ 0 & d_1d_2-d_2d_1 \end{pmatrix}$$

$$[BF]=\begin{pmatrix} 0 & ab-bd \\ dc-ca & 0 \end{pmatrix}.$$

The bracket of two F's is an anticommutator (Jordan product),

$$[F_1F_2]=\begin{pmatrix} b_1c_2+b_2c_1 & 0 \\ 0 & c_1b_2+c_2b_1 \end{pmatrix}.$$

The dimension of SL($m|n$) is $(m+n)^2-1$. For $m\neq n$, SL($m|n$) is a simple GLA, with Bose sector a direct sum of a one-dimensional algebra and the ordinary Lie algebra SL(m)×SL(n). For $m=n$ (we call this the "balanced" case) the identity unit matrix lies in SL($m|n$) and constitutes its one-dimensional center. To get a simple algebra one must divide by the center, getting an algebra PSL($m|m$) of dimension $4m^2-2$. (The "P" stands for projective). For $m\geq 2$, PSL ($m|m$) is simple [its Bose sector is SL(m)×SL(m)]. The case $m=1$ is exceptional and SL(1|1) is nilpotent.

Special linear GLA's are also describable by creation and annihilation operators $a_i^\epsilon(\epsilon=\pm,\ i=1,\ldots,m+n)$. The first m pairs are fermionic, while the remaining n pairs are bosonic. The canonical quantization relations are

$$[a_i^\epsilon, a_j^{-\epsilon'}]=\delta_{\epsilon\epsilon'}\delta_{ij}.$$

As above, this bracket denotes commutator except when both operators are fermionic in which case it is an anticommutator.

Introduce $(m+n)^2$ operators

$$G_{ij}=a_i^+a_j^-. \tag{2.2}$$

Of these, m^2+n^2 are bosonic (those with $i,j\leq m$ and those with $i,j>m$) and the remaining $2mn$ are fermio-fermionic. The G's close under bracketing and span a GLA. This GLA corresponds to the GLA of all $(m+n)\times(m+n)$ matrices in the block form given above (and not just those of trace 0). To implement the correspondence, make G_{ij} correspond to the usual matrix unit e_{ij}.

3. ORTHOSYMPLECTIC GLA's

In describing orthosymplectic GLA's one starts with a graded vector space V with a fermionic component of even dimension $2r$ and bosonic component of dimension s. A metric is imposed which is antisymmetric on the Fermi sector and symmetric on the Bose sector. This metric can be brought to the form

$$G=\begin{matrix}2r\\ s\end{matrix}\begin{pmatrix} C & 0 \\ 0 & I \end{pmatrix}\begin{matrix}2r\\ s\end{matrix}, \tag{3.1a}$$

where C is the antisymmetric $2r\times 2r$ matrix

$$C=\begin{pmatrix} 0 & 1 & & & & \\ -1 & 0 & & & & \\ & & 0 & 1 & & \\ & & -1 & 0 & & \\ & & & & \ddots & \\ & & & & 0 & 1 \\ & & & & -1 & 0 \end{pmatrix}, \quad C^2=-1, \quad C^T=-C, \tag{3.1b}$$

and I is the $s\times s$ identity matrix. The inner product of two vectors x and y of the graded vector space is given by

$$(x,y)=x^TGy, \tag{3.2}$$

x and y being treated as column vectors.

Consider the linear transformations on V. It is meaningful to grade them, and in particular a linear transformation may be Bose or Fermi (even or odd). We impose on U the condition of "antisymmetry." This means

$$(Ux,y)+(x,Uy)=0, \tag{3.3a}$$

except when x and U are both fermionic when it is changed to

$$(Ux,y)-(x,Uy)=0, \tag{3.3b}$$

since the passage from the first to the second term in (3.3) involves an interchange of x and U.

Write U in matrix block form as

$$U=\begin{matrix}2r\\ s\end{matrix}\begin{pmatrix} a & p \\ q & b \end{pmatrix}\begin{matrix}2r\\ s\end{matrix}. \tag{3.4}$$

Then Eqs. (3.3a) and (3.3b) become

$$a^T=CaC, \quad b^T=-b, \quad q=-p^TC \tag{3.5}$$

so that the a's are $2r\times 2r$ skew-symplectic (the Lie algebra of the symplectic group), the b's are $s\times s$ skew-symmetric (the Lie algebra of the orthogonal group), and p is an arbitrary $2r\times s$ matrix which completely determines the $s\times 2r$ matrix q.

The set of all such U's forms a graded Lie algebra. We call it *orthosymplectic* and write it $\mathrm{OSp}(2r|s)$. The dimension of $\mathrm{OSp}(2r|s)$ is $r(2r+1)+s(s-1)/2+2rs$ and it is simple for every r and s. Here is a table of the first few dimensions:

r	1	1	1	2	1	2	1	2	3	1
s	1	2	3	1	4	2	5	3	1	6
dim	5	8	12	14	17	19	23	25	27	30
	2	3	1	2	3	4	1	2	3	4
	4	2	7	5	3	1	8	6	4	2
	32	34	38	40	42	44	47	49	51	53

The Bose sector of $\mathrm{OSp}(2r|s)$ is the ordinary Lie algebra $\mathrm{Sp}(2r)\times \mathrm{O}(s)$. The GLA's $\mathrm{OSp}(2|2)$ and $\mathrm{SL}(2|1)$ are isomorphic; there are no other "accidental" isomorphisms.

4. THE KILLING FORM AND FORMS ARISING FROM OTHER REPRESENTATIONS

The Killing form on a graded Lie algebra with basis G_i is defined by

$$g_{ij}\equiv(G_i,G_j)=\mathrm{Tr}(\mathrm{Ad}(G_i)\,\mathrm{Ad}(G_j)).$$

Here $\mathrm{Ad}(G_i)$ is the adjoint representation, and, as in Sec. 2, the trace is taken positive on the Bose sector, negative on the Fermi sector. For the above algebras the Killing form is nondegenerate except in what we call the "balanced" case, i.e., $m=n$ for $\mathrm{SL}(m|n)$ and $2r+2=s$ for $\mathrm{OSp}(2r|s)$; in the balanced cases the Killing form is identically 0.

The Killing form generalizes to any representation S by using $\mathrm{Tr}(S(G_i)S(G_j))$ and it is advantageous to do so. For example, in the case of a balanced orthosymplectic GLA the matrices occurring in the definition furnish a representation and for this representation the induced form is nondegenerate.

As remarked above, the balanced cases $\mathrm{SL}(m|m)$ of the special linear GLA's are not simple and it is necessary to divide by the center so as to get $\mathrm{PSL}(m|m)$. Now there is no longer a useful representation in sight. But this is easily remedied. As in Ref. 5 we introduce *projective* representations, this meaning homomorphisms into an algebra $\mathrm{PSL}(m|m)$. Attached to any projective representation there is a form, for we can unambiguously calculate the form modulo the center.

We propose to call a GLA *classical* if it is simple and possesses an ordinary or projective representation for which the induced form is nondegenerate. The algebras $\mathrm{SL}(m|n)$ [to be replaced by $\mathrm{PSL}(m|m)$ if $n=m$] and $\mathrm{OSp}(2r|s)$ are classical. One of us (I.K.) is studying the classification of classical simple GLA's. At this writing the evidence suggests that in addition to the special linear and orthosymplectic GLA's there may be an infinite family of 17-dimensional ones, a 31-dimensional one, and a 40-dimensional one. The prospective 31-dimensional one has $G_2\times \mathrm{SL}(2)$ as its Bose sector. The other exceptional ordinary Lie algebras (F_4, E_6, E_7, E_8) do not make an appearance.

In addition there exist nonclassical simple GLA's. The most obvious of these is the algebra of all derivations of a Grassman algebra (or in other words the algebra of general coordinate transformations in a space with fermionic dimensions only). This is incidentally a perfect analogue of the first of Cartan's four families of infinite pseudogroups,[6] and the situation shows a remarkable resemblance to what is known and conjectured in characteristic p.[7] Warren Nichols is investigating nonclassical simple GLA's.

Pais and Rittenberg[8] study GLA's subject to the strong restriction that the Killing form is nonsingular both for the whole algebra and for its boson part (this is their "definition" of semisimplicity). After developing the theory to a certain point they make the further tacit

assumption that the boson part is simple (and not just semisimple). It is as a consequence of this extra assumption that they find only the algebras $OSp(2r|1)$. Indeed of all the classical algebras these are the only ones that meet all the restrictions of Pais and Rittenberg.

The nature of the representations of classical GLA's remains to be investigated. It would be desirable to determine all the irreducible representations and settle for which simple GLA's the general representation is completely reducible (it is easy to see that this is true for some and false for others).

5. REAL FORMS

It must be emphasized that all the above discussion referred to algebras over the complex numbers. A study of the real forms of these complex algebras has yet to be made.

Two real forms deserve special mention. The first are designated *special unitary* and written $SU(m|n)$ [not to be confused with the ordinary Lie algebras $SU(m,n)$]. In matrix form (with $\dagger$ denoting complex conjugate transpose), $SU(m|n)$ consists of all matrices

$$\begin{matrix} m & n \\ \begin{pmatrix} a & b \\ ib^\dagger & d \end{pmatrix} & \begin{matrix} m, \\ n \end{matrix} \end{matrix}$$

where a and d are skew Hermitian ($a^\dagger = -a$, $d^\dagger = -d$) and $\mathrm{tr}(a-d)=0$. The Bose sector of $SU(m|n)$ is $SU(m)\times SU(n)\times U(1)$ for $m \neq n$. On extending the coefficients from the real numbers to the complex numbers, one converts $SU(m|n)$ to $SL(m|n)$. For $m \neq n$, $SU(m|n)$ is simple, but for $m=n$, $SU(m|m)$ must—like $SL(m|m)$—be divided by a one-dimensional center to achieve simplicity.

In the second real form we change the orthogonal part of an orthosymplectic GLA so as to allow an indefinite metric and leave the symplectic part unchanged. Write $s=s_1+s_2$ and let g be the diagonal matrix with s_1 one's and s_2 minus one's down the diagonal. The matrix G giving our metric is now

$$\begin{matrix} 2r & s \\ \begin{pmatrix} C & 0 \\ 0 & g \end{pmatrix} & \begin{matrix} 2r \\ s \end{matrix} \end{matrix}$$

and the GLA $OSp(2r|s_1,s_2)$ consists of all

$$\begin{pmatrix} a & p \\ q & b \end{pmatrix}$$

with

$$a^T = -CaC,$$

$$b^T = -gbg,$$

$$q = -gp^TC.$$

Its Bose sector is $Sp(2r)\times O(s_1,s_2)$. All these real algebras become $OSp(2r|s)$ over the complex numbers.

6. GRADING BY THE INTEGERS

It is to be noted that we have been discussing only GLA's which are graded "mod two." It is desirable also to study GLA's which are graded by the integers. Such an algebra L is a union of subspaces L_i where i ranges over all the integers (positive, zero, and negative). A member of L_i is said to have degree i. The bracket of an element of degree i with one of degree j has degree $i+j$.

Of course from a GLA graded in this fashion we can derive one graded mod two by grouping together all even i to form the bosonic part, and all odd i to form the fermionic part. But this does not answer all questions, and at some future time it is hoped to make a comprehensive study.

7. THE GLA's OF PHYSICS AS CLASSICAL GLA's

The GLA's thus far encountered in physics are (after complexification) special linear or orthosymplectic or else they are Inönü—Wigner contractions of these GLA's.

First of all, the five-dimensional GLA discovered by Neveu, Schwartz and Ramond[1] is precisely $OSp(2|1)$ after complexification.

Wess and Zumino[2] studied a 24-dimensional GLA with a 16-dimensional Bose sector spanned by the 15 generators of the conformal transformations $SO(4,2)$ and by the $U(1)$ of γ_5-transformations. Now $SO(4,2)$ (the special orthogonal group on a six-dimensional space carrying a metric with 4 one's and 2 minus one's) becomes on complexification the algebra of 4×4 matrices of trace 0. Taking the direct sum with a one-dimensional algebra gives us the Bose sector of $SL(4|1)$, and indeed the whole Wess—Zumino algebra becomes $SL(4|1)$ on complexification.

The more popular 14-dimensional GLA of Volkov, Akulov, Wess, and Zumino[2,3] has the 10-dimensional Poincaré algebra as its Bose sector and is not semisimple. It can be obtained however from the simple 14-dimensional orthosymplectic algebra $OSp(4|1)$ by Inönü—Wigner contraction. Indeed the Bose sector $Sp(4)$ of $OSp(4|1)$ is isomorphic after complexification to the de Sitter algebra[9] $SO(3,2)$. If, from the de Sitter generators $M_{ab}(=-M_{ba})$, $a, b=1,\ldots,5$ we define new generators $\bar{M}_{ab}=M_{ab}$ for $a,b=1,\ldots,4$, $\bar{M}_{a5}\equiv\lambda M_{a5}$, the algebra of the $\bar{M}_{ab}$ $(a,b=1,\ldots,5)$ will have λ-dependent structure constants and in the limit $\lambda\to 0$ the Poincaré algebra is obtained. If at the same time one defines barred Fermi generators $\bar{F}_\alpha=\sqrt{\lambda}F_\alpha$, $\alpha=1,\ldots,4$, corresponding to the four Fermi generators of $OSp(4|1)$ then in the limit $\lambda\to 0$ the ten barred Bose and four barred Fermi generators span precisely the nonsemisimple Volkov—Akulov—Wess—Zumino GLA.

Finally, in the "graded" Riemannian geometry[10] of supergauge theories one considers manifolds that can be locally mapped onto a flat graded space with orthosymplectic metric.

We thus see that the GLA's occurring in physics thus far are either classical GLA's or contractions thereof.

8. CONCLUSIONS

We have defined two families of simple GLA's,

special linear and *orthosymplectic*. Every GLA which has hitherto arisen in physics is either one of these algebras or a contraction of one of them.

We call a GLA *classical* if it is simple and possesses an ordinary or projective representation whose induced trace form is nondegenerate. The classification of classical GLA's is under investigation. Evidence points to the possible existence of additional classical GLA's with dimensions 17, 31, and 40.

In addition to these "exceptional" GLA's there exist still other ("hyperexceptional") simple GLA's which are not classical at all, i.e., for them the trace form attached to every representation is identically 0. There appears to be a connection between these hyperexceptional algebras and Cartan's infinite pseudogroups, of the same kind as has been observed for ordinary Lie algebras of characteristic p.

Note added in proof (Dec. 26, 1975): There have been numerous further developments, notably a comprehensive announcement by V.G. Kac, Functional Anal. Appl. 9, 91 (1975). Several GLA "newsletters" have been prepared in an attempt to keep interested people informed. They are available from the authors.

ACKNOWLEDGMENT

We wish to thank Professor I.N. Herstein for conversations that helped bring about this collaboration.

*Work supported in part by the National Science Foundation: Contract Nos. MPS74-08833 and GP-37575 X.

[1]P. Ramond, Phys. Rev. D 3, 2415 (1971); A. Neveu and J. Schwarz, Nucl. Phys. B 31, 86 (1971).
[2]J. Wess and B. Zumino, Nucl. Phys. B 70, 39 (1974).
[3]D.V. Volkov and V.P. Akulov, Phys. Lett. B 46, 109 (1973).
[4]See the review articles of B. Zumino, *Proceedings of the XVII International Conference on High Energy Physics*, The Science Research Council 1974, London, July, 1974; A Salam and J. Strathdee, Phys. Rev. D 11, 1521 (1975); see also L. Corwin, Y. Ne'eman and S. Sternberg, Rev. Mod. Phys. 47, 573 (1975) where some of the algebras discussed here and some (in our terminology) hyperexceptional GLA's are considered.
[5]I. Kaplansky, *Lie Algebras and Locally Compact Groups* (University of Chicago Press, Chicago, 1971).
[6]I.M. Singer and S. Sternberg, J. Analyse Math. 15, 1 (1965).
[7]A.I. Kostrikin and I.R. Shafarevich, Dokl. Akad. Nauk SSSR 168, 740 (1968) [Sov. Math. Dokl. 7, 715 (1966)].
[8]A. Pais and V. Rittenberg, J. Math. Phys. 16, 2062 (1975).
[9]See, e.g., F Gürsey, in *Group-Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, N.Y., 1964).
[10]P. Nath and R. Arnowitt, Phys. Lett. B 56, 177 (1975); Y.M. Cho and P.G.O. Freund, Phys. Rev. D 12, 1711 (1975).

Approach of the statistical theory of light scattering to the phenomenological theory*

V. Varatharajulu† and D. J. Vezzetti

Department of Physics, University of Illinois at Chicago Circle, Chicago, Illinois 60680
(Received 27 June 1975)

A fully statistical treatment of the spectrum of light scattered by a simple fluid is given. The results are shown to be in close accord with the phenomenological theory of the same process.

1. INTRODUCTION

Expressions for the spectral density of light scattered by a simple fluid may be obtained from either a phenomenological or a statistical treatment of the problem.

The phenomenological theory dates to the work of von Smoluchowski[1] and of Einstein.[2] Later and fuller treatments have been given by Benedek[3] and van Kampen[4] among others. A distinguishing feature of these treatments is that the refractive index (or the dielectric tensor) of the medium appears explicitly in the expression for the spectrum.

Carrying over the work of van Hove,[5] Komarov and Fisher[6] were the first to give a statistical treatment of the inelastic scattering in simple fluids. They used first Born approximation in the external field. This work was later extended by Tanaka.[7] In these works, the dependence of the spectrum on the refractive index is not clearly exposed. This dependence has been, at least in connection with the corresponding elastic problem, a subject of much discussion.[8]

In the recent literature, two works bear directly on this connection between the statistical and phenomenological theories. First, Bedeaux and Mazur[9] have given a detailed statistical treatment of the dielectric tensor of a simple fluid. Their work, which is not restricted to dilute systems, has, from our point of view, the advantage that their choice for the effective field acting on a molecule of the medium gives the basic equations of the statistical theory a much simpler form than in previous works. They show that wave propagation in the fluid can be described as propagation between density fluctuations through a medium characterized by the Lorenz–Lorentz value of the dielectric tensor. This is achieved by what amounts to a resummation of the basic series describing propagation between particles through a vacuum and represents the first step in the approach of the statistical theory to the phenomenological theory.

In the second recent paper, Felderhof[10] has reformulated and extended aspects of Ref. 9. He has determined in a fully statistical way the differential elastic scattering cross section and turbidity for optical waves incident on a simple fluid and has shown that these results agree precisely with the phenomenological expressions for a medium characterized by the Lorenz–Lorentz dielectric tensor.

In the present paper, we are concerned with the inelastic scattering of light from simple fluids. We shall not discuss any aspects of the absorption or turbidity problem. In Secs. 2, 3, and 4, we again reformulate the statistical problem and compute the spectral density to the same order as in Ref. 10. This is shown to agree with the phenomenological result. These sections are a more or less direct extension of Felderhof's work to the inelastic case. The basic result of these sections is Eq. (2.25) which expresses the fluctuations of the electric field as a series in orders of density fluctuations. This series allows, in principle, the evaluation of all higher order terms in the spectral density in terms of the correlation functions of the density fluctuations. The "driving" term in this series is essentially the Lorentz effective field.

In Secs. 5 and 6 we take into account the effects on the spectrum of some of the higher order correlation functions. This is achieved by making a Gaussian approximation to the distribution function of the density fluctuations and summing a certain subclass of diagrams. The resulting spectrum is shown to be of the same form as that obtained in Sec. 4 but with an improved value of the refractive index, namely, the refractive index within the Gaussian approximation. This demonstrates, on a completely statistical basis, the justification of the phenomenological approach in which the refractive index is introduced from the outset.

In order to conserve space, we refer frequently to Ref. 9 where derivations of several of the expressions used in this paper are given in detail. Wherever possible, we use the notation of Ref. 9.

2. FLUCTUATIONS OF THE ELECTRIC FIELD

For a system of particles in vacuuo with charge and current densities ρ_{el} and $\mathbf{J}$, the Maxwell equations read, in cgs units,

$$\begin{aligned} &\nabla\cdot\mathbf{E}=4\pi\rho_{el},\\ &\nabla\times\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t},\\ &\nabla\times\mathbf{B}=\frac{4\pi\mathbf{J}}{c}+\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t},\\ &\nabla\cdot\mathbf{B}=0. \end{aligned} \tag{2.1}$$

For a molecular system, the charge and current densities may be expanded in multipole moments as shown, e.g., by de Groot.[11] Retaining only the electric dipole contribution to the current and neglecting a small term due to the convective motion of the dipoles, the current density and electric polarization $\mathbf{P}$ of the medium are related by

$$\mathbf{J}(\mathbf{R}, t) = \frac{\partial \mathbf{P}(\mathbf{R}, t)}{\partial t} = \frac{\partial}{\partial t} \sum_i \mathbf{p}_i(t)\, \delta(\mathbf{R} - \mathbf{R}_i(t)), \tag{2.2}$$

where $\mathbf{p}_i(t)$ and $\mathbf{R}_i(t)$ are, respectively, the induced dipole moment and center of mass coordinate of the ith molecule at time t. Using Eqs. (1) and (2) together with the equation of continuity, we have, after Fourier transformation,[12]

$$\mathbf{E}(\mathbf{R}, \omega) = \mathbf{E}_0(\mathbf{R}, \omega) - \int \mathbf{F}(\mathbf{R} - \mathbf{R}', \omega) \cdot \mathbf{P}(\mathbf{R}', \omega)\, d^3R', \tag{2.3}$$

where $\mathbf{E}_0$ is the externally applied field and the tensor kernel $\mathbf{F}$ is given by

$$\mathbf{F}(\mathbf{R} - \mathbf{R}', \omega) = -\left(\nabla_{\mathbf{R}}\nabla_{\mathbf{R}} + \frac{\omega^2}{c^2}\right) \frac{\exp[i(\omega/c)\,|\mathbf{R} - \mathbf{R}'|}{|\mathbf{R} - \mathbf{R}'|}. \tag{2.4}$$

In order to obtain closed equations for the fluctuating fields $\mathbf{E}$ and $\mathbf{P}$, one requires an expression for the induced moments in terms of the field acting on a molecule of the medium. In most molecular theories of propagation and scattering one chooses, for the field acting on the ith dipole, the sum of the external field and the fields radiated by all dipoles except the ith. This choice neglects radiation damping. The choice of "effective" field used here will be that given by Bedeaux and Mazur[9] in their recent work on the dielectric function of fluids. This choice includes effects of radiation damping and (more importantly for our purposes) leads to a more symmetric form of the equations since the field of the ith dipole is not excluded. Without going into the details, which may be found in Ref. 9, the choice of effective field is defined by

$$\mathbf{p}_i(t) = \alpha_0\, \mathbf{E}_{\mathrm{eff}}(\mathbf{R}_i(t), t) \tag{2.5}$$

and

$$\mathbf{E}_{\mathrm{eff}}(\mathbf{R}_i(t), t) = \mathbf{E}_0(\mathbf{R}_i(t), t) - \sum_j \int \mathbf{H}(\mathbf{R}_i(t) - \mathbf{R}_j(t'), t - t') \cdot \mathbf{p}_j(t')\, dt'. \tag{2.6}$$

Here α_0 is the polarizability of each molecule, assumed scalar and frequency independent, and the kernel $\mathbf{H}$ is given by

$$\mathbf{H}(\mathbf{R}, t) = \begin{cases} \mathbf{F}(\mathbf{R}, t), & |\mathbf{R}| > a \\ \frac{1}{2}[\mathbf{F}(\mathbf{R}, t) - \mathbf{F}^{\dagger}(\mathbf{R}, t)], & |\mathbf{R}| < a. \end{cases} \tag{2.7}$$

The quantity a is an effective molecular hard core diameter and $\mathbf{F}^{\dagger}$ is the adjoint of $\mathbf{F}$.

Equations (5) and (6) now lead to an equation for the polarization of the medium

$$\mathbf{P}(\mathbf{R}, t) = \alpha_0 \rho(\mathbf{R}, t)(\mathbf{E}_0(\mathbf{R}, t) - \int \mathbf{H}(\mathbf{R} - \mathbf{R}', t - t') \cdot \mathbf{P}(\mathbf{R}', t')\, d^3R'\, dt'), \tag{2.8}$$

where

$$\rho(\mathbf{R}, t) \equiv \sum_i \delta(\mathbf{R} - \mathbf{R}_i(t)) \tag{2.9}$$

is the number density.

If we now combine Eqs. (3) and (8), we obtain an equation for $\mathbf{E}$ which may be iterated in terms of the external field $\mathbf{E}_0$. The utility of the resulting series is somewhat limited since the response of the induced polarization to the external field is described by a long-ranged kernel. For a discussion of this point we may refer to Felderhof.[10] It is therefore preferable to adopt a different procedure analogous to that used in Ref. 9 and 10 which allows us to express the fluctuating field $\mathbf{E}$ in terms of the average or macroscopic field in the medium.

We first define the fluctuation of any statistical quantity f by

$$\Delta f \equiv f - \langle f \rangle, \tag{2.10}$$

where the brackets signify a statistical average over an appropriate ensemble. In particular, we have

$$\Delta\rho(\mathbf{R}, t) = \sum_i \delta(\mathbf{R} - \mathbf{R}_i(t)) - \rho_0(\mathbf{R}, t), \tag{2.11}$$

where ρ_0 is the average number density. Adopting the shorthand notation $x = (\mathbf{R}, t)$ and $d^4x = d^3R\, dt$, we obtain from Eqs. (2.8) and (2.10)

$$\Delta\mathbf{P}(x_1) = \alpha_0 \Delta\rho(x_1)\mathbf{E}_0(x_1) - \alpha_0\rho_0(x_1) \int \mathbf{H}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 - \alpha_0 \Delta\rho(x_1) \int \mathbf{H}(x_1 - x_2) \cdot \mathbf{P}(x_2)\, d^4x_2 + \alpha_0 \langle \Delta\rho(x_1) \int \mathbf{H}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 \rangle, \tag{2.12}$$

while from Eq. (2.3) we have

$$\langle \mathbf{E}(x_1) \rangle = \mathbf{E}_0(x_1) - \int \mathbf{F}(x_1 - x_2) \cdot \langle \mathbf{P}(x_2) \rangle\, d^4x_2 \tag{2.13}$$

and

$$\Delta\mathbf{E}(x_1) = -\int \mathbf{F}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2. \tag{2.14}$$

Between Eqs. (2.12) and (2.13) we may eliminate the external field to obtain

$$\int [\delta(x_1 - x_2) + \alpha_0\rho_0(x_1)\mathbf{H}(x_1 - x_2)] \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 = \alpha_0 \Delta\rho(x_1)\, \mathcal{E}(x_1) - \alpha_0 \Delta\rho(x_1) \int \mathbf{H}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 + \alpha_0 \langle \Delta\rho(x_1) \int \mathbf{H}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 \rangle, \tag{2.15}$$

where

$$\mathcal{E}(x_1) \equiv \langle \mathbf{E}(x_1) \rangle + \int [\mathbf{F}(x_1 - x_2) - \mathbf{H}(x_1 - x_2)] \cdot \langle \mathbf{P}(x_2) \rangle\, d^4x_2 \tag{2.16}$$

is a nonfluctuating field which will be interpreted shortly.

The object now is to invert the kernel appearing on the left side of Eq. (2.15) to obtain an expression for $\Delta\mathbf{P}$ which may be iterated in a straightforward manner. Then with the help of Eq. (2.14) we can obtain an expression for the fluctuations $\Delta\mathbf{E}$ of the electric field in terms of the density fluctuations $\Delta\rho$. If the fluid medium is finite in extent, the average density will be constant inside the medium and zero outside. This makes inversion of the kernel difficult. Therefore, for simplicity, we shall assume the medium to be statistically homogeneous, isotropic and infinite in extent so that $\rho_0(x) = \rho_0$ is strictly a constant. (We will not discuss any diffraction effects.) To solve Eq. (2.15) we define the tensor

$$\mathbf{K}(x_1 - x_2) \equiv \int \mathbf{H}(x_1 - x_3) \cdot (1 + \alpha_0\rho_0\mathbf{H})^{-1}(x_3 - x_2)\, d^4x_3, \tag{2.17}$$

and the fluctuating quantities

$$\mathbf{u}(x_1) \equiv \int \mathbf{H}(x_1 - x_2) \cdot \Delta\mathbf{P}(x_2)\, d^4x_2 \tag{2.18}$$

and

$$\mathbf{s}(x_1, x_2) \equiv \mathbf{K}(x_1 - x_2)\, \alpha_0 \Delta\rho(x_2). \tag{2.19}$$

Then, omitting integration signs, Eq. (2.15) becomes, after multiplication by $\mathbf{K}$,

$$\mathbf{u} = \mathbf{s} \cdot \boldsymbol{\mathcal{E}} - \mathbf{s} \cdot \mathbf{u} + \langle \mathbf{s} \cdot \mathbf{u} \rangle, \tag{2.20}$$

with the solution

$$\mathbf{u} = (1 + \mathbf{s})^{-1} \cdot \mathbf{s} \cdot \boldsymbol{\mathcal{E}} + (1 + \mathbf{s})^{-1} \cdot \langle \mathbf{s} \cdot \mathbf{u} \rangle. \tag{2.21}$$

Multiplying this last expression by $\mathbf{s}$ and averaging, we obtain

$$\langle \mathbf{s} \cdot \mathbf{u} \rangle = [1 - \langle \mathbf{s} \cdot (1 + \mathbf{s})^{-1} \rangle]^{-1} \cdot \langle \mathbf{s} \cdot (1 + \mathbf{s})^{-1} \cdot \mathbf{s} \rangle \cdot \boldsymbol{\mathcal{E}}, \tag{2.22}$$

and using this in Eq. (2.21) gives

$$\begin{aligned} \mathbf{u} &= [\mathbf{s} \cdot (1 + \mathbf{s})^{-1} - \langle \mathbf{s} \cdot (1 + \mathbf{s})^{-1} \rangle] \cdot \langle (1 + \mathbf{s})^{-1} \rangle^{-1} \cdot \boldsymbol{\mathcal{E}} \\ &= \Delta[\mathbf{s} \cdot (1 + \mathbf{s})^{-1}] \cdot \langle (1 + \mathbf{s})^{-1} \rangle^{-1} \cdot \boldsymbol{\mathcal{E}}, \end{aligned} \tag{2.23}$$

where we have used the fact that $\langle \mathbf{s} \rangle = 0$.

Defining a kernel $\mathbf{K}'$ by

$$\mathbf{K}'(x_1 - x_2) \equiv - \int \mathbf{F}(x_1 - x_3) \cdot (1 + \alpha_0 \rho_0 \mathbf{H})^{-1}(x_3 - x_2)\, d^4x_3, \tag{2.24}$$

we obtain from Eqs. (2.23), (2.18), and (2.14) our final expression for the fluctuations of the electric field

$$\begin{aligned} \Delta\mathbf{E} = \mathbf{K}' \cdot [&\alpha_0 \Delta\rho (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \\ &- \alpha_0 \langle \Delta\rho (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \rangle] \cdot \langle (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \rangle^{-1} \cdot \boldsymbol{\mathcal{E}}, \end{aligned} \tag{2.25}$$

or in compact form

$$\Delta\mathbf{E} = \mathbf{K}' \cdot \Delta[\alpha_0 \Delta\rho (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1}] \cdot \langle (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \rangle^{-1} \cdot \boldsymbol{\mathcal{E}}. \tag{2.26}$$

Again, we have omitted integration signs and variables of integration.

3. DISCUSSION OF THE FIELD FLUCTUATIONS

Let us now discuss the important features of Eq. (2.25) for the field fluctuations. We first examine the field $\boldsymbol{\mathcal{E}}$. Fourier analyzing $\boldsymbol{\mathcal{E}}$, $\langle \mathbf{E} \rangle$, and $\langle \mathbf{P} \rangle$, we obtain from Eq. (2.16)

$$\boldsymbol{\mathcal{E}}(\mathbf{k}, \omega) = \langle \mathbf{E} \rangle(\mathbf{k}, \omega) + [\mathbf{F}(\mathbf{k}, \omega) - \mathbf{H}(\mathbf{k}, \omega)] \cdot \langle \mathbf{P} \rangle(\mathbf{k}, \omega). \tag{3.1}$$

Since $\langle \mathbf{E} \rangle$ and $\langle \mathbf{P} \rangle$ are the macroscopic, averaged polarization and electric field, their only Fourier components will be at $\omega = \omega_0$ and $|\mathbf{k}| = |n\omega_0/c|$ where ω_0 is the frequency of the external field and n is the exact refractive index of the medium (evaluated at ω_0). As we are concerned with fields of optical wavelengths, we may safely assume

$$\left| \frac{\omega_0 a}{c} \right|, \ |n\omega_0 a/c| \ll 1, \tag{3.2}$$

where a is the hard core diameter. Now the tensors $\mathbf{F}$ and $\mathbf{H}$ have been evaluated[13] with the result

$$\mathbf{F}(\mathbf{k}, \omega) - \mathbf{H}(\mathbf{k}, \omega) = \tfrac{1}{3} + O((ka)^2, (\omega a/c)^2). \tag{3.3}$$

Thus, for optical fields (or in the limit $a \to 0$) we have from Eq. (1)

$$\boldsymbol{\mathcal{E}}(\mathbf{R}, t) \cong \langle \mathbf{E} \rangle(\mathbf{R}, t) + \tfrac{1}{3} \langle \mathbf{P} \rangle(\mathbf{R}, t). \tag{3.4}$$

This is recognized as Lorentz's expression for the effective field acting on a molecule of the medium. A more precise evaluation of $\boldsymbol{\mathcal{E}}$ can be made but this hardly appears necessary and we shall henceforth use the expression (4). Two further points regarding $\boldsymbol{\mathcal{E}}$ deserve special mention. First, we note that the difference between the exact expression (2.16) and the approximate one (3.4) arises solely because the former includes effects of radiation reaction while the latter does not. That is, suppose in Eq. (2.6) we had used instead of $\mathbf{H}$, the dipole kernel $\mathbf{F}$ as is customary.[10] Then, in order to eliminate the divergence which arises when two dipoles are at the same point, it is conventional to cut out a small spherical domain surrounding the dipole at $\mathbf{R}_i(t)$ and take the limit as the diameter of this domain shrinks to zero. This procedure leads exactly to Eq. (3.4) and thus that approximation neglects the radiation reaction. Secondly, the Lorentz effective field is not the true field acting on a dipole. The true field [given by Eq. (2.6) in this model] includes effects of density fluctuations which are absent in Eq. (3.4). These fluctuations appear explicitly in Eq. (2.25).

Introducing now the dielectric function $\epsilon(\mathbf{k}, \omega)$ of the medium, we have from Eq. (3.4)

$$\boldsymbol{\mathcal{E}}(\mathbf{k}, \omega) = \tfrac{1}{3}(\epsilon(\mathbf{k}, \omega) + 2) \cdot \langle \mathbf{E} \rangle(\mathbf{k}, \omega). \tag{3.5}$$

We now examine the second term in Eq. (2.25) and show that in the approximation given by Eq. (3.2) this term is precisely equal to minus $\langle \mathbf{E} \rangle$. To see this, we first note that in the limit as Eq. (3.2) holds, we have, in Fourier transform,[14]

$$\begin{aligned} &\langle \alpha_0 \Delta\rho (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \rangle \cdot \langle (1 + \mathbf{K}\alpha_0 \Delta\rho)^{-1} \rangle^{-1} \\ &\quad = 3[(\epsilon - 1)(\epsilon + 2)^{-1} - (\epsilon_0 - 1)(\epsilon_0 + 2)^{-1}], \end{aligned} \tag{3.6}$$

where

$$\epsilon_0 = (1 + \tfrac{2}{3}\alpha_0 \rho_0)(1 - \tfrac{1}{3}\alpha_0 \rho_0)^{-1} \tag{3.7}$$

is the Lorenz–Lorentz value of the dielectric function. We further have, again in the limit of Eq. (3.2),[15]

$$\mathbf{K}'(\mathbf{k}, \omega) \cong - \frac{(\epsilon_0 + 2)}{3\epsilon_0} \left[k^2 \hat{k}\hat{k} - \epsilon_0 \left(\frac{\omega}{c} \right)^2 \right] \left[k^2 - \left(\frac{\omega}{c} + i\eta \right)^2 \epsilon_0 \right]^{-1}, \tag{3.8}$$

where η is a positive, infinitesimal quantity. Now, assuming that $\langle \mathbf{E} \rangle$ is purely transverse, i.e., $\hat{k} \cdot \langle \mathbf{E} \rangle(\mathbf{k}, \omega) = 0$, we will require only the transverse parts of the tensors ϵ and $\mathbf{K}'$. We write

$$\epsilon(\mathbf{k}, \omega) = \epsilon_{tr}(k, \omega)(1 - \hat{k}\hat{k}) + \epsilon_l(k, \omega)\, \hat{k}\hat{k}. \tag{3.9}$$

Then for the transverse parts of Eqs. (3.5), (3.6), and (3.8) we have, respectively,

$$\tfrac{1}{3}(\epsilon_{tr}(k, \omega) + 2) \langle \mathbf{E} \rangle(\mathbf{k}, \omega), \tag{3.5t}$$

$$9(\epsilon_{tr}(k, \omega) - \epsilon_0)[(\epsilon_{tr}(k, \omega) + 2)(\epsilon_0 + 2)]^{-1}, \tag{3.6t}$$

and

$$- \left(\frac{\omega}{c} \right)^2 \frac{(\epsilon_0 + 2)}{3} \left\{ k^2 - \left(\frac{\omega}{c} + i\eta \right)^2 \epsilon_0 \right\}^{-1}. \tag{3.8t}$$

Combining these three expressions, we obtain

$$- \mathbf{K}' \cdot \langle \alpha_0 \Delta\rho \, (1 + \mathbf{K}\alpha_0\Delta\rho)^{-1} \rangle \circ \langle (1 + \mathbf{K}\alpha_0\Delta\rho)^{-1} \rangle^{-1} \circ \mathcal{E}$$
$$= - \left(\frac{\omega}{c}\right)^2 (\epsilon_{tr}(k, \omega) - \epsilon_0) \left[k^2 - \left(\frac{\omega}{c} + i\eta\right)^2 \epsilon_0\right]^{-1} \langle \mathbf{E} \rangle(\mathbf{k}, \omega). \quad (3.10)$$

Now $\langle \mathbf{E} \rangle(\mathbf{k}, \omega)$ being the average macroscopic field has Fourier components only at $\omega = \omega_0$ and $|\mathbf{k}| = |n\omega_0/c|$ where the refractive index is defined by

$$n^2 = \epsilon_{tr}\left(\frac{n\omega}{c}, \omega\right). \quad (3.11)$$

At these values, however, expression (3.10) becomes simply minus $\langle \mathbf{E} \rangle$ which proves the assertion. The first term of Eq. (2.25) therefore gives the full field as

$$\mathbf{E} = \mathbf{K}' \cdot \alpha_0\Delta\rho(1 + \mathbf{K}\alpha_0\Delta\rho)^{-1} \cdot \langle (1 + \mathbf{K}\alpha_0\Delta\rho)^{-1} \rangle^{-1} \circ \mathcal{E} \quad (3.12)$$

which may be easily checked by averaging both sides and using Eq. (3.8) together with the fact that $\langle \mathbf{E} \rangle$ is transverse.

When the inverse operators in Eq. (3.12) are expanded, we have the obvious interpretation of the field $\mathcal{E}$ propagating from one density fluctuation to the next via a propagator $\mathbf{K}$ and then to the observation point via a propagator $\mathbf{K}'$. The propagator $\mathbf{K}$ is characteristic of a medium with dielectric function ϵ_0, i.e., a medium in which all density fluctuations are absent. As pointed out in Ref. 9 and 10 this is the first step in the approach of the statistical theory to the phenomenological theory in which the propagation is through a medium with the exact dielectric function of the model. In Sec. 5 of this paper we will take the next step toward the phenomenological theory by assuming a definite distribution for the density fluctuations—a Gaussian distribution.

It might appear that a simpler, more compact expression for $\mathbf{E}$ could be obtained by evaluating the averaged terms in Eq. (3.12), i.e.,

$$\langle (1 + \mathbf{K}\alpha_0\Delta\rho)^{-1} \rangle^{-1} \circ \mathcal{E} \quad (3.13)$$

in terms of the exact dielectric function of the medium. This, however, is not possible since expression (3.13) has a zero at $\omega = \omega_0$, $|\mathbf{k}| = |n\omega_0/c|$ which is compensated by a pole in the factors which precede it in Eq. (3.12). Thus, the two factors cannot be separated. Bearing this in mind, the simplest controlled approximation which one can make to Eq. (3.12) is to expand the right-hand side in increasing orders of density fluctuations. Thus we have

$$\mathbf{E} = \mathbf{K}' \cdot \alpha_0\Delta\rho(1 - \mathbf{K}\alpha_0\Delta\rho + \mathbf{K}\alpha_0\Delta\rho \cdot \mathbf{K}\alpha_0\Delta\rho$$
$$- \langle \mathbf{K}\alpha_0\Delta\rho \cdot \mathbf{K}\alpha_0\Delta\rho \rangle + \cdots) \cdot \mathcal{E}, \quad (3.14)$$

or

$$\Delta\mathbf{E} = \mathbf{K}' \cdot [\alpha_0\Delta\rho - \Delta(\alpha_0\Delta\rho\mathbf{K}\alpha_0\Delta\rho) + \cdots] \cdot \mathcal{E}. \quad (3.15)$$

4. THE SPECTRAL DENSITY

Let us now compute, at a point within the medium, the spectral density of one component of the scattered field. We choose a unit vector $\hat{u}_\alpha$ along the αth direction and compute the spectrum of $\hat{u}_\alpha \circ \Delta\mathbf{E}$. If we observe at points for which $\langle \mathbf{E} \rangle \neq 0$, we would have to include also the spectrum of $\hat{u}_\alpha \circ \langle \mathbf{E} \rangle$. This, however, will have components only at $\omega = \omega_0$.

The spectrum is defined by

$$G_\alpha(\mathbf{R}_0, \omega) \equiv \lim_{T\to\infty} \frac{\langle |\hat{u}_\alpha \cdot \Delta\mathbf{E}(\mathbf{R}_0, \omega)|^2 \rangle}{T}, \quad (4.1)$$

where, as usual,[16] the fields $\Delta\mathbf{E}(\mathbf{R}, t)$ are assumed to be truncated for times $|t| > T/2$. It is evident from the form of Eqs. (4.1) and (3.15) that the spectrum depends on the form of the n-point space–time correlation functions of the density fluctuations for all $n \geq 2$.

To lowest order approximation the spectrum is obtained by retaining only the term in Eq. (3.15) with a single density fluctuation. We thus have

$$G_\alpha(\mathbf{R}_0, \omega) = \lim_{T\to\infty} T^{-1}\alpha_0^2 \int d^3R_1 d^3R_2 \, dt_1 \, dt_2 \exp[i\omega(t_1 - t_2)]$$
$$\times [\hat{u}_\alpha \cdot \mathbf{K}'(\mathbf{R}_0 - \mathbf{R}_1, \omega) \cdot \mathcal{E}(\mathbf{R}_1, t_1)]$$
$$\times [\hat{u}_\alpha \cdot \mathbf{K}'(\mathbf{R}_0 - \mathbf{R}_2, \omega) \cdot \mathcal{E}(\mathbf{R}_2, t_2)]^* \langle \Delta\rho(\mathbf{R}_1, t_1)\Delta\rho(\mathbf{R}_2, t_2) \rangle. \quad (4.2)$$

Now suppose that the external field is a narrow, well collimated monochromatic beam of the form

$$\mathbf{E}_0(\mathbf{R}, t) = \mathbf{E}_0 \cos(\mathbf{k}_0 \cdot \mathbf{R} - \omega_0 t + \phi_0). \quad (4.3)$$

Neglecting diffraction effects at the boundary of the scattering medium, the average field in the medium will again be narrow and monochromatic with frequency ω_0. In general the average field will also be attenuated in the direction of propagation. The spectrum given by Eq. (4.1) thus depends on the depth in the medium at which it is observed. In order to avoid this complication (see, however, Ref. 10), we will consider only the case in which the imaginary part of the refractive index is negligibly small and thus attenuation may be neglected at any finite depth in the medium. Under these conditions, the average field in the medium will be given by

$$\langle \mathbf{E}(\mathbf{R}, t) \rangle = \mathbf{A} \cos(n_R(\omega_0)\mathbf{k}_0 \cdot \mathbf{R} - \omega_0 t + \phi_1), \quad (4.4)$$

where $n_R(\omega_0)$ is the real part of the refractive index of the medium at frequency ω_0 and $\mathbf{A}$ is the real, constant amplitude of the average field. [If refraction at the boundary is considered, the wavevector will have magnitude $n_R(\omega_0)|\mathbf{k}_0|$ and direction determined by the Fresnel refraction formulas.]

The integration points $\mathbf{R}_1$ and $\mathbf{R}_2$ in Eq. (4.1) are restricted to the domain where $\mathcal{E}$ is nonzero, i.e., to the narrow region in which the average field, Eq. (4), is nonzero. The points $\mathbf{R}_1$ and $\mathbf{R}_2$ are further restricted by the following consideration. In an experimental situation, the field $\Delta\mathbf{E}$ is detected by a device (system of lenses, etc.) which is so designed as to receive only those spatial Fourier components (of the radiation arriving at $\mathbf{R}_0$) whose wavevectors lie within a small cone of solid angle σ_0. Within the limits of geometrical optics, this is equivalent to restricting the integration point $\mathbf{R}_1$ (or $\mathbf{R}_2$) to lie within this cone.[17] These two considerations thus restrict $\mathbf{R}_1$ and $\mathbf{R}_2$ to lie within a small domain which is the intersection of the cone of observation and the region occupied by the average field. Let this domain be centered at the origin of coordinates and have volume V.

If we take the observation point $\mathbf{R}_0$ far from the origin so that $|\mathbf{R}_0| \gg |\mathbf{R}_1|, |\mathbf{R}_2|$ (for all $\mathbf{R}_1, \mathbf{R}_2 \in V$), the kernel $\mathbf{K}'$ is easily approximated as

$$\mathbf{K}'(\mathbf{R}_0-\mathbf{R}_1,\omega)\cong\frac{(\epsilon_0+2)}{3}\left(\frac{\omega}{c}\right)^2(4\pi R_0)^{-1}(1-\hat{R}_0\hat{R}_0)$$
$$\times\exp\left(in_0\frac{\omega}{c}R_0-in_0\frac{\omega}{c}\hat{R}_0\cdot\mathbf{R}_1\right)\qquad(4.5)$$

where $n_0=\epsilon_0^{1/2}$. We define the dynamic structure factor of the medium by

$$S(\mathbf{q},\Omega)\equiv\int d^3R\,dt\exp(-i\mathbf{q}\cdot\mathbf{R}+i\Omega t)\langle\Delta\rho(\mathbf{R},t)\Delta\rho(\mathbf{R}'=0,t'=0)\rangle.\qquad(4.6)$$

Finally, if we again assume that absorption in the medium is negligible, we obtain from Eqs. (4.4) and (3.5)

$$\mathcal{E}(\mathbf{R},t)=\frac{[n_R^2(\omega_0)+2]}{3}\mathbf{A}\cos[n_R(\omega_0)\mathbf{k}_0\cdot\mathbf{R}-\omega_0t+\phi_1].\qquad(4.7)$$

Now combining Eqs. (4.4), (4.5), and (4.7) with (4.1) and remembering the restrictions on $\mathbf{R}_1$ and $\mathbf{R}_2$, we have after a simple integration

$$G_\alpha(\mathbf{R}_0,\omega)\cong\left[\alpha_0\frac{(\epsilon_0+2)}{3}\frac{(n_R^2(\omega_0)+2)}{3}\left(\frac{\omega}{c}\right)^2\frac{\hat{u}_\alpha\cdot(1-\hat{R}_0\hat{R}_0)\cdot\mathbf{A}}{4\pi R_0}\right]^2$$
$$\times\frac{V}{4}\left[S\left(n_0\frac{\omega}{c}\hat{R}_0-n_R(\omega_0)\mathbf{k}_0,\omega-\omega_0\right)\right.$$
$$\left.+S\left(n_0\frac{\omega}{c}\hat{R}_0+n_R(\omega_0)\mathbf{k}_0,\omega+\omega_0\right)\right],\qquad(4.8)$$

where we have omitted a term proportional to $\delta(\omega_0)$. In this expression, the factor ω^4 is well known. The appearance of two structure factors at the sum and difference frequencies is a result of using a real average field and insures that the spectrum is an even function of ω as required by its definition. The spectrum is completely polarized with $\Delta\mathbf{E}$ lying in the plane formed by $\hat{R}_0$ and $\mathbf{A}$. Summing over the components $\alpha=1,2,3$, we obtain the usual angular dependence

$$\sum_\alpha[\hat{u}_\alpha\cdot(1-\hat{R}_0\hat{R}_0)\cdot\mathbf{A}]^2=A^2\sin^2\theta,\qquad(4.9)$$

where θ is the angle between $\hat{R}_0$ and $\mathbf{A}$.

The factors involving ϵ_0 and n_R are of interest in comparison of the statistical theory with phenomenological results. This has been discussed recently by Felderhof[10] in the case of elastic scattering and our expression is the extension of his to the inelastic case. To see this, we approximate $n_R(\omega_0)$ by the Lorenz–Lorentz value of the refractive index n_0. We then have

$$\left(\alpha_0\frac{(\epsilon_0+2)}{3}\frac{[n_R^2(\omega_0)+2]}{3}\right)^2\to\left(\frac{\partial\epsilon_0}{\partial\rho_0}\right)^2,\qquad(4.10)$$

giving

$$\sum_\alpha G_\alpha(\mathbf{R}_0,\omega)\to\left(\frac{\omega}{c}\right)^4\frac{(\partial\epsilon_0/\partial\rho_0)^2}{(4\pi R_0)^2}A^2\sin^2\theta$$
$$\times\frac{V}{4}\left\{S\left(n_0\left[\frac{\omega}{c}\hat{R}_0-\mathbf{k}_0\right],\omega-\omega_0\right)\right.$$
$$\left.+S\left(n_0\left[\frac{\omega}{c}\hat{R}_0+\mathbf{k}_0\right],\omega+\omega_0\right)\right\}.\qquad(4.11)$$

The total, time averaged, intensity scattered to a point $\mathbf{R}_0$ is obtained by the usual formula

$$I(\mathbf{R}_0)=T^{-1}\int dt\langle|\Delta\mathbf{E}(\mathbf{R}_0,t)|^2\rangle=\int\frac{d\omega}{2\pi}\sum_\alpha G_\alpha(\mathbf{R}_0,\omega).\qquad(4.12)$$

The two structure factors will be appreciable only in the vicinity of $\omega=\pm\omega_0$. This allows us to replace ω^4 by ω_0^4 in the integrations. We then have, e.g.,

$$\int\frac{d\omega}{2\pi}S\left(n_0\left[\frac{\omega}{c}\hat{R}_0\mp\mathbf{k}_0\right],\omega\mp\omega_0\right)$$
$$=\int d^3R\,dt\,\delta\left(t-n_0\frac{\hat{R}_0\cdot\mathbf{R}}{c}\right)\langle\Delta\rho(\mathbf{R},t)\Delta\rho(\mathbf{0},0)\rangle$$
$$\times\exp[\pm i(n_0\mathbf{k}_0\cdot\mathbf{R}-\omega_0t)]$$
$$=\int d^3R\left\langle\Delta\rho\left(\mathbf{R},t=n_0\frac{\hat{R}_0\cdot\mathbf{R}}{c}\right)\Delta\rho(\mathbf{0},0)\right\rangle$$
$$\times\exp\left[\pm in_0\frac{\omega_0}{c}(\hat{\mathbf{k}}_0-\hat{R}_0)\cdot\mathbf{R}\right].\qquad(4.13)$$

Since the integration point $\mathbf{R}$ is confined to a small volume element, the time $t=n_0\hat{R}_0\cdot\mathbf{R}/c$ is essentially zero, so that Eq. (4.13) becomes

$$\int d^3R\exp\left[\pm in_0\frac{\omega_0}{c}(\hat{\mathbf{k}}_0-\hat{R}_0)\cdot\mathbf{R}\right]\langle\Delta\rho(\mathbf{R},0)\Delta\rho(\mathbf{0},0)\rangle$$
$$=\mathcal{S}\left(n_0\frac{\omega_0}{c}(\hat{\mathbf{k}}_0-\hat{R}_0)\right),\qquad(4.14)$$

where $\mathcal{S}$ is the static structure factor. Combining these results with Eq. (11), we have

$$I(\mathbf{R}_0)\cong\left(\frac{\omega_0}{c}\right)^4\left(\frac{\partial\epsilon_0}{\partial\rho_0}\right)^2(4\pi R_0)^{-2}VI_0\sin^2\theta\,\mathcal{S}\left(n_0\frac{\omega_0}{c}(\hat{\mathbf{k}}_0-\hat{R}_0)\right)\qquad(4.15)$$

where I_0 is the time average intensity of the average field. This is precisely the result of Felderhof[10] and agrees with the phenomenological theory in this order of approximation.

The important features of Eqs. (4.8) and (4.15) are the dependence of the amplitude on the dielectric constant (which is difficult to measure) and the dependence of the structure factors on the refractive index (which is experimentally accessible). In the phenomenological approach, these dependences are inserted in an "ad hoc" manner and the above equations provide the justification for this procedure.

5. A GAUSSIAN MODEL

In the previous sections we have seen that it is possible to describe the propagation of the electric field or its fluctuations as a propagation from one density fluctuation to the next thru a medium characterized by the Lorenz–Lorentz value of the dielectric function. The spectral density, in lowest order, then depends in a natural way on this approximate dielectric function.

It appears reasonable that further rearrangement and resummation of the basic equations (3.14) and (3.15) will lead to a spectrum of the same form but with an improved value of the refractive index replacing n_0. As the true refractive index depends on all orders of correlation functions of the density fluctuations, the next step would be to retain only the two point correlation

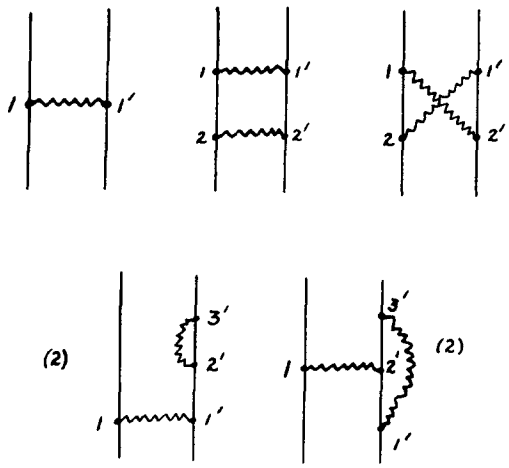

FIG. 1. Diagrams contributing to the spectrum through fourth order in $\Delta\rho$.

function. It would be still better to have reasonable approximations to the higher order correlation functions which would lead to tractable equations. For this purpose, we will adopt a model for the medium in which the density fluctuations have Gaussian distribution with mean value zero and covariance equal to the exact dynamic structure factor S. This approximation becomes exact when the fluctuations are separated into pairs in such a way that each pair is far from every other pair. The approximation is defined by

$$\left\langle \prod_{i=1}^{n} \Delta\rho(x_i) \right\rangle = \begin{cases} 0, & n \text{ odd} \\ \sum \prod_{(\text{pairs})} \langle \Delta\rho(x_i)\Delta\rho(x_j)\rangle & n \text{ even}, \end{cases} \tag{5.1}$$

where the sum extends over all decompositions of the n indices $x_1, \ldots, x_n$ into $n/2$ unordered pairs.

Suppose now that we use the definition (4.1) of the spectral density and insert into it Eq. (2.25) for the field fluctuations. After expanding in orders of $\Delta\rho$ and carrying out the averages by means of Eq. (5.1), the resulting spectrum can be represented by means of a diagrammatic series. For example, through fourth order in density fluctuations the diagrams are shown in Fig. 1. The elements of the diagrams are as follows. A wavy line connecting points n and l (called and S-bond) represents $\alpha_0^2 \langle \Delta\rho(x_n)\Delta\rho(x_l)\rangle$. A straight line connecting points n and l (called a K-bond) represents $\mathbf{K}(x_n - x_l)$. The uppermost line, originating at a point m, represents $\hat{u}_\alpha \cdot \mathbf{K}(\mathbf{R}_0 - \mathbf{R}_m, \omega)\exp(i\omega t_m)$, while the lowermost line, terminating at point 1(or 1'), represents $\mathcal{E}(x_1)$. The points are connected in pairs by S-bonds in all possible ways with two restrictions: (a) there must be at least one S-bond connecting the two lines; and (b) it must must not be possible to cause the diagram to become disconnected by cutting a K-bond at any point below the lowest S-bond connecting the two lines. This is illustrated in Fig. 2, where cutting the K-bonds labelled m, n would violate this condition. All terms in the sum of graphs are taken as positive.

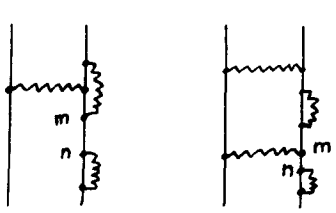

FIG. 2. Diagrams which violate condition (b).

We now consider all graphs which contain exactly one S-bond connecting the two lines. This constitutes the "single scattering" approximation. These graphs fall into four classes, examples of which are shown in Fig. 3. In class C_1, the S-bonds neither overlap nor cross each other. This class can be summed by simple convolution techniques. In class C_2, which includes class C_1, the S-bonds do not cross but may overlap. Class C_3, which includes C_1 and C_2, contains all possible S-bonds which do not cross the one S-bond connecting the two lines. Classes C_1, C_2, and C_3 are simply renormalizations of the K-bonds with C_3 leading to the completely renormalized K-bond, i.e., the exact propagator for the Gaussian model. Class C_4, in which at least one S-bond crosses the S-bond connecting the two lines, may be thought of as corresponding to frequency dependent renormalization of the elementary vertex α_0.

The analysis can be extended to graphs which include two or more S-bonds connecting the two lines. These graphs will have as basic elements the renormalized K-bonds and vertices of the single scattering approximation. The essentially new feature will be the simultaneous renormalization of two or more vertices.

6. RENORMALIZATION OF THE SPECTRUM

Let us now examine the renormalization of the spectral density which arises from taking into account the diagrams of class C_3 of Fig. 3. We use Eq. (2.25) in Eq. (4.1), expand in orders of $\Delta\rho$ and take the average retaining only class C_3 diagrams. We then obtain

$$\begin{aligned} G_\alpha(\mathbf{R}_0, \omega) = \lim_{T\to\infty} \frac{\alpha_0^2}{T} \int d^4x_1 \cdots d^4x_2' \exp[i\omega(t_1 - t_1')] \\ \times [\hat{u}_\alpha \cdot \mathbf{K}'(\mathbf{R}_0 - \mathbf{R}_1, \omega) \circ \psi(x_1 - x_2) \circ \mathcal{E}(x_2)] \\ \times [\hat{u}_\alpha \cdot \mathbf{K}'(\mathbf{R}_0 - \mathbf{R}_1', \omega) \cdot \psi(x_1' - x_2') \circ \mathcal{E}(x_2')]^* \\ \times \langle \Delta\rho(x_2)\Delta\rho(x_2')\rangle. \end{aligned} \tag{6.1}$$

Here, the new tensor ψ is defined by

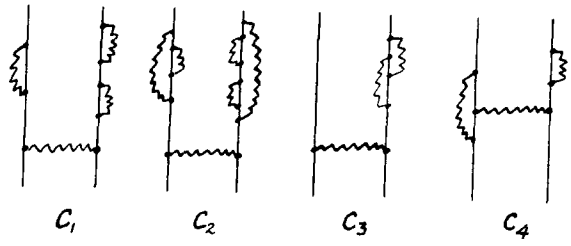

FIG. 3. The four classes of diagrams in single scattering approximation.

$$\psi(\mathbf{k},\omega)\equiv\langle(1+\alpha_0\Delta\rho\mathbf{K})^{-1}\rangle_G(\mathbf{k},\omega)$$
$$=\langle(1+\alpha_0\mathbf{K}\Delta\rho)^{-1}\rangle_G(\mathbf{k},\omega), \tag{6.2}$$

and the subscript G indicates that the average is taken according to the Gaussian prescription given by Eq. (5.1).

If we choose for $\mathcal{E}$ a complex field and again neglect attenuation, we may write

$$\mathcal{E}(\mathbf{R},t)=\frac{[n_R^2(\omega_0)+2]}{3}\,\mathbf{A}\exp[in_R(\omega_0)\,\mathbf{k}_0\cdot\mathbf{R}-i\omega_0 t]. \tag{6.3}$$

The integrations on time and the intermediate space points may be carried out giving

$$G_\alpha(\mathbf{R}_0,\omega)=\frac{\alpha_0^2}{(2\pi)^3}\left(\frac{n_R^2(\omega_0)+2}{3}\right)^2\int d^3R_2 d^3R_2' d^3q$$
$$\times\exp\{i[\mathbf{q}+n_R(\omega_0)\mathbf{k}_0]\cdot(\mathbf{R}_2-\mathbf{R}_2')\}$$
$$\times[\hat{u}_\alpha\cdot\mathbf{T}'(\mathbf{R}_0-\mathbf{R}_2,\omega)\cdot\mathbf{A}]$$
$$\times[\hat{u}_\alpha\cdot\mathbf{T}'(\mathbf{R}_0-\mathbf{R}_2',\omega)\cdot\mathbf{A}]^*\,S(\mathbf{q},\omega-\omega_0), \tag{6.4}$$

where $\mathbf{T}'(\mathbf{R},\omega)$ is the Fourier transform of

$$\mathbf{T}'(\mathbf{k},\omega)\equiv\mathbf{K}'(\mathbf{k},\omega)\cdot\psi(\mathbf{k},\omega). \tag{6.5}$$

In order to carry out the final integration in Eq. (6.4), we require an explicit expression for $\mathbf{T}'$. Whatever this expression, isotropy requires

$$\mathbf{T}'(\mathbf{k},\omega)=T_{tr}'(k,\omega)(1-\hat{k}\hat{k})+T_l'(k,\omega)\,\hat{k}\hat{k}, \tag{6.6}$$

where T_{tr}' and T_l' depend only on the magnitude of $\mathbf{k}$ and on ω. The Fourier transform is then given by

$$\mathbf{T}'(\mathbf{R},\omega)=\frac{1}{2\pi^2}\int_0^\infty k^2\,dk\,\frac{\sin kR}{kR}\,[T_{tr}'(k,\omega)(1-\hat{R}\hat{R})$$
$$+T_l'(k,\omega)\hat{R}\hat{R}]+O\left(\frac{1}{R^2}\right). \tag{6.7}$$

Since the integration points $\mathbf{R}_2$ and $\mathbf{R}_2'$ are restricted to a small region around the origin, we may ignore terms of order $|\mathbf{R}_0-\mathbf{R}_2|^{-2}$, set $|\mathbf{R}_0-\mathbf{R}_2|\sim R_0$ in the denominator and also set $(\mathbf{R}_0-\mathbf{R}_2)/|\mathbf{R}_0-\mathbf{R}_2|\sim\hat{R}_0$. Since the component of the far field parallel to $\hat{R}_0$ is zero, only the term involving T_{tr}' survives. Thus, we obtain

$$G_\alpha(\mathbf{R}_0,\omega)=\frac{1}{2\pi^5}\left(\frac{\alpha_0|u_\alpha\cdot\mathbf{A}|}{4\pi R_0}\,\frac{[n_R^2(\omega_0)+2]}{3}\right)^2$$
$$\times\int d^3R_2 d^3R_2' d^3q\int_0^\infty k\,dk\int_0^\infty k'\,dk'$$
$$\times\exp[i(\mathbf{q}+n_R(\omega_0)\,\mathbf{k}_0)\cdot(\mathbf{R}_2-\mathbf{R}_2')]$$
$$\times T_{tr}'(k,\omega)\,T_{tr}'^*(k',\omega)\sin(k|\mathbf{R}_0-\mathbf{R}_2|)$$
$$\times\sin(k'|\mathbf{R}_0-\mathbf{R}_2'|)\,S(\mathbf{q},\omega-\omega_0). \tag{6.8}$$

To evaluate T_{tr}', we return to Eq. (3.6) and write

$$3(\epsilon-1)(\epsilon+2)^{-1}\equiv\boldsymbol{\gamma}=\alpha_0\rho_0$$
$$+\langle\alpha_0\Delta\rho(1+\mathbf{K}\alpha_0\Delta\rho)^{-1}\rangle\cdot\langle(1+\mathbf{K}\alpha_0\Delta\rho)^{-1}\rangle^{-1}. \tag{6.9}$$

Multiplying this equation by $\mathbf{K}$, we easily find

$$\langle(1+\mathbf{K}\alpha_0\Delta\rho)^{-1}\rangle=[1+\mathbf{K}\cdot(\boldsymbol{\gamma}-\alpha_0\rho_0)]^{-1}. \tag{6.10}$$

If we make the Gaussian approximation to the averages in Eq. (6.9), then Eq. (6.10) becomes

$$\langle(1+\mathbf{K}\alpha_0\Delta\rho)^{-1}\rangle_G=\psi=[1+\mathbf{K}\cdot(\boldsymbol{\gamma}_G-\alpha_0\rho_0)]^{-1}. \tag{6.11}$$

We now decompose the tensors $\mathbf{K}$, $\mathbf{K}'$, and $\boldsymbol{\gamma}_G$ into their transverse and longitudinal parts. In the limit as the hard core diameter becomes zero, we may use[18]

$$\mathbf{K}=\frac{\epsilon_0+2}{9}\left[\left(\frac{\epsilon_0+2}{\epsilon_0}\right)k^2\hat{k}\hat{k}-\left(k^2+2\left(\frac{\omega}{c}\right)^2\right)\right]\left[k^2-\left(\frac{\omega}{c}+i\eta\right)^2\epsilon_0\right]^{-1} \tag{6.12}$$

together with Eq. (3.8) to obtain

$$T_{tr}'(k,\omega)=\frac{\epsilon_{G,tr}+2}{3}\left(\frac{\omega}{c}\right)^2\left[k^2-\left(\frac{\omega}{c}+i\eta\right)^2\epsilon_{G,tr}\right]^{-1} \tag{6.13}$$

where $\epsilon_{G,tr}$ is the Gaussian approximation to the transverse part of the dielectric tensor.

Now, to evaluate the integrals in Eq. (6.8), we observe that the singularities of the integrand occur at the points $k=\pm f(\omega)$ where $f(\omega)$ is the solution of

$$f=\frac{\omega}{c}\,[\epsilon_{G,tr}(f,\omega)]^{1/2} \tag{6.14}$$

and thus

$$f(\omega)=\frac{\omega}{c}\,n_G(\omega) \tag{6.15}$$

where $n_G(\omega)$ is the refractive index (in Gaussian approximation) at frequency ω. Therefore, using this in Eq. (6.8) and carrying out the integrations, we have

$$G_\alpha(\mathbf{R}_0,\omega)=\left(\frac{\alpha_0|\hat{u}_\alpha\cdot\mathbf{A}|}{4\pi R_0}\,\frac{[n_R^2(\omega_0)+2]}{3}\,\frac{[n_G^2(\omega)+2]}{3}\left(\frac{\omega}{c}\right)^2\right)^2$$
$$\times VS\left(\frac{\omega}{c}\,n_G(\omega)\,\hat{R}_0-n_R(\omega_0)\,\mathbf{k}_0,\omega-\omega_0\right). \tag{6.16}$$

Comparing this with Eq. (4.8) (except for the use of a complex field here) we see that the two expressions differ only in that n_G replaces n_0 everywhere. Thus our approximate calculation with the Gaussian model leads, in accordance with the phenomenological theory, to the same spectrum, but with an improved value of the refractive index.

It should be clear that the main feature of the Gaussian approximation is that it enables us to split the correlation between the two fields from the correlations which occur within each field separately. Had we not made Gaussian approximation but, instead, arbitrarily split these correlations, then all the formulas of this section would hold but without the subscript G. In particular, in Eq. (16), the exact refractive index replaces n_G.

APPENDIX

In this appendix, we discuss briefly some uncompleted work on the vertex renormalization of the Gaussian model.

In the "single scattering approximation", the two fields which form the spectrum will be connected by only one S-bond. This means that we may average each of the two fields separately provided we hold one density fluctuation fixed. This fluctuation will than be averaged last with the corresponding fluctuation from the other field. Let us denote the average of $\Delta\mathbf{E}$ with one

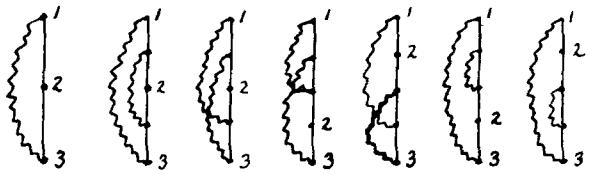

FIG. 4. Lowest order vertex corrections.

fluctuation fixed by $\widetilde{\Delta \mathbf{E}}$. Expanding the expression Eq. (2.25) for $\Delta \mathbf{E}$ and averaging with one fluctuation fixed, we may write $\widetilde{\Delta \mathbf{E}}$ in either of three ways:

$$\widetilde{\Delta \mathbf{E}}(x_0) = \alpha_0 \int \mathbf{T}'(x_0 - x_1) \cdot \mathbf{V}(x_1, x_2, x_3) \cdot \mathcal{E}(x_2)\, \Delta\rho(x_2)\, d^4x_1 \cdots d^4x_3, \tag{A1}$$

or

$$\widetilde{\Delta \mathbf{E}}(x_0) = \alpha_0 \int \mathbf{U}(x_0, x_1, x_2) \cdot \mathcal{E}(x_2)\, \Delta\rho(x_1)\, d^4x_1\, d^4x_2, \tag{A2}$$

or

$$\widetilde{\Delta \mathbf{E}}(x_0) = \alpha_0 \int \mathbf{W}(x_0, x_1, x_2) \langle (1 + \alpha_0 \mathbf{K} \Delta\rho)^{-1} \rangle^{-1} (x_2 - x_3) \cdot \mathcal{E}(x_3)\, \Delta\rho(x_2)\, d^4x_1 \cdots d^4x_3. \tag{A3}$$

In Eq. (A. 1), the tensor $\mathbf{V}$ is defined by

$$\mathbf{V}(x_1, x_2, x_3) = \langle (1 + \alpha_0 \Delta\rho \mathbf{K})^{-1}(x_1, x_2)(1 + \alpha_0 \mathbf{K} \Delta\rho)^{-1}(x_2, x_3) \rangle_{G,L}. \tag{A4}$$

The average is a Gaussian average. The subscript L indicates that only linked diagrams are counted. That is, in the diagram series for $\mathbf{V}$ we discard all diagrams which would become disconnected if any K-bond were cut. Note that the order of $\Delta\rho$ and $\mathbf{K}$ is reversed in the two factors constituting $\mathbf{V}$. This form of $\widetilde{\Delta \mathbf{E}}$ is the most convenient for calculations because it clearly separates the vertex correction. According to (1), the effective field $\mathcal{E}$ undergoes a vertex modification and then propagates to the observation point via the propagator $\mathbf{T}'$ which we have shown is characteristic of the Gaussian medium. The first term in the expansion of $\mathbf{V}$ is simply $\delta(x_1 - x_2)\, \delta(x_2 - x_3)$ [which leads immediately to the spectrum given by Eq. (6. 1)], while the next few terms are shown in Fig. 4.

In Eq. (A2), the tensor $\mathbf{U}$ is defined by

$$\mathbf{U}(x_0, x_1, x_2) = \langle (\mathbf{K}'[1 + \alpha_0 \Delta\rho \mathbf{K}]^{-1})(x_0, x_1)(1 + \alpha_0 \mathbf{K} \Delta\rho)^{-1}(x_1, x_2) \rangle_{G,L'}. \tag{A5}$$

Here the subscript L' indicates that the diagrams are partially linked in that it must not be possible to cause the diagrams to become disconnected by cutting a K-bond between the points x_1 and x_2. Finally, in Eq. (A3), the tensor $\mathbf{W}$ is given by

$$\mathbf{W}(x_0, x_1, x_2) = \langle (\mathbf{K}'[1 + \alpha_0 \Delta\rho \mathbf{K}]^{-1})(x_0, x_1)(1 + \alpha_0 \mathbf{K} \Delta\rho)^{-1}(x_1, x_2) \rangle_G, \tag{A6}$$

where the diagrams may be either linked or unlinked. Clearly, we have the relations

$$\mathbf{U}(x_0, x_1, x_2) = \int d^4y\, \mathbf{T}'(x_0 - y) \cdot \mathbf{V}(y, x_1, x_2) = \int d^4y \mathbf{W}(x_0, x_1, y) \langle (1 + \alpha_0 \mathbf{K} \Delta\rho)^{-1} \rangle^{-1} (y - x_2). \tag{A7}$$

The spectrum can be written formally by using either Eq. (A1), (A2) or (A3), e. g.,

$$G_\alpha(\mathbf{R}_0, \omega) = \lim_{T \to \infty} \frac{\alpha_0^2}{T} \int [\hat{u}_\alpha \cdot \mathbf{T}'(\mathbf{R}_0 - \mathbf{R}_1', \omega) \cdot \mathbf{V}(x_1, x_2, x_3) \cdot \mathcal{E}(x_3)] \times [\hat{u}_\alpha \cdot \mathbf{T}'(\mathbf{R}_0 - \mathbf{R}_1', \omega) \cdot \mathbf{V}(x_1', x_2', x_3') \cdot \mathcal{E}(x_3')]^* \times \exp[i\omega(t_1 - t_1')] \langle \Delta\rho(x_2)\, \Delta\rho(x_2') \rangle\, d^4x_1 \cdots d^4x_3'. \tag{A8}$$

Further evaluation depends upon making specific approximations to the tensor $\mathbf{V}$ and is quite difficult.

Without doing an explicit evaluation we would like to indicate the procedure which should bring Eq. (A8) into line with the phenomenological theory. If one writes the coordinate representation of the tensor $\boldsymbol{\gamma}$ [Eq. (6. 9)] in Gaussian approximation and takes the functional derivative of that expression with respect to $\Delta\rho(y)$ [or what amounts to the same thing, with respect to $\rho(y)$], one has the result

$$\frac{\delta}{\delta\rho(y)}\, \boldsymbol{\gamma}_G(x_1 - x_3) = \alpha_0 \mathbf{V}(x_1, y, x_3). \tag{A9}$$

Now, since $\boldsymbol{\gamma}$ is simply related to the dielectric tensor via Eq. (6. 9), the vertex part $\mathbf{V}$ can be related to the functional derivative of the dielectric tensor. In this manner, one should be able to express the spectrum in single scattering approximation in terms of the dynamic structure factor and the functional derivative of the dielectric tensor. This would be the appropriate generalization of the factor $\partial\epsilon/\partial\rho$ which appears in the phenomenological result.

Finally, let us indicate the general calculation of the spectrum in single scattering approximation in terms of the vertex function $\mathbf{V}$. As the required integrals are difficult to perform rigorously, we shall arbitrarily restrict the spatial integrals to a small volume around the origin and take $\mathbf{R}_0$ in the far zone. We can then write for the transverse part of $\mathbf{T}'$, following the calculation of the previous section,

$$\mathbf{T}'(\mathbf{R}_0 - \mathbf{R}_1, \omega) \cong \frac{1}{4\pi R_0} \frac{n_G^2(\omega) + 2}{3} \left(\frac{\omega}{c}\right)^2 \times \exp\left[i n_G(\omega) \frac{\omega}{c} (R_0 - \hat{R}_0 \cdot \mathbf{R}_1)\right] (1 - \hat{R}_0 \hat{R}_0). \tag{A10}$$

In 4-vector notation, we will write the Fourier transform of $\mathbf{V}$ as

$$\mathbf{V}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{12}} \int \exp[i(k_1 x_1 + k_2 x_2 + k_3 x_3)] \times \mathbf{V}(k_1, k_2, k_3)\, d^4k_1 \cdots d^4k_3. \tag{A11}$$

Then, using Eq. (6. 3) for $\mathcal{E}$ and the notation

$$k^{in} = (n_R(\omega_0)\mathbf{k}_0, \omega_0), \quad k^{out} = \left(n_G(\omega) \frac{\omega}{c} \hat{R}_0, \omega\right), \tag{A12}$$

we can carry out the integrals in Eq. (A8) to obtain

$G_\alpha(\mathbf{R}_0, \omega)$

$$= \lim_{T\to\infty} \frac{1}{T}\left[\frac{\alpha_0}{4\pi R_0}\,\frac{[n_G^2(\omega)+2]}{3}\,\frac{[n_R^2(\omega_0)+2]}{3}\left(\frac{\omega}{c}\right)^2\right]^2$$

$$\times \frac{1}{(2\pi)^4}\int |\hat{u}_\alpha \cdot \mathbf{V}(k^{\text{out}}, -q, -k^{\text{in}})\cdot\mathbf{A}|^2 S(q)\,d^4q. \quad \text{(A13)}$$

In order to check this result, we may use the zeroth order approximation to $\mathbf{V}$, i.e.,

$$\mathbf{V}^{(0)} = \delta(x_1 - x_2)\,\delta(x_2 - x_3). \quad \text{(A14)}$$

We then obtain for the Fourier transform

$$\mathbf{V}^{(0)}(k^{\text{out}}, -q, -k^{\text{in}}) = \int \exp[-i(k^{\text{out}} - q - k^{\text{in}})\cdot x]\,d^4x. \quad \text{(A15)}$$

Thus, for the square of the transform which appears in Eq. (A13), we can write

$$|\hat{u}_\alpha \cdot \mathbf{V}\cdot\mathbf{A}|^2 = |\hat{u}_\alpha\cdot\mathbf{A}|^2 \int \exp[-i(k^{\text{out}} - q - k^{\text{in}})\cdot(x-y)]\,d^4x\,d^4y$$
$$= (2\pi)^4 |\hat{u}_\alpha\cdot\mathbf{A}|^2\, V T\delta(q - k^{\text{out}} - k^{\text{in}}). \quad \text{(A16)}$$

We thus recover the previous result, Eq. (6.16) exactly.

In the more general case, conservation of the "4-momentum" in the vertex part requires that

$$\mathbf{V}(k^{\text{out}}, -q, -k^{\text{in}}) = (2\pi)^4\,\mathbf{v}(q)\,\delta(q - (k^{\text{out}} - k^{\text{in}})). \quad \text{(A17)}$$

Using the same procedure as in Eq. (A16) for the square of the delta function, Eq. (A13) becomes

$$G_\alpha(\mathbf{R}_0, \omega) = \left[\frac{\alpha_0}{4\pi R_0}\,\frac{n_G^2(\omega)+2}{3}\,\frac{n_R^2(\omega_0)+2}{3}\left(\frac{\omega}{c}\right)^2\right]^2$$
$$\times V|\hat{u}_\alpha\cdot\mathbf{v}(k^{\text{out}} - k^{\text{in}})\cdot\mathbf{A}|^2 S(k^{\text{out}} - k^{\text{in}}), \quad \text{(A18)}$$

which is the general result in single scattering approximation.

Note added in proof: Since the completion of this research, the authors learned of recent work along similar lines by Boots, Bedeaux, and Mazur.[19] The reader is referred to this paper for a comparison of techniques and results.

*Based on a dissertation submitted by V.V. to the Department of Physics, University of Illinois at Chicago Circle, November 1974.

†Present address: Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853.

[1] von Smoluchowski, Ann. Phys. **25**, 205 (1908).

[2] A. Einstein, Ann. Phys. **38**, 1275 (1910).

[3] G.B. Benedek, in *Statistical Physics, Phase Transitions and Superfluidity*, edited by M. Chretien, E.P. Gross, and S. Deser (Gordon and Breach, New York, 1968), Vol. II.

[4] N.G. van Kampen in *Proceedings of the International School of Physics: Enrico Fermi*, Course XLII, edited by R.J. Glauber (Academic, New York, 1969).

[5] L. van Hove, Phys. Rev. **110**, 999 (1958).

[6] L.I. Komarov and I.Z. Fisher, Sov. Phys. JETP **16**, 1358 (1963).

[7] M. Tanaka, Prog. Theor. Phys. **40**, 975 (1968).

[8] See B.U. Felderhof, Physica **76**, 486 (1974) for a discussion of this point.

[9] D. Bedeaux and P. Mazur, Physica **67**, 23 (1973).

[10] See Ref. 8.

[11] S.R. de Groot in *Studies in Statistical Physics*, edited by J. de Boer and G.E. Uhlenbeck (North-Holland, Amsterdam, 1969), Vol. IV.

[12] Our convention is

$$f(\mathbf{R}, t) = \frac{1}{(2\pi)^3}\int e^{i\mathbf{k}\cdot\mathbf{R}} f(\mathbf{k}, t)\,d^3k = \frac{1}{(2\pi)^4}\int e^{i\mathbf{k}\cdot\mathbf{R} - i\omega t} f(k,\omega)\,d^3k\,d\omega.$$

[13] See Eqs. (3.15), (4.4), and (4.14) of Ref. 9.

[14] See Eq. (5.1) of Ref. 9.

[15] This is obtained by using Eqs. (2.7) and (4.7) of Ref. 9. together with our defining Eq. (2.24).

[16] See, e.g., M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1965), 3rd revised ed., Chap. X.

[17] The problem is essentially the inverse of that discussed in Sec. 8.8 of Ref. 16.

[18] See Eq. (4.8) of Ref. 9.

[19] H.M.J. Boots, D. Bedeaux, and P. Mazur, Physica A **79**, 397 (1975).

Construction of quantum fields from Euclidean tensor fields

Te Hai Yao*

Mathematics Department, Bedford College, Regent's Park, London NW1, England
(Received 10 June 1975)

We define Euclidean tensor fields over $\mathcal{S}(\mathbb{R}^4)$, from which we construct quantum tensor fields satisfying all the Wightman axioms except the uniqueness of the vacuum. By a process of reduction, it is possible to obtain, from some suitably chosen Euclidean tensor field, a quantum field satisfying all the Wightman axioms except the uniqueness of the vacuum and transforming according to any arbitrarily chosen one-valued finite-dimensional irreducible representation of the restricted Lorentz group $L_+^\uparrow$. We give a Euclidean vector field and a Euclidean tensor field of rank two as examples, leading respectively to the real Proca Wightman field and the free electromagnetic Wightman field.

I. INTRODUCTION

In Ref. 1, Osterwalder and Schrader give a set of necessary and sufficient conditions on Euclidean Green's functions in order that they define a unique Wightman theory of some arbitrary spinor field. In Ref. 2, Nelson defines Euclidean scalar fields over $\mathcal{S}(\mathbb{R}^d)$, for $d \geq 2$, and show that, under certain hermiticity, integrability, and continuity conditions, they lead to scalar quantum fields in d-dimensional space–time satisfying all the Wightman axioms except the uniqueness of the vacuum, assuming a result of Ref. 1. In Ref. 3, a Euclidean covariant (in the sense that the expectation values of the fields are Euclidean covariant) Markov vector random field over $\mathcal{D}(\mathbb{R}^4)$ is found which leads to the real Proca Wightman field. Here we extend the notion of the Euclidean fields of Nelson from the scalar case to tensor cases, for the case of four-dimensional space–time. We find that these Euclidean tensor fields, which are Euclidean covariant (defined in terms of a representation of the full Euclidean group on a probability space by automorphisms of the measure algebra) tensor random fields over $\mathcal{S}(\mathbb{R}^4)$ satisfying a Markov property and a reflection principle, lead to, under certain hermiticity, integrability, continuity, and growth conditions, quantum tensor fields satisfying all the Wightman axioms except the uniqueness of the vacuum. In Secs. II and III we shall define Euclidean tensor fields and give the proof of the above statement (again using a result of Ref. 1). In Sec. IV, we shall consider a reduction procedure whereby quantum fields satisfying all the Wightman axioms except the uniqueness of the vacuum and transforming according to any arbitrary one-valued finite-dimensional irreducible representation of the restricted Lorentz group $L_+^\uparrow$ are obtained. In Sec. V we give examples of Euclidean tensor fields of rank one and two, respectively.

II. EUCLIDEAN TENSOR FIELDS

A Euclidean covariant lth rank tensor random field over $\mathcal{S}(\mathbb{R}^4)$ is a collection of random variables $\phi_{\mu_1\cdots\mu_l}(f)$ on a probability space $(\Omega, \mathcal{B}, \mu)$ [we assume that $\mathcal{B}$ is the smallest σ-algebra with respect to which all $\phi_{\mu_1\cdots\mu_l}(f)$ are measurable] indexed by l indices $\mu_i = 0,1,2,3$, $i = 1, \ldots, l$, and by the elements of $\mathcal{S}(\mathbb{R}^4)$, such that $\phi_{\mu_1\cdots\mu_l}(f)$ is linear and such that $\phi_{\mu_1\cdots\mu_l}(f_\alpha) \to \phi_{\mu_1\cdots\mu_l}(f)$ in measure if $f_\alpha \to f$ in the usual topology of $\mathcal{S}(\mathbb{R}^4)$, and such that there is a representation τ of the full Euclidean group on $\mathbb{R}^4$ on the underlying probability space[2] such that

$$\tau(a,A)\phi_{\mu_1\cdots\mu_l}(f) = \sum_{\nu_1\cdots\nu_l=0}^{3} A^{-1}_{\mu_1\nu_1}\cdots A^{-1}_{\mu_l\nu_l}\phi_{\nu_1\cdots\nu_l}(f_{(a,A)})$$

where $\tau(a,A)$ is the representative of the element (a,A) of the full Euclidean group in the representation τ, with a being a translation and A a rotation [defined here as an arbitrary element of $O(4)$], and $f_{(a,A)}$ is given by

$$f_{(a,A)}(x) = f(A^{-1}(x-a)), \quad x = (x_{(0)}, x_{(1)}, x_{(2)}, x_{(3)}) \in \mathbb{R}^4.$$

If we write $\Phi(f)$ for the set $\phi_{\mu_1\cdots\mu_l}(f)$, $\mu_i = 0,1,2,3$, $i = 1, \ldots l$, where $\phi_{\mu_1\cdots\mu_l}(f)$ is called the $\mu_1 \circ\circ\circ \mu_l$ component of $\Phi(f)$, and if we define an operator $T(a,A)$ on $\Phi(f)$ as

$$\{T(a,A)\Phi(f)\}_{\mu_1\cdots\mu_l} = \sum_{\nu_1\cdots\nu_l=0}^{3} A^{-1}_{\mu_1\nu_1}\cdots A^{-1}_{\mu_l\nu_l}\phi_{\nu_1\cdots\nu_l}(f_{(a,A)}),$$

then the above statement of Euclidean covariance can be written as

$$\tau(a,A)\Phi(f) = T(a,A)\Phi(f)$$

with

$$\{\tau(a,A)\Phi(f)\}_{\mu_1\cdots\mu_l} = \tau(a,A)\phi_{\mu_1\cdots\mu_l}(f).$$

In particular, if we let $(\rho) = (0, A_\rho)$ with

$$A_\rho = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then

$$\{T(\rho)\Phi(f)\}_{\mu_1\cdots\mu_l} = (-1)^{\sum_{j=1}^{l}\delta_{\mu_j,0}}\phi_{\mu_1\cdots\mu_l}(f_\rho).$$

We now define an Euclidean covariant Markov tensor random field of rank l over $\mathcal{S}(\mathbb{R}^4)$. We let U be an open set in $\mathbb{R}^4$, and we let $\sigma(U)$ be the σ-algebra generated by the random variables $\phi_{\mu_1\cdots\mu_l}(f)$ with $\mu_i = 0,1,2,3$, $i = 1, \ldots l$, and supp $f \subset U$. For any arbitrary subset V of $\mathbb{R}^4$ we define $\sigma(V) = \cap\, \sigma(U)$ with the intersection taken over all open sets containing V. Then an Euclidean co-

variant tensor random field $\Phi(f)$ of rank l is called Markov if

$$E(u \mid \sigma(\mathcal{U}')) = E(u \mid \sigma(\partial\mathcal{U}))$$

where $E(\ \mid\ )$ denotes conditional expectation, $\mathcal{U}'$ is the complement of $\mathcal{U}$, and $\partial\mathcal{U}$ the boundary of $\mathcal{U}$, for all positive random variables u measurable with respect to $\sigma(\mathcal{U})$.

The Euclidean covariant tensor random field $\Phi(f)$ is said to satisfy the reflection principle if

$$\tau(\rho)u = u$$

for any u belonging to $L^2(\Omega, \sigma_0, \mu)$, where $\sigma_0 = \sigma(\mathbb{R}^3)$, with $\mathbb{R}^3$ being the hyperplane $x_{(0)} = 0$. We have

$$\tau(\rho)\phi_{\mu_1\cdots\mu_l}(f) = (-1)^{\sum_{j=1}^{l}\delta_{\mu_j,0}}\ \phi_{\mu_1\cdots\mu_l}(f_\rho).$$

We now define an Euclidean tensor field of rank l over $\mathcal{S}(\mathbb{R}^4)$ to be an Euclidean covariant Markov tensor random field of rank l over $\mathcal{S}(\mathbb{R}^4)$ satisfying the above reflection principle. We further introduce the following assumptions:

(i) $\Phi(f)$ is Hermitian, i.e., $\Phi(f)$ is real when f is real.

(ii) For all $f \in \mathcal{S}(\mathbb{R}^4)$, $\phi_{\mu_1\cdots\mu_l}(f)$ is in L^p for $1 \leq p < \infty$ and for $\mu_i = 0, 1, 2, 3$, $i = 1, \ldots, l$, and the mapping

$$f_1 \otimes \cdots \otimes f_n \to E\phi_{\lambda_1}(f_1)\cdots\phi_{\lambda_n}(f_n)$$

is separately continuous, where each λ_k is a set of indices $\mu_1 \cdots \mu_l$, and further we have

$$\left|E\phi_{\lambda_1}(f_1)\cdots\phi_{\lambda_n}(f_n)\right| \leq c_1(n!)^{c_2}\prod_{i=1}^{n}\|f_i\|_s$$

where c_1 and c_2 are some positive numbers and

$$\|f_i\|_s = \sup_{\substack{x\in\mathbb{R}^4 \\ |\alpha|\leq s}} \left|(1+x^2)^{s/2}D^\alpha f(x)\right|$$

for some positive integer s.

We note that (i) guarantees that $\Phi^*(f) = \Phi(f^*)$ for all $f \in \mathcal{S}(\mathbb{R}^4)$.

We now define a Schwinger field $\hat{\Phi}(f)$ with components $\hat{\phi}_{\mu_1\cdots\mu_l}(f)$ defined by

$$\hat{\phi}_{\mu_1\cdots\mu_l}(f) = (-\sqrt{-1})^{\sum_{j=1}^{l}\delta_{\mu_j,0}}\ \phi_{\mu_1\cdots\mu_l}(f).$$

Assumption (ii) guarantees that there exists a tempered distribution $S_{\lambda_1\cdots\lambda_n}$ on $\mathbb{R}^4$ such that

$$E\hat{\phi}_{\lambda_1}(f_1)\cdots\hat{\phi}_{\lambda_n}(f_n)$$
$$= \int\cdots\int dx_1\cdots dx_n S_{\lambda_1\cdots\lambda_n}(x_1, \ldots, x_n)f_1(x_1)\cdots f_n(x_n).$$

We further define $S_0 = 1$.

We now claim that the sequence of distributions S_0, $S_{\lambda_1\cdots\lambda_n}$ satisfies the axioms (E0″)—(E3) of Ref. 1 for arbitrary spinor fields and hence leads to a theory of quantum tensor fields satisfying all the Wightman axioms except the uniqueness of the vacuum.

III. CONSTRUCTION OF QUANTUM TENSOR FIELDS

We now show that the sequence of distributions S_0, $S_{\lambda_1\cdots\lambda_n}$, satisfies the axioms (E0″)–(E3) of Ref. 1 for arbitrary spinor fields. First we note that (E0″) is satisfied by assumption.

To show that (E1) is satisfied, we note that for $R \in \mathrm{SO}(4)$, the maxtrix $\Lambda(U, V)$ of [1] satisfies

$$\Lambda(U, V) = PRP^{-1}$$

where

$$P = \begin{bmatrix} -\sqrt{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Hence we have

$$\Lambda(U^{-1}, V^{-1}) = PR^{-1}P^{-1}.$$

If we let $\Lambda(U^{-1}, V^{-1})^l$ be the direct product of $\Lambda(U^{-1}, V^{-1})$ with itself, taken l times, and if we define P^l, $(R^{-1})^l$, $(P^{-1})^l$ in a similar way, then we have

$$\Lambda(U^{-1}, V^{-1})^l = P^l(R^{-1})^l(P^{-1})^l.$$

We have

$$\hat{\phi}_\lambda(f) = \sum_{\lambda'}(P^l)_{\lambda\lambda'}\phi_{\lambda'}(f)$$

and

$$\tau(a, R)\hat{\phi}_\lambda(f) = \sum_{\lambda'}(\Lambda(U^{-1}, V^{-1})^l)_{\lambda\lambda'}\hat{\phi}_{\lambda'}(f_{(a,R)}).$$

Consequently, we have

$$\tau(a, R)\{\hat{\phi}_{\lambda_1}(f_1)\cdots\hat{\phi}_{\lambda_n}(f_n)\}$$
$$= \sum_{\lambda_1'\cdots\lambda_n'}(\Lambda(U^{-1}, V^{-1})^l)_{\lambda_1\lambda_1'}\cdots(\Lambda(U^{-1}, V^{-1})^l)_{\lambda_n\lambda_n'}$$
$$\hat{\phi}_{\lambda_1'}(f_{(a,R)})\cdots\hat{\phi}_{\lambda_n'}(f_{(a,R)})$$

and therefore

$$S_{\lambda_1\cdots\lambda_n}(x_1, \ldots, x_n)$$
$$= \sum_{\lambda_1'\cdots\lambda_n'}(\Lambda(U^{-1}, V^{-1})^l)_{\lambda_1\lambda_1'}\cdots(\Lambda(U^{-1}, V^{-1})^l)_{\lambda_n\lambda_n'}$$
$$\times S_{\lambda_1'\cdots\lambda_n'}(Rx_1 + a, \ldots, Rx_n + a).$$

To show that (E2) is satisfied, it is sufficient to show that the following inequality

$$\sum_{n,m=0}^{N}\sum_{\lambda,\chi}S_{\lambda_1\cdots\lambda_n x_1\cdots x_m}(\Theta\bar{f}_{n,\lambda_1\cdots\lambda_n}\times f_{m,x_1\cdots x_m}) \geq 0$$

(here $f_0 = 1$) where

$$(\Theta\bar{f}_{n,\lambda_1\cdots\lambda_n})(x_1, \ldots, x_n) = \bar{f}_{n,\lambda_1\cdots\lambda_n}(\theta x_1, \ldots, \theta x_n)$$

with

$$\theta x = (-x_{(0)}, x_{(1)}, x_{(2)}, x_{(3)})$$

is satisfied for

$$f_{n,\lambda_1\cdots\lambda_n}(x_1, \ldots, x_n) = f_{n1,\lambda_1}(x_1)\cdots f_{nn,\lambda_n}(x_n),$$

where

$$f_{n1,\lambda_1}(x_1)\in \mathcal{S}(\mathbb{R}^4) \text{ and } =0 \text{ unless } x_{1(0)}>0.$$

To prove the above inequality we proceed as follows. We define

$$\alpha=\sum_{n=0}^{N}\sum_{\lambda}\hat{\phi}_{\lambda_1}(f_{n1,\lambda_1})\cdots\hat{\phi}_{\lambda_n}(f_{nn,\lambda_n}).$$

(Here the empty product is by convention 1.) Then we have

$$\bar{\alpha}=\sum_{n=0}^{N}\sum_{\lambda}(\sqrt{-1})^{\sum_{j=1}^{l}\delta_{\mu_j(\lambda_1),0}}\cdots(\sqrt{-1})^{\sum_{j=1}^{l}\delta_{\mu_j(\lambda_n),0}}$$
$$\times\phi_{\lambda_1}(\bar{f}_{n1,\lambda_1})\cdots\phi_{\lambda_n}(\bar{f}_{nn,\lambda_n})$$

using hermiticity of $\Phi(f)$. Consequently, we have

$$\tau(\rho)\bar{\alpha}=\sum_{n=0}^{N}\sum_{\lambda}\hat{\phi}_{\lambda_1}(\Theta\bar{f}_{n1,\lambda_1})\cdots\hat{\phi}_{\lambda_n}(\Theta\bar{f}_{nn,\lambda_n})$$

with

$$(\Theta\bar{f}_{n1,\lambda_1})(x)=\bar{f}_{n1,\lambda_1}(\theta x), \text{ etc. .}$$

Hence we have

$$E((\tau(\rho)\bar{\alpha})\alpha)$$
$$=\sum_{n,m=0}^{N}\sum_{\lambda,\chi}S_{\lambda_1\cdots\lambda_n\chi_1\cdots\chi_m}(\Theta\bar{f}_{n,\lambda_1\cdots\lambda_n}\times f_{m,\chi_1\cdots\chi_m}).$$

Following Nelson,[2] we have, using Euclidean covariance, Markov property, reflection property, and assumption (ii), the relation

$$E((\tau(\rho)\bar{\alpha})\alpha)\geq 0.$$

Hence (E2) is proven. (E3) follows from the commutativity of random variables. We note that the quantum tensor field constructed from the Wightman distributions obtained as analytic continuations of the distributions S_0, $S_{\lambda_1\cdots\lambda}$, transform according to the tensor representation $\Lambda(Y,\bar{Y})^l$, i.e., the direct product of the representation $\Lambda(Y,\bar{Y})$ taken l times, of the restricted Lorentz group $L_+^\uparrow$, where Y is one of the two SL(2C) elements corresponding to a given element of $L_+^\uparrow$.

IV. REDUCTION

We now introduce a natural ordering in the set of tensor indices $\mu_1\cdots\mu_l$. Let $\mu_1\cdots\mu_l$ and $\mu_1'\cdots\mu_l'$ be two such tensor indices. We call $(\mu_1\cdots\mu_l)>(\mu_1'\cdots\mu_l')$ if starting from the left, the first unequal subindices μ_i and μ_i' is such that $\mu_i>\mu_i'$. We can then put the set of indices $\mu_1\cdots\mu_l$ into one-to-one correspondence with the set of integers $j=1,2,\ldots,4^l$, with the integer j corresponding to the tensor index $\mu_1\cdots\mu_l$ being greater than the integer j' corresponding to the tensor index $\mu_1'\cdots\mu_l'$ if $(\mu_1\cdots\mu_l)>(\mu_1'\cdots\mu_l')$.

Now since R^l is the direct product taken l times of the element R of SO(4), there exists a real $4^l\times4^l$ nonsingular matrix W which fully reduces R^l to one-valued finite-dimensional irreducible representations $M^{[i]}$, $i=1,\ldots,p$, of SO(4), i.e.,

$$WR^lW^{-1}=\sum_{i=1}^{p}{}^{\oplus}M^{[i]}\equiv M.$$

We define

$$Q_j(f)=\sum_{j'=1}^{4^l}W_{jj'}\phi_{j'}(f),$$

and defining $Q(f)$ to be a column vector with components $Q_j(f)$ and $\Phi(f)$ to be a column vector with components $\phi_j(f)$, we have

$$M^{-1}Q(f)=W(R^{-1})^l\Phi(f).$$

If we define

$$\check{M}=P^lM(P^{-1})^l,$$
$$\check{W}=P^lW(P^{-1})^l,$$
$$\check{Q}(f)=P^lQ(f),$$
$$\check{\Phi}(f)=P^l\Phi(f),$$

we obtain

$$\tau(a,R)\check{Q}(f)=\tau(a;R)\check{W}\check{\Phi}(f)=\check{M}^{-1}\check{Q}(f_{(a,R)}),$$

since

$$\tau(a,R)\Phi(f)=(R^{-1})^l\Phi(f_{(a,R)}).$$

$\check{W}$ is not real. However, it is possible to choose a quasidiagonal matrix

$$B=\sum_{i=1}^{p}{}^{\oplus}B^{[i]}$$

with $B^{[i]}$ nonsingular and of the same order as $M^{[i]}$, such that $\hat{W}\equiv B\check{W}$ is real and

$$\hat{M}\equiv B\check{M}B^{-1}=\sum_{i=1}^{p}{}^{\oplus}\hat{M}^{[i]}$$

with

$$\hat{M}^{[i]}=B^{[i]}\check{M}^{[i]}B^{[i]-1}.$$

Defining

$$\hat{Q}(f)=B\check{Q}(f),$$

we have

$$\tau(a,R)\hat{Q}(f)=\hat{M}^{-1}\hat{Q}(f_{(a,R)}).$$

We also have

$$\check{W}\Lambda(U^{-1},V^{-1})^l\check{W}^{-1}=\check{M}^{-1},$$

and

$$\hat{W}\Lambda(U^{-1},V^{-1})^l\hat{W}^{-1}=\hat{M}^{-1}.$$

We note that

$$\hat{M}^{[i]}=B^{[i]}CM^{[i]}C^{-1}B^{[i]-1}$$

where C is a nonsingular submatrix of P^l, and that $\hat{M}^{[i]}$ is a one-valued finite-dimensional irreducible representation of SO(4) equivalent to $M^{[i]}$.

Since

$$\hat{M}=\sum_{i=1}^{p}{}^{\oplus}\hat{M}^{[i]},$$

$\hat{Q}(f)$ splits up into parts $\hat{Q}^{[i]}(f)$, $i=1,\ldots,p$, each of which has its components transformed into a linear combination of its components under the action of $\hat{M}$. Thus

$$(\hat{M}\hat{Q}(f))^{[i]}{}_{\gamma}=\sum_{\delta}\hat{M}^{[i]}{}_{\gamma\delta}\hat{Q}^{[i]}(f)_{\delta}=(\hat{M}^{[i]}\hat{Q}^{[i]}(f))_{\gamma}.$$

We now assert that a quantum field satisfying all the Wightman axioms except the uniqueness of the vacuum can be constructed from the field $\hat{Q}^{[i]}(f)$, $i=1,\ldots,p$. We define

$$S^{[i]}{}_{\gamma_1\cdots\gamma_n}(f_1,\ldots,f_n)=E\hat{Q}^{[i]}{}_{\gamma_1}(f_1)\cdots Q^{[i]}{}_{\gamma_n}(f_n),$$
$$S^{[i]}{}_{0}\equiv 1.$$

Then the sequence of distributions $S^{[i]}{}_{0}$, $S^{[i]}{}_{\gamma_1\ldots\gamma_n}$, satisfies the axioms (E0″) – (E3) of Ref. 1 for arbitrary spinor fields. Again (E0″) is satisfied as can be easily seen. (E1) is satisfied since we have

$$\tau(a,R)\hat{Q}^{[i]}(f)=\hat{M}^{[i]-1}\hat{Q}^{[i]}(f_{(a,R)})$$

and

$$\tau(a,R)\{\hat{Q}^{[i]}{}_{\gamma_1}(f_1)\cdots\hat{Q}^{[i]}{}_{\gamma_n}(f_n)\}$$
$$=\sum_{\delta_1,\cdots\delta_n}\hat{M}^{[i]-1}{}_{\gamma_1\delta_1}\cdots\hat{M}^{[i]-1}_{\gamma_n\delta_n}\hat{Q}^{[i]}_{\delta_1}(f_{1(a,R)})\cdots\hat{Q}^{[i]}_{\delta_n}(f_{n(a,R)}).$$

Taking expectation values and noting that $\tau(a,R)$ is measure preserving, we have

$$S^{[i]}_{\gamma_1\cdots\gamma_n}(x_1,\ldots,x_n)$$
$$=\sum_{\delta_1,\cdots\delta_n}\hat{M}^{[i]}_{\gamma_1\delta_1}(U^{-1},V^{-1})\cdots\hat{M}^{[i]}_{\gamma_n\delta_n}(U^{-1},V^{-1})$$
$$\times S^{[i]}_{\delta_1\cdots\delta_n}(Rx_1+a,\ldots,Rx_n+a),$$

which is $E(1)$ of Ref. 1 for arbitrary spinor fields.

To show that $(E2)$ is satisfied, we put

$$\alpha=\sum_{n=0}^{N}\sum_{\gamma}\hat{Q}^{[i]}_{\gamma_1}(f_{n1,\gamma_1})\cdots\hat{Q}^{[i]}_{\gamma_n}(f_{nn},\gamma_n).$$
$$=\sum_{n=0}^{N}\sum_{\gamma}[\hat{W}\check{\Phi}(f_{n1,\gamma_1})]^{[i]}_{\gamma_1}\cdots[\hat{W}\check{\Phi}(f_{nn,\gamma_n})]^{[i]}_{\gamma_n}.$$

(Here again the empty product is by convention 1.) Then we have

$$\tau(\rho)\bar{\alpha}=\sum_{n=0}^{N}\sum_{\gamma}\hat{Q}^{[i]}_{\gamma_1}(\Theta\bar{f}_{n1,\gamma_1})\cdots\hat{Q}^{[i]}_{\gamma_n}(\Theta\bar{f}_{nn,\gamma_n}).$$

Consequently, we have (where, again, $f_0=1$)

$$\sum_{n,m=0}^{N}\sum_{\gamma,\delta}S^{[i]}_{\gamma_1\cdots\gamma_n\delta_1\cdots\delta_m}(\Theta\bar{f}_{n,\gamma_1\cdots\gamma_n}\times f_{m,\delta_1\ldots\delta_m})$$
$$=E((\tau(\rho)\bar{\alpha})\alpha)\geq 0,$$

where

$$f_{n,\gamma_1\cdots\gamma_n}(x_1,\ldots,x_n)=f_{n1,\gamma_1}(x_1)\circ\cdots f_{nn,\gamma_n}(x_n),$$

with

$$f_{ni,\gamma_i}(x_i)\in\mathcal{S}(\mathbb{R}^4)\text{ and }=0\text{ unless }x_{i_{(0)}}>0.$$

Hence, by the same reasoning as before, (E2) is satisfied. (E3) is satisfied since $\hat{Q}^{[i]}_{\gamma}(f)$ are random variables. Hence we get a quantum field theory satisfying all the Wightman axioms except the uniqueness of the vacuum, from the random field $\hat{Q}^{[i]}(f)$ over $\mathcal{S}(\mathbb{R}^4)$, for each $i=1,\ldots,p$.

Now any of the one-valued finite-dimensional irreducible representations of SO(4) [i.e., any of the $\tilde{D}^{j,k}$ class of SO(4) with $j,k\geq 0$ and both integral or half-integral] is equivalent to some $M^{[i]}$ obtained in the complete reduction of a tensor representation R^l of SO(4) with suitable l, and conversely, any $M^{[i]}$ obtained in the complete reduction of any tensor representation R^l of SO(4) is equivalent to some one-valued finite-dimensional irreducible representation of SO(4). According to Ref. 1, the quantum field constructed from $\hat{Q}^{[i]}(f)$ transforms according to some one-valued finite-dimensional irreducible representation of the restricted Lorentz group $L_+^\uparrow$ (i.e., equivalent to $D^{j,k}$ of $L_+^\uparrow$ with $j,k\geq 0$ and both integral or half-integral). Further, by this process of reduction, we always get a field $\hat{Q}^{[i]}(f)$ which gives rise to a quantum field satisfying all the Wightman axioms except the uniqueness of the vacuum and transforming according to some arbitrary one-valued finite-dimensional irreducible representation of the restricted Lorentz group $L_+^\uparrow$.

V. EXAMPLES

A. A Euclidean vector field

We consider the vector random field $\Phi(f)$ over $\mathcal{S}(\mathbb{R}^4)$ obtained by extension of the Gaussian vector random field $\Phi(f)$ over $\mathcal{S}_r(\mathbb{R}^4)$, the space of real elements of $\mathcal{S}(\mathbb{R}^4)$, defined by

$$E\phi_\mu(f)=0,$$
$$E\phi_\mu(f)\phi_\nu(g)=\left\langle f,\frac{\delta_{\mu\nu}-(1/m^2)\partial_\mu\partial_\nu}{-\Delta+m^2}g\right\rangle_{L^2(\mathbb{R}^4)}$$

where $\phi_\mu(f)$ and $\phi_\nu(g)$ are the components of $\Phi(f)$ and $\Phi(g)$, respectively, with $\mu,\nu=0,1,2,3$, and where Δ is the four-dimensional Laplacian operator and $\langle\ ,\ \rangle_{L^2(\mathbb{R}^4)}$ is the scalar product in $L^2(\mathbb{R}^4)$.

The vector random field $\Phi(f)$ over $\mathcal{S}(\mathbb{R}^4)$ can be shown to be Markov in a way exactly parallel to what is done in Ref. 3. Further, assumptions (i) and (ii) of Sec. II are satisfied as a consequence respectively of the Gaussian nature of the field when restricted to $\mathcal{S}_r(\mathbb{R}^4)$ and of Wick's theorem for jointly Gaussian random variables. We now demonstrate that the vector field $\Phi(f)$ satisfies Euclidean covariance in the sense of Sec. II and the reflection principle.

We consider the Hilbert space $\mathcal{H}$ obtained by completing the pre-Hilbert space $\mathcal{H}'=\mathcal{S}_{4r}(\mathbb{R}^4)=\mathcal{S}_r(\mathbb{R}^4)\times\mathcal{S}_r(\mathbb{R}^4)\times\mathcal{S}_r(\mathbb{R}^4)\times\mathcal{S}_r(\mathbb{R}^4)$ equipped with the scalar product

$$\langle F,G\rangle=\sum_{\mu,\nu=0}^{3}\int dp\,\tilde{f}_\mu(p)\tilde{g}_\nu(p)\frac{\delta_{\mu\nu}+(1/m^2)p_\mu p_\nu}{p^2+m^2}$$

where $F=\{f_0,f_1,f_2,f_3\}\in\mathcal{H}'$ and $G=\{g_0,g_1,g_2,g_3\}\in\mathcal{H}'$, and the tilde denotes Fourier transform. We let $J(a,A)$ denote the transformation in

$$J(a,A)\mathcal{J}=A\mathcal{J}_{(a,A)},\quad\mathcal{J}\in\mathcal{H}$$

with

$$\mathcal{J}_{(a,A)}=\mathcal{J}(A^{-1}(x-a)),$$

then we have

$$\langle J(a,A)\mathcal{J},\ J(a,A)\mathcal{J}\rangle=\langle\mathcal{J},\mathcal{J}\rangle.$$

Hence $J(a,A)$ is an orthogonal transformation in $\mathcal{H}$.

Since H is real, and since β is the σ-algebra generated by $\phi(H)$, the Gaussian process over H defined by

$$\phi(F)=\phi_0(f_0)+\phi_1(f_1)+\phi_2(f_2)+\phi_3(f_3)$$

for $F\in H'$, and $\phi(\mathcal{J})=$ the fundamental sequence $\{\phi(F_1), \phi(F_2),\cdots\}\in L^2(\Omega,\beta,\mu)$ when $\mathcal{J}$ is a fundamental sequence $\{F_1,F_2,\cdots\}$ with $F_i\in H'$, there exists an automorphism $\tau(a,A)$ of the measure algebra of (Ω,β,μ) such that

$$\tau(a,A)\phi(\mathcal{J})=\phi(J(a,A)\mathcal{J}).$$

Consequently, we have

$$\tau(a,A)\left(\sum_{\mu=0}^{3}\phi_\mu(\mathcal{J}_\mu)\right)=\sum_{\mu,\nu=0}^{3}A^{-1}_{\mu\nu}\phi_\nu(\mathcal{J}_{\mu,(a,A)})$$

where $\mathcal{J}_\mu$ are the components of $\mathcal{J}$ and

$$\mathcal{J}_{\mu,(a,A)}(x)=\mathcal{J}_\mu(A^{-1}(x-a)).$$

Hence we have, for $f\in\mathcal{S}_r(\mathbb{R}^4)$, the relation

$$\tau(a,A)\phi_\mu(f)=\sum_{\nu=0}^{3}A^{-1}_{\mu\nu}\phi_\nu(f_{(a,A)}).$$

The same relation holds for $f\in\mathcal{S}(\mathbb{R}^4)$. Further, τ is a representation of the full Euclidean group on $\mathbb{R}^4$ on (Ω,β,μ). Hence we have demonstrated Euclidean covariance.

To show that $\Phi(f)$ satisfies the reflection principle, we let

$$H_0=\bigcap_\beta H_\beta$$

where H_β is the completion of the pre-Hilbert space $H'_\beta=\{F\in H', \text{supp } F\subset U_\beta, U_\beta$ being any arbitrary open set containing the hyperplane $\mathbb{R}^3\}$, equipped with the scalar product $\langle\ ,\ \rangle$. We know that H_β and H_0 are real Hilbert spaces. We have

$$\text{Fock space over } H_0=\bigcap_\beta(\text{Fock space over } H_\beta).$$

Hence, if we let $\sigma_0=\sigma$-algebra generated by the Gaussian process $\phi(H_0)$ over H_0, and $\sigma_\beta=\sigma$-algebra generated by the Gaussian process $\phi(H_\beta)$ over H_β, then by Segal isomorphism we have

$$L^2(\Omega,\sigma_0,\mu)=\bigcap_\beta L^2(\Omega,\sigma_\beta,\mu)$$

since H_β and H_0 are subspaces of H. Hence we have

$$L^2(\Omega,\sigma_0,\mu)=\bigcap_\beta L^2(\Omega,\sigma_\beta,\mu)=\bigcap_\beta L^2(\Omega,\sigma_\beta,\mu)=L^2(\Omega,\sigma_0,\mu)$$

where σ_β is the σ-algebra generated by $\phi_\mu(f)$, $\mu=0,1,2,3$, $f\in\mathcal{S}(\mathbb{R}^4)$, supp $f\subset U_\beta$. Hence $L^2(\Omega,\sigma_0,\mu)$ is Segal isomorphic to the Fock space over H_0. Now the space H_0 consists of certain distributions over $\mathcal{S}_{4r}(\mathbb{R}^4)$ with support on the hyperplane $\mathbb{R}^3$. They are elements of H of the form

$$\mathcal{J}=\{0,f_1\otimes\delta,f_2\otimes\delta,f_3\otimes\delta\}$$

where $\tilde{f}_i$ belongs to a subspace of $L^2(\mathbb{R}^3)$, with the tilde denoting Fourier transform.

If we let T_ρ be the operator on H_0 defined by

$$(T_\rho\mathcal{J})_\mu=(-1)^{\delta_{\mu,0}}(\mathcal{J}_-)_\mu$$

with

$$\mathcal{J}_-(x_{(0)},x_{(1)},x_{(2)},x_{(3)})=\mathcal{J}(-x_{(0)},x_{(1)},x_{(2)},x_{(3)})$$

for

$$\mathcal{J}\in H_0,$$

then we have

$$\begin{aligned}\phi(T_\rho\mathcal{J})&=\sum_{\mu=0}^{3}\phi_\mu((T_\rho\mathcal{J})_\mu)\\&=\sum_{\mu=0}^{3}(-1)^{\delta_{\mu,0}}\phi_\mu((\mathcal{J}_-)_\mu)\\&=\tau(\rho)\left(\sum_{\mu=0}^{3}\phi_\mu(\mathcal{J}_\mu)\right)\\&=\tau(\rho)\phi(\mathcal{J})\end{aligned}$$

by Euclidean covariance.

Since T_ρ leaves H_0 pointwise invariant, we have

$$\tau(\rho)\phi(\mathcal{J})=\phi(\mathcal{J})\quad\text{for } \mathcal{J}\in H_0.$$

Hence $\tau(\rho)=\Gamma(T_\rho)$ on $L^2(\Omega,\sigma_0,\mu)$, where $\Gamma(T_\rho)$ is the operator on $L^2(\Omega,\sigma_0,\mu)$ corresponding to the operator T_ρ on H_0 via Segal isomorphism. Consequently, we have

$$\tau(\rho)u=u$$

for all $u\in L^2(\Omega,\sigma_0,\mu)$. Hence we have proved that $\Phi(f)$ satisfies the reflection property. We note that this Euclidean vector field leads to the real Proca Wightman field via the procedure discussed in Sec. II.

B. A Euclidean tensor field of rank two

Here we construct an Euclidean tensor field of rank two which leads to the free electromagnetic Wightman field via the procedure of Sec. II.

To do so, we first construct a pre-Hilbert space H' of antisymmetric 4×4 matrices F whose elements $F_{\mu_1\mu_2}$, $\mu_1,\mu_2=0,1,2,3$, belong to $\mathcal{S}_r(\mathbb{R}^4)$ with scalar product

$$\langle F,G\rangle=(1/4)\sum_{\substack{\mu_1,\mu_2\\ \nu_1,\nu_2=0}}^{3}\int dp\,\tilde{F}_{\mu_1\mu_2}(p)\tilde{K}_{\mu_1\mu_2,\nu_1\nu_2}(p)\tilde{G}_{\nu_1\nu_2}(p)$$

where $F,G\in H'$, $\tilde{F}_{\mu_1\mu_2}$ and $\tilde{G}_{\nu_1\nu_2}$ are respectively the Fourier transforms of $F_{\mu_1\mu_2}$ and $G_{\nu_1\nu_2}$, and where

$$\begin{aligned}\tilde{K}_{\mu_1\mu_2,\nu_1\nu_2}(p)=(1/p^2)(&p_{\mu_1}p_{\nu_1}\delta_{\mu_2\nu_2}+p_{\mu_2}p_{\nu_2}\delta_{\mu_1\nu_1}-p_{\mu_1}p_{\nu_2}\delta_{\mu_2\nu_1}\\&-p_{\mu_2}p_{\nu_1}\delta_{\mu_1\nu_2}).\end{aligned}$$

We have

$$\langle F,G\rangle=\sum_{\substack{(\mu_1,\mu_2)\in I\\ (\nu_1,\nu_2)\in I}}\int dp\,\tilde{F}_{\mu_1\mu_2}(p)\tilde{K}_{\mu_1\mu_2,\nu_1\nu_2}(p)\tilde{G}_{\nu_1\nu_2}(p)$$

where $I=\{(0,1),(0,2),(0,3),(1,2),(1,3),(2,3)\}$, and where the matrix $\tilde{K}_{\mu_1\mu_2,\nu_1\nu_2}(p)$, for (μ_1,μ_2) and $(\nu_1,\nu_2)\in I$, is of the following form:

$$
(\nu_1,\nu_2)= \begin{matrix} (0,1) & (0,2) & (0,3) & (1,2) & (1,3) & (2,3) \end{matrix} \quad (\mu_1,\mu_2)=
$$

$$
(1/p^2)\begin{bmatrix}
p_0^2+p_1^2 & p_1p_2 & p_1p_3 & -p_0p_2 & -p_0p_3 & 0 \\
p_1p_2 & p_0^2+p_2^2 & p_2p_3 & p_0p_1 & 0 & -p_0p_3 \\
p_1p_3 & p_2p_3 & p_0^2+p_3^2 & 0 & p_0p_1 & p_0p_2 \\
-p_0p_2 & p_0p_1 & 0 & p_1^2+p_2^2 & p_2p_3 & -p_1p_3 \\
-p_0p_3 & 0 & p_0p_1 & p_2p_3 & p_1^2+p_3^2 & p_1p_2 \\
0 & -p_0p_3 & p_0p_2 & -p_1p_3 & p_1p_2 & p_2^2+p_3^2
\end{bmatrix}
\begin{matrix} (0,1) \\ (0,2) \\ (0,3) \\ (1,2) \\ (1,3) \\ (2,3) \end{matrix}
$$

We let H denote the completion of H', and we define a Gaussian process $\phi(H)$ over H with mean zero and covariance given by

$$E\phi(\mathcal{F})\phi(\mathcal{G})=\langle \mathcal{F},\mathcal{G}\rangle$$

for $\mathcal{F},\mathcal{G}\in H$, such that $\phi(\mathcal{F}_\alpha)\to\phi(\mathcal{F})$ in measure if $\mathcal{F}_\alpha \to \mathcal{F}$ in H. For $f\in \mathcal{S}_r(\mathbb{R}^4)$, we define

$$\phi_{\mu_1\mu_2}(f)=\phi(\mathcal{F})$$

where $\mathcal{F}$ is an element of H with

$$\mathcal{F}_{\lambda_1\lambda_2}=0 \text{ for } (\lambda_1,\lambda_2)\neq(\mu_1,\mu_2),\ (\lambda_1,\lambda_2)\neq(\mu_2,\mu_1),$$

$$\mathcal{F}_{\mu_1\mu_2}=f,$$

$$\mathcal{F}_{\mu_2\mu_1}=-f.$$

Then $\phi_{\mu_1\mu_2}(f)$, for $\mu_1,\mu_2=0,1,2,3$, $f\in\mathcal{S}_r(\mathbb{R}^4)$, are jointly Gaussian random variables and further we have $\phi_{\mu_1\mu_2}(f_\alpha)\to\phi_{\mu_1\mu_2}(f)$ in measure if $f_\alpha\to f$ in $\mathcal{S}_r(\mathbb{R}^4)$. For $f\in\mathcal{S}(\mathbb{R}^4)$, we define $\phi_{\mu_1\mu_2}(f)$ through linearity and we have the same continuity property.

We define an operator $J(a,A)$ on H by

$$J(a,A)\mathcal{F}=(A\times A)\mathcal{F}_{(a,A)}$$

with

$$\mathcal{F}_{(a,A)}(x)=\mathcal{F}(A^{-1}(x-a)).$$

Then we have

$$\langle J(a,A)\mathcal{F},J(a,A)\mathcal{F}\rangle=\langle\mathcal{F},\mathcal{F}\rangle.$$

Consequently, $J(a,A)$ is an orthogonal transformation in H. Hence there exists a representation τ of the full Euclidean group on $\mathbb{R}^4$ on $(\Omega,\mathcal{B},\mu)$ such that

$$\tau(a,A)\phi(\mathcal{F})=\phi(J(a,A)\mathcal{F})$$

and hence

$$\tau(a,A)\phi_{\mu_1\mu_2}(f)=\sum_{\nu_1,\nu_2=0}^{3}A^{-1}_{\mu_1\nu_1}A^{-1}_{\mu_2\nu_2}\phi_{\nu_1\nu_2}(f_{(a,A)})$$

for $f\in\mathcal{S}_r(\mathbb{R}^4)$ and hence also for $f\in\mathcal{S}(\mathbb{R}^4)$. Hence we have shown that $\phi_{\mu_1\mu_2}(f)$, for $f\in\mathcal{S}(\mathbb{R}^4)$, are components of an Euclidean covariant tensor random field $\Phi(f)$ over $\mathcal{S}(\mathbb{R}^4)$ of rank two.

We now let U be an open set in $\mathbb{R}^4$, and let U' be its complement, ∂U its boundary. We let

$$M=\{\mathcal{F}\in H,\ \text{supp}\,\mathcal{F}\subset U'\},$$

$$N=\{\mathcal{F}\in H,\ \text{supp}\,\mathcal{F}\subset \partial U\}.$$

We let $\mathcal{F}$ be an element of H with $\text{supp}\,\mathcal{F}\subset U$ and $\mathcal{F}_{\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(0,1)$ and $(1,0)$. Then the orthogonal projection $\mathcal{F}_M$ of $\mathcal{F}$ onto M satisfies $\mathcal{F}_{M,\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(0,1)$ and $(1,0)$ and $\mathcal{F}_{M,10}=-\mathcal{F}_{M,01}$. We let $\mathcal{G}\in H$ with $\mathcal{G}_{\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(0,1)$ and $(1,0)$ and $\mathcal{G}_{01}\in\mathcal{D}_r(U'^0)$, the space of real infinitely differentiable functions of compact support with support contained in U'^0, the interior of U'. Then we have

$$\langle\mathcal{G},\mathcal{F}_M\rangle=\langle\mathcal{G},\mathcal{F}\rangle\ ,$$

i.e.,

$$\int dx\,\mathcal{G}_{01}(x)\frac{(\partial_0^2+\partial_1^2)}{-\Delta}\mathcal{F}_{M,01}(x)=\int dx\,\mathcal{G}_{01}(x)\frac{(\partial_0^2+\partial_1^2)}{-\Delta}\mathcal{F}_{01}(x).$$

Consequently, we have

$$\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}_{M,01}(x)=\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}_{01}(x)$$

for any $g\in\mathcal{D}_r(U'^0)$.

Similarly, we let $\mathcal{F}'$ be an element of H with $\text{supp}\,\mathcal{F}'\subset U$ and $\mathcal{F}'_{\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(2,3)$ and $(3,2)$. Then the orthogonal projection $\mathcal{F}'_M$ of $\mathcal{F}'$ onto M satisfies $\mathcal{F}'_{M,\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(2,3)$ and $(3,2)$, and $\mathcal{F}'_{M,32}=-\mathcal{F}'_{M,23}$. We let $\mathcal{G}'\in H$ with $\mathcal{G}'_{\mu_1\mu_2}=0$ except for $(\mu_1,\mu_2)=(2,3)$ and $(3,2)$ and $\mathcal{G}'_{23}\in\mathcal{D}_r(U'^0)$. Then we have

$$\langle\mathcal{G}',\mathcal{F}'_M\rangle=\langle\mathcal{G}',\mathcal{F}'\rangle,$$

i.e.,

$$\int dx\,\mathcal{G}'_{23}(x)\frac{(\partial_2^2+\partial_3^2)}{-\Delta}\mathcal{F}'_{M,23}(x)=\int dx\,\mathcal{G}'_{23}(x)\frac{(\partial_2^2+\partial_3^2)}{-\Delta}\mathcal{F}'_{23}(x).$$

Consequently, we have

$$\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}'_{M,23}(x)=\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}'_{23}(x)$$

for any $g\in\mathcal{D}_r(U'^0)$.

If we choose $\mathcal{F}'_{23}=\mathcal{F}_{01}$, we have, for any $g\in\mathcal{D}_r(U'^0)$, and consequently for any $g\in\mathcal{D}(U'^0)$, the space of infinitely differentiable functions with compact support with support contained in U'^0, the equality

$$\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}_{M,01}(x)=\int dx\,g(x)(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}'_{M,23}(x).$$

Hence we have

$$(\partial_0^2+\partial_1^2)(\partial_2^2+\partial_3^2)\mathcal{F}_{M,01}(x)$$

$$= (\partial_0^2 + \partial_1^2)(\partial_2^2 + \partial_3^2)\mathcal{F}'_{\mathcal{M},23}(x)$$

as distributions on $\mathcal{U}'^0$. Consequently, we have

$$\mathcal{F}_{\mathcal{M},01} = \mathcal{F}'_{\mathcal{M},23} + K$$

where $K(x)$ is a distributional solution of

$$(\partial_0^2 + \partial_1^2)(\partial_2^2 + \partial_3^2)K(x) = 0$$

in $\mathcal{U}'^0$, and $K(x)$ can be written [one way to see this is to notice that every distributional solution $K^{(23)}(x_{(2)}, x_{(3)})$ of the Laplace's equation

$$(\partial_2^2 + \partial_3^2)K^{(23)}(x_{(2)}, x_{(3)}) = 0$$

in $\mathcal{U}'^0$ is a C^∞ solution of the equation in the same region[4] as

$$K(x) = \sum_{i=1}^{\infty} k^{(01)}(x_{(0)}, x_{(1)}) k_i^{(23)}(x_{(2)}, x_{(3)})$$

in some sufficiently small neighbourhood of any point in $\mathcal{U}'^0$, where $k_i^{(23)}(x_{(2)}, x_{(3)})$ satisfies

$$(\partial_2^2 + \partial_3^2)k_i^{(23)}(x_{(2)}, x_{(3)}) = 0$$

in that neighbourhood. Choosing $\mathcal{G}_{01}$ and $\mathcal{G}'_{23} = g$ for some $g \in \mathcal{D}_r(\mathcal{U}'^0)$, we have

$$\int dx\, g(x) \frac{(\partial_0^2 + \partial_1^2)}{-\Delta} \mathcal{F}_{\mathcal{M},01}(x)$$

$$= \int dx\, g(x) \frac{(\partial_0^2 + \partial_1^2)}{-\Delta} \mathcal{F}_{01}(x)$$

and

$$\int dx\, g(x) \frac{(\partial_2^2 + \partial_3^2)}{-\Delta} \mathcal{F}_{\mathcal{M},01}(x)$$

$$= \int dx\, g(x) \frac{(\partial_2^2 + \partial_3^2)}{-\Delta} \{\mathcal{F}'_{\mathcal{M},23}(x) + K(x)\}$$

$$= \int dx\, g(x) \frac{(\partial_2^2 + \partial_3^2)}{-\Delta} \mathcal{F}'_{\mathcal{M},23}(x)$$

$$= \int dx\, g(x) \frac{(\partial_2^2 + \partial_3^2)}{-\Delta} \mathcal{F}_{01}(x).$$

Hence we have

$$\int dx\, g(x) \mathcal{F}_{\mathcal{M},01}(x) = \int dx\, g(x) \mathcal{F}_{01}(x) = 0$$

for any $g \in \mathcal{D}_r(\mathcal{U}'^0)$ and hence any $g \in \mathcal{D}(\mathcal{U}'^0)$. Consequently, $\mathcal{F}_{\mathcal{M},01}$ has support on $\partial\mathcal{U}$, i.e., $\mathcal{F}_{\mathcal{M}} \in \mathcal{N}$.

By an extension of this argument, we can show that the orthogonal projection $\mathcal{F}_{\mathcal{M}}$ of any $\mathcal{F} \in \mathcal{H}$ with supp $\mathcal{F} \subset \mathcal{U}$, onto $\mathcal{M}$, is in $\mathcal{N}$. Hence, following Nelson,[2] we can show that the Euclidean covariant tensor random field $\Phi(f)$ over $\mathcal{S}(\mathbb{R}^4)$ is Markov.

The proof that the field $\Phi(f)$ satisfies the reflection principle is similar to that of the previous example in Sec. V.A, noting that the corresponding subspace H_0 here consists of elements of the form

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & f_{12} \otimes \delta & f_{13} \otimes \delta \\ 0 & -f_{12} \otimes \delta & 0 & f_{23} \otimes \delta \\ 0 & -f_{13} \otimes \delta & -f_{23} \otimes \delta & 0 \end{bmatrix}$$

where $\tilde{f}_{ij}$ belongs to a subspace of $L^2(\mathbb{R}^3)$, with $\tilde{f}_{ij}$ being the Fourier transform of f_{ij}.

Assumptions (i) and (ii) of Sec. II are satisfied in this case for the same reason as they are satisfied in the previous example.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. F. Streater, G. Ekhaguere, S.C. Lim, and B. Skagerstam for fruitful conversations. He also wishes to thank Bedford College for hospitality and the Science Research Council for financial support.

*Present address: Institute for Advanced Studies, School of Theoretical Physics, 10 Burlington Road, Dublin 4, Ireland.

[1] K. Osterwalder and R. Schrader, Commun. Math. Phys. 31, 83 (1973); "Axioms for Euclidean Green's Functions II," Harvard University Preprint (1974).

[2] E. Nelson, "Probability Theory and Euclidean Field Theory," in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973).

[3] T. H. Yao, Commun. Math. Phys. 41, 267 (1975).

[4] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968), Vol. II.

Superposition of states and the structure of quantum logics

Platon C. Deliyannis

Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 14 April 1975)

A derivation of the classical Hilbert space model of quantum theory is given (including superselection rules) based (i) on various axioms on the behavior of events and (ii) an explicit formula for the superposition of two pure states in terms of transition probabilities.

1. INTRODUCTION

The purpose of this paper is to present a derivation of the classical representation scheme for quantum logics: a family of pairwise orthogonal Hilbert spaces (one for each atom in the center of the logic) the rays of which correspond to the pure states, while the decomposable self-adjoint operators correspond to the observables.

This derivation is based on a set of several plausible axioms concerning the behavior of events and states, and also on the use of an explicit formula for the superposition of pure states. In this respect we have gone beyond the ideas studied, e.g., in Refs. 1 and 2, by providing specific properties of superpositions rather than a general qualitative definition. In particular it appeared desirable to differentiate between mixtures and superpositions. The former is a strictly classical concept arrived at by considerations regarding ensembles; the concept of superposition, however, involves a sense of interaction between two pure states, apparently due to the nonzero probability of spontaneous (noncausal) transition between pure states. We have tried to make this interaction clear in our formulation.

Section 2 lays the general background on which the derivation is based. In Secs. 3, 4, and 5 we develop the first consequences of our hypotheses on the behavior of superpositions. In Sec. 6 we obtain the structure of the pure state space and in 7 the structure of the logic itself. The last section gives a geometrical criterion for the validity of our superposition axioms.

2. GENERAL BACKGROUND

The class of logics $\mathcal{L}$ we are concened with shall be assumed to satisfy the properties listed below. We shall refer to the elements of $\mathcal{L}$ as "events." The undefined terms shall be, as usual, the relation $\leq$ which we interpret as implication between two events, and the operation $'$ of orthocomplementation which we interpret as negation.

Axiom (i): (a) For all A, B, $C \in \mathcal{L}$ we have: $A \leq A$; $A \leq B$ and $B \leq A$ imply $A = B$; and $A \leq B$, $B \leq C$ imply $A \leq C$. There exist events 0, $I \in \mathcal{L}$ such that $0 \leq A \leq I$ for $A \in \mathcal{L}$.

(b) Furthermore, $(A')' = A$, $A \leq B$ implies $B' \leq A'$ and $A \wedge A' = 0$, $A \vee A' = I$, where $\wedge, \vee$ denote infimum and supremum relative to the partial order $\leq$.

We are not assuming $A \wedge B$, $A \vee B$ to exist universally; we shall be able to establish this later on.

An *atom* of $\mathcal{L}$ is an event A such that $0 \leq B \leq A$ implies $B = 0$ or $B = A$.

Axiom (ii): (a) If $A \leq B$, then $A' \wedge B$ exists and $B = A \vee (A' \wedge B)$.

(b) Furthermore, if $B = A \vee C$ with A, C atoms, then $A' \wedge B$ is also an atom.

The first part is the classical orthomodularity, while the second is a weak form of Piron's covering axiom.[3]

Notation: In case $A \leq B'$ (or equivalently $B \leq A'$) we write $A \perp B$ and say that A, B are disjoint.

Axiom (iii): If $A_1, A_2, \cdots$ is a (finite or infinite) sequence of pairwise disjoint events, then their supremum exists.

Notation: For traditional reasons we shall write this as $\sum A_i$ rather than $\bigvee A_i$, or as $A_1 + A_2 + \cdots + A_n$ in case of a finite sequence.

A *state* of $\mathcal{L}$ is a map $m: \mathcal{L} \to [0,1]$ such that (a) $m(0) = 0$, $m(I) = 1$, (b) $m(\sum A_i) = \sum m(A_i)$. The interpretation is, of course, that $m(A)$ is the probability of occurrence of A, while the "system" is in the state m. We shall write $\mathcal{M}$ for the set of all states.

Axiom (iv): If for each state m for which $m(A) = 1$ we also have $m(B) = 1$, then $A \leq B$.

The physical meaning of this is quite obvious and gives a precise contant to the otherwise vague word "implication."

A state m is a *mixture* of the states $\{m_x\}_{x \in X}$ if there is a positive measure μ on X of total mass 1 such that $m(A) = \int_X m_x(A)\, d\mu(x)$. We shall abbreviate this to $m = \int_X m_x\, d\mu(x)$.

A state is *pure* if it is not a mixture; we shall write $\mathcal{M}_p$ for the set of pure states.

Axiom (v): The pure states generate all states by mixtures. Specifically, there exists a measurable space X such that for every $m \in \mathcal{M}$ we have a family $\{m_x\}_{x \in X}$ with $m_x \in \mathcal{M}_p$ for all $x \in X$ and a positive measure μ on X with $\mu X = 1$ such that $m = \int_X m_x\, d\mu(x)$.

There is certainly no need to defend the plausibility of this hypothesis on physical grounds.

Axiom (vi): If, for some pure state m, all events A_i (any family) occur with certainty, then the infimum A of the family $\{A_i\}$ exists and occurs with certainty in m.

Again, the physical interpretation of this is immediate; indeed one is tempted to assume this for any state, not just for the pure states. We do not really need to go that far. One of the consequences of these axioms shall be the completeness of $\mathcal{L}$.

We shall now present our first conclusions; any proofs omitted can be found in Ref. 4.

Proposition 1: For any $A \neq 0$ in $\mathcal{L}$, there exists a pure state m with $mA = 1$. Furthermore, the set $\{m \in \mathcal{M}_p; m(A) = 1\}$ determines the event A.

Corollary: $\mathcal{L}$ is a complete lattice.

Proof: Consider any family $\{A_i\}$ in $\mathcal{L}$. If the only event $\leq$ each A_i is the 0, then the infimum of $\{A_i\}$ is 0. So let $0 \neq A \leq A_i$ for all i; then there exists a pure state m with $mA = 1$. Hence we also have $m(A_i) = 1$ for all i and *Axiom* (vi) provides us with the desired infimum.

Notation: For a given $m \in \mathcal{M}_p$ the set $\{A: m(A) = 1\}$ has an infimum which we shall write as L_m; note that we have $m(L_m) = 1$ also.

By our interpretation of $\leq$ as implication, the event L_m is the ultimate cause of any event that occurs in the state m. Thus, for any state n, the number $n(L_m)$ is the probability of our system switching "spontaneously" from the state n to m. We therefore call $n(L_m)$ the *probability of transition* from n to m and write it as $\langle n \mid m \rangle$. The hypothesized absence of causal elements in such transitions leads us to:

Axiom (vii): For any pure states n, m we have $\langle n \mid m \rangle = \langle m \mid n \rangle$.

Now consider two states n, $m \in \mathcal{M}_p$ with $\langle n \mid m \rangle = 1$; there is then no way of distinguishing them, since our system is forced to switch back and forth between them. An alternate argument is that any event that occurs with certainty in one of them also occurs with certainty in the other. We shall thus accept:

Axiom (viii): The set of all events occurring with certainty in some pure state determines this state.

We shall use this in the form: $\langle n \mid m \rangle = 1$ implies $n = m$ (for $n, m \in \mathcal{M}_p$). Equivalent statements are: $L_n = L_m$ implies $n = m$, or $\langle m \mid m_1 \rangle = \langle m \mid m_2 \rangle$ for all $m \in \mathcal{M}_p$ implies $m_1 = m_2$.

Proposition 2: For any pure state m, the event L_m is an atom of $\mathcal{L}$; conversely, for each atom A of $\mathcal{L}$ there is a (unique) pure state m such that $A = L_m$. Furthermore, any event A is a disjoint union of atoms L_{m_i} and hence for any pure state we have $m(A) = \sum \langle m \mid m_i \rangle$.

We shall only establish this last point, namely that if $A = \sup\{A_i\}$, A_i any family of pairwise disjoint events, and if m is pure, then $m(A) = \sum m(A_i)$. Note that for any finite set $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ we have $\sum_k m(A_{i_k}) \leq 1$, hence that $\sum m(A_i) \leq 1$ too. Thus all but a countable number of $m(A_i)$ are nonzero; let those A_i's with nonzero probabilities be denoted by B_j, and the rest by C_k. Since $m(C_k') = 1$ for all k, we have $m(\inf C_k') = 1$, and hence $mC = 0$, where $C = \sup C_k$. Since $\{B_j\}$ is countable and suprema associate, we have $A = \sup\{\sum B_j, C\}$; but $B_j \leq C_k'$ for all k, hence $\sum B_j \perp C$, and thus

$$\begin{aligned} m(A) &= m(\textstyle\sum B_j) + mC \\ &= \textstyle\sum m(B_j) \\ &= \textstyle\sum m(A_i). \end{aligned}$$

Finally a word about the center of $\mathcal{L}$ which is the set of all $A \in \mathcal{L}$ such that A commutes with all events in $\mathcal{L}$. Here we say that A, B commute if there exist A_1, B_1, C pairwise disjoint so that $A = A_1 + C$, $B = B_1 + C$.

Proposition 3: A is the center of $\mathcal{L}$ iff for each pure state m we have $m(A) = 0$ or $m(A) = 1$.

Proof: If A is in the center and m is pure with L_m not contained in A, then $L_m \wedge A = 0$; since they commute, we shall also have $L_m \perp A$, and thus $m(L_m) + m(A) \leq 1$, and hence $m(A) = 0$. If $L_m \leq A$, evidently $m(A) = 1$; since L_m is is an atom, these are the only possible cases. Conversely, assume $m(A) = 0$ or 1 for every pure state m. Take any $B \in \mathcal{L}$; we want to show that A, B commute, i.e., $A \wedge (A \wedge B)' \perp B \wedge (A \wedge B)'$. This means $B \wedge (A \wedge B)' \leq A' \vee (A$ $B)$. We shall use Axiom (iv), which is easily seen to hold even for pure states [with the aid of Axiom (v)]. Thus let m be a pure state with $m(B \wedge (A \wedge B)') = 1$, so that $m(B) = 1$ and $m(A \wedge B) = 0$; since $m(A) = 1$ or 0, we see that $m(A)$ must be 0, since otherwise we obtain from Axiom (iv) that $m(A \wedge B) = 1$. But then $m(A') = 1$ and hence $m(A' \vee (A \wedge B)) = 1$ also.

3. SUPERPOSITION OF PURE STATES

A mixture of two states m, n is given by the formula $A \to \lambda m(A) + \mu n(A)$, where $\lambda + \mu = 1$ and $\lambda, \mu \geq 0$. Rewriting this in the form $[a^2 m(A) + b^2 n(A)]/(a^2 + b^2)$, where not both a, b are zero will help to motivate our formula for superpositions. The idea is to change this a little, by the the addition of a term representing the coupling of m, n, in such a way that it shall reduce to a mixture iff this coupling term vanishes.

We shall assume that two pure states m, n can be superposed in any given "proportion" a, b to produce a state $am \oplus bn$ given by:

$$A \to [a^2 m(A) + b^2 n(A) + 2ab(m \times n)(A)]/(a^2 + b^2 + 2ab[m \mid n]).$$

Here $[m \mid n]$ is just $(m \times n)(I)$, and is needed in the denominator to make $am \oplus bn$ have value 1 at the certain event. The term $(m \times n)(A)$ is (by necessity) a countably additive functional in A and represents, as mentioned above, the coupling of the two states m, n. We shall assume $m \times n = n \times m$, since we should have symmetry for the whole superposition: $am \oplus bn = bn \oplus am$. We do not, however, need to assume any explicit form for $m \times n$; the axiom to follow provides us with all necessary structure.

So we have assumed, for $m \neq n$, that all superpositions $am \oplus bn$ are defined except in case both a, b are zero. For $m = n$ we should not expect to obtain any new state, since there is nothing for m to "interact" with; thus we assume $am \oplus bm = m$ for all a, b except in case $a + b = 0$ (since then the two terms should "cancel out").

Now consider $m \neq n$ and the formula for $am \oplus bn$; since the denominator does not vanish for any a, b we must have $|[m \mid n]| < 1$, and so its sign is always +. Hence the numerator is also nonnegative for all a, b, and so $|(m \times n)(A)|^2 \leq (mA)(nA)$. In case $m = n$, we obtain at once $(m \times m)(A) = [m \mid m] \cdot (mA)$.

We can now state our basic hypotheses.

Axiom (ix): (a) For any pure states m, n, s there is a proportionality a, b such that the states $am \oplus bn$ and s are orthogonal, i.e., $(am \oplus bn)(L_s) = 0$.

(b) If $m(A)=n(A)=1$, then $(am\oplus bn)(A)=1$ also, for any a,b.

It is perhaps worth mentioning now that part (a) shall be used for the analysis of the state space, while part (b) shall give use the representation for $\mathcal{L}$.

We have at once from this axiom that, given m, n, s, there exist a,b for which

$$a^2\langle m|s\rangle + b^2\langle n|s\rangle + 2ab(m\times n)(L_s)=0.$$

Thus the discriminant, which is anyway ≤ 0, must vanish, i.e., $[(m\times n)(L_s)]^2=\langle m|s\rangle\langle n|s\rangle$.

Now consider the state $p=m\oplus bn$; we have

$$p(L_m)=(1+b^2\langle n|m\rangle + 2b(m\times n)(L_m))\,(1+b^2+2b[m|n])^{-1}.$$

Since p is a state, this is maximized at $b=0$, and by setting its derivative zero, we obtain $(m\times n)(L_m)=[m|n]$. Combining this with the previous equation, we get $([m|n])^2=\langle m|n\rangle$, or $[m|n]=\epsilon(m,n)\sqrt{\langle m|n\rangle}$, where $\epsilon(m,n)=\pm 1$. In particular $[m|m]=\pm 1$, but it cannot be -1, because then the state $m\oplus m$ is not defined; hence $[m|m]=1$.

Going back to $(m\times n)(L_s)$, we see that it has the form $\epsilon(m,n,s)\sqrt{\langle m|s\rangle}\sqrt{\langle n|s\rangle}$ with $\epsilon(m,n,s)=\pm 1$; our final assumption is that the coupling of m,n in $m\times n$ is not too strong in the sense that $\epsilon(m,n,s)$ will be allowed to factor into $\epsilon(m,s)\epsilon(n,s)$. Thus we obtain

$$(m\times n)(L_s)=[m|s][n|s].$$

A useful corollary of this is that if $\{s_i\}$ is a maximal orthogonal family of states in $\mathcal{M}_p$ (i.e., $\langle s_i|s_j\rangle=0$ for $i\neq j$), then

$$[m|n]=\sum[m|s_i][n|s_i].$$

We shall call this the "Parseval relation."

Finally note that for $p=am\oplus bn$ we have

$$p(L_s)=(a[m|s]+b[n|s])^2(a^2+b^2+2ab[m|n])^{-1}.$$

4. PURITY OF SUPERPOSITIONS

The following regularity property is of fundamental importance.

Proposition 4: Let $p=am\oplus bn$ be pure. Then there exists an $\eta=+1$ or -1 such that for all $s\in\mathcal{M}_p$ we have

$$[p|s]=\eta(a[m|s]+b[n|s])(a^2+b^2+2ab[m|n])^{-1/2}.$$

Proof: First note that equality of absolute values has just been shown at the end of Sec. 3. Then observe that by Parseval's relation it suffices to obtain an η for which the desired result holds for all s in a maximal orthogonal set in $\mathcal{M}_p$. We thus have for each $s\in\mathcal{M}_p$, a number $\epsilon(s)=\pm 1$ such that

$$[p|s]=\epsilon(s)(a[m|s]+b[n|s])\cdot(a^2+b^2+2ab[m|n])^{-1/2}.$$

Now choose some maximal orthogonal set $\{s_i\}_{i\in I}$ in $\mathcal{M}_p$ and let $I_+=\{i|\epsilon(s_i)=1\}$, $I_-=\{i|\epsilon(s_i)=-1\}$. We take a $q\in\mathcal{M}_p$ and calculate as follows:

$$\begin{aligned}\epsilon(q)&(a[m|q]+b[n|q])(a^2+b^2+2ab[m|n])^{-1/2}\\
&=[p|q]=\sum[p|s_i][q|s_i]\\
&=(a^2+b^2+2ab[m|n])^{-1/2}\Big\{a\sum_{i\in I_+}[m|s_i][q|s_i]\\
&\quad+b\sum_{i\in I_+}[n|s_i][q|s_i]\\
&\quad-a\sum_{i\in I_-}[m|s_i][q|s_i]\\
&\quad-b\sum_{i\in I_-}[n|s_i][q|s_i]\Big\};\end{aligned}$$

since

$$\begin{aligned}[m|q]&=\sum_{i\in I}[m|s_i][q|s_i]\\
&=\sum_{i\in I_+}[m|s_i][q|s_i]+\sum_{i\in I_-}[m|s_i][q|s_i]\end{aligned}$$

and similarly for $[n|q]$, we obtain

$$\sum_{i\in I_-}(a[m|s_i]+b[n|s_i])[q|s_i]=0 \quad \text{if } \epsilon(q)=1,$$

$$\sum_{i\in I_+}(a[m|s_i]+b[n|s_i])[q|s_i]=0 \quad \text{if } \epsilon(q)=-1.$$

Using these for $q=m$, we obtain

$$a\sum_{i\in I_-}\langle m|s_i\rangle=-b\sum_{i\in I_-}[m|s_i][n|s_i] \quad \text{if } \epsilon(m)=1,$$

$$a\sum_{i\in I_+}\langle m|s_i\rangle=-b\sum_{i\in I_+}[m|s_i][n|s_i] \quad \text{if } \epsilon(m)=-1$$

with analogous conclusions for $\epsilon(n)=\pm 1$.

Thus, in case $\epsilon(m)=\epsilon(n)=1$, the formula for $p=am\oplus bn$ gives

$$\begin{aligned}\sum_{i\in I_-}\langle p|s_i\rangle&=(a^2+b^2+2ab[m|n])^{-1}\\
&\quad\times\Big\{-a^2\sum_{i\in I_-}\langle m|s_i\rangle+b^2\sum_{i\in I_-}\langle n|s_i\rangle\Big\},\end{aligned}$$

and also

$$\begin{aligned}&=(a^2+b^2+2ab[m|n])^{-1}\\
&\quad\times\Big\{a^2\sum_{i\in I_-}\langle m|s_i\rangle-b^2\sum_{i\in I_-}\langle n|s_i\rangle\Big\}.\end{aligned}$$

Hence all $\langle p|s_i\rangle$ with $i\in I_-$ are zero; this means that for $[p|s_i]\neq 0$ we have $i\in I_+$ and η can be chosen $+1$. The case $\epsilon(m)=\epsilon(n)=-1$ is similar, and we turn to $\epsilon(m)=1$, $\epsilon(n)=-1$. We obtain as above that

$$\begin{aligned}\sum_{i\in I_-}\langle p|s_i\rangle&=(a^2+b^2+2ab[m|n])^{-1}\\
&\quad\times\Big\{-a^2\sum_{i\in I_-}\langle m|s_i\rangle+b^2\sum_{i\in I_-}\langle n|s_i\rangle\Big\}\end{aligned}$$

and

$$\begin{aligned}\sum_{i\in I_+}\langle p|s_i\rangle&=(a^2+b^2+2ab[m|n])^{-1}\\
&\quad\times\Big\{a^2\sum_{i\in I_+}\langle m|s_i\rangle-b^2\sum_{i\in I_+}\langle n|s_i\rangle\Big\},\end{aligned}$$

whence

$$\sum_{i\in I_+}\langle p|s_i\rangle-\sum_{i\in I_-}\langle p|s_i\rangle=(a^2+b^2+2ab[m|n])^{-1}(a^2-b^2).$$

On the other hand,

$$\begin{aligned}[p|m]&=\epsilon(m)(a+b[m|n])(a^2+b^2+2ab[m|n])^{-1/2}\\
&=(a+b[m|n])\cdot(a^2+b^2+2ab[m|n])^{-1/2}\end{aligned}$$

and similarly

$$[p|n] = -(a[m|n]+b)(a^2+b^2+2ab[m|n])^{-1/2},$$

which imply

$$1 = [p|p] = \epsilon(p)(a[m|p]+b[n|p])(a^2+b^2+2ab[m|n])^{-1/2}$$
$$= \epsilon(p)(a^2-b^2)(a^2+b^2+2ab[m|n])^{-1/2}.$$

But this means

$$\left|\sum_{i\in I_+}\langle p|s_i\rangle - \sum_{i\in I_-}\langle p|s_i\rangle\right| = 1,$$

and as the sum of these two is also 1 and each is ≥ 0, one must be zero. Thus again all $\epsilon(s_i)$ for which $[p|s_i] \neq 0$ have the same sign and η can be chosen appropriately.

We shall now supply this to the analysis of whether or not $am \oplus bn$ is pure.

Theorem 1: Let m, n be pure and $p = am \oplus bn$. Then p is a mixture iff $\langle m|s\rangle\langle n|s\rangle = 0$ for all $s \in \mathcal{M}_p$. In such a case we have in particular $\langle m|n\rangle = 0$, and also that any superposition of m, n is mixed.

Proof: Let us first consider the case where $\langle m|n\rangle = 0$; we shall reduce the general case to this. Now suppose $p = \int_X p_x\, d\mu(x)$ is a decomposition of p into pure states p_x, and write $A = L_m + L_n$ (since $\langle m|n\rangle = 0$ we have $L_m \perp L_n$). Then $mA = nA = 1$, hence $pA = 1$ also; but this means that $p_x(A) = 1$ for almost all x, and we shall ignore this exceptional null set. Choosing a maximal orthogonal set of atoms L_{s_i} in A', we see that $\{m, n, s_i\}$ is a maximal orthogonal set of pure states, and thus we have

$$[p_x|s] = \sum_i [p_x|s_i][s|s_i] + [p_x|m][s|m] + [p_x|n][s|n];$$

but $p_x(L_{s_i}) = 0$ since $p_x(A) = 1$ and therefore

$$[p_x|s] = [p_x|m][s|m] + [p_x|n][s|n],$$

or $p_x = f(x)m \oplus g(x)n$, where we set $f(x) = [p_x|m]$, $g(x) = [p_x|n]$. Expanding $\langle p|s\rangle$ in two ways we obtain that

$$(a^2\langle m|s\rangle + b^2\langle n|s\rangle + 2ab[m|s][n|s])(a^2+b^2)^{-1}$$
$$= \int_X \{[f(x)]^2\langle m|s\rangle + [g(x)]^2\langle n|s\rangle + 2f(x)g(x)[m|s][n|s]\}\, d\mu(x)$$

for all $s \in \mathcal{M}_p$. Replacing s by m and n, we obtain

$$a^2/(a^2+b^2) = \int_X f(x)^2\, d\mu(x), \quad b^2/(a^2+b^2) = \int_X g(x)^2\, d\mu(x)$$

and

$$\{ab - \int_X f(x)g(x)\, d\mu(x)\}[m|s][n|s] = 0,$$

for all $s \in \mathcal{M}_p$. Thus, if $[m|s][n|s] \neq 0$ for a single $s \in \mathcal{M}_p$, we get

$$\left(\int_X f(x)^2\, d\mu(x)\right)\left(\int_X g(x)^2\, d\mu(x)\right) = \left(\int_X f(x)g(x)\, d\mu(x)\right)^2,$$

which implies $f(x) = kg(x)$ a.e. for some constant k. But since $p_x = f(x)m \oplus g(x)n$, this means that almost all p_x are identical to $km \oplus n$ and hence p is not a mixture! Hence, in case $\langle m|n\rangle = 0$, a superposition will be pure, unless $\langle m|s\rangle\langle n|s\rangle = 0$ for all $s \in \mathcal{M}_p$. Note that, then, any superposition shall be mixed, by the very formula for superposition. Now consider $\langle m|n\rangle \neq 0$. By Axiom (ii) (b) there exists a state q with $L_m \vee L_n = L_m + L_q$, and by the above argument we have $n = [n|m]m \oplus [n|q]q$. A direct calculation gives $p = (a + b[m|n])m \oplus (b[n|q])q$, but since $\langle m|q\rangle = 0$ and p is assumed mixed, any other superposition of m, q is also mixed; in particular n is mixed—which is absurd. Thus p is pure whenever $\langle m|n\rangle \neq 0$.

5. DECOMPOSITION OF THE PURE STATE SPACE

We shall write $m \sim n$ if some, and hence all superpositions of m, n are pure.

Proposition 5: The relation $\sim$ is an equivalence; the equivalence classes are mutually orthogonal and closed under superpositions.

Proof: The conditions $m \sim m$ and $m \sim n$ implies $n \sim m$ are trivial; so let $m \sim n$, $n \sim q$ and assume $m \oplus q$ is a mixture. As a first step we shall show that for a suitable choice of a, b we have $m \oplus q = \{a(m \oplus n)\} \oplus \{b(-n \oplus q)\}$; note that by hypothesis $m \oplus n$ and $-n \oplus q$ are pure, and the right-hand side does make sense. Now for any a, b we have that

$$\langle\{a(m \oplus n) \oplus b(-n \oplus q)\}|s\rangle = (a^2+b^2+2ab[m \oplus n|-n \oplus q])^{-1}$$
$$\cdot (a[m \oplus n|s] + b[-n \oplus q|s])^2.$$

By Proposition 4, we have $\epsilon, \eta (= \pm 1)$ such that

$$[m \oplus n|s] = \epsilon(2+2[m|n])^{-1/2}([m|s]+[n|s])$$

and

$$[-n \oplus q|s] = \eta(2-2[n|q])^{-1/2}(-n[n|s]+[q|s]).$$

So, if we choose $a = \epsilon(2+2[m|n])^{1/2}$, $b = \eta(2-2[n|q])^{1/2}$ we have at once that

$$a[m \oplus q|s] + b[-n \oplus q|s] = [m|s] + [q|s].$$

Also note that

$$[m \oplus q|-n \oplus q] = \epsilon(2+2[m|n])^{-1/2}([m|-n \oplus q]+[n|-n \oplus q])$$
$$= \epsilon(2+2[m|n])^{-1/2}\eta(2-2[n|q])^{-1/2}$$
$$\times(-[m|n]+[m|q]-1+[n|q]),$$

and so $a^2+b^2+2ab[m \oplus q|-n \oplus q]$ becomes $2+2[m|q]$. Therefore, our choice of a, b will produce

$$\langle\{a(m \oplus n) \oplus b(-n \oplus q)\}|s\rangle$$
$$= (2+2[m|q])^{-1}([m|s]+[q|s])^2 \text{ for all } s.$$

But this means $a(m \oplus n) \oplus b(-n \oplus q) = m \oplus q$. Now since $m \oplus q$ is not pure, we have by Theorem 1 that $[m \oplus n|s][-n \oplus q|s] = 0$ for all s, or by Proposition 4 that

$$-[m|s][n|s] - \langle n|s\rangle + [m|s][q|s] + [n|s][q|s] = 0$$

for all s. Since $m \oplus q$ is not pure, we also have the third term zero; hence $[n|s]([m|s]+[n|s]-[q|s]) = 0$. Take $s = q$ to obtain $[n|q](1-[n|q]) = 0$, i.e., $[n|q] = 0$ or else $n = q$; but $n \neq q$ because $m \oplus n$ is pure and $m \oplus q$ is not. Hence $[n|q] = 0$ and we take $s = n$ to find $[m|n] + 1 = 0$, or $m = n$; again this makes $m \oplus q = n \oplus q$ with the first mixed and the second pure.

Thus the pure state space decomposes into equivalence classes which are mutually orthogonal because of Theorem 1. Now, if $m \sim n$ and $am \oplus bn$ falls in some other class, we have $[am \oplus bn|m] = 0$, $[am \oplus bn|n] = 0$, or $a + b[m|n] = 0$ and $a[m|n] + b = 0$, which makes $\langle m|n\rangle + 1 = 0$ a contradiction.

This decomposition of $\mathcal{M}_p$ is closely related to the center of $\mathcal{L}$. Indeed we have:

Theorem 2: Let λ be one of the equivalence classes in $\mathcal{M}_p$ and $A_\lambda = \sup\{L_s : s \in \lambda\}$. Then A_λ is an atom of the center C of $\mathcal{L}$, C is atomic and each of its atoms has the above form.

Proof: First note that A_λ is defined since $\mathcal{L}$ is complete. Then observe that $m(A_\lambda) = 1$ or 0 according to $m \in \lambda$ or $m \notin \lambda$, because $m \in \lambda$ implies $L_m \leq A_\lambda$ while $m \notin \lambda$ implies $L_m \perp A_\lambda$ (by Theorem 1). Thus by Proposition 3 we have $A_\lambda \in C$. To show that A_λ is actually an atom of C, let $0 \neq B \leq A_\lambda$, $A_\lambda \neq B \in C$; then there exist $L_m \leq B$ and $L_n \leq A_\lambda \wedge B'$. Now, since $B \in C$ we have, for all $s \in \mathcal{M}_p$, that $s(B) = 1$ or 0; thus either $s(L_m)$ or $s(L_n)$ is zero, i.e., $\langle m|s\rangle\langle n|s\rangle = 0$ for all $s \in \mathcal{M}_p$. But this implies that all superpositions of m, n are mixed, which cannot be, since m, $n \in \lambda$. This contradiction implies that A_λ is an atom of C. Finally, note that the supremum of all the A_λ is the unity I of $\mathcal{L}$, since the union of all λ is just $\mathcal{M}_p$. Since C is a Boolean σ-algebra, it must be atomic; and since there is no room for other atoms, they all have the above form.

6. CONSTRUCTION OF THE COMPONENT SPACES

Consider again one of the classes constructed in Sec. 5, say λ, and recall that it consists of pure states; it is closed under superpositions and is maximal as such. We shall associate to this λ a Hilbert space H constructed as follows.

The set H shall consist of all pairs (a, m) with $a \in R$ and $m \in \lambda$, considered distinct, except for the pairs $(0, m)$ which are all identified to each other; we shall write θ for this particular element of H, and abbreviate (a, m) to $a \circ m$. The elements of H shall be called vectors and written as $\phi, \psi, \omega, \cdots$.

For $\psi = a \circ m$ and $c \in R$ we define $c\psi$ to be $(ac) \circ m$; for $\psi = a \circ m$, $\phi = b \circ n$ we define $\psi + \phi$ to be the vector $c \circ p$, where $p = am \oplus bn$ and $c = \eta(a^2 + b^2 + 2ab[m|n])^{1/2}$ ($\eta = \pm 1$ is the sign obtained from Proposition 4, i.e., such that for all $s \in \mathcal{M}_p$ we have

$$[p|s] = \eta(a^2 + b^2 + 2ab[m|n])^{-1/2}(a[m|s] + b[n|s]).$$

There is still left a case to consider: $m = n$ and $a + b = 0$ do not produce a superposition; but $c = 0$ in this case and we set $\psi + \phi = \theta$. Finally we define $(\psi|\phi)$ to be the number $ab[m|n]$.

Theorem 3: The set H is a Hilbert space under the operations defined above.

Proof: The argument is long but straightforward. First we consider addition. It is easy to see that $\psi + \theta = \psi$: Since $\theta = (0, n)$, we have (writting $\psi + \theta = c \circ p$) that $[p|s] = \eta(a/|a|)[m|s]$ (with $\psi = a \circ m$) so that $\langle p|s\rangle = \langle m|s\rangle$ and hence $p = m$; on the other hand, we get $\eta a/|a| = 1$ and since $c = \eta|a|$ we end up with $c = a$, i.e., $\psi + \theta = \psi$. By our definition we also see that for $\psi = a \circ m$ and $\phi = (-a) \cdot m$ we obtain $\psi + \phi = \theta$. The commutative law is trivial, so we now consider the associative: $(\psi + \phi) + \omega = \psi + (\phi + \omega)$, where $\psi = a \circ m$, $\phi = b \circ n$, and $\omega = c \circ q$; we may assume a, b, $c \neq 0$, as in this case our first remark renders the desired result trivial. Now

$$\psi + \phi = \eta(a^2 + b^2 + 2ab[m|n])^{1/2} \circ (am \oplus bn),$$

where η is chosen to make

$$[am \oplus bn|s] = \eta(a^2 + b^2 + 2ab[m|n])^{-1/2}(a[m|s] + b[n|s])$$

valid for all $s \in \mathcal{M}_p$—provided $\psi + \phi \neq \theta$. But $\psi + \phi = \theta$ implies $m = n$ and $a = -b$, and we shall verify associativity first for this case. We have to calculate $\phi + \omega$; if this is also θ, we have at once $\psi = \omega$ and it holds. So let $\phi + \omega \neq \theta$. Since $\phi = (-a) \circ m$, we have

$$\phi + \omega = \epsilon(a^2 + c^2 - 2ac[m|q])^{1/2} \circ ((-a)m \oplus cq),$$

where ϵ is such that

$$\epsilon(a^2 + c^2 - 2ac[m|q])^{1/2}[(-a)m \oplus cq|s] = -a[m|s] + c[q|s]$$

for all s. We must show $\omega = \psi + (\phi + \omega)$, so we calculate this sum; it has the form $d \circ p$, where

$$d = \delta(a^2 + (a^2 + c^2 - 2ac[m|q]) + 2a\epsilon(a^2 + c^2 - 2ac[m|q])^{1/2} \times [(-a)m \oplus q|m])^{1/2}$$

and

$$d[p|s] = a[m|s] + \epsilon(a^2 + c^2 - 2ac[m|q])^{1/2}[(-a)m \oplus cq|s].$$

Using the above to calculate $[(-a)m \oplus cq|m]$, we find that $d = \delta|c|$ and hence $[p|s] = (\delta c/|c|)[q|s]$; so we get again $p = q$, $c/|c| = \delta$ and finally $d = c$, i.e., $\omega = d \circ p$. Now for the general case where $\psi + \phi \neq \theta \neq \phi + \omega$, we have the previous formula for $\psi + \phi$ and also

$$\phi + \omega = \epsilon(b^2 + c^2 + 2bc[n|q])^{1/2}(bn \oplus cq),$$

with

$$[bn \oplus cq|s] = \epsilon(b^2 + c^2 + 2bc[n|q])^{-1/2} \circ (b[n|s] + c[q|s]).$$

Again letting $d \circ p$ equal $\psi + (\phi + \omega)$, we have

$$d = \delta(a^2 + b^2 + c^2 + 2bc[n|q] + 2a\epsilon \times (b^2 + c^2 + 2bc[n|q])^{1/2}[m|bn \oplus cq])^{1/2}$$

and

$$d[p|s] = \delta(a[m|s] + \epsilon(b^2 + c^2 + 2bc[n|q])^{1/2}[bn \oplus cq|s]).$$

Again, as before, we obtain

$$d = \delta(a^2 + b^2 + c^2 + 2bc[n|q] + 2ab[n|m] + 2ac[q|m])^{1/2}$$

and

$$[p|s] = (1/d)(a[m|s] + b[n|s] + c[q|s]).$$

This shows that $\langle p|s\rangle = [p|s]^2$ is symmetric in ψ, ϕ, ω, hence so is p; therefore, d is symmetric and finally δ. But this means $\psi + (\phi + \omega)$ is symmetric in ψ, ϕ, ω, which does it.

We shall now verify that $c(\psi + \phi) = c\psi + c\phi$ and omit the remaining properties of scalar multiplication; they follow analogous patterns. Using the previous notation, we have

$$c\psi + c\phi = (ac) \cdot m + (bc) \circ n$$
$$= \delta(a^2c^2 + b^2c^2 + 2abc^2[m|n])^{1/2} \circ ((ac)m \oplus (bc)n)$$
$$= \delta|c|(a^2 + b^2 + 2ab[m|n])^{1/2}(am \oplus bn),$$

where we have δ chosen so that

$$[(ac)m \oplus (bc)n|s] = \delta(|c|(a^2 + b^2 + 2ab[m|n])^{1/2})^{-1/2}(ac[m|s] + bc[n|s]).$$

But since $(ac)m \oplus (bc)n = am \oplus bn$, we obtain $\delta c/|c| = \eta$, so that finally

$$c\psi + c\phi = c\eta(a^2 + b^2 + 2ab[m\,|\,n])^{1/2}(am \oplus bn) = c(\psi + \phi).$$

For the inner product we see that

$$\begin{aligned}(\psi + \phi\,|\,\omega) &= \eta(a^2 + b^2 + 2ab[m\,|\,n])^{1/2} c[am \oplus bn\,|\,q] \\ &= c(a[m\,|\,q] + b[n\,|\,q]) = ac[m\,|\,q] + bc[n\,|\,q] \\ &= (\psi\,|\,\omega) + (\phi\,|\,\omega),\end{aligned}$$

and similarly for the remaining properties. Also $(\psi\,|\,\psi) = a^2$, so $\|\psi\| = 0$ implies $\psi = \theta$.

Finally we verify that H is complete. It suffices (see, e.g., Ref. 5) to show that any maximal orthonormal set $\{\psi_i\}$ is a basis. Now $\psi_i = a_i \circ m_i$, and since $\|\psi_i\| = 1$ (i.e., $|a_i| = 1$), we can assume $a_i = 1$ without loss. Also, it suffices to expand a vector ψ of the form $1 \cdot m$. First note that $\{m_i\}$ is a maximal orthogonal set of pure states in the given class λ; next, that $\langle n\,|\,m_i\rangle = 0$ for all i implies $\langle n\,|\,m\rangle = 0$ also, because if this is not the case we have $n \sim m$ by Theorem 1. Hence $n \in \lambda$ and thus $\{m_i\}$ is not maximal orthogonal. Thus, if we augment $\{m_i\}$ to a set $\{m_i, n_j\}$, maximal orthogonal in M_p, we have $\langle m\,|\,n_j\rangle = 0$ for all j. Therefore, $\sum \langle m\,|\,m_i\rangle = 1$. Now we can calculate

$$\|\psi - \sum_{i=1}^{n} (\psi\,|\,\psi_i)\psi_i\|^2 = 1 - \sum_{i=1}^{n} \langle m\,|\,m_i\rangle$$

for any finite subset in $\{m_i\}$, and since this last term can be made as small as we like, we have $\psi = \sum_i (\psi\,|\,\psi_i)\psi_i$.

7. THE REPRESENTATION THEOREM

Let Λ be the set of all classes of M_p obtained in Sec. 5, and for each $\lambda \in \Lambda$ let H_λ be the Hilbert space constructed in Sec. 6. By Theorem 2 we have also associated with λ, an atom C_λ in the center of our logic. Now, given $A \in L$ we construct the family $\{A_\lambda\}_{\lambda \in \Lambda}$, where $A_\lambda = A \wedge C_\lambda$, and note that $A = \sum_\lambda A_\lambda$ (since C_λ are in the center); conversely, any family $\{A_\lambda\}$ with $A_\lambda \leq C_\lambda$ produces by summation an element of L. Since the C_λ are the atoms of the center, we obtain at once that L is the orthogonal sum of the logics $L_\lambda = \{A \in L: A \leq C_\lambda\}$. We shall therefore restrict attention to L_λ for some fixed λ, and note that λ is precisely the set of pure states of L_λ. From now on we shall write L and H instead of L_λ and H_λ.

Consider any $A \in L$ and write H_A for $\{a \circ m: m(A) = 1\}$; by our fundamental hypothesis on superpositions [Axiom (xi) (b)] we see that H_A is a linear manifold in H.

Proposition 6: For any $A \in L$, $H_{A'}$ is the orthocomplement of H_A in H.

Proof: Let $\psi \in H_A$ and $\phi \in H_A$, with $\psi = a \circ m$, $\phi = b \circ n$; then $m(A) = 1$, hence $m(A') = 0$ while $L_n \leq A'$, i.e., $m(L_n) = 0$ or $\langle m\,|\,n\rangle = 0$. Thus $(\psi\,|\,\phi) = 0$ and $H_{A'} \subseteq (H_A)^\perp$. fix $\phi \in (H_A)^\perp$; for any $\psi \in H_A$ we have $\langle m\,|\,n\rangle = 0$, or $n(L_m) = 0$, hence $n(A) = 0$ (since A is the disjoint union of atoms L_m). Thus $n(A') = 1$ and $\phi \in H_A$, or $H_{A'} = (H_A)^\perp$.

In particular $H_A = (H_{A'})^\perp$ and is therefore closed. We shall write P_A for the orthogonal projection on H with range H_A. Note that for $\|\psi\| = 1$, i.e., $\psi = (\pm 1) \circ m$ we have $\|P_A\psi\|^2 = m(A)$, because if $\{s_i\}$ is a maximal orthogonal set of states with $s_i(A) = 1$, we have $m(A) = \sum \langle m\,|\,s_i\rangle = \sum [m\,|\,s_i]^2$; so writing $\psi_i = 1 \cdot s_i$, we obtain on the one hand $\{\psi_i\}$ in H_A, and on the other the relation $\sum (\psi\,|\,\psi_i)^2 = \|P_A\psi\|^2$. Since $(\psi\,|\,\psi_i) = \pm [m\,|\,s_i]$ we are done.

Theorem 4: The map $A \to P_A$ is an isomorphism of the logic L onto the lattice of all projections in H.

Proof: First we verify that $A \leq B$ iff $P_A \leq P_B$. The first holds iff $m(A) = 1$ imples $m(B) = 1$ for all pure states, i.e., iff the range of P_A is contained in the range of P_B. In particular, the map $A \to P_A$ is one to one.

We have already seen that $P_{A'} = I - P_A$. Next we show that if $A = \sum A_i$, then $P_A = \sum P_{A_i}$ (in the weak topology of H): For any $\psi \in H$, $\|\psi\| = 1$ [i.e., $\psi = (\pm 1) \circ m$] we have

$$\begin{aligned}\|P_A\psi\|^2 = m(A) &= \sum m(A_i) = \sum \|P_{A_i}\psi\|^2 \\ &= \sum (P_{A_i}\psi\,|\,\psi) = ((\sum P_{A_i})\psi\,|\,\psi) = \|(\sum P_{A_i})\psi\|^2.\end{aligned}$$

We shall use this to show that the map $A \to P_A$ is onto. Consider any subspace K of H and a basis $\{\psi_i\}$ in K; we can assume $\psi_i = 1 \circ m_i$. Write P_i for the projection on the space spanned by ψ_i, so that the projection on K is $\sum P_i$. But the space spanned by ψ_i is H_{A_i}, where $A_i = L_{m_i}$; hence the projection on K is simply P_A with $A = \sum A_i$.

8. A GEOMETRICAL CONDITION

We shall now translate the hypotheses in Sec. 3 concerning superpositions of pure states into properties of the functional $\langle\ |\ \rangle$ on the state space.

To avoid mentioning events, we introduce the following terminology: Two families $\{s_i\}$, $\{t_j\}$ of orthogonal states are "equivalent" if for each m we have $\sum_i \langle m\,|\,s_i\rangle = \sum_j \langle m\,|\,t_j\rangle$. This means that $\sum L_{s_i} = \sum L_{t_j}$, but knowledge of $\langle\ |\ \rangle$ is suffcient to determine equivalence.

Theorem 4: Suppose that to each pair m, n of pure states there corresponds a number $[m\,|\,n]$ in such a way that:

(i) $[m\,|\,n] = [n\,|\,m]$, $[m\,|\,n]^2 = \langle m\,|\,n\rangle$, $[m\,|\,m] = 1$;

(ii) for any pair of equivalent families $\{s_i\}$, $\{t_j\}$ we have $\sum_i [m\,|\,s_i][n\ s_i] = \sum_j [m\,|\,t_j][n\,|\,t_j]$ (for all m, n).

Then the map $m \times n$: $A \to \sum_i [m\,|\,s_i][n\,|\,s_i]$ [where $\{s_i\}$ is a maximal orthogonal set of states with $s_i(A) = 1$] is single valued and countably additive on L. Finally, unless $m = n$ and $a + b = 0$, the map

$$am \oplus bn = (a^2 + b^2 + 2ab[m\,|\,n])^{-1}(a^2 m + b^2 n + 2ab(m \times n))$$

is a state of L for which all hypotheses in Sec. 3 hold.

Proof: Hypothesis (ii) is designed to guarantee the single valued of $m \times n$. However, let us first verify that the series actually converge: We have

$$(\sum |[m\,|\,s_i][n\,|\,s_i]|)^2 \leq (\sum [m\,|\,s_i]^2)(\sum [n\,|\,s_i]^2)$$

by Schwartz; hence we obtain the bound $(\sum \langle m\,|\,s_i\rangle)(\sum \langle n\,|\,s_i\rangle) = m(A)n(A)$. Thus, we also have $|(m \times n)(A)|^2 \leq m(A)n(A)$. To verify countable additivity of $m \times n$, let $A = \sum \{A_j: j \in J\}$ and choose $\{s_{jk}\}_{k \in K_j}$ maximal orthogonal with $s_{jk}(A_j) = 1$; then $\{s_{jk}: k \in K_j, j \in J\}$ is maximal or orthogonal in A, and

$$\begin{aligned}(m \times n)(A) &= \sum_{j,k} [m\,|\,s_{jk}][n\,|\,s_{jk}] \\ &= \sum_{j \in J} \sum_{k \in K_j} [m\,|\,s_{jk}][n\,|\,s_{jk}]\end{aligned}$$

by absolute convergence. But the inner sum is just $(m$

$\times n)(A_j)$. Choosing a maximal orthogonal set in $\mathcal{H}_p$ containing m, we see that $(m\times n)(I)=[m\,|\,m][n\,|\,m]+0=[m\,|\,n]$, and so the value of $am\oplus bn$ at I is 1. Since $|(m\times n)(A)|^2\le m(A)n(A)$, the numerator is ≥ 0, and since $|[m\,|\,n]|\le 1$, the denominator is also, and $am\oplus bn$ finally becomes a state.

Given any s, we have by our definition $(m\times n)(L_s)=[m\,|\,s][n\,|\,s]$, and this gives Axiom (ix)(a) at once. To verify the second part, first note that "Parseval" holds: For any maximal orthogonal set $\{s_i\}$ of states, we have

$$[m\,|\,n]=(m\times n)(I)=\sum[m\,|\,s_i][n\,|\,s_i].$$

Next, observe that $m(A)=1$ implies $\langle m\,|\,t\rangle=0$ for any t for which $L_t\perp A$; thus, choosing $\{s_i\}$ maximal orthogonal in A [i.e., $s_i(A)=1$] and $\{t_j\}$ maximal orthogonal in A' $[t_j(A')=1]$ we argue as follows. Given $m(A)=1$, $n(A)=1$, we have $[m\,|\,t_j]=[n\,|\,t_j]=0$; hence

$$[m\,|\,n]=\sum[m\,|\,s_i][n\,|\,s_i]=(m\times n)(A),$$

which implies

$$(am\oplus bn)(A)=(a^2+b^2+2ab[m\,|\,n])^{-1}\times(a^2m(A)+b^2n(A)+2ab(m\times n)(A))=1.$$

[1]V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, 1968).
[2]S. P. Gudder, J. Math. Phys. 11, 1037 (1970).
[3]C. Piron, Helv. Phys. Acta 37, 439 (1964).
[4]P. C. Deliyannis, J. Math. Phys. 14, 249 (1973).
[5]S. P. Gudder, Am. Math. Monthly 81, 29 (1974).

Renormalization group for a system of continuous spins on a lattice*

K. Subbarao†

Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14850
(Received 2 April 1975)

The renormalization group for a field theory on a lattice, or equivalently, in the language of statistical mechanics, a system of continuous spins on a lattice, described by Wilson, is considered in detail. The conditions under which a class of transformations are renormalization group transformations are studied.

I. INTRODUCTION

The modern formulation of the renormalization group[1] by Wilson offers promise of enabling one to compute, perhaps using numerical approximations, the short distance behavior of a field theory. Various field theories in $4-\epsilon$ dimensions[2] have been solved in an expansion in ϵ; scalar field theory in three dimensions has been solved in a qualitative way.[3] On the other hand, lattice problems like the Ising model have been attacked numerically by renormalization group methods[4–6]; the approximation involved is a restriction to a small number of lattice sites. A lattice formulation of field theory therefore deserves study. Such a formulation has been provided by Wilson.[7] After a brief discussion in Sec. II of renormalization group transformations in continuous space-time, the lattice formulation of Wilson is explained in Sec. III for a boson system, and whether or not a class of transformations is allowed is discussed in detail.

The action principle for a field theory with Euclidean metric can be expressed in terms of a path integral formulated on a lattice.[8] In the language of statistical mechanics, we will be dealing with a system of continuous "spins" on a lattice. The "spin" variable which varies from $-\infty$ to $+\infty$ is just the field variable at a lattice site. The Hamiltonian appearing in the exponent of the Boltzmann factor is, in field theoretic language, the action (integral of the Lagrangian). We will use the terms spin and hamiltonian. (Our remarks here are oversimplified. For a further discussion of the connection between field theory and statistical mechanics, see Wilson and Kogut, Ref. 1, Chap. X.)

II. RENORMALIZATION GROUP IN THE CONTINUUM

For orientation purposes, consider a renormalization transformation in continuous space.[9] It takes a Hamiltonian $H(S_q)$, a function of the spin-variable S_q in momentum space, to a Hamiltonian $H'(S'_{q'})$; the transformation integrates out the degrees of freedom corresponding to high momenta (say, momenta $\geq \exp(-t)$ where t is a continuous variable) while preserving the low momentum behavior. A detailed discussion is found in Ref. 9. An explicit example is given by[10]

$$\exp[H'(S'_{q'})]=\int_s \exp\left(\frac{-1}{2}\int_q \frac{\exp(dt)\{S'_{q'}-S_q\exp[-\beta(q^2,t)]\}\{S'_{-q'}-S_{-q}\exp[-\beta(q^2,t)]\}}{1-\exp[-2\alpha(q^2,t)]}\right)\cdot\exp(H(S_q)), \tag{2.1}$$

where $\int_s$ is functional integration, $q'=qe^t$, and $\alpha(q^2,t)=\beta(q^2,t)+dt/2$. Roughly speaking, $S'_{q'}$ behaves like $S_q\exp[-\beta(q^2,t)]$; thus, we would like $\exp[-\beta(q^2,t)]$ to be small for $q\geq\exp(-t)$, and of the order of unity for $q\leq\exp(-t)$. A particular choice is[11]

$$\beta(q^2,t)=q^2(\exp(2t)-1)-pt \tag{2.2}$$

Here p is a constant which is constrained as noted below. (The pt term is essential to impose a suitable normalization condition on $S'_{q'}$.) In Eq. (2.1) only a linear combination of S' and S appears. For this reason we will call (2.1) a linear renormalization transformation.[12] This linearity leads to a linear relation between the initial and transformed correlation functions. A critical system has infinite correlation length and the correlation function for small momenta falls off like a power[13] say $q^{-2+\eta}$. Imposing this condition on the correlation function before and after the transformation gives[1]

$$p=\frac{d-2+\eta}{2}. \tag{2.3}$$

Equation (2.1) can be written in configuration space as

$$\exp[H'(S')]=\int_s \exp\left[-\tfrac{1}{2}\int_{x,y}F(x-y)[s'(x')-\langle s(x)\rangle]\times[s'(y')-\langle s(y)\rangle]\right]\cdot\exp[H(S)], \tag{2.4}$$

where

$$x'=x\exp(-t),\quad y'=y\exp(-t),$$
$$s(x)=(2\pi)^{-d}\int S_q\exp(-iqx),$$
$$\langle s(x)\rangle=\int_z W_t(x-z)s(z), \tag{2.5}$$

$$\int_x W_t(x)\exp(iqx)=\exp[-\beta(q^2,t)], \tag{2.6}$$

$$F(x)=\int_q \exp(iqx)/\{1-\exp[-2\alpha(q^2,t)]\}. \tag{2.7}$$

$\langle s(x)\rangle$ is a weighted average of s around x with weight function W_t. One can choose this weight function to be constant in a cube of side (e^t-1) and zero outside. This gives

$$\exp(-\beta)=\exp(pt)\prod_{i=1}^{d}\sin\left(\frac{q_i(e^t-1)}{2}\right)\Big/\left(\frac{q_i(e^t-1)}{2}\right). \tag{2.8}$$

Equations analogous to (2.3) and (2.8) will appear in the lattice formulation.

III. RENORMALIZATION GROUP ON A LATTICE

Renormalization group transformations on a lattice formulated by Wilson[7] will now be explained. Consider a lattice in a d-dimensional space with unit spacing. The lattice provides an ultraviolet cutoff: the momentum components lie in the interval $(-\pi,\pi)$. Let S_n be the spin variable at the lattice site n, $-\infty\leq S_n\leq\infty$. Let $H_0(S)$ be the hamiltonian of the system. A renormalization transformation will now be defined.[14] Divide the lattice into cells, each cell being a hypercube with side two units (Fig. 1); the nth cell has 2^d sites, labeled $2\mathbf{n}+\mathbf{m}$ where each component of m has the value 0 or 1. Associate a spin S'_n with the nth cell. Let $H_1(S')$ be the hamiltonian for the new lattice. The transformation taking $H_0(S)$ to $H_1(S')$ is a renormalization transformation,

$$\exp[H_1(S')]=\int\prod_i dS_i\, T_1(S',S)\exp[H_0(S)]. \tag{3.1}$$

The transformation will be required to have the property that the correlation length of the new system (in units of the new lattice spacing) is half that of the old system (in units of old lattice spacing). If the initial system is critical, i.e., if it corresponds to infinite correlation length, then the new system also has infinite correlation length. In this case, by a repeated application of transformation (3.1) one can expect to reach a fixed point of the transformation. Linearized perturbations around this fixed point then give critical exponents or anomalous dimensions in the usual way.[1]

A simple choice of $T_1(S',S)$ is[15]

$$T_1(S',S)=\exp[-\tfrac{1}{2}\sum_n a_1(S'_n-\sum_{m=0}^{1}b_{1,m}S_{2n+m})^2], \tag{3.2}$$

where a_1 and the $b_{1,m}$ are parameters.

If H_k is the hamiltonian obtained from H_0 by applying transformation (3.1) k times, then one can write, for $1\leq k\leq l$,

$$\exp[H_l(S')]=\int\prod_i dS_i\, T_k(S',S)\exp[H_{l-k}(S)]. \tag{3.3}$$

Choice (3.2) makes T_k a simple function, i.e.,

$$T_k(S',S)=\exp[-\tfrac{1}{2}\sum_n a_k(S'_n-\sum_{m=0}^{L-1}b_{k,m}S_{Ln+m})^2], \tag{3.4}$$

where

$$L=2^k. \tag{3.5}$$

Let us look at Eq. (3.3) with $l=k$. Each cell is a cube of side L and the L^d sites of the nth cell are labeled by $L\mathbf{n}+\mathbf{m}$ where each component of m runs from 0 to $L-1$. One can show that

$$a_k^2=a_1^2(1-B)/(1-B^k),$$

where

$$B=\sum_{m=0}^{1}b_{1,m}^2. \tag{3.7}$$

One can solve for $b_{k,m}$ similarly; for brevity the result will be given later only for two special cases of interest. Is transformation (3.1) with (3.2) an acceptable renormalization transformation for all the values of the parameters? To find out, consider the relation between the correlation functions corresponding to H_k and H_0,

$$\Gamma_k(n)=\sum_{m_1,m_2=0}^{L-1}b_{k,m_1}b_{k,m_2}\Gamma_0(Ln+m_1-m_2)+\frac{1}{a_k^2}\delta_{n0}. \tag{3.8}$$

It might seem that for large n, one can replace $Ln+m_1-m_2$ by Ln to conclude that $\Gamma_k(n)\approx B^2\Gamma_0(Ln)$; since $\Gamma_0(Ln)$ behaves like $|Ln|^{-(d-2+\eta)}$, $\Gamma_k(n)$ also seems to have the same n dependence thus preserving long distance properties. This argument is incorrect. The large n limit in the left hand side of (3.8) corresponds to long wavelength limit. However, waves of wavelengths which are submultiples of L on the old lattice look like waves of infinite wavelength on the new lattice. Such wavelengths are present on the right side of Eq. (3.8) and correspond to the various values of m_1 and m_2. In other words, the long wavelength behavior of Γ_k depends not only on the long wavelength behavior of Γ_0 but also on a sequence of smaller wavelengths; however, if the dependence on these shorter wavelengths is weak enough, the long wavelength behavior of Γ_k will be the same as that of Γ_0.

To see this clearly, let us go to momentum space where we introduce the Fourier transform S_q of S_n,

$$S_q=\sum S_n\exp(iqn). \tag{3.9}$$

Equation (3.3) now becomes

$$T_k(S',S)=\exp\left\{-\frac{1}{2}\frac{1}{(2\pi)^d}\int_q a_k^2\left[S'_q-\sum_{l=-L/2+1}^{L/2}U_k\left(\frac{q+2\pi l}{L}\right)S_{(q+2\pi l)/L}\right]\times[q\to -q]\right\}, \tag{3.10}$$

where

$$U_k(p)=\sum_m b_{k,m}\exp(ipm). \tag{3.11}$$

Comparing (3.10) with (2.1) we see that U_k plays a role analogous to that of $\exp(-\beta)$ (k plays the role of t). Roughly speaking, from (3.10) it follows that S'_q behaves like $\sum_l U_k((q+2\pi l)/L)S_{(q+2\pi l)/L}$; thus S'_q depends not only on $S_{q/L}$ but on $S_{(q+2\pi l)/L}$. We would like the transformation to preserve the small momentum properties; this means that S'_q should depend primarily on $S_{q/L}$, and the weight factors associated with $S_{(q+2\pi l)/L}$ for $l\neq 0$ should be small compared to the weight factor for $l=0$. In order to make this quantitative, look at Eq. (3.8) in momentum space (momentum components, whether they refer to the old or the new lattice, vary from $-\pi$ to π),

$$\tilde{\Gamma}_k(q)=2^{-dk}\sum_l\left|U_k\left(\frac{q+2\pi l}{L}\right)\right|^2\tilde{\Gamma}_0\left(\frac{q+2\pi l}{L}\right)+\frac{1}{a_k^2}. \tag{3.12}$$

If the small q behavior of $\tilde{\Gamma}_k$ is the same as that of $\tilde{\Gamma}_0$, then one can look at the limit $k\to\infty$ to reach a fixed point, and obtain information about Γ_0 by studying only the fixed point and its neighborhood. If one starts from a

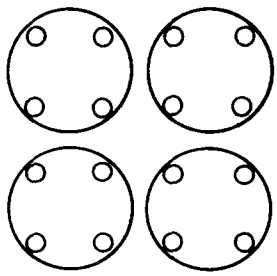

FIG. 1. Part of an infinite two-dimensional square lattice divided into 2×2 blocks.

critical hamiltonian, $\tilde{\Gamma}_0(q)$ behaves like $1/q^{2-\eta}$ for small q where η is a critical exponent. Knowing this, and for any specific choice of $b_{1,m}$ in Eq. (3.2), computing $U_k(p)$, one can take the limit of $k \to \infty$ in Eq. (3.12) and check if $\tilde{\Gamma}_k(q)$ also behaves like $1/q^{2-\eta}$.

We will now consider two specific cases, one of which is a good transformation, the other being unacceptable.

Case 1: $b_{1,m} = b_1$ independent of m. All site spins are given equal weight [see Eq. (3.2)]. One can show that

$$b_{k,m} = b_1^k, \tag{3.13}$$

$$a_k^2 = a_1^2(1 - 2^d b_1^2)/[1 - (2^d b_1^2)^k], \tag{3.14}$$

$$U_k(p) = b_1^k \prod_{i=1}^{d} \exp[ip_i(L-1)/2] \frac{\sin(p_i L/2)}{\sin(p_i/2)}. \tag{3.15}$$

A graph of $\sin(p_i L/2)/\sin(p_i/2)$ for $0 < p_i < \pi$ is sketched in Fig. 2 for $L = 8$. Crudely speaking, transformation (3.10) thins the degrees of freedom[16] in the following way: The degrees of freedom[16] in the interval $(\pi/L, 3\pi/L)$ are folded over the interval $(-\pi/L, \pi/L)$ by the $l = 1$ term; similarly for other intervals. Since U_k is large in the interval $(-\pi/L, \pi/L)$ compared to its value in any other interval, one can expect this folding over not to foul up the small q behavior. Thus this transformation acting on a critical Hamiltonian should lead to a sensible fixed point. In the limit $k \to \infty$, Eq. (3.12) gives

$$\tilde{\Gamma}^*(q) = \left(\prod_i \frac{1}{4} \sin^2 \frac{q_i}{2}\right) \sum_l \prod_i \frac{1}{(q_i + 2\pi l_i)^2} \frac{1}{(\mathbf{q} + 2\pi \mathbf{l})^{2-\eta}} + \frac{1 - 2^{-2+\eta}}{a_1^2} \tag{3.16}$$

It is also necessary to choose[17,18]

$$b_1^2 = 2^{-(d+2-\eta)}. \tag{3.17}$$

For small q, the $l = 0$ term in (3.16) dominates, ensuring that the small q behavior of $\Gamma^*(q)$ is $1/q^{2-\eta}$.

Case 2: $b_{1,m} = b_1$ for one of the 2^d site spins, 0 for the rest. This gives $b_{k,m} = b_1^k$ for one of the 2^{kd} site spins, 0 for the rest.

$$a_k^2 = a_1^2(1 - b_1^2)/(1 - b_1^{2k}). \tag{3.18}$$

Also $|U_k(p)| = b_1^k$ and is independent of p. Hence the folding over of the degrees of freedom should foul up the small momentum behavior leading, if at all, to an uninteresting fixed point. Taking the limit of Eq. (3.12) as $k \to \infty$, and replacing $\tilde{\Gamma}_0((q + 2\pi l)/L)$ by $[L/(q + 2\pi l)]^{2-\eta}$, we get

$$\tilde{\Gamma}_k(q) \sim (b_1^2 2^{-d+2-\eta})^k \sum_{l=-2^{k-1}+1}^{2^{k-1}} \frac{1}{|q + 2\pi l|^{2-\eta}} + \frac{1}{a_k^2}. \tag{3.19}$$

The series diverges as $k \to \infty$ for all cases of interest if $d \geq 2$. Another way to see this trouble is to look at Eq. (3.8) which for this case is

$$\Gamma_k(n) = b_1^{2k} \Gamma_0(Ln) + \frac{1}{a_k^2} \delta_{n_0}. \tag{3.20}$$

Since $\Gamma(Ln)$ behaves like $|Ln|^{-d+2-\eta}$ for large n, one has to choose $b_1^2 = 2^{d-2+\eta}$. For all cases of interest in $d \geq 2$, b_1 is greater than unity. Equation (3.20) for $n = 0$ reads

$$\Gamma_k(0) = b_1^k \Gamma_0(0) + \frac{b_1^{2k} - 1}{b_1^2 - 1} \frac{1}{a_1^2}. \tag{3.21}$$

This diverges as $k \to \infty$.

There are, of course, any number of intermediate situations between Cases 1 and 2. In Case 1, all of the 2^d spins of a block were given equal weight; in Case 2, only one of the 2^d spins was given nonzero weight. For intermediate situations, the above discussion is readily applicable with the obvious modifications. As one decreases the number of spins which are given nonzero weight from 2^d, it takes an increasing number of iterations to get rid of the effect of the overlapping of the high momentum degrees of freedom on the low momentum behavior, resulting in a slower convergence to a fixed point. Eventually a stage will be reached when the overlapping fouls up the low momentum behavior leading, if at all, to an unphysical fixed point.

It is also clear from the above discussion that if one eliminates the overlapping completely, i.e., if one chooses the function U_k such that it vanishes outside the interval $(-\pi/L, \pi/L)$, the convergence to the fixed point is expected to be faster. In this case the block spins will be linear combinations of spins extending beyond the block.

Niemeijer and van Leeuwen[4] have done very successful finite lattice calculations for the two-dimensional Ising model using a nonlinear renormalization transformation. The transformation used in this paper is a linear renormalization transformation. A nonlinear analog of Eq. (3.2) is

$$T_1(S', S) = \exp\{-\tfrac{1}{2} \sum_n a_1(S'_n - b \sum_m S_{2n+m}$$

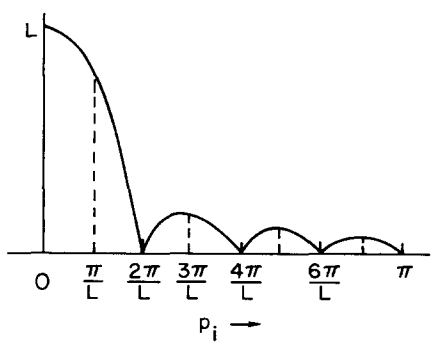

FIG. 2. A graph of the function $\sin(p_i L/2)/\sin(p_i/2)$ for $L = 8$.

$$-c\sum_{m_1,m_2,m_3} S_{2n+m_1}S_{2n+m_2}S_{2n+m_3})^2\}. \tag{3.22}$$

If $c=0$, then b is constrained to be equal to $2^{-(d+2-\eta)/2}$ (Eq. (3.17); one might guess that if c is not zero, then b (as well as c) is unconstrained (just as it presumably happens in the Ising case) at least over some region. (For a discussion of this see Bell and Wilson, Ref. 12). For this reason the above transformation may indeed be very desirable, but it is too complicated for the sort of analysis done for linear transformations in this paper.

ACKNOWLEDGMENT

The author is deeply indebted to Professor Kenneth Wilson for helpful discussions.

*Supported by the National Science Foundation.

†Present address: Division of Natural Sciences, University of California at Santa Cruz, Santa Cruz, California 95064.

[1]K. G. Wilson and J. Kogut, Phys. Rep. C 12, 75 (1974).

[2]K. G. Wilson and M. E. Fisher, Phys. Rev. Lett. 28, 240 (1972); K. G. Wilson, Phys. Rev. D 7, 2911 (1973).

[3]K. G. Wilson, Phys. Rev. D 6, 419 (1972).

[4]Th. Niemeijer and J. M. J. van Leeuwen, Phys. Rev. Lett. 31, 1411 (1973); Physica 71, 17 (1974); also in *Phase Transitions and Critical Phenomena*, Vol. 6, edited C. Domb and M. Green (to be published). The last is a review article which contains additional references.

[5]K. G. Wilson, Rev. Mod. Phys. 47, 773 (1975).

[6]L. P. Kadanoff, Phys. Rev. Lett. 34, 1005 (1975).

[7]K. G. Wilson (unpublished); T. L. Bell and K. G. Wilson, Phys. Rev. B 11, 3431 (1975). The latter has an extensive discussion of finite lattice approximation to the Gaussian model.

[8]For a review see E. S. Abers and B. W. Lee, Phys. Rep. C 9, 1 (1974). For a review of the relation between Euclidean and Minkowskian metrics see K. Symanzik, in *Proceedings of the International School of Physics "Enrico Fermi,"* edited by R. Jost (Academic, New York, 1969), Course XLV.

[9]Ref. 1, especially Chaps 3 and 11.

[10]Eq. (2.1) is obtained by combining Eqs. (XI.7) and (XI.11) of Wilson and Kogut, Ref. 1.

[11]In order to preserve group property, namely successive transformations through t_1 and t_2 to be equivalent to a single transformation through t_1+t_2, it is necessary to have $\beta(q^2,t) = f(q'^2) - f(q^2)$. The choice of (2.2) corresponds to $f(x) = x - \frac{1}{2}p \ln x$.

[12]The transformation is, however, nonlinear on the space of hamiltonians. For a discussion of nonlinear renormalization groups, see T. L. Bell and K. G. Wilson, Phys. Rev. B 10, 3935 (1974).

[13]For a review, see L. P. Kadanoff *et al.*, Rev. Mod. Phys. 39, 395 (1967).

[14]This is quite analogous to the renormalization transformations defined for Ising spins by K. Wilson, Ref. 5, and L. P. Kadanoff and A. Houghton, Phys. Rev. B 11, 377 (1975). See also K. Subbarao, Phys. Rev. B 11, 1165 (1975).

[15]If $a_1 \to \infty$, the transformation reduces to a δ-function.

[16]Thinning the degrees of freedom was the original motivation for introducing the renormalization group. See L. P. Kadanoff, Physics 2, 263 (1966). Also Ref. 1.

[17]It was shown by Wilson (Ref. 1, Chap. XII; Ref. 5) that linear renormalization transformations have such a parameter constrained by the exponent η.

[18]Note that if $H_0(S)$ in Eq. (3.1) is quadratic in the spins, then with the choice (3.2) for T_1, the transformed hamiltonian also is quadratic. Furthermore if $H_0(S) = \int \rho_0(q) S_q S_{-q}$ then the correlation function in momentum space $\Gamma_0(q)$ is $1/\rho_0(q)$. In this case the fixed point hamiltonian is given by $H^*(S) = \frac{1}{2}\int 1/\Gamma^*(q) S_q S_{-q}$ with Γ^* given by Eq. (3.16), η being zero. This Gaussian model has been discussed by Bell and Wilson, Ref. 7

Asymptotic properties of generalized Chaba and Pathria lattice sums

George L. Hall

Department of Physics, North Carolina State University, Raleigh, North Carolina 27607
(Received 8 September 1975)

For an arbitrary Bravais τ lattice in an m-dimensional Euclidean space amd for $0<a<\infty$, we present an extension of the Chaba and Pathria method of evaluating the lattice sums $\Sigma'_\tau \tau^{-2k}\exp(-a\tau^2)$ from integral k values to all positive k values. We use the extension to study the asymptotic properties of these sums as the parameter a approaches zero. The leading term is given by $\pi^{m/2}\,[\Gamma(k)(m/2-k)a^{m/2-k}]^{-1}$ for $0<2k<m$, by $\pi^{m/2}\Gamma(k)^{-1}\ln(1/a)$ for $2k=m$, and by $\Sigma'_\tau\tau^{-2k}$ for $2k>m$. Thus $2k=m$ gives a transition point from structure independence to structure dependence.

1. INTRODUCTION

Let $\{\tau\}$ be a unit Bravais lattice in an m-dimensional Euclidean space with a unit reciprocal lattice $\{\gamma\}$, let $0<a<\infty$, let $0<k$, and then consider the class of lattice sums defined by

$$J_\tau(a,k,m)\equiv\sum'_\tau \tau^{-2k}\exp(-a\tau^2). \tag{1.1}$$

Chaba and Pathria (CP)[1] have recently supplied a procedure for evaluating these sums for the subclass where k is a positive integer. They give detailed expressions for the special case where the τ lattice is m-dimensional simple-cubic, for which the two lattices are identical and some expressions are specialized, but they also give a sufficient outline that permits one to extend any of their equations for the general τ-lattice. In addition, CP discuss earlier related work of Glasser[2-4] and Zucker[5,6] and point out many areas of applications. These areas include Bose–Einstein condensation in finite systems,[7,8] the stability of an array of quantized vortices in Type II superconductors and in rotating superfluid helium,[9] and certain Madelung sums.[10]

We have been studying the larger class of sums for which k is any positive number, and we have determined the asymptotic properties of them as the parameter a approaches zero. Using Ewald's theta function method (TFM),[11-14] we derive expressions for $J_\tau(a,k,m)$ valid for all $0<a<\infty$ and all $0<k$ and cast these in a form similar to those of CP but void of CP's integration constants. Then we use these to determine the asymptotic properties. For three reasons we think these results may be of interest: (1) In the areas of application of CP sums with integral k values mentioned above[7-10] there may be interest in modifying the theoretical models such that nonintegral k values enter; (2) The CP sums may be used to study or evaluate other sums that are expressible in terms of integrals containing the CP sums as a portion of the integrand. (3) The generalized CP sums provide another example of a class of sums whose asymptotic properties are independent of the lattice structure over a certain domain of some parameter and dependent on the lattice structure over the rest of the domain except for the "boundary" point where the properties are somewhat special. This special point acts somewhat like a "transition temperature."[15]

As a approaches zero, the leading term for $J_\tau(a,k,m)$ is simply given by setting $a=0$ in Eq. (1.1) provided $2k>m$, which clearly depends on the structure of the τ lattice. For $0<2k<m$, the leading term diverges inversely as a power of a in a manner that is independent of the type of lattice; for $2k=m$, the leading term diverges logarithmically as $\ln(1/a)$, again in a manner independent of the structure of the lattice.

2. DERIVATION OF THE EXTENDED FORMULA

The TFM suggests writing $J_\tau(a,k,m)$ in one of the various forms,

$$\begin{aligned}\Gamma(k)J_\tau(a,k,m)&=\sum'_\tau\int_0^\infty y^{k-1}\exp[-(a+y)\tau^2]\,dy\\&=\sum'_\tau\int_a^\infty(x-a)^{k-1}\exp(-x\tau^2)\,dx\\&=\int_a^\infty(x-a)^{k-1}\sum'_\tau\exp(-x\tau^2)\,dx,\end{aligned}\tag{2.1}$$

where interchanging the order of summation and integration may be justified by applying the Weierstrass M-test for uniform convergence. The TFM would also lead to a consideration of the quantity

$$\begin{aligned}\Gamma(k)H_\gamma(a,k,m)&\equiv\pi^{m/2}(-1)^{k-1}a^{k-m/2}\\&\quad\times\sum'_\gamma\int_1^\infty(x-1)^{k-1}x^{(m/2-k-1)}\exp(-\pi^2\gamma^2x/a)\,dx\\&=\pi^{m/2}a^{k-m/2}\sum'_\gamma\int_0^1(t-1)^{k-1}t^{-m/2}\\&\quad\times\exp(-\pi^2\gamma^2/at)\,dt.\end{aligned}\tag{2.2}$$

The second expression in Eq. (2.2) may be rewritten as

$$\begin{aligned}&a^k\int_0^1(t-1)^{k-1}(\pi/at)^{m/2}[\textstyle\sum_\gamma\exp(-\pi^2\gamma^2/at)-1]\,dt\\&=\int_0^1[a^k(t-1)^{k-1}\textstyle\sum_\tau\exp(-at\tau^2)-(t-1)^{k-1}(\pi/at)^{m-2}]\,dt\\&=k^{-1}a^k(-1)^{k+1}+\int_0^a(x-a)^{k-1}[\textstyle\sum'_\tau\exp(-x\tau^2)\\&\quad-\int\cdots\int\exp(-x\beta^2)d^m\beta]\,dx,\end{aligned}\tag{2.3}$$

where use has been made of the Poisson summation formula (PSF).[16] We extend CP's results by combining

the above relations so as to yield for $0 < k$,

$$\sum_\tau' \tau^{-2k} \exp(-a\tau^2) + H_\gamma(a, k, m)$$
$$= [k\Gamma(k)]^{-1} d^k (-1)^{k+1} + P_\tau(a, k, m) + A(a, k, m), \quad (2.4)$$

where

$$\Gamma(k) A(a, k, m) \equiv \int_a^b (x-a)^{k-1} (\pi/x)^{m/2}\, dx, \quad (2.5)$$

$$\Gamma(k) P_\tau(a, k, m) \equiv \int_0^1 (x-a)^{k-1} [\sum_\tau' \exp(-x\tau^2) - \int \cdots \int \exp(-x\beta^2)\, d^m\beta]\, dx + \sum_\tau' \int_1^\infty (x-a)^{k-1} \exp(-x\tau^2)\, dx + \int_b^1 (x-a)^{k-1} (\pi/x)^{m/2}\, dx, \quad (2.6)$$

and where

$$b = \begin{cases} \infty, & k < m/2, \\ 1, & k = m/2 \\ 0, & k > m/2. \end{cases} \quad (2.7)$$

For the special case $k = 1$, Eq. (2.4) reduces to

$$\sum_\tau' \tau^{-2} \exp(-a\tau^2) + \pi^{2-m/2} \sum_\gamma' \gamma^{2-m} \Gamma(m/2-1, \pi^2\gamma^2/a)$$
$$= a + P_\tau(a, 1, m) + \int_a^b (\pi/x)^{m/2}\, dx, \quad (2.8)$$

where

$$P_\tau(a, 1, m) = \int_0^1 [\sum_\tau' \exp(-x\tau^2) - \int \cdots \int \exp(-x\beta^2)\, d^m\beta]\, dx + \sum_\tau' \tau^{-2} \exp(-\tau^2) + \int_b^1 (\pi/x)^{m/2}\, dx \quad (2.9)$$

is independent of a and reduces to CP's constant C_m when the τ lattice is simple-cubic.

3. ASYMPTOTIC PROPERTIES

Suppose that $0 < a \ll \pi$ and that $0 < 2k \leq m$. Then Eq. (2.4) yields

$$J_\tau(a, k, m)$$
$$\approx A(a, k, m) \approx \pi^{m/2} \Gamma(k)^{-1} \int_a^b x^{k-m/2-1}\, dx$$
$$= \begin{cases} \pi^{m/2} [\Gamma(k)(m/2-k) a^{m/2-k}]^{-1}, & 0 < 2k < m, \\ \pi^{m/2} \Gamma(k)^{-1} \ln(1/a), & 2k = m. \end{cases} \quad (3.1)$$

Thus for $0 < 2k \leq m$, the asymptotic behavior of $J_\tau(a, k, m)$ is, to first order, independent of the structure of the τ lattice.

Next suppose that $0 < a \ll \pi$ and that $m < 2k$. Then clearly we have

$$J_\tau(a, k, m) \approx \sum_\tau' \tau^{-2k}, \quad (3.2)$$

which is structure dependent.[15]

Clearly, one might have guessed the results of Eq. (3.1) from examining CP's results for integral k values, but it is useful to know for certain the dependence on k (especially for small k) as well as the dependence on a.

4. THE CASE $a \simeq \pi$

If $a = \pi$ we have

$$J_\tau(\pi, k, m) = \Gamma(k)^{-1} \pi^k \sum_\tau' \int_1^\infty (x-1)^{k-1} \exp(-\pi x\tau^2)\, dx, \quad (4.1)$$
$$H_\gamma(\pi, k, m) = (-1)^{k-1} \Gamma(k)^{-1} \pi^k \sum_\gamma' \times \int_1^\infty (x-1)^{k-1} x^{m/2-k-1} \exp(-\pi x\gamma^2)\, dx. \quad (4.2)$$

Thus $J_\tau(\pi, k, m)$ and $H_\gamma(\pi, k, m)$ may be comparable in magnitude, especially if $2k = m - 2$. There are both difficulties and useful features associated with this fact.

A difficulty arises if one wishes to use Eq. (2.4), or CP's method (for integral k-values), to evaluate $J_\tau(a, k, m)$ for $a \approx \pi$, because one may have to sum over many terms in evaluating $H_\gamma(a, k, m)$. However, in this domain of a-values direct use of the defining expression of Eq. (1.1) is feasible provided the dimension of the lattice is not too large.

Even when interest is centered on evaluating $J_\tau(a, k, m)$ for $a \ll \pi$ useful relations may be found by considering $a = \pi$, as CP have illustrated with their Eqs. (9) and (10). With increasing degrees of specialization, consider Eqs. (4.1), (4.2), and (2.4) with $a = \pi$ and first with $2k = m - 2$, second with additionally $k = 1$, and third with the τ lattice simple-cubic. The third stage reproduces CP's Eqs. (9) and (10), the second stage extends CP's Eq. (9) to general Bravais τ lattices, and the first stage obviously leads to simplifications, but we have not shown their usefulness within the general framework of the CP method.

[1] A. N. Chaba and R. K. Pathria, J. Math. Phys. 16, 1457 (1975).
[2] M. L. Glasser, J. Math. Phys. 14, 409 (1973).
[3] M. L. Glasser, J. Math. Phys. 14, 701 (1973).
[4] M. L. Glasser, J. Math. Phys. 15, 188 (1974).
[5] I. J. Zucker, J. Math. Phys. 15, 187 (1974).
[6] I. J. Zucker, J. Phys. A 7, 1568 (1974); I. J. Zucker and M. M. Roberson, J. Phys. A (to be published).
[7] S. Greenspoon and R. K. Pathria, Phys. Rev. A 9, 2103 (1974).
[8] A. N. Chaba and R. K. Pathria, to be published.
[9] A. L. Fetter, P. C. Hohenberg and P. Pincus, Phys. Rev. 147, 140 (1966).
[10] F. E. Harris and H. J. Monkhorst, Phys. Rev. B 2, 4400 (1970).
[11] P. P. Ewald, Ann. Phys. 54, 519, 577 (1917); 64, 253 (1921). Also see the next reference.
[12] M. Born and K. Huang, *Dynanical Theory of Crystal Lattices* (Oxford U. P., Oxford, 1954), pp. 248, 388. See other references given in this book.
[13] J. K. Mackenzie, Proc. Phys. Soc. Lond. A 63, 1370 (1950).
[14] B. M. E. van der Hoff and G. C. Benson, Can. J. Phys. 31, 1087 (1953).
[15] G. L. Hall, Bull. Am. Phys. Soc. **19**, 697 (1974); J. Stat. Phys. **14**, No. 6 (June, 1976) (to be published).
[16] E. M. Stein and G. Weiss, *Fourier Analysis on Euclidean Spaces* (Princeton U. P., Princeton, N. J., 1971), p. 253.

Scalar potentials in the Dirac equation*

Charles L. Critchfield

Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545
(Received 5 September 1975)

The Dirac equation for motion in a central potential is generalized to include scalar potentials proportional to r and to r^{-1}. It is solved by analytic methods. The linear dependence upon radius leads to spectra similar to that of the harmonic oscillator except that the approximately constant level distance applies to E^2 instead of E. A large negative term in the rest mass displaces the equilibrium point of the oscillator to a large radius.

I. INTRODUCTION

A scalar potential in the Dirac equation is equivalent to a dependence of the rest mass upon position. Of particular interest is the case in which the potential is a function of radius about a fixed center. Such potentials appear in the "quasi-independent" models[1,2] and "bag" models[3,4] of hadrons. It is known that the Dirac equation is readily solved when a scalar $1/r$ term is added.[5] It is the purpose of this paper to present the properties of the solutions when the rest mass is generalized to

$$m(r) = \mu + \lambda/r + \kappa^2 r, \tag{1}$$

where μ is the mass of the free particle and λ and κ characterize the scalar potential.

The term in Eq. (1) proportional to r produces eigenstates all of which are bound. Its presence introduces a regular singular point in addition to the one at $r=0$. Consequently the second-order differential equations for the Dirac amplitudes are of the same general type as those for spheroidal harmonics. Methods for finding eigenvalues for such equations are known[6,7] and will be applied to the Dirac equation in Sec. III.

The system of units which we use is that in which

$$\hbar = c = 1.$$

With the customary definition of the Dirac matrices $\boldsymbol{\alpha}$ and β the equation for ψ is

$$(E + \eta/r)\psi = (\boldsymbol{\alpha}, \mathbf{p})\psi + m(r)\beta\psi. \tag{2}$$

The parameter η corresponds to e^2 in the hydrogen atom and is positive for a negative potential. Equation (2) may be separated in spherical coordinates. When $m(r)$ is given by Eq. (1) the two functions of radius,[8] ψ_a and ψ_b called the large and small components respectively, obey

$$\begin{aligned} &[E + (\eta - \lambda)/r - \mu - \kappa^2 r]\psi_a + \psi_b' + [(j+1)/r]\psi_b = 0, \\ &[E + (\eta + \lambda)/r] + \mu + \kappa^2 r]\psi_b - \psi_a' + [(j-1)/r]\psi_a = 0, \end{aligned} \tag{3}$$

where if l is the orbital angular momentum in ψ_a, $j = l+1$ when the total angular momentum is $l+\frac{1}{2}$, and $j = -l$ when the total is $l-\frac{1}{2}$. The prime denotes differentiation by r.

In the following section we review briefly the properties of bound solutions when $\kappa = 0$ for which the general method of Sec. III is sufficient, of course, but not necessary.

II. BOUND STATES FOR $\kappa = 0$

When $\kappa = 0$, Eq. (3) is basically the same as those for the hydrogen atom.[5] The solutions are of the form

$$\left.\begin{aligned} \psi_a &= f_n(r) \\ \psi_b &= g_n(r) \end{aligned}\right\} r^{s-1} \exp[-r(\mu^2 - E^2)^{1/2}], \tag{4}$$

where $f_n(r)$ and $g_n(r)$ are polynomials of degree n. The index is readily found to be

$$s = (j^2 + \lambda^2 - \eta^2)^{1/2}, \tag{5}$$

and the eigenvalues of energy are given by

$$(n+s)(\mu^2 - E^2)^{1/2} - E\eta + \mu\lambda = 0, \tag{6}$$

or

$$E_{n,j} = \frac{\lambda\eta + (n+s)[(n+s)^2 + \eta^2 - \lambda^2]^{1/2}}{(n+s)^2 + \eta^2}\,\mu. \tag{7}$$

Thus levels with the same n and j^2 are degenerate except for $n=0$ for which there are no solutions with negative j.

Setting $\lambda = 0$ in Eq. (7) leads to the well-known formula for hydrogenlike spectra with $\eta = Ze^2$. The strongest binding occurs for $n=0$, i.e.,

$$E_{0,j} = \mu(j^2 - \eta^2)^{1/2}/|j|, \tag{8}$$

which approaches zero as $\eta \to 1$, if $j = 1$. The next lowest level in that limit is doubly degenerate with quantum numbers $n=1$, $j = \pm 1$, viz., $E_{1,1} = \frac{1}{2}\mu\sqrt{2}$.

One sees from Eq. (6) that bound states for $\eta = 0$ occur when λ is negative. The positive eigenvalues are

$$\begin{aligned} &E = \mu[1 - (\lambda/n+s)^2]^{1/2}, \\ &s = (j^2 + \lambda^2)^{1/2}, \quad \lambda < 0. \end{aligned} \tag{9}$$

When λ is near zero the spectrum is the same to second-order as that for the H-like atom with $\lambda^2 = \eta^2$. Strongly bound states arise as $\lambda \to -\infty$; expanding in negative powers of $|\lambda|$ we find

$$E^2 = \mu^2[2n/|\lambda| + (j^2 - 3n^2)/\lambda^2 + \ldots]. \tag{10}$$

Therefore, all states for which n^2 and j^2 are small compared to λ^2 approach zero energy as $\lambda \to -\infty$. Also, except for $n=0$ (the nondegenerates states) the eigenvalues approach independence of j. When $n=0$,

$$E_{0,j} = j\mu/|\lambda|. \tag{11}$$

The resemblance of Eq. (10) to the formula for the

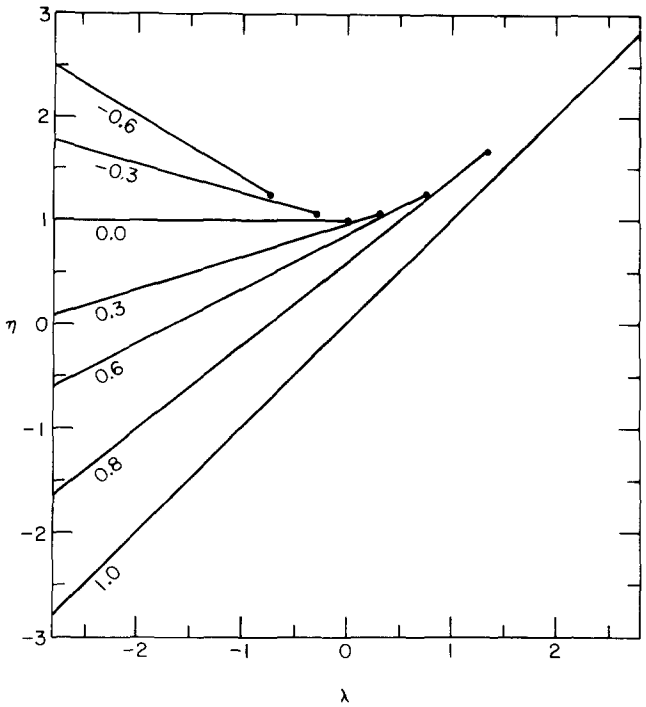

FIG. 1. Curves of equal ground state energy when $\kappa = 0$, in the λ—η plane. They are straight lines terminating on the $\eta > 0$ branch of the hyperbola $\eta^2 = 1 + \lambda^2$. Labels give the values of E/μ.

harmonic oscillator spectrum is a consequence of the fact that when $\lambda < 0$ the effective mass becomes negative inside $r_0 = |\lambda|/\mu$. Therefore at $r < r_0$ the scalar potential creates an outward acceleration. In effect the particle is trapped in the neighborhood of $r = r_0$. This is verified by the nature of the wave function, Eq. (4), which when λ^2 is large, is proportional to the product of a high power of r and a decreasing exponential. Thus the low-lying wave functions are of the "bubble" type.[4] The relation between states concentrated at a large radius and those of the harmonic oscillator is discussed in detail in Sec. IV.

Between the two limiting cases presented above there exists a continuum of values for η and λ that lead to bound states. The necessary conditions on η and λ are that s and $(\mu^2 - E^2)^{1/2}$ in Eq. (4) be real and positive. From Eq. (5) we see that the limitation is most severe when $j^2 = 1$. Therefore, we consider the possible ground states ($n = 0$) for $j = 1$. Let us define the angle θ to be such that

$$E_{0,1} = \mu \sin\theta, \quad (\mu^2 - E_{0,1}^2)^{1/2} = \mu \cos\theta.$$

From Eqs. (5) and (6) one readily finds

$$s = \sin\theta - \lambda\cos\theta, \quad \eta = \lambda \sin\theta + \cos\theta. \tag{12}$$

In order for $(\mu^2 - E_{0,1}{}^2)^{1/2}$ and s to be positive, we need

$$\pi/2 > \theta > -\pi/2, \quad \lambda < \tan\theta. \tag{13}$$

The linear relation between η and λ, Eq. (12), is illustrated in Fig. 1, where the lines are labelled by their value of $E_{0,1}/\mu$. The lines terminate at $\lambda = \tan\theta = E_{0,1}/(\mu^2 - E_{0,1}{}^2)^{1/2}$ on the $\eta > 0$ branch of the hyperbola $\eta^2 = 1 + \lambda^2$.

III. GENERAL SOLUTION

We now consider the general method of solving Eq. (3). Let $f(r)$ and $g(r)$ be series in nonnegative powers of r with leading terms f_0 and g_0, respectively, and write

$$\left.\begin{aligned}\psi_a &= f(r)\\ \psi_b &= g(r)\end{aligned}\right\} r^{s-1} \exp(-\mu r - \tfrac{1}{2}\kappa^2 r^2). \tag{14}$$

The equations for f and g are

$$\begin{aligned}&[E + (\eta - \lambda)/r - \mu - \kappa^2 r]f + g'\\ &\quad + [(s + j)/r - \mu - \kappa^2 r]g = 0,\\ &[E + (\eta + \lambda)/r + \mu + \kappa^2 r]g - f'\\ &\quad - [(s - j)/r - \mu - \kappa^2 r)f = 0.\end{aligned} \tag{15}$$

As $r \to 0$ we must have

$$(\eta - \lambda)f_0 + (s + j)g_0 = 0, \quad -(s - j)f_0 + (\eta + \lambda)g_0 = 0.$$

The vanishing of the determinant of these equations gives us s as expressed in Eq. (5).

In developing the series it is convenient to define

$$P \equiv f + g, \quad Q \equiv f - g, \tag{16}$$

whose differential equations as derived from Eq. (15) are

$$\begin{aligned}&Q' + [(s + \lambda)/r]Q = [E + (\eta + j)/r]P\\ &P' + [(s - \lambda)/r - 2\mu - 2\kappa^2 r)P + [E + (\eta - j)/r]Q = 0.\end{aligned} \tag{17}$$

Substituting

$$P = \sum_0^\infty p_k r^k, \quad Q = \sum_0^\infty q_k r^k \tag{18}$$

into Eq. (17) and applying the latter to the coefficients of r^k, we obtain

$$(k + 1 + s + \lambda)q_{k+1} = Ep_k + (\eta + j)p_{k+1}, \tag{19}$$

$$(k + 1 + s - \lambda)p_{k+1} - 2\mu p_k - 2\kappa^2 p_{k-1} + Eq_k + (\eta - j)q_{k+1} = 0. \tag{20}$$

Eliminating q_k and q_{k+1} from Eq. (20) by using Eq. (19) and multiplying with $(k + s + \lambda)$ $(k + 1 + s + \lambda)$, we get

$$A_k p_{k+1} + B_k p_k + C_k p_{k-1} = 0, \tag{21}$$

with

$$\begin{aligned}A_k &\equiv (k + 1)(k + \lambda + s)(k + 2s + 1),\\ B_k &\equiv E(j + \eta) + 2(k + \lambda + s)[E\eta - \mu(k + 1 + \lambda + s)],\\ C_k &\equiv (k + 1 + s + \lambda)[E^2 - 2(k + \lambda + s)\kappa^2].\end{aligned} \tag{22}$$

Since p_k vanishes for $k < 0$, we see from Eqs. (21) and (22) that p_0 is arbitrary and that

$$A_0 p_1 + B_0 p_0 = 0. \tag{23}$$

Equation (23) determines the eigenvalues E because p_1 is related to all p_k through Eq. (21) and the p_k must form a suitably convergent series for P of Eq. (18).

When $\kappa \neq 0$ and as $k \to \infty$ Eqs. (21) and (22) show that

$$kp_{k+1} \to 2\kappa^2 p_{k-1} + 2\mu p_k.$$

Therefore, in order for ψ_a and ψ_b to be normalizable,

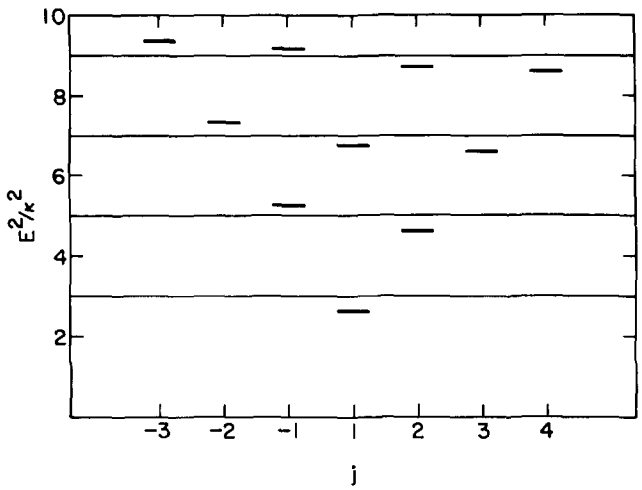

FIG. 2. Spectrum of E^2/κ^2 vs j when $\eta=\lambda=\mu=0$. Lines at odd integers represent the energy eigenvalues for the nonrelativistic isotropic harmonic oscillator.

the series for P must converge at large n like the series

$$\sum_0^\infty (n!)^{-1/2}(-\sqrt{2}\,\kappa r)^n.$$

Thus

$$b_{k+1}=b_k[1+O(1/k)]<0, \tag{24}$$

where

$$b_k \equiv p_k/p_{k-1}. \tag{25}$$

When $p_k \neq 0$, we may write Eq. (21) as

$$A_k b_{k+1}+B_k+C_k/b_k=0, \tag{26}$$

from which we find

$$b_k=-C_k/(A_k b_{k+1}+B_k). \tag{27}$$

The procedure for finding eigenvalues[7] is to choose a trial value for E, select a sufficiently large N and approximate Eq. (24) by $b_{N+1}=b_N$. Equation (26) is then a quadratic equation for b_N with the negative solution

$$b_N=-(1/2A_N)[B_N+(B_N^2-4A_N C_N)^{1/2}]. \tag{28}$$

For $k<N$ each b_k is determined from b_{k+1} by Eq. (27). If the trial value for E is such that Eq. (23) is satisfied, with $p_1=b_1 p_0$ it is an approximation to an eigenvalue and will be designated as E_N. The process involves approximating an infinite continued fraction by a finite one so that the true E is bounded by E_N and E_{N+1}, and as N becomes greater $\frac{1}{2}(E_N+E_{N+1})$ is an increasingly closer approximation to the true value.

The parameters η, λ, μ may be such that the true eigenvalue being sought causes C_k in Eq. (22) to vanish for some $k=\nu$, where $\nu>0$. Equation (27) shows that b_ν vanishes in such instances and hence $p_\nu=0$. But the method entails division by p_ν and thus formally does not apply. However, Eq. (21) for $k=\nu$ shows that when p_ν and C_ν vanish, so does $p_{\nu+1}$ ($A_\nu \neq 0$ for $\nu>0$) and therefore all $p_{\nu+i}=0$, $i>0$. The resulting series for P, Eq. (18), is then a polynomial of degree $\nu-1$. One of the roots for E in the finite continued fraction must agree with one of those of C_ν. Conversely one can find other values of the parameters for the same E and j as roots of polynomials, and use these discrete points to sketch out the continuum of solutions. Some care is required to connect the roots properly.

IV. SOLUTIONS FOR $\kappa \neq 0$

Bound states that are produced by the $1/r$ terms alone in Eq. (3) have been examined in Sec. II. We now take up the investigation of solutions when $\kappa>0$. The asymptotic form of Eq. (14) at large r is dominated by the factor

$$\exp[-(1/2)\kappa^2 r^2].$$

The wavefunctions are therefore related to those of the harmonic oscillator of mass M and frequency ω if we interpret κ as $(M\omega)^{1/2}$.

The parameter κ may be considered as the unit wavenumber, but there remain three free parameters in Eq. (3), viz., η, λ, and μ. It is desirable to restrict study to a few key choices for these parameters. The obvious beginning is to determine the characteristic effect of a finite κ and set

$$\eta=\lambda=\mu=0. \tag{29}$$

Under these conditions we use Eq. (3) to derive the coupled second-order differential equations for ψ_a and ψ_b, i.e.,

$$\begin{aligned}\psi_a''+(2/r)\psi_a'+[E^2-j(j-1)/r^2-\kappa^4 r^2]\psi_a-\kappa^2\psi_b=0,\\ \psi_b''+(2/r)\psi_b'+[E^2-j(j+1)/r^2-\kappa^4 r^2]\psi_b-\kappa^2\psi_a=0.\end{aligned} \tag{30}$$

These are to be compared with the equation for the nonrelativistic, isotropic three-dimensional oscillator whose wavefunction we shall call ψ_0,

$$\psi_0''+(2/r)\psi_0'+[2MW-l(l+1)/r^2-M^2\omega^2 r^2]\psi_0=0, \tag{31}$$

where W is the energy. From the definition of j in Sec. I we have $j(j-1)=l(l+1)$ so that the terms in r^{-2} are the same for ψ_0 and the large component ψ_a. In the nonrelativistic limit we can ignore the small component and the equations for ψ_0 and ψ_a become the same if we set $\kappa=(M\omega)^{1/2}$ as noted above and replace E^2 by $2MW$.

The eigenvalues of Eq. (3) under the conditions set in Eq. (29) were determined by the general method of Sec. III. Results for E^2/κ^2 in the ten lowest levels are plotted against j in Fig. 2. The levels are represented by short heavy bars. The corresponding eigenvalues of Eq. (31) are the horizontal lines representing

$$2W/\omega=2n+3. \tag{32}$$

The similarity of the two spectra is evident. Spin doubles the number of levels when $l>0$ and there is a slight spin-orbit splitting which, of course, is "anomalous." Numerical values are presented in Table I where n is half the integer nearest $(E^2/\kappa^2)-3$. This is the same n as in Eq. (32). Any positive E which belongs to a given j also represents a state of energy $-E$ belonging to $-j$.

These relativistic "harmonic oscillator" states are

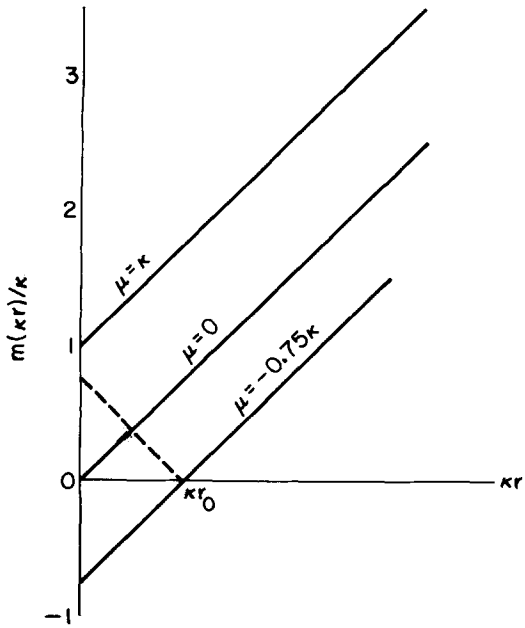

FIG. 3. Three examples of the linear scalar potential. The dashed line shows the apparent potential energy inside r_0 when $\mu < 0$.

quite different from those derived by postulating a linear restoring force.[9,10] The results presented here are more closely related to the classical motion of a particle whose mass is a linear function of position. In one dimension let x be the position and $\dot{x}$ the velocity of a "free" particle whose mass is $\gamma^2 x$. Its energy is then

$$H = \gamma^2 x(1 - \dot{x}^2)^{-1/2}, \tag{33}$$

with an obvious solution in harmonic motion,

$$x = H\gamma^{-2}\sin(H^{-1}\gamma^2 t).$$

Although the force is everywhere negative, when $x < 0$ the mass is also negative and the acceleration becomes positive. Note that the time averaged mass is zero and that adding a constant in the numerator of Eq. (33) does not change the average mass but shifts the equilibrium point of oscillation.

These considerations suggest that the next step in the investigation is to keep $\eta = \lambda = 0$ but take $\mu \neq 0$. Three typical possibilities are illustrated in Fig. 3. The spectrum presented in Table I is that for the curve $\mu = 0$. A positive μ leads to a potential like the curve for $\mu = \kappa$. Obviously its spectrum will be very similar to that for $\mu = 0$ with some fraction of μ added to each term-value for E. As μ increases, the motion becomes less relativistic and the fraction of μ which is added approaches unity. Thus a positive μ acts like a true addition to the rest mass. Results of calculation for some of the lower levels appear in that part of Fig. 4 corresponding to $\mu/\kappa > 0$.

TABLE I. Eigenvalues of the ten lowest states for $\eta = \lambda = \mu = 0$. The quantum number n is the integer nearest $\frac{1}{2}(E^2/\kappa^2 - 3)$.

n	j	E/κ	E^2/κ^2
0	1	1.6194	2.6226
1	−1	2.2940	5.2626
	2	2.1465	4.6076
2	−2	2.7044	7.3138
	1	2.6026	6.7737
	3	2.5693	6.6012
3	−3	3.0560	9.3390
	−1	3.0310	9.1871
	2	2.9520	8.7141
	4	2.9322	8.5977

The situation that arises for negative μ is illustrated in Fig. 3 by the curve labelled $\mu = -0.75\ \kappa$. Inside the radius r_0 at which $m(r_0) = \mu + \kappa^2 r_0 = 0$, viz.,

$$r_0 = -\mu\kappa^{-2}, \tag{34}$$

the scalar potential produces an outward acceleration. The apparent potential energy inside r_0 is shown by the dashed line. Just as in the classical example $-\mu$ acts as a displacement of the neutral point of oscillation. In the limit of large $-\mu$ the motion approaches that of an oscillator in one dimension; the average mass approaches zero and the states are highly relativistic. The states that arise in the SLAC bag model[4] are of the same type but not identical because in that model the mass is not a simple linear function of radius. Also in this limit the radial function ψ_b becomes almost equal to $-\psi_a$.

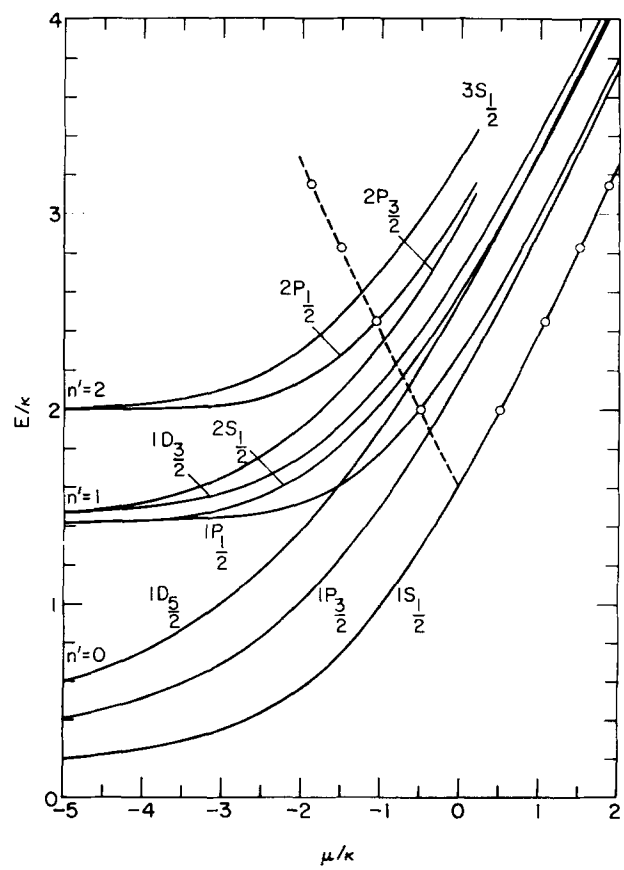

FIG. 4. The lower energy levels as functions of μ when $\eta = \lambda = 0$. The dashed line is one example of the spurious results from the method of infinite continued fractions when $\mu < 0$. The open circles give a few results when the continued fraction is finite.

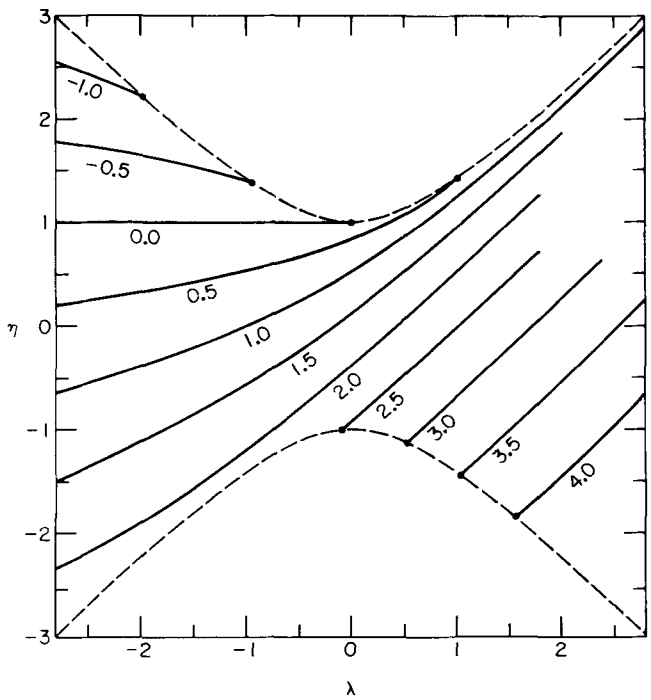

FIG. 5. Curves of equal ground state energy when $\mu = 0$, in the λ—η plane. Labels give the values of E/κ. The dashed lines show the limiting hyperbola, $\eta^2 = 1 + \lambda^2$. Lines for E/κ between 1 and 3.5 are not shown to the edge of the graph because of the very large N (> 2000) required for convergence in that region.

Let us put

$$y = (\psi_a - \psi_b)r,$$

and derive from Eq. (3), with $\eta = \lambda = 0$, the second-order equation for y,

$$y'' + [E^2 + \kappa^2 - j^2/r^2 - (\mu + \kappa^2 r)^2]y + (j/r)(\psi_a + \psi_b) = 0. \tag{35}$$

When $-\mu \gg \kappa$ the wave function is concentrated about r_0 which, from Eq. (34), is much larger than κ^{-1} and we may approximate r^{-2} in Eq. (35) by r_0^{-2}. The inhomogeneous term in Eq. (35) is the product of two small quantities and will be ignored to obtain the equation for the asymptotic form of y which is denoted by ϕ, i.e.,

$$\phi'' + [E^2 + \kappa^2 - j^2 r_0^{-2} - \kappa^4 (r - r_0)^2]\phi = 0. \tag{36}$$

Equation (36) has the same form as that for the one-dimensional oscillator and may be solved in the familiar way to obtain the eigenvalues,

$$E^2 = (2n' + j^2 \kappa^2 \mu^{-2})\kappa^2. \tag{37}$$

Here n' is a nonnegative integer which may be expressed in terms of n and j of Table I as

$$n' = \tfrac{1}{4}[2(n - |j|) + 3 - j/|j|]. \tag{38}$$

The close relationship between Eq. (37) and Eq. (10) is quite evident. Both arise from the effect of a negative mass inside a large radius. It will be noted that the ground state spectra ($n = n' = 0$) are formally the same when μ^2/λ^2 of Eq. (10) and κ^4/μ^2 of Eq. (37) are considered to be equivalent. Also, from Fig. 4, we see that states belonging to $n' = 0$ all have total angular momentum $l + \frac{1}{2}$, i.e., $j > 0$. This agrees with the results for $1/r$ potentials in which the only solutions for $n = 0$ belong to positive j. The weak dependence upon j^2 in the asymptotic limit is the result of the motion becoming essentially one-dimensional.

The curves in Fig. 4 for $\mu < -3\kappa$ were calculated using Eq. (37). Between $\mu = -3\kappa$ and $\mu = 0$ the general method of Sec. III cannot be applied because for all $\mu < 0$ that method converges to eigenvalues for nonphysical states. More specifically the curves obtained for E vs μ, $\mu < 0$ have negative slope and that is not acceptable in states of positive energy for which the expectation value of the operator β cannot be negative. One sees from Eq. (2) that the average values of β and of $\partial E/\partial \mu$ are the same. A single example is given in Fig. 4 by the dashed line which is a locus of eigenvalues for $j = -1$ and is the mirror image (about $\mu = 0$) of the acceptable locus for $j = 1$ and $\mu > 0$.

The gap between $\mu = -3\kappa$ and $\mu = 0$ was filled in by several means. For one thing some of the polynomial solutions fall in that gap as shown in Fig. 4 by open circles (only a few such solutions are shown). Secondly, for the lowest states of each j the calculus of variations was applied to the expectation value of E^2. In fact, minimizing the expectation value of E produces very good results when f and g are approximated as quadratic functions but then one has no guarantee that the result is an upper bound. Mostly the curves in this region of μ were calculated by numerical methods on finite matrices.[11]

Also shown in Fig. 4 are a few examples of the polynomial solutions for P. When $\lambda = \eta = 0$ and $j = \pm 1$, $s = 1$, C_k of Eq. (22) becomes

$$C_k = (k + 2)[E^2 - 2(k + 1)\kappa^2].$$

Since C_0 never enters the calculation there is a polynomial solution whenever E^2/κ^2 is an even integer greater than 2. Thus when $C_1 = 0$ the positive eigenvalue is $E = 2\kappa$ and $p_1 = 0$. Equation (23), by using Eq. (22), is then

$$B_0 = Ej - 4\mu = 0, \quad j = \pm 1.$$

Hence $\mu = \frac{1}{2}\kappa$ for $j = 1$, i.e., the $1S_{1/2}$ state and $\mu = -\frac{1}{2}\kappa$ for $j = -1$, which is the $1P_{1/2}$ state. Both solutions appear in Fig. 4 as the lowest pair of open circles. When $C_2 = 0$, Eqs. (27) and (22) show that

$$b_1 = -C_1/B_1 = 3(E^2 - 4\kappa^2)/(Ej - 12\mu),$$

which when used in Eq. (23) produces a quadratic for μ. The larger solution $\mu = 1.0537\ \kappa$ appears as the second point on $1S_{1/2}$ and its negative is shown on $2P_{1/2}$. The other solution, $\mu = -0.2372\ \kappa$, falls on the curve for $2S_{1/2}$ and its negative on $1P_{1/2}$ but these are not shown in Fig. 4. The selection of such points to be displayed was made in order to emphasize the fact that polynomial solutions also fall on curves for spurious eigenvalues as they do here on the dashed line and that one must take care in using them to construct energy curves.

In Fig. 5 are shown the loci of admissible ground states as functions of η and λ for $\mu = 0$. The curves are labelled with the value of E/κ to which each belongs. The limiting hyperbola is shown by the dashed curves.

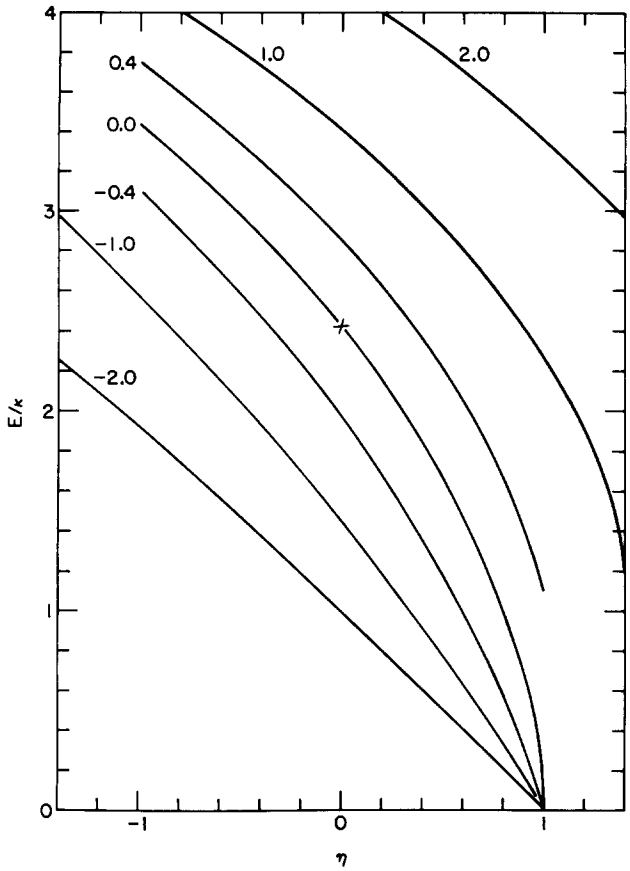

FIG. 6. Positive ground state spectra for E/κ as a function of η when $\mu=\kappa$. The curves are labelled with their value of λ. Note that all those for nonpositive λ converge at $E=0$, $\eta=1$. The solution which appears in Fig. 4 is marked with a cross.

The loci are no longer the straight lines of Fig. 1 for $\kappa=0$ except for $\eta=1$ which persists as a singular locus. This arises from the fact that when $\eta=1$, $s=|\lambda|$ so that $s+\lambda=0$ when $\lambda<0$. Then, in Eq. (22), $A_0=B_0=C_0=E=0$ for all negative λ. That $E=0$ is the limiting value under these conditions is shown in a different way in Fig. 6, where the loci of positive ground states for $\mu=\kappa$ are shown in the $E-\eta$ plane. The curves are labelled with their values of λ. It is to be noted also that $B_0=0$ for all negative λ whenever $\eta=-j$. This merely implies that one must avoid such input data.

V. CONCLUSION

Energy eigenvalues for the Dirac equation in which the mass has been generalized as in Eq. (1) are readily found as roots of infinite continued fractions. In practice one finds the roots of finite continued fractions. The results converge upon physical solutions, however, only when μ, the constant term in the mass, is not negative. Other means of computation are needed when $\mu<0$ (except for isolated polynomial solutions which arise when the continued fraction is truly finite).

A positive linear scalar potential in the Dirac equation appears in the Klein–Gordon form as the quadratic potential for harmonic motion in the nonrelativistic limit. As a result the eigenvalues for E^2 bear a close resemblance to those for energy of the three-dimensional oscillator, although the physical origin as a radially dependent mass is quite different from the ordinary prescription. The isotropic oscillatorlike spectrum dominates when μ is not negative.

Large negative values of μ, or a large negative $1/r$ scalar potential, lead to highly relativistic, one-dimensional (radial) harmonic motion about a finite radius. The corresponding wave functions are concentrated about that radius and are thus of the "bubble" type. When both scalar and vector $1/r$ attractive potentials occur all ground states have zero energy for $\eta=1$.

Detailed consideration has been limited to states in which the average value of Dirac's β operator is positive. Corresponding solutions for states in which $\langle\beta\rangle<0$ (the negative energy states for a free particle) are readily deduced from the obvious symmetry of Eq. (3) under interchanging ψ_a with ψ_b and reversing the signs of E, j, and η.

ACKNOWLEDGMENT

It is a pleasure to thank M. Bolsterli for his continued interest in this study, for the use of his computer code for solving the Dirac eqaution which was needed in certain regimes of negative rest mass, and for his assistance in using that code.

*Work supported by the Energy Research and Development Administration.

[1]P. N. Bogoliubov, Ann. Inst. Henri Poincaré 8, 163 (1968).
[2]C. L. Critchfield, Phys. Rev. D 12, 923 (1975).
[3]A. Chodos, R. L. Jaffee, K. Johnson, C. B. Thorn, and V. F. Weisskoff, Phys. Rev. D 9, 3471 (1974).
[4]W. A. Bardeen, M. S. Chanowitz, S. D. Drell, M. Weinstein and T.-M. Yan, Phys. Rev. D 11, 1094 (1975).
[5]J. Vasconcelos, Rev. Bras. Fis. 1, 441 (1971).
[6]R. L. McLaurin, Trans. Cambridge Philos. Soc. 17, 33 (1898).
[7]M. F. Manning, Phys. Rev. 48, 161 (1935).
[8]P. A. M. Dirac, *Quantum Mechanics* (Oxford U. P., 1958), 4th ed., p. 269 *et seq.*
[9]C. Møller, Mat. Fys. Medd. 30, 10 (1955), p. 16.
[10]R. Penfield and H. Zatzkis, J. Franklin Inst. 262, 121 (1956).
[11]M. Bolsterli, private communication, cf. Sec. VI.

Bilinear quantum field theories and their coherent states*

B. DeFacio† and C. L. Hammer

Ames Laboratory-ERDA and Department of Physics, Iowa State University, Ames, Iowa 50010
(Received 18 July 1975)

The quantization method developed by Hammer and Tucker, which is based upon a set of equations of motion and their conserved currents rather than a canonical formalism, is extended to interacting systems. The operators of the theory are bilinear and are essentially self-adjoint on a dense domain which is spanned by a suitably chosen subset of the coherent states. Both proper and improper gauge transformations of the second kind are discussed. For the proper case, the connection is given between these transformations and coherent states, which are discussed in detail. One interesting result is that a "smeared" Fock space can be constructed for a system where the particles have the same *average* quantum numbers. For the improper case, the gauge transformation of the second kind is related to the purely absolutely continuous measure. The formalism is applied to two examples. One is a Dirac field minimally coupled to a massive vector field, and the other is Klauder's ultralocal models.

1. INTRODUCTION

A basic problem in quantum field theories is the formulation of a theory which can, at least eventually, lead to a nontrivial scattering matrix. While field theories based upon a Lagrangian formalism may have this property, there are a number of mathematical reasons for seeking an alternative approach. In fact, for some rigorous soluble field theories no Lagrangian nor canonical theory exists.

However, the formalism originally developed by Takahashi[1] and Hammer and Tucker[2] (HT) which is based upon free field equations and the conserved currents which are derivable from them does not suffer from these same difficulties. Furthermore, the symmetry operators derivable from the conserved currents are bilinear forms. This has important mathematical consequences. Reeh[3] and Kastler, Robinson, and Swieca[4] have shown that any symmetry operator must diverge at least as fast as r, $r \to \infty$, in the weak limit. This suggests investigating the coherent states as a possible domain for symmetry operators for bras and kets of matrix elements. In his study, Klauder[5] showed that the most general self-adjoint field operators are simple *bilinear* expressions in terms of creation and annihilation operators. These methods and techniques are directly applicable to the HT formalism.

In this paper the HT formalism is extended to include interacting systems. Bilinear objects, called generalized currents, including q-number—q-number, q-number—c-number, and c-number—c-number will be used throughout. Various symmetry operators are constructed from these forms and their properties are investigated. Of particular importance is the gauge transformation of the second kind

$$U_\varphi(x_0) = \exp[\Sigma_\varphi(x_0)], \tag{1.1}$$

where Σ_φ is a q-number—c-number bilinear form. This transformation can be used to construct coherent states from the vacuum whenever $\varphi \in L^2(R^S, d\mu)$. For suitable non-$L^2$ functions, $\hat{\varphi}$, the transformation is related to a purely absolutely continuous measure.

The organization of this paper is the following. In Sec. 2, the Hammer—Tucker quantization procedure is generalized to include interactions. Then in Sec. 3 this procedure is illustrated by the example of a Dirac field minimal coupled to a massive vector field. This example is nontrivial because no Lagrangian exists which gives all of the field equations. In Sec. 4 the gauge transformation of the second kind is introduced and coherent states including their relationship to c-number—c-number forms are discussed in some detail. A smeared Fock space is developed which has a discrete spectrum of average quantum numbers, rather than the usual "sharp" eigenvalues. In Sec. 5 we discuss the c-number Hilbert space, operator extensions, and a distinguished family of non-L^2 solutions to the field equations. These solutions give a representation of the fields which is unitarily inequivalent to the Fock representation. In Sec. 6 we use the extended HT method to reproduce the ultralocal model solutions of Klauder.[6–10] In Sec. 7 our conclusions are listed.

2. METHOD OF QUANTIZATION

One essential feature of including nontrivial interactions is that a general field operator must be at least bilinear in the Fock-representation operators $\psi^\dagger(x)$, $\psi(x)$. Since the conserved current is a naturally occurring bilinear, the method of quantization first proposed by Takahashi[1] and Hammer and Tucker[2] (HT hereafter), which is based upon conserved currents, will be used here. This quantization method has been discussed primarily within the context of free fields, excepting the arbitrary spin electrodynamics constructed by Tucker and Hammer.[11] The purpose of this section is to review the HT formalism and to extend it to the interacting case.

Let ψ denote an operator valued solution to the configuration space partial differential equations,

$$D(\partial)\psi = j(x), \quad \text{and} \quad \bar{\psi}\tilde{D}(-\partial) = \bar{j}(x), \tag{2.1}$$

where $j(x)$ is a source term, $D(\partial)$ is a configuration space differential operator, and

$$\bar{\psi}(x) = [\gamma_4\psi(x)]^\dagger, \quad \bar{j}(x) = [\gamma_4 j(x)]^\dagger, \tag{2.2}$$

and

$$D^\dagger(\partial)\gamma_4 = -\gamma_4^\dagger D(-\partial).$$

The quantity γ_4 is the $[2(2s+1)\times 2(2s+1)]$-dimensional generalization of the Dirac matrix γ_4 only when ψ is a field constructed from symmetric spinors of like indices. Otherwise it is a matrix which is chosen to insure the Lorentz transformation properties of the conserved current. A general four-vector A_μ is written in terms of real quantities A_i $(i=1, 2, 3)$ and A_0 as

$$A_\mu = \begin{pmatrix} A_i \\ iA_0 \end{pmatrix}, \quad \text{and} \quad \partial_\mu = \begin{pmatrix} \partial_i \\ -i\partial_0 \end{pmatrix} \equiv \begin{pmatrix} \nabla \\ -i\dfrac{\partial}{\partial t} \end{pmatrix}. \tag{2.4}$$

The functions $u_r(\mathbf{p}, x)$ and $v_r(\mathbf{p}, x)$ are homogeneous c-number solutions to

$$D(\partial)u_r(\mathbf{p}, x) = 0 \tag{2.5}$$

and

$$D(\partial)v_r(\mathbf{p}, x) = 0, \tag{2.6}$$

where r is a set of discrete quantum numbers, p is the three-momentum, and x is a space–time point. Their charge conjugation properties are

$$u_r(\mathbf{p}, x) = Cv_r^*(\mathbf{p}, x), \tag{2.7}$$

$$v_r(\mathbf{p}, x) = Cu_r^*(\mathbf{p}, x), \tag{2.8}$$

where $*$ denotes complex conjugation and C is a c-number matrix which satisfies $C^2 = 1$. We consider that class of theories for which a conserved current $J_\mu(\psi_1, \psi_2)$ exists with

$$\partial_\mu J_\mu(\psi_1, \psi_2) = 0, \tag{2.9}$$

and which is bilinear in the operators ψ_1, ψ_2. In general, ψ_1 and ψ_2 are two independent solutions to Eqs. (2.1) and (2.2), and factors of i are chosen in J_μ such that

$$J_i^\dagger(\psi_1, \psi_2) = J_i(\psi_2, \psi_1), \tag{2.10}$$

$$J_0^\dagger(\psi_1, \psi_2) = J_0(\psi_2, \psi_1). \tag{2.11}$$

The Hermitian adjoint of any c-number operator q is defined as

$$J_\mu(\psi_1, q\psi_2) \equiv J_I(q^\dagger\psi_1, \psi_2). \tag{2.12}$$

We know of no general proof that such a conserved current must exist although all known successful theories do have this property.

We further assume the existence of a free limit with a *free* conserved current, J_μ^f, which satisfies

$$\partial_\mu J_\mu^f(\psi_1, \psi_2) = \{\bar{\psi}_1^{\text{in(out)}}\vec{D}(\partial)\psi_2^{\text{in(out)}} - \bar{\psi}_1^{\text{in(out)}}\overleftarrow{D}(-\partial)\psi_2^{\text{in(out)}}\} = 0. \tag{2.13}$$

Note that γ_4 must be chosen such that J_μ^f transforms as a four-vector under Lorentz transformations.

The creation and destruction operators for particles A, $A^\dagger$ and antiparticles B, $B^\dagger$ are defined in terms of the zeroth component of J_μ^f according to

$$A_i(\mathbf{p}, x_0) = \int d\mathbf{x} J_0^f(u_i(\mathbf{p}, x), \psi(x)),$$

$$A_i^\dagger(\mathbf{p}, x_0) = \int d\mathbf{x} J_0^f(\psi(x), u_i(\mathbf{p}, x)),$$

$$B_i^\dagger(\mathbf{p}, x_0) = \int d\mathbf{x} J_0^f(v_i(\mathbf{p}, x), \psi(x)),$$

and

$$B_i(\mathbf{p}, x_0) = \int d\mathbf{x} J_0^f(\psi(x), v_i(\mathbf{p}, x)).$$

The reader is especially reminded of an *implied* Hermitian conjugation of operators (complex conjugation of c-number solution functions) in all left arguments of $J_0^f(\ ,\ )$ or $J_\mu(\ ,\ )$. Also, we call attention to the fact that the particular bilinears in Eq. (2.14) all consist of mixed pairs of one q-number and one c-number. The "in" and "out" creation and destruction operators for particles and antiparticles are defined as

$$\begin{aligned} z^{1/2}a_i^{\text{in(out)}}(\mathbf{p}) &= \text{w-}\lim_{x_0 \to -\infty(\infty)} [A_i(\mathbf{p}, x_0)], \\ z^{1/2}[a_i^{\text{in(out)}}(\mathbf{p})]^\dagger &= \text{w-}\lim_{x_0 \to -\infty(\infty)} [A_i^\dagger(\mathbf{p}, x_0)], \\ z^{1/2}b_i^{\text{in(out)}}(\mathbf{p}) &= \text{w-}\lim_{x_0 \to -\infty(\infty)} [B_i(\mathbf{p}, x_0)], \end{aligned} \tag{2.15}$$

and

$$z^{1/2}[b_i^{\text{in(out)}}(\mathbf{p})]^\dagger = \text{w-}\lim_{x_0 \to -\infty(\infty)} [B_i^\dagger(\mathbf{p}, x_0)],$$

where $z^{1/2}$ is a renormalization constant.

The HT quantization postulates for Bosons $(-)$ and Fermions $(+)$ are

$$\int d\sigma_\mu(x) J_\mu^f(u_i(\mathbf{p}, x), u_m(q, x)) = -[a_i^{\text{in(out)}}(\mathbf{p}), a_m^{\text{in(out)}\dagger}(\mathbf{q})]_\pm, \tag{2.16}$$

and

$$\int d\sigma_\mu(x) J_\mu^f(v_i(\mathbf{p}, x), v_m(\mathbf{q}, x)) = -[b_i^{\text{in(out)}}(\mathbf{p}), b_m^{\text{in(out)}\dagger}(\mathbf{q})]_\pm, \tag{2.17}$$

where $d\sigma_\mu(x)$ is an integration variable over a spacelike surface $\sigma_\mu(x)$ which contains the point x. In this construction as $x_0 \to \mp\infty$, an operator ψ has the limits

$$\psi(x) = z^{1/2}\psi^{\text{in}}(x) + \int d^4y G_R(x-y, m_0^2)j(x), \tag{2.18}$$

$$\psi(x) = z^{1/2}\psi^{\text{out}}(x) + \int d^4y G_A(x-y, m_0^2)j(x), \tag{2.19}$$

where

$$D(\partial)\psi^{\text{in(out)}}(x) = 0, \tag{2.20}$$

and where G_R and G_A are the retarded and advanced Green's functions for $D(\partial)$, respectively.

Remark: The sets of functions $\{u_I, u_I^*\}$ and $\{v_I, v_I^*\}$ *need not* be complete, orthonormal sets. For example, they could be wave packet solutions to Eqs. (2.5) and (2.6) or c-number coherent states. In the special case where the u's and v's are a complete set of plane wave functions, the quantization postulates Eqs. (2.16) and (2.17) become

$$\int d\sigma_\mu(x) J_\mu^f(u_i(\mathbf{p}, x), u_m(\mathbf{q}, x)) = -\rho(E_p)\delta_{im}\delta(\mathbf{p}-\mathbf{q}), \tag{2.21}$$

and

$$\int d\sigma_\mu(x) J_\mu^f(v_i(\mathbf{p}, x), v_m(\mathbf{q}, x)) = -\rho(E_p)\delta_{im}\delta(\mathbf{p}-\mathbf{q}), \tag{2.22}$$

where $\rho(E_p)$ is an energy density of states factor, and HT reduces to the Yang–Feldman[12] formalism. Also, when the u's and v's are not a complete orthonormal set, Eqs. (2.14) and (2.15) cannot be inverted to give

ψ and $\bar{\psi}$ in terms of the A's and B's although a complete theory exists in terms of the bilinear $J_\mu^f(\ ,\ )$.

Some nonvanishing commutation relations which follow from the quantization postulate include

$$[a_l^{\text{in(out)}}(p),\ \bar{\psi}^{\text{in(out)}}(x)]_\pm = \bar{u}_l(\mathbf{p},\ x), \tag{2.23}$$

$$[b_l^{\text{in(out)}}(\mathbf{p}),\ \psi^{\text{in(out)}}(x)]_\pm = v_l(\mathbf{p},\ x),$$

and

$$[\psi^{\text{in(out)}}(x),\ \bar{\psi}^{\text{in(out)}}(y)]_\pm = G(x-y,\ m_0^2), \tag{2.24}$$

where $G(x-y, m_0^2) = \Sigma_l u_l(x)\bar{u}_l(y)$. If the sum on l is over a complete set of states, then $G(z) = G_A(z) - G_R(z)$.

Following HT, a q-number operator Q_r is defined in terms of a corresponding c-number operator q_r as the q-number–q-number form

$$Q_r^{\text{in(out)}}(x_0) = \int d\mathbf{x} J_0^f(\psi^{\text{in(out)}},\ q_r\psi^{\text{in(out)}}). \tag{2.25}$$

If, furthermore, $q_r = s_r$ is a self-adjoint, c-number symmetry operator, then

$$[s_r,\ \mathcal{D}(\partial)] = 0, \tag{2.26}$$

and the corresponding q-number operator $S_r^{\text{in(out)}}$ is time independent. It follows from Theorem 3 of HT that

$$[\psi^{\text{in(out)}}(x),\ S_r^{\text{in(out)}}] = s_r\psi^{\text{in(out)}}(x), \tag{2.27}$$

and

$$[S_r^{\text{in(out)}},\ S_m^{\text{in(out)}}] = \int \mathbf{dx} J_0^f(\psi^{\text{in(out)}},\ [s_r,\ s_m]\psi^{\text{in(out)}}. \tag{2.28}$$

Therefore, $S_r^{\text{in(out)}}$, s_r have the same algebraic properties and they must generate corresponding symmetry transformations in their respective Hilbert spaces. These equations also imply that any $\psi^{\text{in(out)}}(x)$ with a dense set of analytic vectors transforms finitely as

$$e^{-i\alpha S^{\text{in(out)}}}\psi^{\text{in(out)}}(x)e^{i\alpha S^{\text{in(out)}}} = e^{i\alpha s}\psi^{\text{in(out)}}(x), \tag{2.29}$$

where α is a real parameter. Since $S^{\text{in(out)}}$ is independent of x, this transformation can be used to show that

$$e^{-\alpha S^{\text{in(out)}}}Q_r^{\text{in(out)}}(x_0)e^{i\alpha S^{\text{in(out)}}} = \int \mathbf{dx} J_0^f(e^{-i\alpha S^{\text{in(out)}}}\psi^{\text{in(out)}}(x)e^{i\alpha S^{\text{in(out)}}},\ q_r e^{-i\alpha S^{\text{in(out)}}}\psi^{\text{in(out)}}(x)e^{i\alpha S^{\text{in(out)}}}) \tag{2.30}$$

or that, infinitesmally,

$$[Q_r^{\text{in(out)}},\ S_m^{\text{in(out)}}] = \int \mathbf{dx} J_0^f(\psi^{\text{in(out)}}, [q_r,\ s_m]\psi^{\text{in(out)}}). \tag{2.31}$$

Thus, $Q_r^{\text{in(out)}}$ has the same tensor properties under the symmetry generated by the $\{S_m^{\text{in(out)}}\}$'s as q_r does with the $\{s_m\}$'s.

It is important to note that the commutation relations between general operators $Q_r^{\text{in(out)}}(x_0)$ are fixed by the form given by Eq. (2.25) once the q_r are defined. Conversely, a commutator algebra for the $Q_r^{\text{in(out)}}(x_0)$ suffices to define the operators q_r.

For example, if $J_0^f(\psi^{\text{in(out)}},\ q_r\psi^{\text{in(out)}})$ can be put into the "canonical form" described by Schweber[13] as

$$J_0^f(\psi^{\text{in(out)}},\ q_r\psi^{\text{in(out)}}) = \sum_i \pi_i^{\text{in(out)}} q_r \psi_i^{\text{in(out)}}, \tag{2.32}$$

where $\psi_i^{\text{in(out)}}$ and $\pi_i^{\text{in(out)}}$ are the independent fields and their canonical conjugates, then it is easy to show that

$$[Q_r^{\text{in(out)}}(x_0),\ Q_m^{\text{in(out)}}(x_0)] = \int \mathbf{dx} J_0^f(\psi^{\text{in(out)}},\ [q_r,\ q_m]\psi^{\text{in(out)}}. \tag{2.33}$$

These equations can be used to define q_r if the q-number algebra is given. The position operators of Fleming[14] and Sankaranarayanan and Good,[15] where q_r is not the simple spatial variable $\mathbf{x}$, are typical examples. However, a q_r satisfying Eqs (2.32) or (2.33) may not exist for a J_0^f of arbitrary form so that, in general, q_r and Q_r may have different commutation properties.

The free field Fock space generators of the Poincaré group such as the energy-momentum $P_\mu^{\text{in(out)}}$, the angular momentum-rapidity $M_{\mu\nu}^{\text{in(out)}}$, and the charge operato $Q^{\text{in(out)}}$, are given by

$$P_\mu^{\text{in(out)}} = \int \mathbf{dx} J_0(\psi^{\text{in(out)}},\ p_\mu\psi^{\text{in(out)}}),$$

$$M_{\rho\nu}^{\text{in(out)}} = \int \mathbf{dx} J_0^f(\psi^{\text{in(out)}},\ m_{\mu\nu}\psi^{\text{in(out)}}),$$

and

$$Q^{\text{in(out)}} = \int d\mathbf{x} J_0^f(\psi^{\text{in(out)}},\ \psi^{\text{in(out)}}). \tag{2.34}$$

These imply the corresponding equations of motion:

$$[\psi^{\text{in(out)}}(x),\ P_\mu^{\text{in(out)}}] = p_\mu\psi^{\text{in(out)}}(x),$$

$$[\psi^{\text{in(out)}}(x),\ M_{\mu\nu}^{\text{in(out)}}] = m_{\mu\nu}\psi^{\text{in(out)}}(x),$$

$$[\psi^{\text{in(out)}}(x),\ Q^{\text{in(out)}}] = q_0\psi^{\text{in(out)}}(x), \tag{2.35}$$

where p_μ and $m_{\mu\nu}$ are the configuration space, c-number generators of the Poincaré group and q_0 is the renormalized or observed electric charge. For example, p_μ is given by

$$p_\mu = \left(-i\nabla,\ -\frac{\partial}{\partial t}\right). \tag{2.36}$$

The generalization to the interacting case is straightforward. In parallel to Eq. (2.25), an operator is defined in terms of an interacting (not free!) unrenormalized conserved current J_μ according to

$$Q_r(x_0) = \int \mathbf{dx} J_0(\psi,\ q_r\psi). \tag{2.37}$$

Then if $(q_r\psi)$ is also a solution to Eq. (2.1), $Q_r = S_r$ is an exact symmetry operator for the system. Since S_r is time independent, J_0 can be evaluated in the w-limit at $t \to \pm\infty$, giving

$$S_r = KS_r^{\text{in(out)}}, \tag{2.38}$$

where the proportionality constant K is some function of the renormalization constants.

For a general interacting current J_μ, it is necessary to *postulate* that the fields ψ transform according to

$$e^{-i\alpha S}\psi e^{i\alpha S} = e^{i\alpha S}\psi(x), \tag{2.39}$$

where S (and s) is a self-adjoint, exact symmetry operator and α is a real parameter. Infinitesmally, one has that

$$[\psi,\ S] = s\psi. \tag{2.40}$$

The transformation property of an operator $Q_r(x_0)$ under the symmetry operator S is then

$$e^{-i\alpha S}Q_r(x_0)e^{i\alpha S} = \int \mathbf{dx} J_0(\psi,\ e^{-i\alpha s}q_r e^{i\alpha s}\psi). \tag{2.41}$$

This result can be used to show

$$[Q_r,\ S_t] = \int \mathbf{dx} J_0(\psi,\ [q_r,\ s_t]\psi). \tag{2.42}$$

Just as for the free particle case, the tensorial properties of Q_r follow from those of S.

Note that Schweber's canonical form holds for interacting fields as well as noninteracting fields so that the discussion in the paragraph which includes Eqs. (2.32) and (2.33) also applies to the interacting case. Thus, the tensorial and algebraic properties of $Q_r(x_0)$ are then in one-to-one correspondence with those of $Q_r^{\text{in(out)}}$, since

$$\text{w-}\lim_{x_0 \to \pm\infty} [Q_r(x_0)] = KQ_r^{\text{in(out)}}, \tag{2.43}$$

where again K is some function of the renormalization constants. The unrenormalized interacting Poincare generators and the charge operator become

$$\mathbf{P} = \int \mathbf{dx} J_0(\psi,\ -i\nabla\psi),$$

$$H = \int \mathbf{dx} J_0\left(\psi,\ i\frac{\partial}{\partial t}\psi\right),$$

$$M_{\mu\nu} = \int \mathbf{dx} J_0(\psi,\ m_{\mu\nu}\psi),$$

and

$$Q = q\int \mathbf{dx} J_0(\psi,\ \psi), \tag{2.44}$$

which is completely parallel to Eq. (2.29). The requirement H be self-adjoint implies that substitution from Eq. (2.1) can be made such that

$$H = \int \mathbf{dx} J_0\left(\left(i\frac{\partial}{\partial t}\psi\right)(x),\ \psi(x)\right).$$

While this discussion is adequate for the purposes of this paper, it is oversimplified to the extent that an interacting system may have more than one field equation and more than one conserved current. Then the various generators can be defined in terms of the sum of the independent currents. However, for the particular case of the system Hamiltonian operator, care must be taken to not multiply count the various interactions. See especially Tucker and Hammer[11] on the electrodynamics of general spin vector mesons for an example of this.

The commutation rules, such as Eq. (2.40), which symmetries impose on the system, require that

$$\int \mathbf{dy}[\psi(\mathbf{x},\ t),\ J_0(\psi(\mathbf{y}, t),\ s\psi(\mathbf{y}, t))]_{\pm} = s\psi(\mathbf{x}, t). \tag{2.46}$$

If we further assume that the independent fields are *local*,

$$[\psi_i(\mathbf{x},\ x_0),\ \psi_j(\mathbf{y},\ x_0)]_{\pm} = 0 \tag{2.47}$$

or *quasilocal*,

$$[\psi_i(\mathbf{x},\ x_0),\ \psi_j(\mathbf{y},\ x_0)]_{\pm} \to 0, \tag{2.48}$$

as $|\mathbf{x}-\mathbf{y}| \to \infty$, then one expects that Eq. (2.46) is sufficient to derive the remaining equal time commutators. While no general proof for this exists, it is certainly true for canonical theories such as those which satisfy Eq. (2.32). Then Eq. (2.46) dictates, for the local case,

$$[\psi_i(\mathbf{x},\ x_0),\ \pi_j(\mathbf{y},\ x_0)]_{\pm} = i\delta_{ij}\delta(\mathbf{x}-\mathbf{y}), \tag{2.49}$$

the usual canonical rule.

In summary, the form of $J_\mu(\ ,\ )$ is determined by the equations of motion Eq. (2.1), and any relativistic field theory requires Eq. (2.39) or (2.40). These equations together with Eq. (2.37), which defines the q-number operators, ensure the covariance of the theory. These requirements give necessity conditions for the equal time commutation relations, if any. The locality or quasilocality postulate is a separate and distinct assumption which is not required. Therefore, our approach does not requre the existence of a Lagrangian, a canonical structure nor locality in general, although it can, of course, accomodate any of these notions.

3. EXAMPLE, A COUPLED THEORY, DIRAC-MASSIVE VECTOR FIELDS

For a nontrivial illustration of our formalism, we present a coupled system comprised of a Dirac field and a massive vector field. In particular, consider the system as specified by the field equations

$$\gamma_\mu(p_\mu - qA_\mu)\psi(x) = im\psi(x), \tag{3.1}$$

$$(\Box + M^2)A_\mu(x) = iq\bar{\psi}(x)\gamma_\mu\psi(x), \tag{3.2}$$

with the "Lorentz condition"

$$\partial_\mu A_\mu(x) = 0, \tag{3.3}$$

where it is clear from the Hermitian conjugation properties of the right-hand side of Eq. (3.2) that

$$A_i = (A_i)^\dagger,\ A_0 = A_0^\dagger, \tag{3.4}$$

and where

$$\Box = \partial_0^2 - \nabla^2. \tag{3.5}$$

This system poses some formal difficulties for a Lagrangian formalism, even for the free field case because not all the components A_μ are independent. If Eqs. (3.2) and (3.3) are recovered from a Lagrangian, then A_0 has no canonical partner. This makes the procedure for obtaining the Hamiltonian and other physical operators unclear, and at best, ad hoc.

If the Lagrangian is defined with an auxiliary field which is the canonical partner of A_0, then Eq. (3.2) is not the field equation for A_μ. Equations (3.1)—(3.3) are recovered only after a series of complicated transformations.

The free field conserved currents are given by

$$J_\mu^f(\psi_1^{\text{in(out)}},\ \psi_2^{\text{in(out)}}) = i\bar{\psi}_1^{\text{in(out)}}\gamma_\mu\psi_2^{\text{in(out)}} \tag{3.6}$$

for the Dirac particle, and

$$K_\mu^f(A_1^{\text{in(out)}}(x),\ A_2^{\text{in(out)}}(x)) = -i\ [\bar{A}_{1\nu}^{\text{in(out)}}(\partial_\mu A_{2\nu}^{\text{in(out)}}) - (\partial_\mu \bar{A}_{1\nu}^{\text{in(out)}})A_{2\nu}^{\text{in(out)}}] = -i\bar{A}_{1\nu}^{\text{in(out)}}(x)\overleftrightarrow{\partial}_\mu A_{2\nu}^{\text{in(out)}}(x) \tag{3.7}$$

for the massive vector particle. The Lorentz transformation properties of K_μ require the inner product $(\bar{A}_1 \cdot A_2)$ to be a Lorentz scalar for any two solutions to the free field equations. Thus, both $\bar{A}_{1\mu}$ and $A_{2\mu}$ must be four vectors. This, along with the definition for $\bar{A}_1$ implied by Eq. (2.2) requires that

$$L = \gamma_4^{-1}L^*\gamma_4, \tag{3.8}$$

where L is the Lorentz transformation

$$A_1'(x') = LA_1(x). \tag{3.9}$$

A suitable choice for γ_4 is the metric

$$g=(g_{\mu\nu})=\mathrm{diag}(+1,\ +1,\ +1,\ -1), \tag{3.10}$$

which defines $\bar{A}_1$ as

$$\bar{A}_1=gA_1^\dagger. \tag{3.11}$$

In either case, whether $A_\mu(x)$ is a free or interacting field operator, Eq. (3.4) implies additionally that

$$\bar{A}_\mu=A_\mu. \tag{3.12}$$

Thus A_μ is self-adjoint. However, the plane wave solution to the free massive vector particle equation,

$$(\Box+M^2)\varphi_{\lambda\mu}(\mathbf{p},\ x)=0, \tag{3.13}$$

with the Lorentz condition

$$\partial_\mu\varphi_{\lambda\mu}(\mathbf{p},\ x)=0, \tag{3.14}$$

is given by

$$\varphi_{\lambda\mu}(\mathbf{p},\ x)=(2\pi)^{-3/2}\epsilon_\mu(\lambda)e^{ip\cdot x},\quad(\lambda=1,\ 2,\ 3), \tag{3.15}$$

with

$$(p\cdot\epsilon)=0, \tag{3.16}$$

where $p_\mu=(\mathbf{p},\ iE_p)$, $E_p=(\mathbf{p}^2+M^2)^{1/2}$, wnereas the adjoint solution is

$$\begin{aligned}\bar{\varphi}_{\lambda\mu}(\mathbf{p},\ x)&=g_{\mu\nu}\varphi^*_{\lambda\nu}=(2\pi)^{-3/2}g_{\mu\nu}\epsilon^*_\nu(\lambda)e^{-ip\cdot x}\\&\equiv(2\pi)^{-3/2}\bar{\epsilon}_\mu(\lambda)e^{-ip\cdot x},\end{aligned} \tag{3.17}$$

where $(p\cdot\bar{\epsilon})=0$. The Lorentz condition and normalization,

$$2E\delta_{\lambda\lambda'}\delta(\mathbf{p}-q)=\int d\mathbf{x}K_0^f(\varphi_{\lambda\mu}(\mathbf{p},\ x),\ \varphi_{\lambda'\mu}(q,\ x)), \tag{3.18}$$

imply that the polarization vectors satisfy

$$\bar{\epsilon}(\lambda)\cdot\epsilon(\lambda')=\delta_{\lambda\lambda'},$$

and

$$\sum_\lambda\epsilon_\mu(\lambda)\bar{\epsilon}_\nu(\lambda)-M^{-2}p_\mu p_\nu=\delta_{\mu\nu}, \tag{3.19}$$

for each λ, $\lambda'=1,2,3$ and for each μ, $\nu=1,2,3,4$. Note that $\epsilon=\bar{\epsilon}$ only holds for the case of plane polarization.

Additional information can be obtained by examining the charge conjugation properties of the fields. If the charge conjugation of the Dirac field is chosen as

$$C\psi(x)C^{-1}=C\bar{\psi}^{\mathrm{t}}=\gamma_2\psi^{\mathrm{t}}(x), \tag{3.20}$$

then the requirement that Eqs. (3.1)–(3.3) be invariant under this transformation imposes the condition

$$CAC^{-1}=CA^{\mathrm{t}}=-A \tag{3.21}$$

on the vector field. Comparing this to Eq. (3.11) shows that

$$C=-g \tag{3.22}$$

is a suitable choice for the c-number charge conjugation matrix C. Thus, the charge conjugation plane wave solutions become

$$\begin{aligned}\varphi^C_{\lambda\mu}(\mathbf{p},\ x)&=C\varphi^*_{\lambda\mu}(\mathbf{p},\ x)=-(2\pi)^{-3/2}\bar{\epsilon}_\mu(\lambda)e^{-ip\cdot x}\\&=-\bar{\varphi}_{\lambda\mu}(\mathbf{p},\ x),\end{aligned} \tag{3.23}$$

and

$$\bar{\varphi}^C_{\lambda\mu}(\mathbf{p},\ x)=(2\pi)^{-3/2}\epsilon_\mu(\lambda)e^{ip\cdot x}=\varphi_{\lambda\mu}(\mathbf{p},\ x). \tag{3.24}$$

It is easily verified that

$$\int d\mathbf{x}K_0^f(\varphi^C_{\lambda\mu}(\mathbf{p},\ x),\quad\varphi^C_{\lambda'}(\mathbf{q},\ x))=-2E\delta_{\lambda\lambda'}\delta(\mathbf{p}-\mathbf{q}),$$

$$\int d\mathbf{x}K_0^f(\varphi^C_{\lambda\mu}(\mathbf{p},\ x),\ \varphi_{\lambda'\mu}(\mathbf{q},\ x))=0,$$

and

$$\int d\mathbf{x}K_0^f(\varphi_{\lambda\mu}(\mathbf{p},\ x),\ \varphi^C_{\lambda'\mu}(\mathbf{q},x))=0. \tag{3.25}$$

The solutions $\varphi_{\lambda\mu}(\mathbf{p},x)$ and $\varphi^C_{\lambda\mu}(\mathbf{q},x)$ would form a complete set except for the missing (timelike) polarization direction. Thus, the homogeneous Green's function, which follows from the discussion below Eq. (2.24), is

$$\begin{aligned}G(x-y,\ M^2)&=\sum_{\lambda=1}^{3}\int\frac{d\mathbf{p}}{2E_p}[\varphi_{\lambda\mu}(\mathbf{p},\ x)\bar{\varphi}_{\lambda\nu}(\mathbf{p},\ y)\\&\quad-\varphi^C_{\lambda\mu}(\mathbf{p},\ x)\ \bar{\varphi}^C_{\lambda\nu}(\mathbf{p},y)]\\&=-i[\delta_{\mu\nu}-M^{-2}\frac{\partial^2}{\partial x_\mu\partial x_\nu}]\Delta(x-y,\ M^2),\end{aligned} \tag{3.26}$$

where

$$\Delta(x,M^2)=(2\pi)^{-3}\int\frac{d\mathbf{p}}{E_p}e^{i\mathbf{p}\cdot\mathbf{x}}\sin(E_px_0). \tag{3.27}$$

The commutation properties for the free field vector operators now follow from Eqs. (2.15), (2.16) and (2.24) as

$$\begin{aligned}&[A_\mu^{\mathrm{in(out)}}(x),\ A_\nu^{\mathrm{in(out)}}(y)]\\&=-i[\delta_{\mu\nu}-M^{-2}\frac{\partial^2}{\partial x_\mu\partial x_\nu}]\Delta(x-y,\ M^2).\end{aligned} \tag{3.28}$$

Finally, the in(out) limits for the vector field are defined as

$$\begin{aligned}z_3^{1/2}A_\lambda^{\mathrm{in(out)}}(\mathbf{p})&=\underset{x_0\to-\infty(\infty)}{\text{w-lim}}\int d\mathbf{x}K_0^f(\varphi_\lambda(\mathbf{p},\ x),A(x)),\\z_3^{1/2}[A_\lambda^{\mathrm{in(out)}}(\mathbf{p})]^\dagger&=\underset{x_0\to-\infty(\infty)}{\text{w-lim}}\int d\mathbf{x}K_0^f(\varphi^C_\lambda(\mathbf{p},\ x),\ A(x))\\&=\underset{x_0\to-\infty(\infty)}{\text{w-lim}}\int d\mathbf{x}K_0^f(A(x),\ \varphi_\lambda(\mathbf{p},\ x)),\end{aligned} \tag{3.29}$$

with

$$\begin{aligned}[A_\lambda^{\mathrm{in(out)}}(\mathbf{p}),\ A_{\lambda'}^{\dagger,\mathrm{in(out)}}(\mathbf{q})]&=\int d\mathbf{x}K_0^f(\varphi_\lambda(\mathbf{p},\ x),\ \varphi_{\lambda'}(\mathbf{q},\ x))\\&=2E_p\delta_{\lambda\lambda'}\delta(\mathbf{p}-\mathbf{q}).\end{aligned} \tag{3.30}$$

Note that Eq. (3.26) can be used to invert Eq. (3.29) to give the usual result for $A_\mu(x)$ and that it is unnecessary to define a timelike destruction operator for A_0 in the present formalism.

The Dirac field is quantized similarly. The u_λ and v_λ denote the usual plane wave Dirac spinors

$$u_\lambda(\mathbf{p},\ x)=(2\pi)^{-3/2}u_\lambda(\mathbf{p})e^{ip\cdot x}, \tag{3.31}$$

and

$$v_\lambda(\mathbf{p},\ x)=(2\pi)^{-3/2}v_\lambda(\mathbf{p})e^{-ipx}=\gamma_2u^*_\lambda(\mathbf{p},\ x). \tag{3.32}$$

Then the destruction operators are defined as

$$z_1^{1/2}a_\lambda^{\mathrm{in(out)}}(\mathbf{p})=\underset{t\to-\infty(\infty)}{\text{w-lim}}\int d\mathbf{x}u_\lambda^{\mathrm{t}}(\mathbf{p},\ x)\psi(x),$$

and

$$z_1^{1/2}b_\lambda^{\mathrm{in(out)}}(\mathbf{p})=\underset{t\to-\infty(\infty)}{\text{w-lim}}\int d\mathbf{x}\bar{\psi}^{\mathrm{t}}(x)v_\lambda(\mathbf{p},\ x). \tag{3.33}$$

The corresponding anticommutation relations become

$$\{a_\lambda^{\text{in(out)}}(\mathbf{p}),\ a_{\lambda'}^{\dagger\,\text{in(out)}}(\mathbf{p})\}$$
$$= \int d\mathbf{x}\, u_\lambda^\dagger(\mathbf{p},\ x) u_{\lambda'}(\mathbf{q},\ x)$$
$$= \frac{E_p}{m}\ \delta_{\lambda\lambda'}\delta(\mathbf{p}-\mathbf{q}),$$
$$\{b_\lambda^{\text{in(out)}}(\mathbf{p}),\ b_\lambda^{\dagger\,\text{in(out)}}(\mathbf{q})\} \tag{3.34}$$
$$= \int d\mathbf{x} v_\lambda^\dagger(\mathbf{p},\ x) v_{\lambda'}(\mathbf{q},\ x)$$
$$= \frac{E_p}{m}\delta_{\lambda\lambda'}\delta(\mathbf{p}-\mathbf{q}),$$

with all other anticommutation relations zero. The anticommutation relation for the Dirac fields is the usual

$$\{\psi^{\text{in(out)}}(x), \bar{\psi}^{\text{in(out)}}(y)\} = \tilde{S}(x-y,\ m), \tag{3.35}$$

where because the u's and v's form a complete set, $\tilde{S} = \tilde{S}_A - \tilde{S}_R$. The free field charge, momentum, and Hamiltonian follow Eq. (2.34) as

$$Q^{\text{in(out)}} = q_0 \int d\mathbf{x} \psi^{\dagger\,\text{in(out)}}(x)\, \psi^{\text{in(out)}}(x) \tag{3.36}$$

$$\mathbf{P}^{\text{in(out)}} = \int d\mathbf{x} \psi^{\dagger\,\text{in(out)}}(x)\, (-i\nabla\psi)^{\text{in(out)}}$$
$$+ (i/2)\int d\mathbf{x}\{A_\nu^{\text{in(out)}}(-i\nabla)\dot{A}_\nu^{\text{in(out)}} \tag{3.37}$$
$$- \dot{A}_\nu^{\text{in(out)}}(-i\nabla)A_\nu^{\text{in(out)}}\},$$

$$H^{\text{in(out)}} = \int d\mathbf{x} \psi^{\dagger\,\text{in(out)}}(i\partial_0)\psi^{\text{in(out)}}$$
$$+ (i/2)\int d\mathbf{x}\{A_\nu^{\text{in(out)}}(i\partial_0)\dot{A}_\nu^{\text{in(out)}} - \dot{A}_\nu^{\text{in(out)}}(i\partial_0)A_\nu^{\text{in(out)}}\}. \tag{3.38}$$

The extra factor of $\frac{1}{2}$ in the vector particle part of Eqs. (3.38) and (3.39) is required because A_μ is self-adjoint. These can be brought into more recognizable form by substituting the values of $(i\dot{\psi}^{\text{in(out)}}(x))$ and $\ddot{A}_\mu^{\text{in(out)}}(x))$ from the free particle equations. For example, the Hamiltonian can be rewritten as

$$H^{\text{in(out)}} = \int d\mathbf{x} \psi^{\dagger\,\text{in(out)}}(x)\{\boldsymbol{\alpha}\cdot\mathbf{p} + m\beta\}\psi(x)^{\text{in(out)}}$$
$$+ \tfrac{1}{2}\int d\mathbf{x}[(\nabla A_\nu^{\text{in(out)}}(x))\cdot(\nabla A_\nu^{\text{in(out)}}(x))$$
$$+ \dot{A}_\nu^{\text{in(out)}}(x)\dot{A}_\nu^{\text{in(out)}}(x) + M^2 A_\nu^{\text{in (out)}}(x)A_\nu^{\text{in(out)}}(x)\}. \tag{3.39}$$

For the interacting case, the only obviously nonvanishing conserved current is

$$J_\mu(\psi,\ \psi) = i\bar{\psi}(x)\gamma_\mu\psi(x), \tag{3.40}$$

which is derivable from Eq. (3.1), and which guarantees conservation of the unrenormalized charge operator

$$Q = q\int d\mathbf{x} J_0(\psi,\ \psi) = q\int d\mathbf{x}\psi^\dagger(x)\psi(x). \tag{3.41}$$

The requirement that

$$[\psi,\ Q] = q\psi, \tag{3.42}$$

together with the locality assumption

$$\{\psi(x),\ \psi(y)\}_{x_0=y_0} = 0, \tag{3.43}$$

and Eq. (3.41), then imply that

$$\{\psi(x),\ \psi^\dagger(y)\}_{x_0=y_0} = \delta(\mathbf{x}-\mathbf{y}), \tag{3.44}$$

which coincides with the equal-time specialization of Eq. (3.35).

For the massive vector field it is more complicated to work out the equal time commutators because the Lorentz condition, Eq. (3.3), relates the various components of A_μ. There are three independent components and we follow the usual convention by choosing them as the three spatial components $\mathbf{A}_i$ $(i = 1, 2, 3)$. Thus for *local* fields the locality assumption is

$$[A_0(x),\ A_0(y)]_{x_0=y_0} = 0, \tag{3.45}$$
$$[A_i(x),\ A_j(y)]_{x_0=y_0} = 0, \tag{3.46}$$

and

$$[A_i(x),\ \psi(y)]_{x_0=y_0} = 0. \tag{3.47}$$

Combining the Lorentz condition with Eqs. (3.46) and (3.47) yields

$$[A_i(x),\ \dot{A}_0(y)]_{x_0=y_0} = 0, \tag{3.48}$$

and

$$[\psi(x),\ \dot{A}_0(y)]_{x_0=y_0} = 0. \tag{3.49}$$

The other equal time commutators are obtained by requiring that the Poincaré generators satisfy Eq. (2.46). For example, Eq. (2.44) defines the momentum and Hamiltonian generators as

$$\mathbf{P} = \int d\mathbf{x}\psi^\dagger(-i\nabla)\psi + (i/2)\int d\mathbf{x}\{A_\nu(-i\nabla)\dot{A}_\nu - \dot{A}_\nu(-i\nabla)A_\nu\}, \tag{3.50}$$

and

$$H = \int d\mathbf{x}\,\psi^\dagger(i\partial_0)\psi + 1/2\int d\mathbf{x}\{(\nabla A_\nu)\cdot(\nabla A_\nu) + \dot{A}_\nu\dot{A}_\nu + M^2 A_\nu A_\nu\} \tag{3.51}$$

$$= \int d\mathbf{x}\ \psi^\dagger\{\boldsymbol{\alpha}\cdot(\mathbf{p}-q\mathbf{A}) + A_0 + m\beta\}\psi$$
$$+ \tfrac{1}{2}\int d\mathbf{x}\{\nabla A_\nu)\cdot(\nabla A_\nu) + \dot{A}_\nu\dot{A}_\nu + M^2 A_\nu A_\nu\}. \tag{3.52}$$

(We always assume normal ordering of operator expressions.) A natural way to proceed next is to postulate the free field equal time commutation relations and to see if they generate Eqs. (3.1)–(3.3) and the necessary invariants. But this procedure fails for the present example and it is necessary to use the specialization of Eq. (2.46), i. e.,

$$[\psi(x),\ H] = i\partial_0\psi(x), \tag{3.53}$$

and

$$[A_\mu(x),\ H] = i\partial_0 A_\mu(x), \tag{3.53}$$

in order to derive the remaining equal time commutation relations. For the spatial components to the vector field, Eqs. (3.45)–(3.49) and Eqs. (3.2) and (3.3) can be used to show

$$[A_i(x),\ H] = \int d\mathbf{y}\dot{A}_j(y)[A_i(x),\ (\dot{A}_j(y) + \nabla_j A_0(y))]_{x_0=y_0}. \tag{3.55}$$

Comparing this to Eq. (3.54) requires

$$[A_i(x),\ (\dot{A}_j(y) + \nabla_j A_0(y))]_{x_0=y_0} = i\delta_{ij}\delta(\mathbf{x}=\mathbf{y}), \tag{3.56}$$

so that

$$\pi_j(y) = \dot{A}_j(y) + \nabla_j A_0(y) \tag{3.57}$$

is the canonical partner to A_j. The Lorentz condition and Eqs. (3.2) and (3.47) can be used to recast Eq. (3.56) in the form

$$[A_i(x),\ (\nabla^2 A_0 - \ddot{A}_0)]_{x_0=y_0} = -i\nabla_{x_i}\delta(\mathbf{x}-\mathbf{y}), \tag{3.58a}$$

or

$$[A_i(x),\ A_0(y)]_{x_0=y_0} = -iM^{-2}\nabla_{x_i}\delta(\mathbf{x}-\mathbf{y}). \tag{3.58b}$$

Putting Eq. (3.58b) into Eq. (3.56) gives

$$[A_i(x),\ \dot{A}_j(y)]_{x_0=y_0} = i(\delta_{ij} - M^{-2}\nabla_{x_i}\nabla_{x_j})\delta(\mathbf{x}-\mathbf{y}), \tag{3.59a}$$

and the Lorentz condition together with Eq. (3.58b) gives

$$[A_0(x),\ \dot{A}_0(y)]_{x_0=y_0} = -iM^{-2}\nabla_x^2\,\delta(\mathbf{x}-\mathbf{y}). \tag{3.59b}$$

Equation (3.53) can be used to show

$$[\psi(x),\ \pi_i(y)]_{x_0=y_0} = 0. \tag{3.60}$$

Note that our locality assumption was made only for the independent fields. It is therefore gratifying that this expression occurs for the independent canonical partners without additional assumptions. Combining Eq. (3.60), the field equations and the Lorentz condition implies that

$$[\psi(x),\ A_0(y)]_{x_0=y_0} = qM^{-2}\psi(x)\delta(\mathbf{x}-\mathbf{y}) \tag{3.61}$$

and

$$[\psi(x),\ \dot{A}_i(y)]_{x_0=y_0} = qM^{-2}\psi(x)\nabla_{x_i}\delta(\mathbf{x}-\mathbf{y}), \tag{3.62}$$

both of which vanish in the free field limit ($q=0$). Two useful commutators which follow from these results include

$$[A_0(x),\ \psi^\dagger(y)\hat{O}(y)\psi(y)]_{x_0=y_0} = -qM^{-2}\psi^\dagger(y)[\hat{O}(y),\ \delta(\mathbf{x}-\mathbf{y})]\psi(y), \tag{3.63}$$

and

$$[A_\mu(x),\ J_\nu(y)]_{x_0=y_0} = 0. \tag{3.64}$$

This completes the list of equal time commutators and it is a simple exercise to verify that

$$\begin{aligned}[A_0(x),\ H]_{x_0=y_0} &= i\dot{A}_0(x),\\ [A_\mu(x),\ \mathbf{P}]_{x_0=y_0} &= -i\nabla A_\mu(x),\end{aligned} \tag{3.65}$$

and

$$[\psi(x),\ \mathbf{P}]_{x_0=y_0} = -i\nabla\psi(x). \tag{3.66}$$

4. COHERENT STATES AND GAUGE TRANSFORMATIONS

The conserved current approach is also useful for discussing gauge transformations of the second kind, and for relating them to the coherent states. Consider the bilinear form $J_0(\varphi,\ \psi)$ where φ is a c-number function (to be specified) and ψ is an operator valued solution to Eq. (2.1). Specifically, when $\varphi \in L^2(R^n)$ we have the case discussed by Klauder,[16] and when $\varphi \notin L^2$ something new and useful occurs. Until we state differently, φ will be taken to be $L^2(R^n)$. We begin by defining an operator

$$\Sigma_\varphi(x_0) = \sigma^\dagger(\varphi)\eta_A - \bar{\eta}_A\sigma(\varphi), \tag{4.1}$$

where

$$\sigma(\varphi) = \int d\mathbf{x} J_0(\varphi,\ \psi), \tag{4.2}$$

where for fermions

$$\{\eta_F,\ \bar{\eta}_F\} = \{\bar{\eta}_F,\ \bar{\eta}_F\} = \{\eta_F,\ \eta_F\} = \{\eta_F,\ \psi\} = \{\bar{\eta}_F,\ \psi\} = 0, \tag{4.3}$$

and for bosons, $\eta_B = 1$. The necessity for two *independent* objects η_F and $\bar{\eta}_F$ is well known and full construction of these quantities can be found in Rzewuski.[17] Since $(\bar{\eta}\ \eta)$ is real and commutes with η, $\bar{\eta}$, and ψ, it is a real c-number. We choose conventions such that

$$(\bar{\eta}_F\ \eta_F)^n = (-\eta_F\ \bar{\eta}_F)^n = \delta_{n1}, \tag{4.4}$$

for all integers n, $n \geq 1$. The c-number objects η_F, $\bar{\eta}_F$, which also appear in heuristic functional approaches to field theories, are required for both $\sigma^\dagger(\varphi)\eta_A$ and $\bar{\eta}_A\sigma(\varphi)$ to satisfy the same commutation relationships regardless of whether they represent boson or fermion fields.

Note that Σ_φ is a conserved quantity and is time independent only if $\varphi(x)$ satisfies Eq. (2.1). If a c-number solution exists only in the free particle limit, only $\Sigma_\varphi^{\text{in(out)}}$ is conserved. It is then convenient to expand φ in terms of plane wave states,

$$\varphi(x) = \sum_\lambda \int \frac{d\mathbf{p}}{\rho(E_p)} K_\lambda(\mathbf{p}) u_\lambda(\mathbf{p},\ x). \tag{4.5}$$

For an antiparticle c-number state, one replaces $u_\lambda(\mathbf{p},\ x)$ by $v_\lambda(\mathbf{p}, x)$. Substitution for φ from Eq. (4.5) into Eq. (4.2) gives, in the weak limit,

$$\begin{aligned}\sigma^{\text{in(out)}}(\varphi) &= \int d\mathbf{x} J_0^f(\varphi,\ \psi^{\text{in(out)}})\\ &= \sum_\lambda \int \frac{d\mathbf{p}}{\rho(E_p)} K_\lambda^*(\mathbf{p}) a_\lambda^{\text{in(out)}}(\mathbf{p}),\end{aligned} \tag{4.6}$$

where we have used the definitions of $a_\lambda^{\text{in(out)}}(\mathbf{p})$ from Eq. (2.15).

In some field theoretical models, $\varphi(x)$ can be taken as a c-number solution to Eq. (2.1). In these cases, $\sigma(\varphi)$ is conserved and Eq. (4.6) can be taken as an expansion in terms of a complete set of energy eigenstates $\{u_\lambda(\mathbf{p},\ x)\}$, each of which is also a solution to Eq. (2.1). Correspondingly,

$$\sigma(\varphi) = \sum_\lambda \int \frac{d\mathbf{p}}{\rho(E_p)} K_\lambda^*(\mathbf{p}) a_\lambda(\mathbf{p}),$$

with

$$a_\lambda(\mathbf{p}) = \int d\mathbf{x} J_0(u_\lambda(\mathbf{p},\ x),\ \psi(x)), \tag{4.7}$$

where $a_\lambda(\mathbf{p})$ destroys a particle with quantum numbers $\mathbf{p}$ and λ and satisfies the commutation or anticommutation relations of Eqs. (2.16) and (2.17). In general, $\sigma(\varphi)$ is a particle or antiparticle destruction or creation operator if φ is a c-number particle (antiparticle) function. This follows from the fact that in any given Lorentz frame it is the momentum operator

$$\mathbf{p} = \mathbf{p}^{\text{in(out)}}, \tag{4.8}$$

which determines the $\mathbf{x}$ dependence of $\psi(x)$. Since it is the $\mathbf{x}$ dependence that determines the orthogonality of φ with ψ, φ always acts to project onto either particle or antiparticle portions of ψ.

If Σ_φ is not conserved, we assume that $J_\mu(\ ,\ )$ has

the form of Eq. (2.32). Then whether or not Σ_φ is conserved, it satisfies the commutation relation

$$[\psi(x),\ \Sigma_\varphi(x_0)]=\eta_A\varphi(x). \tag{4.9}$$

This in turn implies a gauge transformation of the second kind,

$$e^{-\Sigma_\varphi(x_0)}\psi(x)e^{\Sigma_\varphi(x_0)}=\psi(x)+\eta_A\varphi(x), \tag{4.10}$$

directly from Eqs. (4.1)—(4.5). If S is a time independent self-adjoint symmetry operator as defined in Eqs. (2.26)—(2.31) and (2.37)—(2.42), then it is easy to show that

$$e^{i\beta S}\Sigma_\varphi(x_0)e^{-i\beta S}=\Sigma_{\varphi'}(x_0), \tag{4.11}$$

where

$$\varphi'(x)=e^{i\beta s}\varphi(x). \tag{4.12}$$

Two additional properties of this gauge transformation which follow from the commutation relations and Baker—Hausdorf—Campbell theorem are

$$e^{\Sigma_{\varphi_1}(x_0)}=e^{\sigma^\dagger\eta_A}e^{-\bar\eta_A\sigma}=\exp[-\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}J_0(\varphi,\ \varphi)], \tag{4.13}$$

and the multiplication rule

$$e^{\Sigma_{\varphi_1}(x_0)}e^{\Sigma_{\varphi_2}(x_0)}=e^{\Sigma_{\varphi_1+\varphi_2}(x_0)}\exp\{\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}\,[J_0(\varphi_2,\ \varphi_1)-J_0(\varphi_1,\ \varphi_2)]\}. \tag{4.14}$$

Note that if φ_1 and φ_2 are orthogonal,

$$\int d\mathbf{x}J_0(\varphi_2,\ \varphi_1)=0 \tag{4.15}$$

and

$$e^{\sum\varphi_1(x_0)}e^{\sum\varphi_2(x_0)}=e^{\sum\varphi_{1+}\varphi_2(x_0)}. \tag{4.16}$$

This is just a superposition theorem since if φ is expanded in terms of an orthonormal set of functions $\{u_\lambda\}$

$$\varphi=\sum_\lambda k_\lambda u_\lambda, \tag{4.17}$$

with

$$\int d\mathbf{x}J_0(u_\lambda,\ u_{\lambda'})=\delta_{\lambda\lambda'} \tag{4.18}$$

then a consequence of Eq. (4.16) is

$$e^{\sum\varphi(x_0)}=\prod_\lambda e^{\sum k_\lambda u_\lambda(x_0)}. \tag{4.19}$$

This is of particular importance for functional integration purposes where one defines

$$\int(\ )\mathcal{D}\varphi=\int d^2k_{\lambda_1}d^2k_{\lambda_2}\cdots d^2k_{\lambda_n}(\ ). \tag{4.20}$$

The $d^2k_{\lambda_i}$ integrations are over $d[\mathrm{Re}(k_{\lambda_i})]$ and $d[\mathrm{Im}(k_{\lambda_i})]$, i.e., the real and imaginary parts of each k_{λ_i}.

The gauge transformation of the second kind also generates the coherent state which is defined as

$$|\eta_A\varphi\rangle\equiv e^{\Sigma\varphi}|\Omega\rangle,$$

$$\langle\eta_A\varphi|\equiv\langle\Omega|e^{-\Sigma\varphi}, \tag{4.21}$$

where $|\Omega\rangle$ is the physical vacuum defined by

$$P_\mu|\Omega\rangle=0. \tag{4.22}$$

Because of Eq. (4.10) the average value of $\psi(x)$ taken over a coherent state is the usual result

$$\langle\eta_A\varphi|\psi(x)|\eta_A\varphi\rangle=\eta_A\varphi(x), \tag{4.23}$$

and if $\psi_+(x)$ is a pure particle destruction operator part of ψ, then

$$\psi_+(x)|\eta_A\varphi\rangle=\eta_A\varphi(x)|\eta_A\varphi\rangle. \tag{4.24}$$

Similarly, if $\varphi(x)$ is an antiparticle c-number L^2 function

$$\psi_-^\dagger(x)|\eta_A\varphi\rangle=\eta_A\varphi \tag{4.25}$$

where $\psi_-^\dagger$ is the antiparticle destruction operator. The inner product of two coherent states follows directly from Eqs. (4.13) and (4.14), i.e.,

$$\begin{aligned}\langle\eta_A\varphi_1|\eta_A\varphi_2\rangle&=\langle\Omega|e^{\Sigma\varphi_2-\varphi_1(x_0)}|\Omega\rangle\\&\times\exp\{-\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}[J_0(\varphi_2,\ \varphi_1)-J_0(\varphi_1,\ \varphi_2)]\}\\&=\exp[-\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}\,J_0(\varphi_2-\varphi_1,\ \varphi_2-\varphi_1)]\\&\quad\times\exp\{-\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}[J_0(\varphi_2,\ \varphi_1)-J_0(\varphi_1,\ \varphi_2)]\}\\&=\exp\{-\tfrac{1}{2}\bar\eta_A\eta_A\int d\mathbf{x}[J_0(\varphi_2,\ \varphi_2)+J_0(\varphi_1,\ \varphi_1)\\&\quad-2J_0(\varphi_1,\ \varphi_2)]\}.\end{aligned} \tag{4.26}$$

The inner product is normalized to unity for $\varphi_1=\varphi_2$ as it must, and never vanishes for $\varphi_1,\ \varphi_2\in L^2$.

The transformation properties of the coherent states are obtained from Eqs. (4.11) and (4.13), i.e.,

$$e^{i\alpha S}|\eta_A\varphi\rangle=|\eta_A\varphi'\rangle=|\eta_Ae^{i\alpha s}\varphi\rangle \tag{4.27}$$

and Eq. (4.26) can be used to show that

$$\langle\eta_A\varphi_1|e^{i\alpha S}|\eta_A\varphi_2\rangle=\exp[-\bar\eta_A\eta_A\int d\mathbf{x}J_0(\varphi_1,\ (1-e^{i\alpha s})\varphi_2)]\langle\eta_A\varphi_1|\eta_A\varphi_2\rangle. \tag{4.28}$$

For small real α this reduces to

$$\langle\eta_A\varphi_1|S|\eta_A\varphi_2\rangle=\int d\mathbf{x}J_0(\varphi_1,\ s\varphi_2)\langle\eta_A\varphi_1|\eta_A\varphi_2\rangle, \tag{4.29}$$

so that there is a one-to-one correspondence between coherent state matrix elements and the appropriate c-number matrix elements. This is an important observation since this means the domain of the q-number operator S is restricted by the c-number operator s.

If S is nonzero and

$$[\Sigma_\varphi,\ S]=0, \tag{4.30}$$

then

$$S|\varphi\rangle=e^{\Sigma\varphi}S|\Omega\rangle. \tag{4.31}$$

Then if $|\Omega\rangle$ is the physical vacuum and S a physical observable, $S|\Omega\rangle=0$, and

$$S|\varphi\rangle=0. \tag{4.32}$$

This implies that $|\varphi\rangle$ is degenerate with the vacuum.

An example of this is the gauge invariance of the second kind for the photon field. All physical operators are invariant under the gauge transformation generated by

$$\Sigma_{\partial_\mu\chi} = (i/2)\int d\mathbf{x}(\partial_\mu\chi \overleftrightarrow{\partial}_0 A_\mu), \tag{4.33}$$

or equivalently

$$\Sigma_\chi = (i/2)\int d\mathbf{x}(\chi \overleftrightarrow{\partial}_0 \partial_\mu A_\mu). \tag{4.34}$$

The coherent state $|\chi\rangle$ is a coherent state of unphysical photons degenerate with the vacuum.[18]

The coherent state $|\chi\rangle$ is a coherent state of unphysical by an argument due to Klauder[19] and Klauder, McKenna, and Currie[20] who discussed coherent state representations for the CCR and the density operator. Let all states be considered discrete for simplicity purposes and let $A_1^\dagger(x_0)$ be the creation operator defined by Eq. (2.14). Then from Eqs. (4.13) and (4.19) one finds

$$e^{\Sigma_\varphi(x_0)} = \prod_\lambda \{e^{k_\lambda A_\lambda^\dagger \eta_A} e^{-k_\lambda^* \bar{\eta}_A A_\lambda} \exp(-\tfrac{1}{2}\bar{\eta}_A\eta_A|k_\lambda|^2)\}, \tag{4.35}$$

$$\int |\varphi\rangle\langle\varphi| \mathcal{D}\varphi = \prod_\lambda \int \frac{d^2k_\lambda}{\pi} e^{-\bar{\eta}_A\eta_A|k_\lambda|^2} e^{k_\lambda A_\lambda^\dagger \eta_A} |\Omega\rangle\langle\Omega| e^{k_\lambda \bar{\eta}_A A_\lambda}$$

$$= \sum_{\{n_i\}} \left[\frac{(A_1^\dagger\eta_A)^{n_1}}{\sqrt{n_1!}} \frac{(A_1^\dagger\eta_A)^{n_2}}{\sqrt{n_2!}} \cdots |\Omega\rangle\langle\Omega| \frac{(\bar{\eta}_A A_\lambda)^{n_1}}{\sqrt{n_1!}} \frac{(\bar{\eta}_A A_\lambda)^{n_2}}{\sqrt{n_2!}} \cdots\right]. \tag{4.36}$$

This argument is completed by observing that the states

$$(1/\sqrt{n_1!})(A_1^\dagger(x_0))^{n_1}|\Omega\rangle = |n_1, x_0\rangle \tag{4.37}$$

are complete. In this argument and in subsequent ones to follow we consider $J_0(\varphi, \varphi)$ as a c-number. This is not true in general for interacting fields since J_0 may contain some other q-number field besides ψ, e.g., the fourth component of the photon field in a charged, scalar electrodynamics. However, such terms acting upon $|\Omega\rangle$ leave a vacuum for ψ, i.e., serve as a "dressing transformation" for the physical vacuum $|\Omega\rangle$. Thus, we consider $J_0(\varphi, \varphi)$ as a c-number with respect to the physical vacuum $|\Omega\rangle$. The mathematics required is just that given by Simon[21] in his discussion of projecting onto a unique physical vacuum.

In general, from Eqs. (4.13) and (4.21) a coherent state can be written as

$$|\eta_A\varphi\rangle = e^{\sigma^\dagger\eta_A}|\Omega\rangle \exp[-\tfrac{1}{2}\bar{\eta}_A\eta_A\int d\mathbf{y}J_0(\varphi, \varphi)]$$

$$= \sum_{n=0}^{\infty} \frac{(\sigma^\dagger\eta_A)^n}{\sqrt{n!}} |\Omega\rangle \exp[-\tfrac{1}{2}\bar{\eta}_A\eta_A\langle N\rangle], \tag{4.38}$$

where by Eq. (4.29)

$$\langle N\rangle = \langle\eta_A\varphi|N|\eta_A\varphi\rangle = \int d\mathbf{x}J_0(\varphi, \varphi), \tag{4.39}$$

and the number (charge) operator is obtained by setting $s=1$. Because φ projects out only particle (or antiparticle) states, Eq. (4.39) always gives the average number of particles (or antiparticles) in the coherent state.

For bosons, $\eta_B = 1$. When $\Sigma_\varphi(x_0)$ is conserved (time independent) Eqs. (4.6) and (4.7) can be used to reduce Eq. (4.38) to

$$|\varphi\rangle = e^{-(1/2)\langle N\rangle}\sum_n \frac{1}{\sqrt{n!}} \prod_{l=1}^n \sum_{\lambda_l}\int \frac{d\mathbf{k}_l}{\rho(E_l)} K_{\lambda_l}(\mathbf{k}_l)|k_1, \lambda_1, \ldots, k_n, \lambda_n\rangle \tag{4.40}$$

where

$$\langle N\rangle = \int d\mathbf{x}J_0(\varphi, \varphi) = \sum_\lambda \int \frac{d\mathbf{k}}{\rho(E_k)} |K_\lambda(\mathbf{k})|^2, \tag{4.41}$$

and

$$|k_1, \lambda_1, \ldots, k_n, \lambda_n\rangle = \frac{1}{\sqrt{n!}} a_{\lambda_1}^\dagger(\mathbf{k}_1)\cdots a_{\lambda_n}^\dagger(\mathbf{k}_n)|\Omega\rangle. \tag{4.42}$$

This is the usual result and was clearly known by Klauder, *et al.*[19-20]

The connection to the coherent states described by Glauber[22] is made by observing that Eq. (4.6) or (4.7) can be rewritten as

$$z^*a = \sum_\lambda \int \frac{d\mathbf{k}}{\rho(E_k)} K_\lambda^\dagger(\mathbf{k})a_\lambda(\mathbf{k})$$

$$= \int d\mathbf{x}J_0(\varphi, \psi), \tag{4.43}$$

where z^* is a c-number and a is a q-number destruction operator. If z is normalized according to

$$|z|^2 = \sum_\lambda \int \frac{d\mathbf{k}}{\rho(E_k)} |K_\lambda(\mathbf{k})|^2 = \int d\mathbf{x}J_0(\varphi, \varphi) = \langle N\rangle, \tag{4.44}$$

then the commutation relations for $a_\lambda(\mathbf{k})$ which follow from Eqs. (2.16) and (2.17) and which apply for any conserved current determine that

$$[a, a^\dagger] = 1 \tag{4.45}$$

and

$$[N, a^\dagger] = a^\dagger,$$

where

$$N = \int d\mathbf{x}J_0(\psi_+, \psi_+) = \sum_\lambda \int \frac{d\mathbf{k}}{\rho(E_k)} a_\lambda^\dagger(\mathbf{k})a_\lambda(\mathbf{k}). \tag{4.47}$$

Thus, an n-particle state can be constructed according to

$$(1/\sqrt{n!})(a^\dagger)^n|\Omega\rangle = |n\rangle,$$

with

$$N|n\rangle = n|n\rangle. \tag{4.48}$$

In terms of n-particle states the coherent state $|\varphi\rangle$ is given by

$$|\varphi\rangle = \exp[-\tfrac{1}{2}|z|^2]\sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle. \tag{4.49}$$

This is the Glauber form for a coherent state for which the completeness relationship is

$$\int \frac{d^2z}{\pi} |\varphi\rangle\langle\varphi| = 1,$$

where

$$d^2z = dxdy$$

and

$$z = x + iy. \tag{4.50}$$

However, in the Glauber formalism the states $|n\rangle$ are states of n particles, each of which have the same quantum numbers. Here the $|n\rangle$ particle states are states of n-particles each of which have the same *average* quantum numbers. To see this, consider the matrix element of a q-number symmetry operator S

$$\langle m|\int d\mathbf{x} J_0(\psi,\ s\psi)|n\rangle =$$
$$(1/\sqrt{n!}\ \sqrt{m!}) \times \int d\mathbf{x}\langle\Omega|J_0([a^m,\ \psi],\ s[\psi,\ a^{\dagger n}])|\Omega\rangle, \tag{4.51}$$

where J_0 is assumed to be normal ordered. The definitions of a and $a^\dagger$ can be used to show that

$$[a,\ a^\dagger_\lambda(\mathbf{k})] = (z^*)^{-1}K^*_\lambda(\mathbf{k}),$$

and

$$[\psi,\ a^\dagger] = (z)^{-1}\varphi(x). \tag{4.52}$$

These commutation rules applied to Eq. (4.51) give that

$$\langle m|S|n\rangle = \frac{mn}{\sqrt{mn}\,|z|^2}\int d\mathbf{x} J_0(\varphi,\ s\varphi)\ \langle m-1|n-1\rangle$$
$$= \frac{m\delta_{mn}\int dx J_0(\varphi,\ s\varphi)}{\int d\mathbf{x} J_0(\varphi,\ \varphi)}$$
$$= m\langle s\rangle\delta_{mn}. \tag{4.53}$$

For fermions, $\eta_F^2 = 0$ and Eq. (4.38) reduces to the result

$$|\eta_F\varphi\rangle = \exp[-(1/2)|z|^2\bar{\eta}_F\eta_F]\{|\Omega\rangle + \eta_F z|1\rangle\}. \tag{4.54}$$

Thus, these coherent states cannot contain two or more fermions with the same average quantum numbers. However, there may be more than one fermion in a state if some of the quantum numbers have different averages. Such a state can be constructed by taking a direct product of the different copies of Eq. (4.54). This type of coherent state can be constructed from the formalism leading to Eq. (4.54) by the replacement

$$\eta_F\varphi \to \sum_l \eta_{Fl}K_l u_l,$$
$$za \to \sum_l z_l a_l, \tag{4.55}$$

where now $(\bar{\eta}_{Fl}\ \eta_{Fl})^n = \delta_{n1}$ for all integers $n \geq 1$, but now $\eta_{Fl}\ \eta_{Fm} \neq 0$ when $l \neq m$.

All this is for the case where φ is an L^2 c-number function. However, a useful property can be obtained when φ is *not* L^2. As Klauder has shown, the measure dm of a field theory can be written as

$$\mathrm{dm}(\ ) = \mathrm{dm}_F(\ ) \oplus \mathrm{dm}_{AC}(\ ) \tag{4.56}$$

where $\mathrm{dm}_F(\ )$ is the free measure and $\mathrm{dm}_{AC}(\ )$ is the pure absolutely continuous measure. Nontrivial interactions "live on" $\mathrm{dm}_{AC}(\ )$. For the translation given in Eq. (4.10),

$$e^{-\Sigma_\varphi}\psi(x)e^{-\Sigma_\varphi} = \psi(x) + \eta_A\varphi(x), \tag{4.10'}$$

with $\varphi \in L^2$, we remain in the free (noninteracting!) measure $\mathrm{dm}_F(\ )$. To escape L^2, we must translate according to a c-number function, specifically designated as $c(\ )$, which becomes non-L^2 as a function of a singularity parameter γ. Then, it must be emphasized, Σ_c generates a transformation $e^{\Sigma c}$ which is *improper* rather than unitary. In the present approach, the *same current* J_μ or J^f_μ generates both the unitary $(e^{\Sigma\varphi})$ and improper $(e^{\Sigma c})$ gauge transformation of the second kind. The improper transformation will be discussed in the next two sections.

5. HILBERT SPACE, OPERATOR EXTENSIONS AND BANACH SPACE

So far, a good deal of heuristic formalism has been presented without mathematical rigor. In this section, we study the Hilbert space properties operators and solutions, including essential self-adjointness and domains. This will be based on the work of Reeh[3] who gave a rigorous discussion of symmetries and symmetry breaking in the presence of a conserved current such as our Eq. (2.9), and of Klauder[5] who gave a rigorous discussion of exponential Hilbert spaces and bilinear operators. However, in contrast to Reeh[3] and Klauder[5] we have a differential equation which determines the conserved current $J_\mu(\ ,\ )$ and thereby the Lebesque measure $d\mu$ of $L^2(\ ,\ d\mu)$.

The differential equations may also have solutions which are not L^2, such as for example virtual or resonance eigenfunctions which have complex eigenvalues. Equivalently, the differential operator and the operators constructed from the current $J_0(\ ,\ )$, as in Eq. (2.37), may be non-self-adjoint. These operators "live on" a Banach space $\mathcal{B}$ and *not* a Hilbert space $\mathcal{H}$. Some properties of these operators will also be discussed in this section.

One of the first tasks of constructing a rigorous quantized field theory is to take the "classical field ansatz" for a Hamiltonian and symmetry operators, such as Eqs. (2.37) and (2.44), and find self-adjoint extensions with a common dense domain. The pioneering work on essential self-adjointness[23] for the operators appearing in quantum mechanics was by Kato.[24] Two recent studies by Faris and Lavine[25] and Schroeck[26] are especially relevent to our work. Faris and Lavine[25] have extended the essential self-adjointness proofs to operators which are not semibounded, which is necessary for charge operators with positve and negative eignevalues. Schroeck[26] has extended the essential self-adjointness proofs to the purely absolutely continuous measure. This is necessary for interacting systems[7] and for irreducible representations of the Weyl algebra of the fields.[27] A number of other references can be traced through the bibliographes of these studies.[23-26] We refer the interested reader to their references for further discussion of these issues.

In our approach the emphasis is placed on the study of the c-number differential equation, Eq. (2.1), which simplifies matters considerably. Thus, whenever an operator s or h is self-adjoint on a dense domain, Eq. (4.29) implies that S or H is self-adjoint on C_0^∞. Furthermore, Eq. (4.53) with the existence of a unique vacuum implies that the Fock representation exists and that the

eigenvalues of S and H are *average* quantum numbers. Thus, if the partial differential operator $D(\partial)$ is well enough behaved, wave packets with continuous spectra can be treated as well as point and absolutely continuous spectra.

Although *nonlocality* will not be investigated in this work, our analysis seems to include this possibility since S and H can be smeared over the spectrum of s and h without affecting our results.

It follows from Eq. (4.29) that the existence of a classical c-number set of solutions is absolutely required for the existence of a quantized theory. One can ask what this implies about quantum electrodynamics, which is certainly one of the most successful second quantized theories at making accurate experimental predictions. Happily, thanks to the nice work by Gross[28] the answer is that a classical solution exists. And from a study of Gross's work one expects that given enough strength and patience, his study could be generalized to the case $M \neq 0$ as in our example in Sec. 3. However, that is a worthy project for (possible) future study and we simply conjecture that Sec. 3 has an underlying classical theory and leave matters there, for the present.

Next, let us turn to the non-L^2 solutions. Let $\mathcal{H}_u$ denote the c-number space of L^2 solutions $\{u_n\}$, and let $\mathcal{B}$ denote the Banach space of all solutions $\{u_n,\ c\}$ to Eqs. (2.1), not all of which are L^2. Thus $\mathcal{H}_u \subset \mathcal{B}$ properly. Define $\mathcal{C}_0$ as the dual to $\mathcal{B}$ in terms of the conserved current, Eq. (2.12), so that $\mathcal{C}_0 \subset \mathcal{H}_u$ properly. For this case, there exists a one-one, continuous inclusion mapping

$$\mathcal{C}_0 \subset \mathcal{H}_u \subset \mathcal{B}, \tag{5.1}$$

both algebraically and topologically. This structure is called a rigged Hilbert space.[29] Thus, there will exist solutions $c \in \mathcal{B}$, $c \notin \mathcal{H}_u$, and at least one L^2 solution, say u_0, for which

$$\int d\mathbf{x} J_0(u_0,\ c) \to \infty. \tag{5.2}$$

For example, this describes the case when the coupling constant reaches a value such that the pole corresponding to one of the u_n's is driven into the unphysical sheet. Equivalently, the solution is driven out of $\mathcal{H}_u$ into $\mathcal{B}$.

A general c-number function $\varphi \in \mathcal{B}$ can be written as

$$\hat{\varphi} = \sum_n \beta_n u_n + c, \tag{5.3}$$

where the β_n's are constant, $u_n \in \mathcal{H}_u$, $c \subset \mathcal{B}$ and $c \notin \mathcal{H}_u$ and a transformation

$$U_{\hat{\varphi}} = e^{\Sigma_{\hat{\varphi}}(x_0)}, \tag{5.4}$$

can be defined in parallel to Eq. (4.13). This transformation is now improper since by Eqs. (4.7), (2.16), (2.17), and (5.2),

$$[\sigma(\hat{\varphi}),\ \sigma^\dagger(\hat{\varphi})] = \int d\mathbf{x} J_0(\hat{\varphi},\ \hat{\varphi}) \to \infty, \tag{5.5}$$

and is at best an isometry since by Eqs. (4.39), (4.40)

$$\langle \hat{\varphi} | \Omega \rangle = 0, \tag{5.6}$$

thereby making $|\hat{\varphi}\rangle$ unitarily inequivalent to the Fock space $\{|m\rangle\}$. In fact, it is easy to show that

$$\langle \hat{\varphi} | \varphi \rangle = 0, \tag{5.7}$$

whenever

$$|\int d\mathbf{x} J_0(c,\ c)| / |d\mathbf{x} J_0(c,\ \varphi)| \to \infty, \tag{5.8}$$

so that all of the $|\hat{\varphi}\rangle$'s are disjoint from $\mathcal{H}_F'$, the Fock representation Hilbert space.

It is also clear that $|\hat{\varphi}\rangle$ is not in a Fock representation, from the fact that[30]

$$\langle \hat{\varphi} | N | \hat{\varphi} \rangle / \langle \hat{\varphi} | \hat{\varphi} \rangle = \int d\mathbf{x} J_0(\hat{\varphi},\ \hat{\varphi}) \to \infty \tag{5.9}$$

by Eqs. (4.39) and (5.2). However, other operators S as in Eq. (4.29), or even a current algebra among the S_i's may be defined on the $|\hat{\varphi}\rangle$'s whenever $J_0(\hat{\varphi},\ s\hat{\varphi})$ or $J_0(\hat{\varphi},\ [s_i, s_j]\hat{\varphi})$ exists. For example, given

$$D(\partial_x) - j(x) = -\frac{d^2}{dx^2} + \frac{\gamma(\gamma+1)}{x^2} - 2\epsilon_n + \omega^2 x^2, \tag{5.10}$$

and a solution

$$\hat{\varphi}(x) = \beta(1 + 1/|x|^\gamma) e^{-\omega x^2/2} \tag{5.11}$$

with

$$J_0(\hat{\varphi},\ \hat{\varphi}) = \hat{\varphi}^*(x)\hat{\varphi}(x), \tag{5.12}$$

where β, ω, γ are positive, real parameters, Eq. (5.9) clearly implies that the number operator is ill-defined for $\gamma \geq \frac{1}{2}$. Equally clearly, for large enough real numbers r,

$$\int dx J_0(\hat{\varphi},\ x^r\hat{\varphi}) < \infty. \tag{5.13}$$

Thus, operators and algebras of operators can be perfectly well defined in such a non-Fock representation. Such constructs are prominent in Klauder's ultralocal model field theories which we discuss in the next section.

6. ULTRALOCAL MODEL QUANTUM FIELD THEORIES

The ultralocal quantum field theories which Klauder[5-10] has recently formulated and solved can be contructed in terms of bilinear forms derivable from differential equations in an internal, space-time independent, variable λ. Ultralocal models are non-covariant mode field theories in which the spatial gradients have all been dropped from a covariant model. Since Klauder's results[5-10] directly apply to our case, we shall omit the many proofs which make his work (and ours) mathematically sound.

To establish the ultralocal analog of Eq. (2.1), let $A(x,\ \lambda)$ be the field operator which satisfies the field equation

$$D(\partial_\lambda,\ \partial_0,\ v(\lambda))A(x,\ \lambda) = 0, \tag{6.1}$$

where $v(\lambda)$ is the c-number scalar "interaction potential" to be determined, $x = (\mathbf{x},\ ix_0)$, and λ is a space time independent variable. We assume that a total set of stationary, L^2, c-number states exists which satisfy

$$D(\partial_\lambda,\ -i\epsilon_n,\ v(\lambda))u_n(\lambda) = 0, \tag{6.2}$$

$$u_n(\lambda,\ x_0) = u_n(\lambda)\exp(-i\epsilon_n x_0), \tag{6.3}$$

and

$$\int d\lambda J_0(u_n(\lambda,\ x_0),\ u_m(\lambda,\ x_0)) = \delta_{nm}, \tag{6.4}$$

and J_0 is the timelike component of the conserved

current in the variables $(\lambda,\ x_0)$. If $\{f_n(\mathbf{x})\}$'s are a total set of c-number solutions for the x-dependence, then a general c-number solution is of the form

$$\varphi(x,\ \lambda)=\sum_n \beta_n f_n(\mathbf{x})u_n(\lambda,\ x_0), \tag{6.5}$$

where the β_n's are constants and n is (possibly continuum cardinality) the index set of allowed quantum numbers.

As shown in Sec. 2, time independent q-number annihilation operators can be defined as

$$a_n(\mathbf{x})=\int d\lambda\, J_0(u_n(\lambda,\ x_0),\ A(x,\ \lambda)), \tag{6.6}$$

and

$$\begin{aligned}\sigma(\varphi)&=\int d\mathbf{x}\int d\lambda\, J_0(\varphi(x,\ \lambda),\ A(x,\ \lambda))\\ &=\sum_n \int d\mathbf{x}\bar{f}_n(\mathbf{x})a_n(\mathbf{x}).\end{aligned} \tag{6.7}$$

Following the basic quantization postulate of Eq. (2.16), we require that

$$[\sigma(\varphi_1),\ \sigma(\varphi_2)]=0 \tag{6.8}$$

and

$$[\sigma(\varphi_1),\ \sigma^\dagger(\varphi_2)]=\int d\mathbf{x}\int d\lambda\, J_0(\varphi_1(x,\ \lambda),\ \varphi_2(x,\ \lambda)). \tag{6.9}$$

The unequal time commutators which follow from Eq. (6.9), using the technique described in HT, are

$$[A(x,\ \lambda),\ \sigma(\varphi)]=0, \tag{6.10}$$

$$[A(x,\ \lambda),\ \sigma^\dagger(\varphi)]=\varphi(\mathbf{x},\ \lambda), \tag{6.11}$$

and

$$[A(x,\ \lambda),\ A^\dagger(y,\ \lambda)]=\sum_n f_n(\mathbf{x})u_n(\lambda,\ x_0)\bar{u}_n(\lambda',\ y_0)\bar{f}_n(\mathbf{y}). \tag{6.12}$$

A Hamiltonian operator H and a momentum operator $\mathbf{P}$ can be defined as

$$H=\int d\mathbf{x}\int d\lambda\, J_0\left(A(x,\ \lambda),\ \left(i\frac{\partial A}{\partial t}\right)(x,\ \lambda)\right), \tag{6.13a}$$

and

$$\mathbf{P}=\int d\mathbf{x}\int d\lambda\, J_0(A(x,\ \lambda),\ (-i\nabla_{\mathbf{x}}A)(x,\ \lambda)). \tag{6.13b}$$

Since H is to be essentially self-adjoint, substitutions from Eq. (6.1) into Eq. (6.13a) can be made such that

$$H=\int d\mathbf{x}\int d\lambda\, J_0\left(\left(i\frac{\partial A}{\partial t}\right)(x,\ \lambda),\ A(x,\ \lambda)\right) \tag{6.14}$$

on a dense domain. In fact if any operator S is closed and semibounded then by Theorem VIII. 15 of Reed and Simon[23] it has a unique self-adjoint quadratic form. Since S must be semibounded for a stable vacuum, we need only to show closure on the coherent states to show that Eqs. (6.13) and (6.14) are uniquely self-adjoint.

As a result of Eq. (6.11), a gauge transformation of the second kind may be defined for these fields as

$$e^{-\Sigma_\varphi}A(x,\ \lambda)e^{-\Sigma_\varphi}=A(x,\ \lambda)+\varphi(x,\ \lambda), \tag{6.15}$$

where

$$\Sigma_\varphi=\sigma^\dagger(\varphi)-\sigma(\varphi), \tag{6.16}$$

with $\varphi(x,\ \lambda)$ given by Eq. (6.5). As discussed in Sec. 4, the coherent states $|\varphi\rangle$, $\varphi(\ )\in L^2(\ ,\ d\mu)$, can be written as

$$|\varphi\rangle=e^{\Sigma_\varphi}|\Omega\rangle. \tag{6.17}$$

Now following Klauder,[6] we introduce an improper transformation $e^{\Sigma_{\hat{\varphi}}}$ as in Eq. (5.4) with

$$\hat{\varphi}(x,\ \lambda)=\varphi(x,\ \lambda)+c(\lambda), \tag{6.18}$$

where c satisfies the condition of Eq. (5.2) with $c\in\mathcal{B}$ but $c\notin\mathcal{H}_c$. The operators

$$\begin{aligned}B(x,\ \lambda)&=e^{-\Sigma_{\hat{\varphi}}}A(x,\ \lambda)e^{\Sigma_{\hat{\varphi}}}\\ &=A(x,\ \lambda)+\hat{\varphi}(x,\ \lambda),\end{aligned} \tag{6.19}$$

can be defined which also satisfy Eqs. (6.1) and (6.12). Although $B(x,\ \lambda)$ satisfies the same differential equation and commutation rules as $A(x,\ \lambda)$, it cannot be used to generate a Fock representation because it is unitarily inequivalent to $A(x,\ \lambda)$. Equivalently, a number operator

$$N_B=e^{-\Sigma_\varphi}\int d\mathbf{x}d\lambda\, J_0(A(x,\ \lambda),\ A(x,\ \lambda))e^{\Sigma_\varphi}, \tag{6.20}$$

does not exist.

To reproduce Klauder's formulation of ultralocal models,[6-10] choose J_0 with the canonical form of Eq. (2.32), specifically he chooses

$$J_0(A(x,\ \lambda),\ q(\lambda,\ \partial_\lambda)A(x,\lambda))=A^\dagger(x,\ \lambda)q(\lambda,\ \partial_\lambda)A(x,\ \lambda), \tag{6.21}$$

and assume $c(\lambda)$ satisfies the time independent equation, Eq. (6.2), with $\epsilon_n=0$

$$D(\partial_\lambda,\ 0,\ v(\lambda))c(\lambda)=0. \tag{6.22}$$

In technical terms,[6] c must be real, even and nowhere vanishing with

$$\int_{-\infty}^{\infty} c^2(\lambda)d\lambda=\infty, \tag{6.23}$$

and

$$\int_{-\infty}^{\infty} \lambda^2c^2(\lambda)d\lambda/(1+\lambda^2)<\infty. \tag{6.24}$$

Thus, $c(\lambda)\sim|\lambda|^{-\gamma}$ with γ real and $\frac{1}{2}\leq\gamma<\frac{3}{2}$ near the origin in λ. A useful subclass of Eqs. (6.23) and (6.24) is Klauder's family of model functions,[6] i.e.,

$$c(\lambda)=e^{-y(\lambda)}/|\lambda|^\gamma, \tag{6.25}$$

where $y(\lambda)$ is an even polynominal with degree $(y)\geq 2$.

Clearly, Eq. (6.22) determines the interaction potential $v(\lambda)$ once $c(\lambda)$ is given. This is an interesting aspect of Klauder's work on ultralocal models. Any model function $c(\lambda)$ can be choosen providing it satisfies Eqs. (6.23), and (6.24). This choice determines the energy spectrum $\{\epsilon_m\}$ and L^2 eigenfunctions $\{u_n\}$. This together with Eq. (4.29) assures the closure needed to establish that S is unique and self-adjoint as a quadratic form. Note that the action of driving a state out of L^2 is accomplished by continuous variation of γ so that $\sum_\varphi\to\sum_{\hat{\varphi}}$ becomes singular as $\gamma>\frac{1}{2}$.

Of particular interest is the transformation generated $\sum_c(\tilde{\varphi}=c)$ because the time independence of $c(\lambda)$ and Eqs. (6.13) and (6.14) insure the invariance of the Hamiltonian and momentum under

$$H = e^{-\Sigma_c} H e^{\Sigma_c} = \int d\mathbf{x} \int d\lambda\, J_0\left(B(x,\lambda), \left(i\frac{\partial B}{\partial t}\right)(x,\lambda)\right)$$
$$= \int d\mathbf{x} \int d\lambda\, J_0\left(A(x,\lambda), \left(i\frac{\partial A}{\partial t}\right)(x,\lambda)\right), \tag{6.26a}$$

and

$$\mathbf{P} = e^{-\Sigma_c}\mathbf{P}\, e^{\Sigma_c} = \int d\mathbf{x} \int d\lambda\, J_0(B(x,\lambda), (-i\nabla B)(x,\lambda))$$
$$= \int d\mathbf{x} \int d\lambda\, J_0(A(x,\lambda), (-i\nabla A)(x,\lambda)). \tag{6.26b}$$

Because the operators A and B satisfy the same differential equation, J_0 has the same form for both sets of operators. Thus one can formally define operators as in Sec. 2,

$$\mathcal{Q}_B(x) = \int d\lambda\, J_0(B(x,\lambda), qB(x,\lambda)),$$
$$Q_B(x_0) = \int d\mathbf{x}\, \mathcal{Q}_B(\mathbf{x}), \tag{6.27}$$

and, since J_0 has the canonical form,

$$[\mathcal{Q}_{B1}(x), \mathcal{Q}_{B2}(y)]_{x_0=y_0}$$
$$= \delta(\mathbf{x}-\mathbf{y}) \int d\lambda\, J_0(B(x,\lambda), [q_1, q_2]B(x,\lambda)). \tag{6.28}$$

For the form of J_0 used by Klauder, Eq. (6.21), two such formal operators are the ultralocal field

$$\Phi(x) = \int d\lambda\, J_0(B(x,\lambda), \lambda B(x,\lambda)), \tag{6.29}$$

and its "canonical partner"

$$\pi(x) = \int d\lambda\, J_0\left(B(x,\lambda), \left(-i\frac{\partial B}{\partial \lambda}\right)(x,\lambda)\right). \tag{6.30}$$

Because of Eqs. (6.23) and (6.24), and the fact that

$$\int_{-\infty}^{\infty} \lambda c^2(\lambda) d\lambda = 0, \tag{6.31}$$

$\Phi(x)$ is well behaved but $\pi(x)$ is not defined. Furthermore,

$$[\Phi(x), \pi(y)]_{x_0=y_0} = i\delta(\mathbf{x}-\mathbf{y}) N_B, \tag{6.32}$$

is not defined because of N_B which contains the infinite field strength renormalization, $\int c^2(\lambda) d\lambda = \infty$. Interestingly, Hegerfeldt and Klauder[31] have shown that no partner to Φ can exist on a dense domain and therefore no canonical partner to Φ exists. Since their study was based upon the expectation functional, they excluded *all possible* canonical partners not just those suggested by dynamical considerations, such as Eq. (6.30).

It is straightforward to show Klauder's result[32] for a renormalized bilinear F_R,

$$F_R = Z^{-1} f(Z\Phi(x))$$
$$= \int dx\, J_0(B(x,\lambda), f(\lambda)B(x,\lambda)), \tag{6.33}$$

where formally $Z^{-1} = \delta(0)$ holds for all J_0 of the canonical form of Eq. (2.32) if $f(\lambda)$ is expandable in powers of λ. This result is useful for discussing the ultralocal interaction potential because the form of Eq. (6.1) used by Klauder is

$$D(\partial_\lambda, \partial_0, v(\lambda)) = h(\partial_\lambda, \lambda) - i\frac{\partial}{\partial t}, \tag{6.34}$$
$$h = -\frac{\partial^2}{\partial \lambda^2} + v(\lambda). \tag{6.35}$$

The Hamiltonian of Eq. (6.26) then becomes

$$H = \int d\mathbf{x} \int d\lambda B^\dagger(x,\lambda) h B(x,\lambda). \tag{6.36}$$

It is therefore reasonable to define the renormalized ultralocal interaction potential V_R

$$V_R = Z^{-1} v(Z\Phi(\mathbf{x})) = \int d\lambda B^\dagger(x,\lambda) v(\lambda) B(x,\lambda). \tag{6.37}$$

This illustrates the fact that the choice of $c(\lambda)$ determines both the dynamics [through $v(\lambda)$] and the representation of the field Φ (through $\exp\sum_c$), linking the dynamics and the field representation Equations (6.28) and (6.33) are also useful for obtaining the ultralocal form of equations of motion. A direct calculation using Eq. (6.34) gives

$$[\Phi, H] = i\dot{\Phi} = i\pi(x), \tag{6.38}$$

which is undefined although

$$[[\Phi, H], H] = -\ddot{\Phi}$$
$$= Z^{-1} V_R'(Z\Phi) \tag{6.39}$$

is perfectly well defined as a quadratic form, where V_R' is the functional derivative of V_R with respect to $(Z\Phi)$. The absence of spatial gradients in Eq. (6.39) is the characteristics feature of ultralocal fields.

The-form equation of motion displayed in Eq. (6.39) is inadequate for dynamics, because the construction of a solution requires powers of the field which is more information than the *form* contains. Klauder[8] has solved this problem by constructing *operator* solutions. He found that a generalized field operator

$$\Phi_R^\theta(x) = \int d\lambda\, J_0(B(x,\lambda), \lambda^\theta B(x,\lambda)) \tag{6.40}$$

is needed, with $\theta = p + 2\gamma$, p a strictly positive integer $\{1, 2, 3, \cdots\}$, and γ the singularity parameter. When θ is a noninteger, λ^θ is defined as the odd extension,

$$\lambda^\theta = \begin{cases} |\lambda|^\theta, & \lambda > 0 \\ -|\lambda|^\theta, & \lambda < 0. \end{cases} \tag{6.41}$$

Although it is far from obvious, Φ_R^θ has an *operator* equation of motion,

$$\ddot{\Phi}_R^\theta = -[H, [H, \Phi_R^\theta]]$$
$$= 4H(x) - \sum_n w_{2n}[\Phi_R^{(2n)}(x) - \langle\Omega|\Phi_R^{(2n)}(x)|\Omega\rangle]. \tag{6.42}$$

The operator equation of motion specified by Eq. (6.42), although complicated, is trouble-free and specifies viable dynamics. We will not repeat Klauder's argument here, but refer to his paper[8] for details. Clearly, ultralocality is preserved under renormalization and time evolution. Thus, our formalism reproduces Klauder's ultralocal solutions[6-10] when we require that our conserved current $J_0(\ ,\ )$ has the form of Eq. (6.21). Presumably with the bilinear formalism presented here, Klauder's ultralocal models can be extended to include $D(\partial)$ other than Eqs. (6.34) and (6.35). In general, this will require a different $J_\mu(x)$ which can then be used to exploit the symmetries of the theory.

Klauder[7] has shown that the set of all

$$|f\rangle = e^{i\Phi(f)}|\Omega\rangle; \quad \Phi(f) = \int d\mathbf{x} f(x)\Phi(x), \tag{6.43}$$

is a non-Fock representation which spans $\mathcal{H}$ and is in the domain of $H^{3/4-(1/2)\gamma-\epsilon}$ $(\epsilon > 0)$. This is to be contrasted with the set of all

$$|\varphi\rangle = e^{\Sigma_\varphi}|\Omega\rangle, \tag{6.44}$$

which is a Fock representation of $\mathcal{H}$ and is in the domain of H. Observe that

$$\langle c | \varphi \rangle = 0 \tag{6.45}$$

as well as

$$\langle c | f \rangle = 0, \tag{6.46}$$

so that $|c\rangle$ is disjoint from both the Fock and non-Fock representation of $\mathcal{H}$. Thus the disjoint state $|c\rangle$, which has the same quantum numbers as the vacuum, *condenses* out of $\mathcal{H}$. This is suggestive of some sort of "collective exitation." "Higher exitations" can be formed from other singular functions c_n which are solutions to

$$hc_n = \epsilon_n' c_n. \tag{6.47}$$

7. CONCLUSIONS

We have extended the formalism of Hammer and Tucker[2] to interacting cases. Within this formalism, coherent states are discussed including symmetry operators on coherent state domains. The concept of a "smeared" Fock representation with spectrum given by average eigenvalues is developed. Singular or improper gauge transformations of the second kind, which follow from non-L^2 solutions to the field equations, are shown to generate representations unitarity inequivalent to the smeared Fock space. Such representations are necessary to support interactions.

Two completely different examples were discussed, (i) a Dirac field minimally coupled to a massive vector field, and, (ii) J. R. Klauder's ultralocal models, as examples of our formalism. Neither of these examples have a canonical Lagrangian formulation. However, in each case the field equations determine very different conserved currents which give the structure of the theory.

In discussing essential self-adjointness and domains, q-number operators S, H are discussed in terms of underlying "internal symmetry" operators s, h. This suggests an "induced self-adjointness," and may possibly be related to methods developed by Dashen, Hasslacher and Neveu.[33]

Some key concepts used in this work are taken from Klauder's studies[5] of representations of CCR on an exponential Hilbert space. In that paper, bilinear "fields" occur naturally on a coherent state domain and *infinite divisibility* is shown to provide several interesting examples. We show that these ideas can be used to generalize (and add rigor to) Hammer and Tucker[2] as well. We therefore suggest that (noncurrent algebra) bilinear formalisms and ultralocal model quantum field theories still have a good deal to teach us about "quantized fields." We hope to explore this conjecture in future works.

The formalism developed here may provide an alternate way to study many-body systems since analogs to both the L^2 gauge transformation $e^{\Sigma\varphi}$, and the non-L^2 gauge transformation $e^{\Sigma\Phi}$ play a key role in the *Boson method* which Umezawa and collaborators[34-36] are presently developing with functional integration techniques.

ACKNOWLEDGMENTS

One of us (B. D. F.) wishes to thank the ISU Physics Department of Ames Laboratory-ERDA for their hospitality during 1974—1975 and the University of Missouri-Columbia for a sabbatical leave. Also, we both thank Dr. J. R. Klauder, Dr. Bing-lin Young, Dr. R. H. Tucker and Professor H. Umezawa for useful and interesting comments and helpful criticism.

*Prepared for the Energy Research and Development Administration under Contract No. W-7405-Eng-82.

†Permanent address: Department of Physics, University of Missouri-Columbia, Columbia, Missouri 65201.

[1]Y. Takahashi and H. Umezawa, Prog. Theor. Phys. **9**, 14 (1953); Nucl. Phys. **51**, 193 (1964). However, we especially recommend the monograph, Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, London, 1969) which is much more complete.
[2]C. L. Hammer and R. H. Tucker, J. Math. Phys. **12**, 1327 (1971).
[3]H. Reeh, Fortsch. Phys. **16**, 687 (1968).
[4]D. Kastler, D. W. Robinson, and J. A. Swieca, Commun. Math. Phys. **2**, 108 (1966).
[5]J. R. Klauder, J. Math. Phys. **11**, 609 (1970).
[6]J. R. Klauder, Commun. Math. Phys. **18**, 307 (1970).
[7]J. R. Klauder, Acta. Phys. Austriaca Suppl. **8**, 227 (1971).
[8]J. R. Klauder, "Heisenberg Equations of Motion," *Proc. Sym. Basic Question Elem. Part. Phys.*, Münich, June 8-18, (1971).
[9]J. R. Klauder, in *Lectures in Theoretical Physics, Mathematical Methods in Theoretical Physics, Vol. XIV. B.* edited by W. Brittin (Colorado Associated Universities Press, Boulder, 1973), pp. 329—421.
[10]J. R. Klauder, Ann. Phys. (NY) **79**, 111 (1973).
[11]R. H. Tucker and C. L. Hammer, Phys. Rev. D **3**, 2448 (1971).
[12]C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).
[13]S. S. Schweber, *An Introduction to Relativistic Field Theory*, (Harper and Row, New York, 1961), See esp. Eq. (3.64), p. 211.
[14]G. N. Fleming, Phys. Rev. **137**, B188 (1965).
[15]A. Sankaranarayanan and R. H. Good, Jr., Phys. Rev. **140**, B509 (1965).
[16]Ref. 9. See esp. pp. 350, 351.
[17]J. Rzewuski, *Field Theory, Vol.* 2 (Iliffe Books, London, and PWN Polish Scientific Publishers, Warsaw, 1969). See esp. pp. 252—55. We thank Dr. Klauder for pointing this out.
[18]Ref. 13. See esp. p. 250.
[19]J. R. Klauder, J. Math. Phys. **4**, 1055 (1963).
[20]J. R. Klauder, J. McKenna and D. G. Currie, J. Math. Phys. **6**, 734 (1965).
[21]B. Simon, *The $P(\varphi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, New Jersey, 1974). See esp. pp. 348—57.
[22]R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[23]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. 1. Functional Analysis* (Academic, New York, 1972). See esp. the discussion on pp. 249—317 for the best introductory discussion available on unbounded operators, essential self-adjointness, and quadratic forms.
[24]T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[25]W. Faris and R. B. Lavine, Commun. Math. Phys. **35**, 39 (1974).
[26]F. E. Schroeck, Jr., "Cores for Operators in Quantum Mechanics," J. Math. Phys. (accepted for publication).

[27]G. Goldin, J. Math. Phys. **12**, 462 (1971).
[28]L. Gross, Commun. Pure Appl. Math. **19**, 1 (1966).
[29]I.M. Gel'Fand and N.Ya. Vilenkin, *Generalized Functions, Vol. 4, Applications of Harmonic Analysis* (Translated by Amiel Feinstein) (Academic, New York, 1964). See esp. pp. 103 *et seq.*
[30]For an infinite particle number operator the quantity z in Eq. (4.43) is reminescent of an infinite field strength renormalization constant. In this case, Eq. (4.53) gives a renormalized matrix element and for more general c-number functions $\hat{\varphi}$, only the most singular terms survive.
[31]G.C. Hegerfeldt and J.R. Klauder, Nuovo Cimento A **10**, 723 (1972).
[32]See esp. p. 317 of Ref. 6.
[33]R.F. Dashen, B. Hasslacher and A. Neveu, Phys. Rev. D **10**, 4114, 4130, and 4138 (1974); and Princeton Institute for Advanced Study Preprint (to be published). These authors and other workers on quark confinement have studied the quantization of nonlinear classical systems in two-dimensional space-time.
[34]H. Umezawa, in *Renormalization and Invariance in Quantum Theory*, edited by E.R. Caianiello (Plenum, New York, 1974).
[35]L. Leplae, H. Umezawa, and F. Mancini, Phys. Rep. 10C, 152 (1974).
[36]H. Matsumoto, N.J. Papastamatiou, and H. Umezawa, *Infrared Effect of the Goldstone Boson and the Order Parameter* (University of Wisconsin-Milwaukee Preprint UWM-4867-75-5, May 1975).

The Weyl tensor and stationary electrovac space-times

F. Paul Esposito

Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221

E. N. Glass*

Department of Physics, University of Windsor, Windsor, Ontario N9B 3P4, Canada
(Received 14 July 1975; revised manuscript received 29 September 1975)

It is shown that static electrovac spaces admit only static electric type Maxwell fields, and it is proved that a necessary and sufficient condition for a stationary electrovac space-time to be static is that the Weyl tensor and the Maxwell field tensor both be pure electric type.

In recent work[1,2] the connection between the magnetic part of the Weyl tensor and the rotational properties of stationary vacuum space-times and perfect fluid space–times has been discussed. Both classes of space–times have a preferred timelike vector field with respect to which the magnetic and electric parts of the Weyl tensor are defined. It has been proved that a necessary and sufficient condition for a stationary vacuum space-time to be static[1] is that the Weyl tensor be pure electric type; also a shear-free perfect fluid space–time is irrotational[2] if and only if the Weyl tensor is pure electric type.

This note extends the theorem on stationary vacuum spaces to include source-free electromagnetic fields. All equations here numbered with the prefix A refer to kinematic equations for the Weyl tensor which are collected in an appendix of Ref. 2.

Theorem: A necessary and sufficient condition for a stationary electrovac space–time to be static is that the Weyl tensor and the Maxwell field tensor both be pure electric type.

Proof: Consider a stationary electrovac space–time with the Einstein–Maxwell field equations[3]

$$R_{\mu\nu}=\kappa[F_{\mu\alpha}F_{\nu}{}^{\alpha}-\tfrac{1}{4}g_{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}], \tag{1}$$

$$\nabla_{\nu}F^{\mu\nu}=\nabla_{\nu}\overset{*}{F}{}^{\mu\nu}=0. \tag{2}$$

The space–time has a timelike Killing vector ξ^{α} with norm $\phi^2:=\xi^{\alpha}\xi_{\alpha}$. The unit vector along ξ^{α} is defined by

$$u^{\alpha}:=\phi^{-1}\xi^{\alpha}.$$

The Weyl tensor and Ricci tensor comprise the algebraically independent parts of the curvature tensor,

$$R^{\mu\nu}{}_{\alpha\beta}=C^{\mu\nu}{}_{\alpha\beta}-2\delta^{[\mu}{}_{[\alpha}R^{\nu]}{}_{\beta]}+\tfrac{1}{3}\delta^{\mu}{}_{[\alpha}\delta^{\nu}{}_{\beta]}R,$$

and in this work $R=0$.

The electric and magnetic parts of the Weyl tensor and Maxwell field tensor are defined with respect to the unit Killing vector as follows:

$$E_{\mu}:=F_{\mu\nu}u^{\nu}, \tag{3a}$$

$$B_{\mu}:=\overset{*}{F}{}_{\mu\nu}u^{\nu}, \tag{3b}$$

$$E_{\alpha\beta}:=C_{\alpha\mu\beta\nu}u^{\mu}u^{\nu}, \tag{4a}$$

$$B_{\alpha\beta}:=\overset{*}{C}{}_{\alpha\mu\beta\nu}u^{\mu}u^{\nu}. \tag{4b}$$

All electric and magnetic vectors and tensors are orthogonal to u^{α}. In addition, the electric and magnetic parts of the Weyl tensor are symmetric and trace-free.

The Ricci tensor expressed in terms of electric and magnetic fields is given by

$$R_{\mu\nu}=\kappa[-\tfrac{1}{2}(E^2+B^2)u_{\mu}u_{\nu}+2u_{(\mu}\eta_{\nu)\alpha\beta}E^{\alpha}B^{\beta}+E_{\mu}E_{\nu}+B_{\mu}B_{\nu}+\tfrac{1}{2}\gamma_{\mu\nu}(E^2+B^2)], \tag{5}$$

where $E^2:=-E^{\alpha}E_{\alpha}$, $B^2:=-B^{\alpha}B_{\alpha}$, $\gamma_{\mu\nu}:=g_{\mu\nu}-u_{\mu}u_{\nu}$, and γ^{μ}_{ν} acts as a projection operator onto the quotient 3-spaces to u^{α}. The operation of projection onto these 3-spaces will be denoted by $\perp$ which projects all free indices. Maxwell's equations become

$$\nabla_{\mu}B^{\mu}+a_{\mu}B^{\mu}+2\omega^{\mu}E_{\mu}=0, \tag{6a}$$

$$\nabla_{\mu}E^{\mu}+a_{\mu}E^{\mu}-2\omega^{\mu}B_{\mu}=0, \tag{6b}$$

$$\perp\dot{E}^{\mu}-\eta^{\mu\alpha\beta}(\omega_{\alpha}E_{\beta}+\nabla_{\alpha}B_{\beta}-a_{\alpha}B_{\beta})=0, \tag{7a}$$

$$\perp\dot{B}^{\mu}-\eta^{\mu\alpha\beta}(\omega_{\alpha}B_{\beta}-\nabla_{\alpha}E_{\beta}+a_{\alpha}E_{\beta})=0, \tag{7b}$$

where

$$a_{\alpha}:=u^{\beta}\nabla_{\beta}u_{\alpha}=\phi^{-2}\xi^{\beta}\nabla_{\beta}\xi_{\alpha}=-\phi^{-1}\nabla_{\alpha}\phi,$$

$$\omega^{\mu}:=\tfrac{1}{2}\eta^{\mu\nu\alpha\beta}u_{\nu}u_{\alpha;\beta}=(\phi^{-2})\tfrac{1}{2}\eta^{\mu\nu\alpha\beta}\xi_{\nu}\xi_{\alpha;\beta},$$

and a dot denotes the covariant derivative in the u^{α} direction.

The necessary part of the theorem follows when the twist of the Killing vector is equated to zero, i.e., $\omega^{\alpha}=0$. Equation (A4) shows that $B_{\mu\nu}=0$ independent of field equations when the space–time admits a twist-free timelike Killing vector. In fact, the reflection symmetry of static spaces implies that $B_{\mu\nu}$ and B_{μ} must both be zero directly from their definitions.[4] Maxwell's equation (7a) then requires that the electric field be constant. Thus the electromagnetic field in a static electrovac space must always be static electric type.

To prove sufficiency, consider the divergence equation (A8):

$$\nabla^{\mu}B_{\mu\alpha}=(\kappa/2)[E_{\alpha;\beta}B^{\beta}-E_{\alpha}B^{\beta}{}_{;\beta}+a_{\beta}(E_{\alpha}B^{\beta}-B_{\alpha}E^{\beta})-\omega_{\alpha}(E^2+B^2)-2u_{\alpha}\omega^{\mu}\eta_{\mu\rho\sigma}E^{\rho}B^{\sigma}]-a^{\beta}B_{\alpha\beta}-3\omega^{\beta}E_{\alpha\beta}, \tag{8}$$

where Eq. (5) and the shear-free property of the Killing vector have been used. With $B_{\mu\nu}=B_{\mu}=0$, Eq. (8) yields

$$\omega^\beta[E_{\alpha\beta} + (\kappa/6)E^2\gamma_{\alpha\beta}] = 0. \tag{9}$$

When the Weyl tensor is either pure electric or pure magnetic type it must be[5] Petrov type I, D, or O. $\omega_\alpha = 0$ follows immediately for type O. For types I and D

$$\det[E_{\alpha\beta} + (\kappa/6)E^2\gamma_{\alpha\beta}] \neq 0, \tag{10}$$

since a vanishing determinant implies $E_{\alpha\beta}$ would have three equal eigenvalues, which is not possible. Thus $\omega^\alpha = 0$. ■

ACKNOWLEDGMENTS

The authors are grateful for the hospitality of the Aspen Center for Physics where this work was completed, and the valuable suggestions of the referee.

*Supported in part by NRC of Canada grant A9343.

[1]E. N. Glass, J. Math. Phys. **15**, 1930 (1974).

[2]E. N. Glass, J. Math. Phys. (to appear).

[3]Covariant derivatives are denoted by semicolons and ∇_μ. Parentheses around indices denote symmetrization and brackets around indices denote antisymmetrization. The dual operator is $\frac{1}{2}\eta^{\mu\nu\alpha\beta}$ and the dual is denoted by an asterisk. The alternating tensor in the quotient space is $\eta_{\mu\nu\alpha} := \eta_{\mu\nu\alpha\beta}\delta^\beta$.

[4]Static systems are time reversal invariant and imply vanishing magnetic Maxwell field only in the absence of magnetic monopoles. Such "pseudoscalar" sources are incompatible with the idea that static systems should have no preferred time orientation.

[5]P. Jordan, J. Ehlers, and W. Kundt, Akad. Wiss. Lit. Mainz Abh. Math. Naturwiss. Kl., No. 2 (1960).

Erratum: Gauge transformations and normal states of the CCR [J. Math. Phys. 16, 2086 (1975)]

M. Fannes and A. Verbeure

Katholieke Universiteit Leuven, Belgium
(Received 3 October 1975)

As pointed out by F. Rocca, Theorem II.4 fails as it stands. The correct statement should be:

Theorem II.4: Let ω_ρ be a normal state on $\overline{\Delta(H,\sigma)}$ such that there exists a dense set $\mathcal{D} \subset \mathcal{L}_2(H, d\tilde{\psi})$ with the property $[1/\omega(W_0)]\mathcal{D} \times \mathcal{D} \subset \mathcal{F}\mathcal{L}_1(H, d\tilde{\psi})$, where $f \times g$ stands for the σ-convolution and $\mathcal{F}$ for the σ-Fourier transform. Then the linear hull of the set $\{\omega_\rho \circ \tau_\chi \mid \chi \in K\}$ is norm dense in the set of normal states on $\overline{\Delta(H,\sigma)}$.

In spite of the less strong character of the conditions, the proof can be performed along the same lines as that of F. Rocca in Ref. 3. It is easy to check that all states ω such that $\omega(W_\psi) \neq 0$ and infinitely differentiable satisfy the condition.

In Lemma II.2, A should be a Hilbert–Schmidt operator and the state ω should satisfy $\omega(W_\psi) \neq 0$ for all $\psi \in H$.

By those changes Lemma II.3 proves that span $\{\tau_\chi \rho \mid \chi \in K\}$ is dense in the set of Hilbert–Schmidt operators.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*

